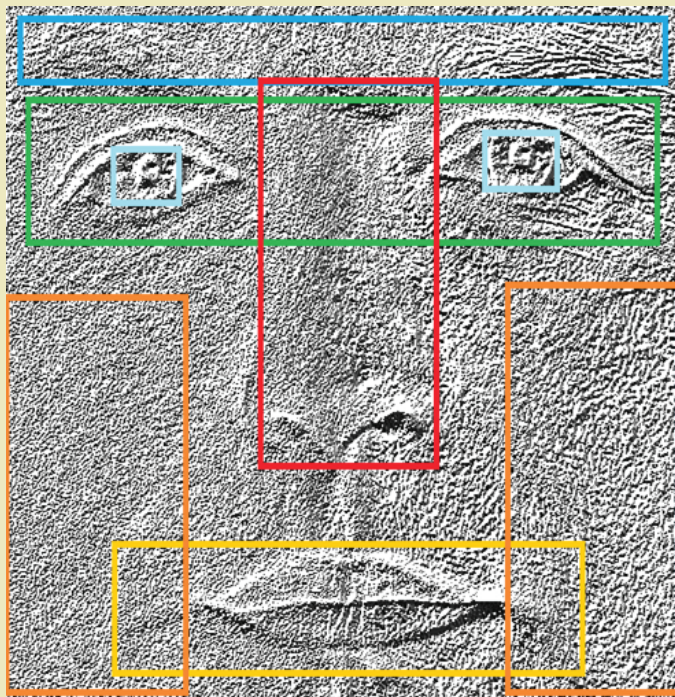




Paweł Karczmarek

Selected Problems of Face Recognition and Decision-Making Theory



M
O
N
O
G
R
A
F
I
E

Lublin 2018

To Kinga and Piotr

Monografie – Politechnika Lubelska



Politechnika Lubelska
Wydział Elektrotechniki i Informatyki
ul. Nadbystrzycka 38A
20-618 Lublin

Paweł Karczmarek

Selected Problems of Face Recognition and Decision-Making Theory



Wydawnictwo Politechniki Lubelskiej
Lublin 2018

Recenzent:

dr hab. Zenon Sosnowski, prof. Politechniki Białostockiej

Publikacja wydana za zgodą Rektora Politechniki Lubelskiej

© Copyright by Politechnika Lubelska 2018

ISBN: 978-83-7947-332-8

Wydawca: Wydawnictwo Politechniki Lubelskiej

www.biblioteka.pollub.pl/wydawnictwa

ul. Nadbystrzycka 36C, 20-618 Lublin

tel. (81) 538-46-59

Druk: TOP Agencja Reklamowa Agnieszka Łuczak

www.agencjatorp.pl

Elektroniczna wersja książki dostępna w Bibliotece Cyfrowej PL www.bc.pollub.pl

Nakład: 100 egz.

Table of Contents

Preface	7
References.....	11
1. Preliminaries of Computational Intelligence and Decision-Making Theory..	15
1.1. Analytic Hierarchy Process.....	15
1.2. Particle Swarm Optimization.....	17
1.3. Fuzzy Measure and Choquet Integral	17
1.4. T-norms.....	19
1.5. Aggregation Functions.....	19
1.6. Face Image Datasets	20
1.7. Conclusions.....	21
References.....	21
2. Local Descriptors in Face Recognition.....	25
2.1. Introduction.....	25
2.2. Main Concept of Local Descriptors	28
2.3. Gabor filters	32
2.4. Local descriptors in an application to Gabor wavelet images.....	33
2.5. Distance measures.....	34
2.6. Experimental results.....	34
2.7. Chain Code-Based Local Descriptor and Its Extension	41
2.8. Experimental Results	45
2.9. Conclusions.....	57
References.....	58
3. An Application of Linguistic Descriptors to the Face Recognition Problems	65
3.1. Introductory Notes	65
3.2. Facial Features Saliency	70
3.3. Observer's Classification Process Confidence	73
3.4. Experimental Studies	75
3.5. Classification Based on the Linguistic Descriptors	83
3.6. Linguistic and Numeric Information Fusion.....	84
3.7. The Process of Classification.....	88
3.8. Experimental Studies	89
3.9. Voting and Linguistic Descriptors for Face Recognition	91
3.10. Experimental Results	93
3.11. Conclusions.....	98
References.....	98
4. User Centric Graphic Enhancements of Methods of Decision-Making.....	105
4.1. Enhancements of Analytic Hierarchy Process	105
4.2. Non-linear Transformation of the Reciprocal Matrix	108
4.3. Minimization of Inconsistency.....	109
4.4. Experimental Results	110
4.5. Fuzzy Extensions of the Analytic Hierarchy Process	118

4.6. Experimental Results	119
4.7. An Application of Graphical Interface to the Biometric Features Description.....	120
4.8. Conclusions.....	124
References.....	125
5. Fuzzy Measures and Facial Features Saliency.....	128
5.1. Introduction.....	128
5.2. A General Scheme of the Aggregation Process	131
5.3. Fuzzy Measure and Choquet Integral Interpretation.....	132
5.4. Experimental studies.....	133
5.5. Sugeno Fuzzy Measure Densities Evaluation	146
5.6. Experimental Results	148
5.7. Conclusions.....	157
References.....	158
6. Aggregation Mechanism in Face Recognition.....	164
6.1. Introduction.....	164
6.2. Aggregation Functions for Face Recognition	166
6.3. Similarity Measures	168
6.4. Experimental Studies	169
6.5. General Results	170
6.6. Generalizations of Aggregation Functions	183
6.7. Experimental Results	184
6.8. Further Generalizations of Choquet Integral	189
6.9. Experimental Results	190
6.10. Conclusions.....	193
References.....	193
Conclusions and Future Work	199

Preface

Face recognition has been one of the most challenging research problems for over last three decades. The main reasons are its vast applications such as missing persons or criminal identification, access control, driving license, ID card, or passport verification, and many others. Other causes of the growth of interest in the face recognition techniques are the huge computational power of modern CPUs and GPUs, the spread of personal computers and security or surveillance systems. These areas of applications need their own dedicated systems related to the problems which may occur. They are partial occlusion of a face, distance to the camera, illumination, pose, age, tattoos, or emotion expression changes, quality of an image, and system requirements. Admittedly, in contrast to other biometric solutions such as fingerprints or iris recognition, facial recognition is a non-invasive method, i.e., there is no need to involve a verified person in the data acquisition process, however, problems with the personal data protection are one of the factors inhibiting its development.

Over the years, many studies on face recognition have been published. Let us refer to the most important ones. Kanade (1977) proposed the method based on geometric dependencies between the facial parts. Turk & Pentland (1991) and Belhumeur et al. (1997) proposed two significant and still enhanced methods of dimensionality reduction, namely Eigenfaces (based on Principal Component Analysis, PCA) and Fisherfaces (based on PCA followed by Linear Discriminant Analysis, LDA). Other important methods proposed in the 90's were Support Vector Machines (SVM, Phillips et al. 1998) or Elastic Bunch Graph Matching (EBGM, Wiskott et al. 1997). Ahonen et al. (2004) introduced to face recognition the theory of local descriptors, particularly the well-known Local Binary Pattern based on the local descriptions of specific pixels' neighborhoods and building histograms of such descriptions. Kwak & Pedrycz (2005) worked on information aggregation (fusion) on a basis of classification processes conducted for the whole images or their parts. Recently, the most popular have been two branches of methods. The first is a group of sparse representation-based approaches of dimensionality reduction, see Wright et al. (2009). The second, and giving accurate results, is the field of deep learning methods, see Huang et al. 2012, Sun et al. 2014, etc. An interested reader can find many comparative analyses and reviews of the face recognition approaches such as Forczmański & Kukharev (2007), Hassaballah & Aly (2015), Siwek & Osowski (2018), Wójcik et al. (2018), or Zhao et al. (2003).

Obviously, many of the methods of facial recognition have their origins and utilize more general techniques of image analysis (Tadeusiewicz & Korohoda 1997), classification (Stapor 2011), fuzzy sets (Zadeh 1965), rough sets (Pawlak 1991, Pawlak & Skowron 2007), Computational Intelligence (Kacprzyk & Pedrycz 2015, Rutkowski 2008), neural networks (Tadeusiewicz 1993, Duch et

al. 2000, Goodfellow et al. 2016), Granular Computing (Bargiela & Pedrycz 2001, Pedrycz 2013, Pedrycz & Homenda 2013), or software libraries (e.g., Rafajłowicz et al. 2009).

The classification problems are strictly related to the tasks of decision-making theory, particularly multi-criteria decision-making theory (see, e.g., Słowiński 1998, Ishizaka & Nemery 2013). One of the connectors is the problem of aggregation of the classifiers results which can be solved by using aggregation functions (Beliakov et al. 2007, Grabisch et al. 2009, Gągolewski 2015). The examples of modern aggregation functions are t-norms (Klement & Mesiar 2005), Ordered Weighted Averaging operators (OWA, Yager & Kacprzyk 2012), or Choquet integral (Sugeno 1974). Newer studies show that the special conditions on aggregation operators can be weakened leading to the increase of the final classification result, see, e.g., Lucca et al. 2015. Moreover, one of the advantages of using the decision-making theory is that it can be a helpful tool when making final decision about classification/misclassification of the subject in biometrical systems by its operator. It is worth noting that in biometric systems, particularly in forensic applications, the final decision about the criminal identification must be made by an expert. Hence, the methods such as Analytic Hierarchy Process (Saaty 1980, Saaty & Vargas 2012) can be successfully applied to the problems of describing the suspect, group decision on the classification, etc. It is worth to note that the systems of classification, when incorporated in the processes of linguistic modeling reflecting the way people describe other people using the decision-making tools, may significantly improve their performance. Moreover, there are still the problems of decision-making theory which need solving. One of them is developing an interface allowing the expert to become fully independent from the limitations resulting from the use of numbers and language descriptions.

In this book, we present the solutions of several problems appearing in face classification and related decision-making theory. Despite the area of interest presented in this study is wide, all issues discussed have some common features. The first is that all the techniques can be used to solve the classification or identification problems, even the AHP method. The second common feature is that local and linguistic descriptors, AHP, and fuzzy measure can serve as vehicles to extract or describe the features (e.g., biometric facial features) or their importance. Next, in case of all the methods considered here, the presence of an expert making a decision on classification or the reflection of the way human describe others is essential. Therefore, the proposed methods correspond to the paradigm of human-centric computing. Finally, at the base of the majority of considered methods are fuzzy techniques or the proposed models can be easily extended by using fuzzy approach. The organization of the material reflects the main objectives of the study.

In chapter 2, we present the common and most frequently appearing techniques in the next parts of the book. They are, in particular, Analytic

Hierarchy Process and its variants, Particle Swarm Optimization (Kennedy et al. 2001), being one of the most efficient and fast optimization tools. Moreover, we discuss the λ -fuzzy measure and Choquet integral, playing a pivotal role in the aggregation tasks. Finally, the general concept of t-norms and aggregation operators are recalled.

In chapter 3, we study the important problem of aging and age differences in face recognition which often appears when compared are the photographs taken in large intervals. We thoroughly examine the efficiency and potential application of many classical local descriptors such as Local Binary Pattern or Multi-Block LBP (Chan et al. 2007). Using the specialized FG-NET dataset with maximal age difference between the times of taking the photographs of one subject being 54 years we demonstrate the ability of local descriptors and local descriptors applied to Gabor wavelet images to cover the aging problem. Moreover, analyzed are various measures of similarity between the histogram features. At the same chapter, we propose a novel and original local descriptor which, unlike other local descriptors, is built on chain codes describing the local cross-neighborhood of a pixel. We analyze its two variants (i.e., simple and block-based) in relation to the problems of illumination, pose, or age on a basis of different public image datasets. Moreover, its robustness to noise and face occlusion is discussed.

Chapter 4 is devoted to the application of linguistic descriptors to face identification algorithms. People are strongly efficient in recognizing faces. Moreover, the knowledge coming from human recognition processes observation and opinion yielded from experts' assessments can be an invaluable input to the computational facial recognition systems and methods. First, we propose a method of facial features saliency evaluation based on the Analytic Hierarchy Process. The technique is equipped in novel method of confirmation of experts' confidence based on the entropy measure. On a basis of the knowledge coming from experienced practitioners in the fields of psychology and criminology we determine the importance of the facial features and their groups in the processes of classification. Discussed are also potential constraints of the method and possible ways of its enhancement. Next, we propose a method of incorporating the results of AHP process carried for specific faces in the tasks of classification of the faces both by humans as well as with an application of computational methods related to real measures (lengths) of facial features. Presented are also processes based on classification of faces with an application of the voting and importance of the features determined by experts using AHP. It is worth noting that an application of PSO leads to the significant improvement of consistency of experts' answers.

In chapter 5, we partially depart from the face recognition problems to the more general idea of designing the graphical interface to the AHP method. People often do not understand the scales (both linguistic and numeric) associated with various decision-making tools. Moreover, they feel somehow

uncomfortable with such constraints. Therefore, we propose two simple graphical components, namely a slider and a dial arc, to improve the level of consistency of the reciprocal matrices produced in the AHP processes. In a series of experiments with crisp and fuzzy variants of the AHP we demonstrate the effectiveness of our proposal. Finally, the method is applied to the description of chosen facial features. We prove that the usage of a slider significantly improve the correctness of face parts' description by an expert.

Chapter 6 deals with an application of the Sugeno fuzzy measure and Choquet integral to the problems of face recognition. We are interested in an investigation of the abilities of the fuzzy measure to reflect the saliency of the information contained in specific facial regions. Moreover, we evaluate the role of the regions and their combinations in the facial recognition processes. In this context, we compare the results of experiments obtained with the presence of subjects and realized by computational methods of face recognition. Finally, we design the Sugeno fuzzy measure on a basis of the psychological experiments reported in the literature and related to the saliences of facial cues and propose a novel model of the face identification mechanism. In the second part of the chapter, we propose a method of finding the optimal values of the fuzzy measure densities (Choquet integral weights) on a basis of so-called *positive* and *negative* optimization reflecting the maximization and minimization of the Choquet integral values in the processes of aggregation of classification results, respectively.

The last chapter of this book is directly devoted to the aggregation mechanism in face recognition problems. On a basis of the most popular classifiers, namely Eigenfaces and Fisherfaces, we examine various aggregation techniques in the context of their efficiency. The results of computational experiments suggest that Choquet integral, voting, median, and modified Hamacher t-norm are one of the best aggregation operators when the classifiers are based on the facial parts. Next, we analyze the generalizations and modifications of the Choquet integral based on the replacement of the product t-norm under the integral sign by other t-norms and operators. The proposal has demonstrated a very good performance in a comparison with other methods studied in the literature.

Chapter 2 is partially based on the ideas presented in Bereta et al. 2013 where the local descriptors in according to the aging problems were discussed and Karczmarek et al. 2016b; 2017c where the CCBLD local descriptor and its enhanced version were introduced, respectively. The material contained in the third chapter is partially based on the papers by Karczmarek et al. 2017d, where the saliency of facial features was discussed and by Karczmarek et al. 2019b, where an application of the AHP method to the face identification problem was proposed, and work (Karczmarek et al. 2016a) containing the discussion on voting with an application of facial weights. Chapter 4 deals with the material presented in an article by Karczmarek et al. (2018a) where a proposition of

describing facial features using graphical inputs to the AHP method was widely discussed and in the paper by Karczmarek et al. (2018c) where an application of graphical tools to the decision-making process was proposed. The content of the fifth chapter uses some ideas presented in the paper by Karczmarek et al. (2014), where the discussion on the fuzzy measure in a context of face recognition was presented, and works by Karczmarek et al. (2017a; 2017b), where the evaluation of the fuzzy measure densities was performed. The last chapter is partially based on the material contained in Karczmarek et al. (2019a), where a comparison of various aggregation techniques in an application to face recognition was presented and Karczmarek et al. (2018b), where an analysis of a generalized Choquet integral as an aggregation operator was carried.

I would like to thank Prof. Witold Pedrycz for invaluable discussions, in-depth comments on the manuscript, and motivating me to convey the research. Moreover, I would like to thank Dr. Adam Kiersztyn who was the first reader of the book.

The study was cofounded by National Science Centre, Poland (grant no. 2014/13/D/ST6/03244).

References

- Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with Local Binary Patterns. In: Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science **3021**, pp. 469–481 (2004)
- Bargiela, A., Pedrycz, W.: Granular Computing: An introduction. Springer, New York (2001)
- 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)
- Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: A guide for practitioners. Springer–Verlag, Berlin Heidelberg (2007)
- Bereta, M., Karczmarek, P., Pedrycz, W., Reformat, M.: Local descriptors in application to the aging problem in face recognition, *Pattern Recognit.* **46**, 2634–2646 (2013)
- Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern histograms for face recognition. In: Lee, S.-W., Li, S. Z. (Eds.): *ICB 2007*, Lecture Notes in Computer Science **4642**, pp. 809–818 (2007)
- Duch, W., Korbicz, J., Rutkowski, L., Tadeusiewicz, R.: Neural networks. Exit, Warszawa (2000)
- Forcmański, P., Kukharev, G.: Comparative analysis of simple facial features extractors. *J. Real-Time Image Proc.* **1** (4), 239–255 (2007)

- Gągolewski, M.: Data fusion. Theory, methods, and applications. Institute of Computer Science, Polish Academy of Sciences, Warsaw (2015)
- Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. The MIT Press, Cambridge (2016)
- Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: Aggregation functions. Cambridge University Press, Cambridge (2009)
- Hassaballah, M., Aly, S.: Face recognition: challenges, achievements and future directions. *IET Comput. Vis.* **9** (4), 614–626 (2015)
- 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)
- Ishizaka, A., Nemery, P.: Multi-criteria decision analysis: Methods and software. Wiley (2013)
- Kacprzyk, J., Pedrycz, W.: Springer handbook of Computational Intelligence. Springer-Verlag, Berlin Heidelberg (2015)
- Kanade, T.: Computer recognition of human faces. Birkhauser, Basel (1977)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W.: An application of graphic tools and analytic hierarchy process to the description of biometric features. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J. (Eds.): *Artificial Intelligence and Soft Computing. ICAISC 2018. Lecture Notes in Computer Science* **10842**, 137–147 (2018a)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W.: An evaluation of fuzzy measure for face recognition. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L., Zurada J. (Eds.): *Artificial Intelligence and Soft Computing. ICAISC 2017. Lecture Notes in Computer Science* **10245**, pp. 668–676 (2017a)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W.: Generalized Choquet integral for face recognition. *Int. J. Fuzzy Syst.* **20** (3), 1047–1055 (2018b)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W.: On developing Sugeno fuzzy measure densities in problems of face recognition. *Int. J. Mach. Intell. Sensory Signal Proc.* **2** (1), 80–96 (2017b)
- 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 (2017c)
- Karczmarek P., Kiersztyn A., Pedrycz W., Dolecki M.: Linguistic descriptors and analytic hierarchy process in face recognition realized by humans. In: Rutkowski L., Korytkowski M., Scherer R., Tadeusiewicz R., Zadeh L., Zurada

- J. (Eds.): *Artificial Intelligence and Soft Computing. ICAISC 2016. Lecture Notes in Computer Science* **9692**, pp. 584–596 (2016a)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W., Dolecki, M.: Linguistic descriptors in face recognition. *Int. J. Fuzzy Syst.* (2019b) [in press]
- Karczmarek, P., Kiersztyn, A., Pedrycz, W., Rutka, P.: A study in facial features saliency in face recognition: An analytic hierarchy process approach. *Soft Comput.* **21** (24), 7503–7517 (2017d)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W., Rutka, P.: Chain code-based local descriptor for face recognition. In: Burduk, R., Jackowski, K., Kurzyński, M., Woźniak, M., Żołnierek A. (Eds.): *Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015*, pp. 307–316 (2016b)
- Karczmarek, P., Pedrycz, W., Kiersztyn, A.: Graphic Interface to Analytic Hierarchy Process and Its Optimization, *Eur. J. Oper. Res.* [submitted] (2018c)
- Karczmarek, P., Pedrycz, W., Kiersztyn, A., Dolecki, M.: A comprehensive experimental comparison of the aggregation techniques for face recognition. *Irani. J. Fuzzy Syst.* (2019a) [in press]
- Karczmarek, P., Pedrycz, W., Reformat, M., Akhoundi, E.: A study in facial regions saliency: A fuzzy measure approach. *Soft Comput.* **18**, 379–391 (2014)
- Kennedy, J. F., Eberhart, R. C., Shi, Y.: *Swarm intelligence*. Academic Press, San Diego (2001)
- Klement, E. P., Mesiar, R.: *Logical, algebraic, analytic, and probabilistic aspects of triangular norms*. Elsevier, Amsterdam (2005)
- Kwak, K.-C., Pedrycz, W.: Face recognition: A study in information fusion using fuzzy integral. *Pattern Recognit. Lett.* **26**, 719–733 (2005)
- Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., Mesiar, R., Kolesárová, A., Bustince, H.: The notion of pre-aggregation function. In: Torra, V., Narakawa Y. (Eds.): *MDAI 2015, LNAI 9321*, pp. 33–41 (2015)
- Pawlak, Z.: Imprecise categories, approximations and rough sets. In: *Rough Sets. Theory and Decision Library (Series D: System Theory, Knowledge Engineering and Problem Solving)*, vol 9. Springer, Dordrecht, pp. 9–32 (1991)
- Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Inf. Sci.* **177** (1), 3–27 (2007)
- Pedrycz, W.: *Granular Computing. Analysis and design of intelligent systems*. CRC Press, Boca Raton (2013)
- Pedrycz, W., Homenda, W.: Building the fundamentals of granular computing: a principle of justifiable granularity. *Appl. Soft Comput.* **13** (10), 4209–4218 (2013)
- Phillips, P. J.: Support vector machines applied to face recognition. *Adv. Neural Inf. Process. Syst.* **11**, 803–809 (1998)

- Rafajłowicz, E., Rafajłowicz, W., Rusiecki, A.: Algorithms of image processing and introduction to the work with OpenCV library. Wrocław University of Technology Press (2009)
- Rutkowski, L.: Computational intelligence: Methods and techniques. Springer Science & Business Media, Berlin Heidelberg (2008)
- Saaty, T. L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
- Saaty, T. L., Vargas, L. G.: Models, methods, concepts & applications of the analytic hierarchy process. Springer, New York (2012)
- Siwek, K., Osowski, S.: Deep neural networks and classical approach to face recognition – comparative analysis. *Electrotech. Rev.* **94** (4), 1–4 (2018)
- Słowiński, R.: Decision analysis, operations research and statistics. Handbooks of Fuzzy Sets Series, vol. 5. Kluwer Academic Publishers, Boston (1998)
- Stapor, K.: Classification methods in computer vision. PWN, Warszawa (2011)
- Sugeno, M.: Theory of fuzzy integral and its applications. Dissertation. Tokyo Institute of Technology, Tokyo (1974)
- Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, pp. 1891–1898 (2014)
- Tadeusiewicz, R.: Neural networks. Academic Press, Warszawa (1993)
- Tadeusiewicz, R., Korohoda, P.: Computer analysis and image processing. Progress of Telecommunication Foundation, Kraków (1997)
- Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1991)
- Wiskott, L., Fellous, J.-M., Krüger, N., von der Malsburg, C.: Face recognition by Elastic Bunch Graph Matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 775–779 (1997)
- Wójcik, W., Gromaszek, K., Junisbekov, M.: Face recognition: Issues, methods and alternative applications. In: Ramakrishnan, S. (Ed.): Face Recognition. Semisupervised Classification, Subspace Projection and Evaluation Methods. IntechOpen, pp. 1–22 (2016)
- 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)
- Yager, R. R., Kacprzyk, J.: The ordered weighted averaging operators: Theory and Applications. Springer Science+Business Media, New York (2012)
- Zadeh, L. Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
- Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surv.* **35**, 399–458 (2003)

1. Preliminaries of Computational Intelligence and Decision-Making Theory

In this chapter, we recall the most important tools and frameworks of Computational Intelligence and decision-making theory. We briefly discuss the concepts of Analytic Hierarchy Process, Particle Swarm Optimization, fuzzy measure, Choquet integral, and t-norms. Moreover, we discuss the general concept of aggregation functions. At the end of this chapter, we place short description of various image datasets used in this book along with the sources.

1.1. Analytic Hierarchy Process

Analytic Hierarchy Process is one of the most important multi-criteria decision-making theory techniques which can be successfully applied both when one as well as a group of experts is engaged in a decision-making process. It was introduced by Saaty (Saaty 1980; 1988). Let us briefly discuss its most essential algorithmic and conceptual aspects. The method can be widely applied to generate the decisions related with choice, prioritization, ranking, order, evaluation, and other objectives at many levels of hierarchy. This may be understood as follows. At the beginning, one has to outline the hierarchy of concepts appearing in the problem. The top element is the goal. Lower are placed the criteria, and finally, at the bottom of the hierarchy placed is the set of alternatives. Assume that there are n alternatives (elements of interest). They can be organized in the form of a so-called reciprocal matrix R of the size $n \times n$. The entities of the matrix are the resulting values of pairwise comparisons between the alternatives. To estimate the values one frequently uses the following scale, see Saaty (1980), Saaty & Vargas (2012):

- equal importance (1),
- weak importance (2),
- moderate importance (3),
- moderate plus importance (4),
- strong (essential) importance (5),
- strong plus importance (6),
- very strong (demonstrated) importance (7),
- very, very strong importance (8),
- and extreme importance (9)

or preference of one feature over the second. The full question can be formulated as: “To which extent the alternative A is preferred over the alternative B ?”. It means that an expert (a group of experts) has to make $n(n - 1)/2$ pairwise comparisons. Of course, in the literature there are many different crisp versions of the above scale (1, 2, ..., 7; 1, 3, 5, 7, 9; 1, 3, 5, 7; 1, 1.1, 1.2, ..., 1.9, 2, 3, ..., 9, etc., or other scales, see Saaty (1977), Harker & Vargas (1987), Ishizaka &

Labib (2011) as well as built of the interval, fuzzy, or granular numbers, see, e.g., van Laarhoven & Pedrycz (1983), Pedrycz (2013), Kubler et al. (2016). The matrix R is called reciprocal since it satisfies the condition that for each entry r_{ij} ($i, j = 1, 2, \dots, n$) we have $r_{ij} = 1/r_{ji}$. It implies that the diagonal elements are $r_{ii} = 1$.

When working with real life problems one is interested in preserving possible high level of consistency of the pairwise comparison process results. This task is realized with using a so-called inconsistency index and a consistency ratio. The first of these parameters is defined as $\nu = \frac{\lambda_{\max} - n}{n - 1}$. Here, $\lambda_{\max} \geq n$ is a maximal eigenvalue of R . The second parameter, namely the consistency ratio μ reads as $\mu = \nu/r$. The value r was found empirically by Saaty & Mariano (1979) as the average consistency index of 500 random reciprocal matrices. Its values are as follows: 0, 0, 0.52, 0.89, 1.11, 1.25, 1.35, 1.40, 1.45, 1.49 for $n = 1, 2, \dots, 10$. For higher n the r -values were discussed by, for instance, Saaty (2000) or Alonso & Lamata (2006).

Note that a common assumption here is that the value of consistency ratio μ should not be greater than 0.1 to satisfy the decision makers (Saaty & Vargas 2012). Such level of satisfaction from the process can be difficult to reach, in particular, in the case when intangible features (subjective ideas, psychological concepts, etc.) are evaluated. However, one can use various techniques of optimization to rebuild the reciprocal matrix. Here, we observe one of the main properties of AHP, namely a natural feedback mechanism expressing the level of consistency of the expert decision which is one of the pillars of its concepts, see Saaty & Vargas (2012).

The importance (the degree of preference) of particular alternatives is obtained as the values of the eigenvector associated with the eigenvalue λ_{\max} . The way of presentation of its values is typically two-fold. They are either normalized to the sum of the entries or to their maximal value.

An interesting question appears when there are two or more experts engaged in the decision-making process using AHP. The final result of their independent (i.e., conducted individually) pairwise comparison processes can be obtained in two ways. The first is an arithmetic mean of the eigenvectors obtained on a basis of their work. The average may be weighted using the weights related to the inconsistency index (or consistency ratio). The second manner is a geometric mean of the reciprocal matrices' entries which preserves the reciprocity property of the resulting matrix, see Aczél & Roberts 1989, Aczél & Saaty 1983, Forman & Peniwati 1998, etc.

Finally, it is worth noting that AHP can be generalized by Analytic Network Process, where the criteria and alternatives depend on each other, see Saaty & Vargas (2012), Saaty (2005).

1.2. Particle Swarm Optimization

Particle Swarm Optimization (Kennedy et al. 2001) is a socially-inspired method offering very fast and effective way of solving the optimization problems according to the assumed criteria. In particular, it is inspired by observation of behavior of flocks of birds or fish schools. The solutions of the considered optimization problems are represented by so-called particles. The set of particles creates the swarm. The particles during the execution of the method move through the search space and communicate their positions to the particles from the neighborhood. In each of the iterations, the positions are updated according to the parameters such as velocity and the differences between their best positions and actual positions, and the global best position and the actual position. To be more precise, assume that the i th particle's velocity is denoted as \mathbf{v}_i , its position is given by \mathbf{x}_i , personal best and global best positions by \mathbf{p}_i and \mathbf{p}_g , respectively, see Kacprzyk & Pedrycz (2015). These two parameters are the values giving the best optimization result in the previous iterations of the method. The velocities \mathbf{v}_i and positions \mathbf{x}_i are updated as follows.

$$\mathbf{v}_i = \mathbf{v}_i + a_1 \mathbf{R}_1 \otimes (\mathbf{p}_i - \mathbf{x}_i) + a_2 \mathbf{R}_2 \otimes (\mathbf{p}_g - \mathbf{x}_i) \quad (1.1)$$

$$\mathbf{x}_i = \mathbf{x}_i + \mathbf{v}_i \quad (1.2)$$

Here, \mathbf{R}_j ($j = 1, 2$) are random values from the range $[0, 1]$, a_j ($j = 1, 2$) are so-called acceleration coefficients, and \otimes denotes point-wise multiplication of vectors. The force that pulls the particles towards their own best positions is represented by a so-called cognitive term $a_1 \mathbf{R}_1 \otimes (\mathbf{p}_i - \mathbf{x}_i)$ while the social part is represented by $a_2 \mathbf{R}_2 \otimes (\mathbf{p}_g - \mathbf{x}_i)$.

Note that the most common number of particles is 40. However, the number and criteria of stopping the algorithm vary according to the field of application. It may be the limiting number of iterations or reaching a solution of the optimization problem at a satisfying level. One of the most often modifications of the PSO algorithm is its inertia weight version with a parameter w standing by the first term of the formula (1):

$$\mathbf{v}_i = w \mathbf{v}_i + a_1 \mathbf{R}_1 \otimes (\mathbf{p}_i - \mathbf{x}_i) + a_2 \mathbf{R}_2 \otimes (\mathbf{p}_g - \mathbf{x}_i) \quad (1.3)$$

The parameter w balances the local and global search (Shi & Eberhart 1998). It can significantly reduce the time of the execution reaching satisfying level of convergence after few generations of the algorithm, see Eberhart & Shi (2001).

1.3. Fuzzy Measure and Choquet Integral

The fuzzy measure introduced by Sugeno (1974) generalizes the concept of classical measure. It replaces the typical additivity condition with the condition of monotonicity. Let X be a set and $P(X) = 2^X$ be a family of all subsets of the

set X . Of course, such a family is σ -field. Fuzzy measure is a set function $g: 2^X \rightarrow [0, 1]$ fulfilling the following conditions:

$$g(\emptyset) = 0 \quad (1.4)$$

$$g(X) = 1 \quad (1.5)$$

$$g(A) \leq g(B) \text{ for } A \subset B, A, B \in P(X) \quad (1.6)$$

The last condition can be substituted by the following limit condition

$$\lim_{n \rightarrow \infty} g(A_n) = g\left(\lim_{n \rightarrow \infty} A_n\right) \quad (1.7)$$

Here, the series $\{A_n\}$ ($n=1, 2, \dots$) is an increasing series of measurable sets.

Also Sugeno (1974) introduced the following parametrized fuzzy measure version (an aggregation scheme):

$$g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B), \lambda > -1 \quad (1.8)$$

Here, A and B are disjoint sets. The value of parameter λ describes the dependency between sets. If $\lambda < 0$ the measure is sub-additive. It means that the satisfaction which arises from one evidence source entails the satisfaction of the second one. It leads to the conclusion that they are in redundancy (competition) and that a combination of sets is not efficient. If $\lambda > 0$ then the synergy effect is present and the evidence sources efficiently support each other, see Grabisch 1995 or Pedrycz & Gomide 1998. The value of λ can be yielded in a unique form ($\lambda > -1, \lambda \neq 0$) from the equation (Sugeno 1974)

$$1 + \lambda = \prod_{i=1}^n (1 + \lambda g_i), g_i = g(\{x_i\}) \quad (1.9)$$

Here, similarly, x_1, \dots, x_n are non-overlapping sets, g_i are the fuzzy measure densities. If we denote $A_i = \{x_1, \dots, x_i\}, A_{i+1} = \{x_1, \dots, x_i, x_{i+1}\}$, the fuzzy measure over the area being the combination of the sets is determined recursively as

$$g(A_{i+1}) = g(A_i) + g_{i+1} + \lambda g(A_i)g_{i+1}, g(A_1) = g_1 \quad (1.10)$$

Now, assume that the values of function h , namely $h(x_i), i = 1, 2, \dots, n$, are sorted non-increasingly, and, moreover, we assume that

$$h(x_{n+1}) = 0 \quad (1.11)$$

Then one can define the Choquet integral of a function h with respect to a measure g as follows:

$$\int h \circ g = \sum_{1 \leq i \leq n} \left((h(x_i) - h(x_{i+1}))g(A_i) \right) \quad (1.12)$$

Let us illustrate the concept with the help of the following example: Assume that the importance of particular sensors is as follows: $g_1 = 0.4, g_2 = 0.35, g_3 = 0.15, g_4 = 0.32, g_5 = 0.5$ and the related values measured by the sensors are $h(x_1) = 0.45, h(x_2) = 0.2, h(x_3) = 0.5, h(x_4) = 0.15, h(x_5) = 0.8$. After sorting the $h(x_i)$ values and reordering the associated weights g_i one can obtain the value of parameter $\lambda \approx -0.91, g(A_1) = 0.5, g(A_2) = 0.59, g(A_3) = 0.8, g(A_4) = 0.92, g(A_5) = 1$. The value of Choquet integral is 0.57. Note that the fuzzy Choquet integral can be interpreted as specific median of the form (see Pedrycz & Gomide 1998)

$$\int h \circ g = \text{med}(h(x_1), \dots, h(x_n), g(A_1), \dots, g(A_n)) \quad (1.13)$$

1.4. T-norms

T-norms (triangular norms) are a generalization of a concept of probabilistic metric spaces (Menger 1942, Schweizer & Sklar 1983). These functions play a pivotal role when realizing the intersection and union operations on fuzzy sets, guaranteeing the properties of commutativity, associativity, and monotonicity. They can be formally defined as follows.

A function $t: [0,1]^2 \rightarrow [0,1]$ is said to be t-norm if it satisfies the following conditions of commutativity, associativity, monotonicity, and boundary conditions, i.e.,

$$t(x, y) = t(y, x) \quad (1.14)$$

$$t(x, t(y, z)) = t(t(x, y), z) \quad (1.15)$$

$$t(x, y) \leq t(z, w) \text{ for } x \leq z \text{ and } y \leq w \quad (1.16)$$

and

$$t(0, x) = 0 \text{ and } t(1, x) = x \quad (1.17)$$

respectively. Typical examples of t-norms are product

$$T_P(a, b) = ab \quad (1.18)$$

Łukasiewicz t-norm

$$T_L(a, b) = \max(0, a + b - 1) \quad (1.19)$$

drastic product

$$T_{DP}(a, b) = \begin{cases} b & \text{for } a = 1 \\ a & \text{for } b = 1 \\ 0 & \text{for } a, b \neq 1 \end{cases} \quad (1.20)$$

and Hamacher product

$$T_H(a, b) = \frac{ab}{a+b-ab} \text{ for } a, b \neq 0 \text{ and } 0 \text{ otherwise} \quad (1.21)$$

1.5. Aggregation Functions

The problem of aggregation (or, in general, fusion) of information appears in many areas of life. It can be, for instance, finding a champion on a basis of a series of sports competitions, decision-making on a basis of many experts' evaluations, classification when more than one classification methods (classifiers) are merged, specifically, when the results must be aggregated. Formally, from the mathematical point of view, an aggregation function is n -argument function $f: [0,1]^n \rightarrow [0,1]$ having the properties as follows (see, Beliakov et al. 2007)

$$f(0, 0, \dots, 0) = 0 \quad (1.22)$$

$$f(1, 1, \dots, 1) = 1 \quad (1.23)$$

and

$$f(\mathbf{x}) \leq f(\mathbf{y}) \text{ for } \mathbf{x} \leq \mathbf{y}, \mathbf{x}, \mathbf{y} \in [0,1]^n \quad (1.24)$$

Typical examples of aggregation functions are minimum, maximum, product, arithmetic, geometric, and harmonic means, or more advanced operators like

Choquet integral or Ordered Weighted Averaging operators (OWA, Yager & Kacprzyk 2012, Gadomer & Sosnowski 2018). The main properties of aggregation functions which can be as follows:

- averaging aggregation

$$\min \mathbf{x} \leq f(\mathbf{x}) \leq \max \mathbf{x} \quad (1.25)$$

- conjunctive aggregation

$$f(\mathbf{x}) \leq \min \mathbf{x} \quad (1.26)$$

- disjunctive aggregation

$$\max \mathbf{x} \leq f(\mathbf{x}) \quad (1.27)$$

- idempotency:

$$f(x, x, \dots, x) = x \quad (1.28)$$

- symmetry (the values are not dependent on the permutation of elements)

- neutral element

$$f(e, \dots, e, x, e, \dots, e) = x \quad (1.29)$$

and many others, see (Beliakov et al. 2007).

1.6. Face Image Datasets

In the book we use the following facial image databases.

- AT&T, formerly called ORL (AT&T Laboratories Cambridge). It is one of the best known and exploited databases of facial images consisting of 400 images of 40 people (10 images per individual). The subjects present different pose, emotion, or mimic and the light can differ. The size of the images is 112×92 pixels.
- The Facial Recognition Technology (FERET). It contains images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office (Phillips et al. 1998; 2000). The most used here subset of images is built of 600 images (grouped in sets called *ba*, *bk*, and *bj*) of 200 people with different illumination and expression conditions and size 256×384 pixels. Moreover, we discuss the set called ColorFERET dataset. We select its 2722 photographs of 994 individuals. Each of them has from 2 to 22 images taken with different distances to the camera. Another subset of FERET considered in this book is its grayscale part with subsets called *fa*, *fb*, *ba*, *bk*, and *bj*. The total number of images is 3880.
- Face and Gesture Aging Database (FG-NET, Lanitis 2008). The set is consisted of Caucasian individuals' images (1002 facial images of 82 people). Their age varies in-between 0 and 69. Moreover, the pose, expression, and illumination vary. The image number per subject is in-between 6 and 18. Maximum time gap is 54 years. Moreover, the dataset includes 640 children images (aged < 18 years).

- Yale Face Database (Belhumeur et al. 1997). It consists of 165 images of 15 individuals with various expression and configuration (e.g., presence of glasses, light, etc.).
- CAS-PEAL (Gao et al. 2008). It is a Chinese dataset with 9029 photographs of 1040 people. There are from 3 to 49 images per individual. The images are taken with various pose, expression, lighting and accessories (glasses or caps).
- Essex Collection of Facial Images. Here, we work with the subset called faces94 contained in built of 20 images of 153 people. Their size is 180×200 px.
- PUT (Poznań Univerisity of Technology) Face Database (Kasiński et al. 2008). An example of a face image is presented in Fig. 4.16. Its subset of 1100 frontal images (11 per individual) has been used.
- Labelled Faces in the Wild (LFW, Huang et al. 2007) images, in particular, its version by Sanderson & Lovell (2009) containing 12233 cropped images of 5748 people.
- MUCT (Milborrow et al. 2010) containing 3755 faces with 76 manual landmarks.

1.7. Conclusions

In this chapter, we have briefly discussed the most important and used tools and mathematical models from the fields of Computational Intelligence and decision-making theory. In particular, we have recalled the definitions or descriptions of Analytic Hierarchy Process, Particle Swarm Optimization, fuzzy measure and fuzzy Choquet integral, aggregation functions, and t-norms. Moreover, we have presented various datasets appearing in the sets of experiments reported in this book.

References

- Aczél, J, Roberts, F. S.: On the possible merging functions. *Math. Soc. Sci.* **17**, 205–243 (1989)
- Aczél, J., Saaty, T. L.: Procedures for synthesizing ratio judgements. *J. Math. Psychol.* **27**, 93–102 (1983)
- Alonso, J. A., Lamata, M. T.: Consistency in the analytic hierarchy process: A new approach. *Int. J. Uncertain. Fuzz.* **14**, 445–459 (2006)
- AT&T Laboratories Cambridge. [online] The Database of Faces, <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html> (Accessed 11 July 2018)

- 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)
- Beliakov, G., Pradera, A., Calvo, T.: *Aggregation functions: A guide for practitioners*. Springer-Verlag, Berlin Heidelberg (2007)
- Eberhart, R. C., Shi, Y.: Tracking and optimizing dynamic systems with particle swarms. In: *Proceedings of the 2001 IEEE Congress on Evolutionary Computation*, Vol. 1, pp.94–100 (2001)
- Essex Collection of Facial Images. [online] <https://cswww.essex.ac.uk/mv/allfaces/index.html> (Accessed 26 May 2016)
- Forman, E., Peniwati, K.: Aggregating individual judgments and priorities with the analytic hierarchy process. *Eur. J. Oper. Res.* **108**, 165–169 (1998)
- Gadomer, Ł., Sosnowski, Z. A.: Knowledge aggregation in decision-making process with C-fuzzy random forest using OWA operators. *Soft Comput.* doi: 10.1007/s00500-018-3036-x (2018)
- Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., Zhao, D.: The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Trans. System Man Cybern. A* **38**, 149–161 (2008)
- Grabisch, M.: Fuzzy integral in multicriteria decision-making. *Fuzzy Set. Syst.* **69**, 279–298 (1995)
- Harker, P. T., Vargas, L. G.: The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Manag. Sci.* **33** (11), 1383–1403 (1987)
- 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)
- Ishizaka, A., Labib, A.: Review of the main developments in the analytic hierarchy process. *Expert Syst. Appl.* **38** (11), 14336–14345 (2011)
- Kacprzyk, J., Pedrycz, W.: *Springer handbook of Computational Intelligence*. Springer-Verlag, Berlin Heidelberg (2015)
- Kasiński, A., Florek, A., Schmidt, A.: The PUT face database. *Image Process. Commun.* **13**, 59–64 (2008)
- Kennedy, J. F., Eberhart, R. C., Shi, Y.: *Swarm intelligence*. Academic Press, San Diego (2001)
- Kubler, S., Robert, J, Derigent, W., Voisin, A., Le Traon, Y.: A state-of-the-art survey & testbed of fuzzy AHP (FAHP) applications. *Expert Syst. Appl.* **65**, 398–422 (2016)
- Lanitis, A.: Comparative evaluation of automatic age-progression methodologies. *EURASIP J. Adv. Signal Process.* 2008, doi: 10.1155/2008/239480 (2008)

- Menger, K.: Statistical metric spaces. *Proc. Nat. Acad. Sci. (USA)* **28**, 535–537 (1942)
- Milborrow, S., Morkel, J., Nicolls, F.: The MUCT Landmarked Face Database. *Pattern Recognit. Assoc. South Afr.* (2010)
- Pedrycz, W., Gomide, F.: An introduction to fuzzy sets: Analysis and design. The MIT Press, Cambridge (1998)
- Pedrycz, W.: Granular Computing. Analysis and design of intelligent systems. CRC Press, Boca Raton (2013)
- Phillips, P. J., Moon, H., Rizvi, S. A., Rauss, P. J.: The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1090–1104 (2000)
- 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)
- Saaty, T. L.: A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **15** (3), 234–281 (1977)
- Saaty, T. L.: Fundamentals of decision-making and priority theory with the analytic hierarchy process. Vol. 6., Analytic hierarchy process series. RWS Publications, Pittsburgh (2000)
- Saaty, T. L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
- Saaty, T. L.: Theory and applications of the analytic network process: Decision-making with benefits, opportunities, costs, and risks. RWS Publications, Pittsburgh (2005)
- Saaty, T. L.: What is the analytic hierarchy process? In: Mitra, G. (Ed.): Mathematical models for decision support. NATO ASI Series, F48. Springer, Berlin, pp. 109–121 (1988)
- Saaty, T. L., Mariano, R. S.: Rationing energy to industries: Priorities and input-output dependence. *Energy Syst. Policy* **3** (1), 85–111 (1979)
- Saaty, T. L., Vargas, L. G.: Models, methods, concepts & applications of the Analytic Hierarchy Process. Springer, New York (2012)
- Sanderson, C., Lovell, B. C.: Multi-region probabilistic histograms for robust and scalable identity inference. In: International Conference on Biometrics 2009, LNCS 5558, pp. 199–208 (2009)
- Schweizer, B., Sklar, A.: Probabilistic metric spaces. North Holland, Amsterdam (1983)
- Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence*, Anchorage, pp. 69–73 (1998)

Sugeno, M.: Theory of fuzzy integral and its applications. Dissertation, Tokyo Institute of Technology, Tokyo (1974)

van Laarhoven P. J. M., Pedrycz, W.: A fuzzy extension of Saaty's priority theory. *Fuzzy Sets Syst.* **11** (1–3), 199–227 (1983)

Yager, R. R., Kacprzyk, J.: The ordered weighted averaging operators: Theory and applications. Springer Science+Business Media, New York (2012)

2. Local Descriptors in Face Recognition

In this chapter, we present an in-depth view into local descriptors used in the area of face recognition. First, we propose a comparison of the quality of the main local descriptors related to Local Binary Pattern and its variations in an application to the problem of aging in face recognition. Next, we propose a so-called Chain-Code Based Local Descriptor which describes both local and global facial features using the mechanism of chain codes.

2.1. Introduction

Typically, the studies on face recognition consider the problems of expression, pose, or illumination. The problem of aging is, in general, not commonly discussed. However, the number of works discussing this topic has been still increasing. Among the reasons are wide area of applications such as, for instance, missing people classification or an appearing of public datasets focusing the studies involving age factor, namely MORPH (Ricanek & Tesafaye 2006) and Face and Gesture Aging Database (FG-NET).

In general, the age-invariant face recognition techniques comprise two sets of methods. The first group embraces the methods focused on estimation of the age and simulation of the process of aging, see Lanitis et al. (2002), Park et al. (2010), Ramanathan & Chellappa (2010), Wang et al. 2006, Luu et al. (2011). These techniques result in new pictures using the models of aging compensating age effects. Such artificial facial images are matched. However, the process of aging is still difficult to simulate. One of the reasons is that different people age in a different manner since the process can be impacted by the factors like lifestyle, health, climate, place of living, etc.

The second set of methods utilizes the features which are robust or relatively robust to the progression of age. An example is GOP (gradient orientation pyramid) combined with SVM (support vector machine), see Ling et al. 2010. The authors tested the algorithm using two private databases of passports. The errors were reported for various age gaps between the training set and testing set. Eigenfaces (Turk & Pentland 1991), Elastic Bunch Graph Matching (Wiskott et al. 1997) alone and in combination (also with soft biometric features) were discussed by Guo et al. (2010). The results were obtained for images of people presenting significant spans of age. The authors have shown that the accuracy does not decrease linearly with respect to the age differences. Moreover, when the age gaps are bigger than 15 years the recognition rate is drastically lower than for the differences lower than 15 years. Meng et al. (2010) compared PCA, LBP by Ahonen et al. (2004), GOP, and Gabor wavelets. LBP with Chebyshev norm has exhibited well performance for the age gaps in the scope 7–9 years and 10–12 years. PCA was analyzed, among others, by Ricanek & Boone (2005) and

Sethuram et al. (2009). Multi-feature discriminant analysis was examined by Li et al. (2011). Periocular area was thoroughly examined by Juefei-Xu et al. (2011) through the algorithm called Walsh-Hadamard transform encoded LBP and unsupervised discriminant projection (Yang et al. 2007). For more in-depth analysis of the literature an interested reader can see the survey papers by Panis & Lanitis (2015), Jindal et al. (2015), Nimbarte & Bhoyar (2016), or Osman & Viriri (2018) as well as recent methods by Yang et al. (2014), Du et al. (2015), Nagpal et al. (2015), Boussaad et al. (2016), Pontes et al. (2016; 2017), Becerra-Riera et al. (2017), Belver et al. (2017), Deb et al. (2017), Best-Rowden & Jain (2018), Georgopoulos et al. (2018), Li et al. (2018), Nimbarte & Bhoyar (2018a; 2018b), or Punyani et al. (2018).

The main goal of the study presented here is to evaluate and compare the efficiency and potential of the most important local descriptors along with their use in the facial recognition problems related to aging (age differences). Local descriptors (Local Binary Pattern and its variations) have been one of the dominating trends in computational face recognition literature. It is well known that the local approach is, in general, robust to expression, pose, or illumination changes. Here, we are interested in an in-depth examination of their robustness (efficiency) to the age changes. Moreover, we examine not only the wide class of local descriptors but we also analyze Gabor filters. Therefore, we offer the insight into the effectiveness and role of the approaches based on local descriptors in age invariant facial recognition. We evaluate the dependence of the accuracy on the age differences between the training and testing set as well as on the groups of age. Moreover, an evaluation of the dependence of the recognition rate on the subjects' age appears as an interesting area of investigation. It should answer the question on the applicability of local descriptor-based methods as parts of more sophisticated face recognition systems. In addition, the methods incorporating local descriptors can play a key role in the problems related with age estimation. Finally, we bring forward an in-depth comparison of local descriptors-based methods. We test the following descriptors: LBP (Local Binary Pattern), CSLBP (center-symmetric LBP), DLTP (differential local ternary pattern), ILBP (improved Local Binary Pattern), LGPBP (local Gabor phase binary pattern), LXGP (local Gabor XOR pattern), LXP (local XOR pattern), MBLBP (Multi-Scale Block LBP), TPLBP (three patch LBP), and WLD (a simplified Weber local descriptor). Moreover, chosen descriptors are tested in combination with Gabor magnitude images. We have selected such a set of local descriptors since they come with very promising accuracies in comparison with the results obtained on a basis of other local descriptors with no age gap. Note that here, we analyze only algorithms based on local descriptors and do not compare them with other techniques. This analysis would be interesting per se. However, it goes beyond the merit of this work. We concentrate on the aging problem in the context of its influence on the local

descriptors' performance. The quantification of this relation will determine the best local approach and it will be a good input to combine with other methods.

To complete our analysis of local descriptors we use a nearest neighbor classifier which is based on finding the distance measure between the vectors representing images. Therefore, a proper choice of this function is an important step. Therefore, to find the suitable measure, we conduct a series of experiments with various versions of measures of dissimilarity/similarity. The compared distances are as follows: Bray-Curtis, Canberra, Chebyshev, chi square statistics, correlation, cosine, histogram intersection, Euclidean, log-likelihood statistics, and Manhattan.

Moreover, it is worth to stress that the local descriptors have been still intensively explored and developed as an interesting branch of methods in face recognition, see, for instance, works by Dornaika et al. (2014), Girish et al. (2014), Liao (2014), Reddy (2015), Ren et al. (2015; 2016), Smiatacz & Rumiński (2015), Yang et al. (2017b; 2018), Abid et al. (2018), Memiş (2018), and many others. A survey of local descriptors is presented in details by Bereta et al. (2013).

In particular, an interesting local approach was presented by Chan et al. (2015). The local descriptor is called Full Ranking (FR). The method is based on the use of the available word dictionary. The technique directly incorporates the paradigm called BoVW (Bag-of-Visual-Words, see Sivic & Zisserman 2003). Similarly to the Local Binary Pattern, the neighboring pixels are under consideration. They are ordered and their indexes create a *word*. This word is substituted by another word being the most similar to the words contained in the dictionary. The word histograms are then the inputs to the classification processes. It is worth noting that Chan et al (2015) proposed circular and square neighborhoods of the central pixel having the dimensions 3×3 and 5×5 px.

Here, we propose a descriptor called Chain Code-Based Local Descriptor (CCBLD). This novel and original local descriptor is based on an application of *words* but in a slightly different manner than by Chan et al. (2015). Here, we do not establish only one form of neighborhood, namely square or circular. On the contrary, we build the chain codes which start from the analyzed pixel. Such codes are built using, for instance, the consecutive maxima or minima of the gray scale pixel values in the cross-neighborhood. In this way, we replace the values from the range 0-255 by words which are built from four letters only. The result of the procedure is the local description of a facial image. Note that the codes containing more number of letters contain the information on the greater areas (being some kind of global structures) than the arbitrary shapes (squares or circles). Therefore, another main objective of the study detailed in this chapter is to present the Chain Code-Based Local Descriptor and its extended forms realized by an application of pixel blocks instead of singletons. Moreover, we analyze the dependency of the recognition accuracy on the length of the dictionary and the number of subregions which are the inputs to the histograms

describing the image parts. We also demonstrate the applicability, usefulness, and robustness of our CCBLD local descriptor to various conditions, age, and environments on a basis of experiments carried for the datasets such as AT&T, CAS-PEAL, ColorFERET, Essex, FERET, FG-NET, and Yale. It is worth to add that chain codes are very interesting artifacts per se. They are widely used to describe contours, see, for instance, McKee & Aggrawal 1977, Mehtre et al. 1997, Briabiesca 1999, Bartyzel 2010, Yang et al. 2017a.

2.2. Main Concept of Local Descriptors

The origins of local descriptors are in texture analysis. Relatively recently, they have appeared in the studies on face recognition. The general concept of local descriptors-based techniques can be outlined as follows. The image features are described using the neighborhood (neighboring pixels). Next, they form a description of the whole image. The vectors created in this manner can be compared in the processes of classification of people. In the most cases, the vectors contain the histograms of features bins' values for the image blocks (regions). Such histograms are built of pixels' descriptions (labels). A pixel description is calculated on a basis of its neighborhood. The specific blocks' descriptions can be concatenated giving rise to the geometry of the whole face. Note that such kind of approach does not need a training stage and seems to be relatively robust to the pose, luminance, or expression changes. However, the shortcomings lay, among others, in the length (size) of the compared vectors. The general concept of the descriptors of local features is depicted in Fig. 2.1. Note that the image fiducial points no. $j, j = 1, \dots, n$, can be described by the vectors \mathbf{v}_j . These vectors are then concatenated into the one vector describing the whole image.

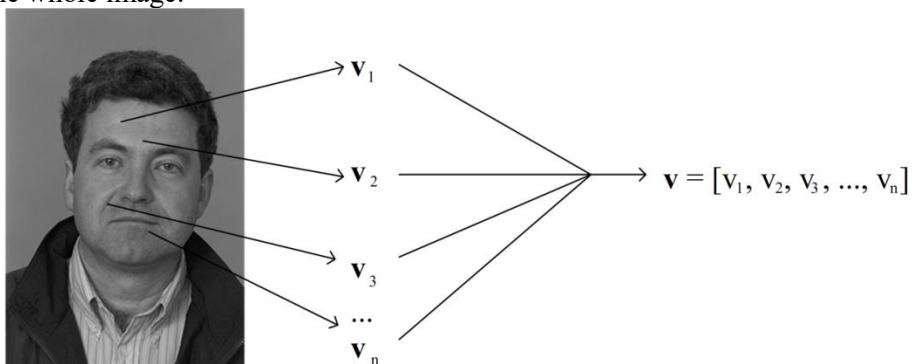


Fig. 2.1 The concept of local descriptors.

Now, we will discuss various examples of local descriptors. Local Binary Pattern (LBP) is one of the best known. Ahonen et al. (2004) proposed its

application to the facial recognition methods. The formula describing a given central pixel p_c reads as

$$LBP(p_c) = \sum_{j=0}^7 s(p_j - p_c)2^j \quad (2.1)$$

Here,

$$s(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \quad (2.2)$$

From now, the variables p_c and p_j ($i = 0, \dots, 7$) denote the greyscale-level values of the center pixel and its neighboring pixels, respectively.

An example of the LBP descriptor is shown in Fig. 2.2. The elements of the matrix are thresholded by the function s with respect to the central pixel value. The zeros and ones obtained in this way are the input to the binary number built on a basis of a concatenation in a clockwise direction. Then the final decimal label of a pixel is found. An example face image coming from the FERET set and its LBP transformation are shown in Fig. 2.3.

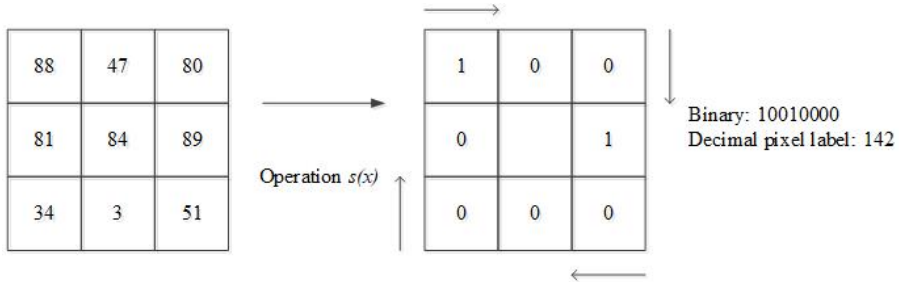


Fig. 2.2 LBP transform.

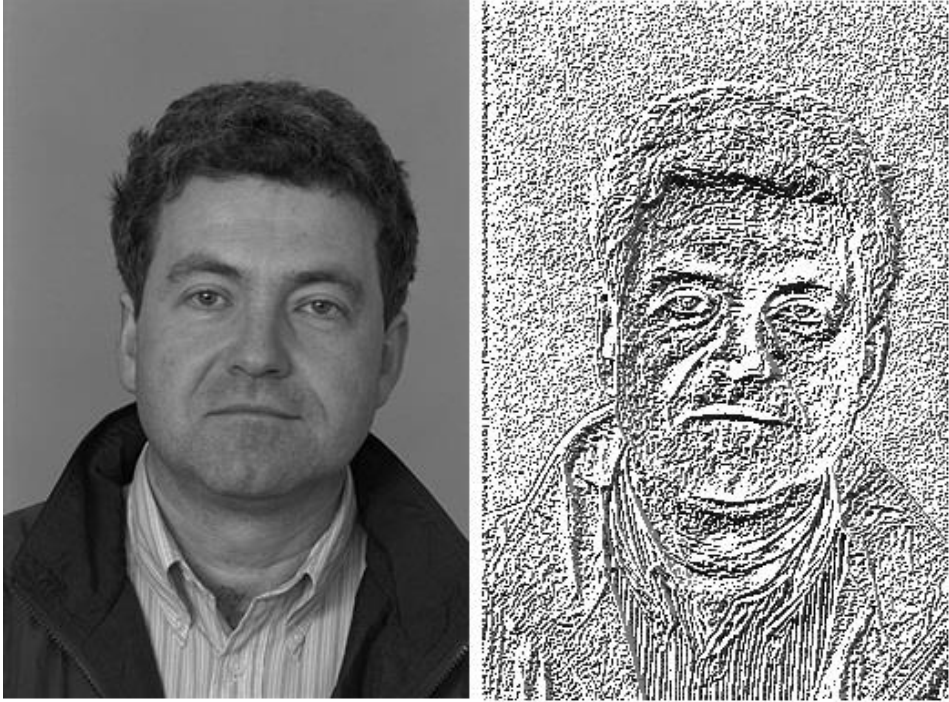


Fig. 2.3 Basic LBP transformation example: Original image (left) and the image after LBP transformation (right).

An interesting modification of LBP is its circular version (Ahonen et al. 2004), i.e., the modification based on the circular neighborhood of the pixel and bilinear interpolation at equidistant points positioned on the circle with center at p_c . The advantages of this method are its robustness to the changes in illumination, and noise in flat areas, low computational cost, and rotation invariance. Here, we discuss its version proposed by Heikkilä et al. (2006; 2009) in the following form:

$$CSLBP(p_c) = \sum_{j=0}^{\frac{P}{2}-1} s\left(p_j - p_{j+\frac{P}{2}}\right) 2^j \quad (2.3)$$

As previously, $s(x)$ denotes a thresholding procedure and P is a neighbor's number.

The next considered descriptors are improved LBP, see Jin et al. (2004), and multi-scale block LBP (Liao et al. 2007). The first of them is built using comparisons of all the pixels in the considered neighborhood with their mean (including the central value). ILBP is considered as the method well utilizing the information about texture and local shapes. Hence, it is robust to the illumination changes.

MBLBP is built in the following way: The average values of pixel blocks are used instead of values of single values. Each of these blocks is a square with pixels. One can group these blocks in larger blocks or place them equidistantly on a circle which surrounds the block placed in the center. One of the advantages of the MBLBP is that it enables encoding more global image structures than LBP in its basic form. MBLBP can be computed with using so-called integral images (Viola & Jones 2004). MBLBP was also considered by Chan et al. (2007). An example of the calculations is presented in Fig. 2.4.

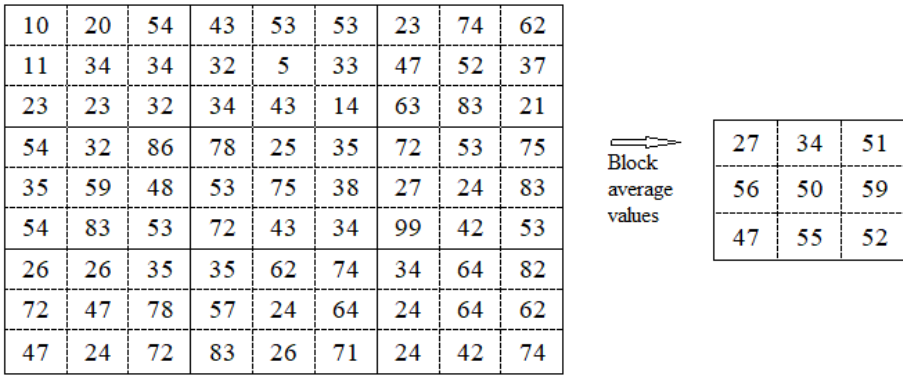


Fig. 2.4 An example of MBLBP with 9 blocks. Average values over 3x3 inner blocks are found. The block on the right is next an input to the simple LBP operation.

Of course, it is possible to build the circular version of the MBLBP on a basis of the averages of patches positioned equidistantly on a given circle.

The next considered local descriptor is LTP (local ternary pattern, Tan & Triggs 2007). The version presented here is called DLTP, differential LTP. It is based on the following transformation:

$$p'_j = \begin{cases} -1 & \text{for } p_j \leq p_c - z \\ 0 & \text{for } |p_j - p_c| < z \\ 1 & \text{for } p_j \geq p_c + z \end{cases} \quad (2.4)$$

Here, p_c, p_i, p'_i are related to the values of central, neighboring, and the new value of neighboring pixels, respectively. z is an arbitrary value. Next, if all the -1 values are replaced by 0, a so-called LTPU channel is formed. Similarly, if all the -1 values are replaced by 1 and 1 by 0, a channel called LTPL is built. Next, LBP is applied to these decimal values $LTPU(p_c)$ and $LTPL(p_c)$. The final result is

$$DLTP(p_c) = |LTPU(p_c) - LTPL(p_c)| \quad (2.5)$$

see (Bendada & Akhloufi 2010). The advantages of the descriptor are robustness to noise, expression and illumination changes.

The next compared descriptor is TPLBP, three patch local ternary pattern. It is based on the transformation as follows (Wolf et al. 2008):

$$TPLBP(p_c) = \sum_{j=0}^{P-1} s \left(d(C_j, C_{p_c}) - d(C_{j+\alpha \bmod P}, C_{p_c}) \right) 2^j \quad (2.6)$$

where P denotes a number of patches of a size $w \times w$ being equidistantly located on the circle with the center in p_c . The parameter α is a distance between the patches positioned on a circle (here, one takes into account their order). The function $d(\cdot, \cdot)$ is any distance while $s(\cdot)$ stands for a threshold operation. The symbols $C_j, C_{j+\alpha \bmod P}$ ($j = 0, \dots, P-1$) relate to the patches along the ring while the central patch is denoted as C_{p_c} .

An interesting descriptor is WLD (Weber local descriptor, Chen et al. 2010). Its name comes from the psychological law which quantifies the change perception in a given stimulus. It states that the size of *just noticeable difference* between stimuli reads as a constant ratio of the value of original stimulus. WLD illustrates this concept. The components of WLD are orientation and differential excitation. The last one is dependent on the differences of intensities between a given pixel and its neighbors and on this pixel intensity. This can be formally written by the following equality:

$$WLD(p_c) = \arctan \sum_{j=0}^{P-1} \frac{p_j - p_c}{p_c} \quad (2.7)$$

Here, p_c stands for the central pixel. The parameter j goes through its neighborhood. The advantages of WLD are, among others, noise and illumination robustness and the high ability to represent textures.

2.3. Gabor filters

Here, we discuss a role of Gabor wavelets in a combination with local descriptors. Moreover, note that the Gabor wavelets can be treated as local descriptors per se. An $I(x, y)$ image representation with a Gabor wavelet is the convolution

$$G_{m,n}(x, y) = I(x, y) * \psi_{m,n}(x, y) \quad (2.8)$$

Here (Wiskott et al. 1997, Xie et al. 2010),

$$\psi_{m,n}(x, y) = \frac{\|k_{m,n}\|^2}{\sigma^2} e^{-\frac{\|k_{m,n}\|^2 \|z\|^2}{2\sigma^2}} \left(e^{ik_{m,n}} - e^{-\frac{\sigma^2}{2}} \right), z = (x, y) \quad (2.9)$$

and $\|\cdot\|$ is a norm operation while $k_{m,n} = \begin{pmatrix} k_{j,x} \\ k_{j,y} \end{pmatrix} = \begin{pmatrix} k_n \cos \phi_m \\ k_n \sin \phi_m \end{pmatrix}$, $k_n = \frac{f_{max}}{2^2} \frac{n}{2^2}$ ($m = 0, \dots, m_{max} - 1$, $n = 0, \dots, n_{max} - 1$) are to parameterize the Gabor filter orientation and frequency, respectively. In the most of the applications $\phi_m = \frac{m\pi}{8}$, $m_{max} = 8$, and $n_{max} = 5$. σ denotes the ratio of the Gaussian window to the wavelength. The parameter f_{max} stands for the maximum frequency. Gabor magnitude is a modulus of $G_{m,n}(x, y)$ and Gabor phase is its argument.

Gabor wavelets are used to model cells in visual cortex of the brains of mammals (Lee 1996). Therefore, they have found applications in image analysis. Moreover, they provide detailed information about the regions of a face (Zhang et al. 2009). Therefore, they were successfully used in cooperation with the algorithms such as EBGm or LBP or improved by jet selection (see Perez et al. 2011, Serrano et al. 2010, Shen & Bai 2006). The shortcoming of Gabor filters usage is that they generate features with relatively long descriptions and of high dimensionality. One can rectify this problem using, for instance, Eigenfaces (Turk & Pentland 1991), Fisherfaces (Belhumeur et al. 1997), their simplified version (SGWs, Choi et al. 2008) and other techniques, see (Shylaja et al. 2011).

2.4. Local descriptors in an application to Gabor wavelet images

The methods based on combination of Gabor filters and local descriptions produce surprisingly well results. Examples are LGBPHS (local Gabor binary pattern histogram, Zhang et al. 2005) or MULGBP (multi resolution uniform local Gabor binary patterns, Jun et al. 2009). Here, we work with the detailed in the previous sections local descriptors being fused with the feature being the magnitude of Gabor wavelet. It is caused by the fact that local descriptors and Gabor filters are complementary. In addition, we discuss here three descriptors based on Gabor phase feature. They are LXP (local XOR pattern, Zhang et al. 2007), LGPDP (local Gabor phase difference pattern, Guo & Xu 2008), and LGXP (local Gabor XOR pattern, Xie et al. 2010).

The method called LXP first encodes the complex response of Gabor filter phase information by two bits. This process is realized on a basis of a quadrant bit coding. Two bits encode the quarter of the plane in which the considered complex response is placed. These two bits produce two binary maps, namely imaginary and real. The process results in 80 binary maps produced for standard collection of 40 Gabor filters. Next, for each pixel considered as a central point, XOR operation is carried to all pairs built of this pixel and its neighboring ones. The final result is the binary description in a decimal form.

The LGPDP descriptor encodes the difference of Gabor phases between the central and neighboring pixels at each orientation and scale of Gabor wavelets. The concept of the method is that the differences are reformed to one-bit numbers on a basis of the following rule: In case of the absolute value belonging to $[0, \pi/2]$ the resulting value is 1. In an opposite case the returned bit value is 0. Next, the decimal values of the numbers coming from a neighborhood are found. Finally, the spatial histograms at each orientation and scale are concatenated. This procedure results in both global and local form of information about the image.

The last method considered here is LGXP. In this case, two phases belonging to the same interval are quantized within the same range. Therefore, if the phases of the neighbor pixel and the center pixel have similar phases (i.e., belonging to

the same interval), the number 1 is assigned to the neighboring pixel. In an opposite case, the value 0 is set. At the end of the process the binary results are merged as the local pattern (Xie et al. 2010).

A general processing scheme of the classification process is depicted in Fig. 2.5. At the beginning of the process a facial image is preprocessed (gray scale conversion, cropping, scaling). Next, it can be proceeded in one of the three manners: (i) using local descriptors, (ii) using Gabor filters and then local descriptors, or (iii) using Gabor filters only. The result is then compared with the images contained in a dataset and classified.

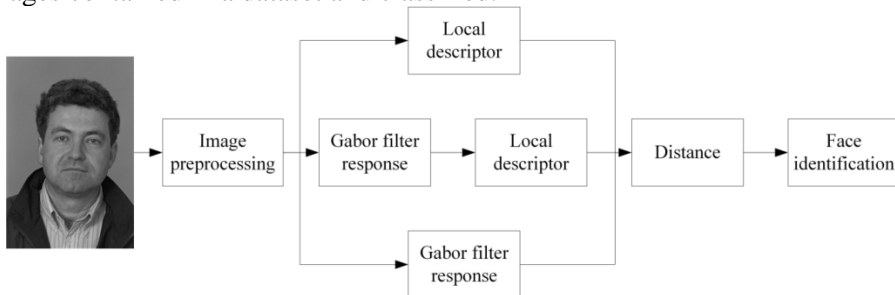


Fig. 2.5 A general scheme of processing with local descriptors.

2.5. Distance measures

In the phase of recognition (classification) one has to determine a distance between two vectors representing facial images which belong to the training and testing set, respectively. A proper choice of a distance function plays a pivotal role here (Perlibakas 2004, Xue et al. 2007, Naveena et al. 2010, Bharkad & Kokare 2011, Smiatacz 2016). In this study, we compare various distances, or generally speaking, measures of similarity/dissimilarity, in an application to nearest neighbor classifier. They are Bray-Curtis, Canberra, Chebyshev, correlation, cosine, Manhattan, and three functions of the form

$$D(A, B) = \sum_j \min(A_j, B_j) \quad (2.10)$$

$$\chi^2(A, B) = \sum_j \frac{(A_j - B_j)^2}{A_j + B_j} \quad (2.11)$$

$$D(A, B) = -\sum_j A_j \log(B_j) \quad (2.12)$$

(i.e., histogram intersection, chi square statistics, log-likelihood statistics, respectively, see Ahonen et al. 2004). Here, A and B denote histograms, j means the bin number, A_i and B_i are the bin values.

2.6. Experimental results

To assess the efficiency of local descriptors in an application to aging problem in face recognition we use the FG-NET Aging Database. In the series of

experiments, we preprocess the photographs using gray scale conversion, cropping, and scaling procedures.

We have arranged the experiments into 3 categories as follows: (i) age differences between training and testing images, (ii) recognition accuracy in various age groups, (iii) leave-one-out experiments for the whole dataset.

In the first experimental series, we extracted the training set containing 82 images of 82 people. The age of each individual is possibly nearest to 18. The testing sets are built of the rest images according to the age differences between images contained in training set and testing set, respectively. The results of this partition are six sets containing the faces with age differences in between 0 and 5 years and 6-10, 11-15, 16-20, 21-30, and 30+ years. The number of images in each of the sets is 270, 239, 221, 118, 47, and 25, respectively. The example faces (after preprocessing) are presented in Fig. 2.6. In the first row, presented is a training image. In the second row presented are two example images (the youngest and the oldest photographs) correctly classified while in the third row there are two analogical images missclassified with MBLBP local descriptor combined with the Gabor wavelet. Table 2.1 enlists the percentage values of rank-5, rank-10, and rank-15 recognition rates with respect to the above age difference groups for the local descriptors with no combination with Gabor filter images while Table 2.2 lists similar standing with recognition rates for the combinations of descriptors and Gabor wavelets. The distance function used in NN-classifier was the Euclidean one. The descriptors are represented by the best ones according to their settings across the age difference groups.

Table 2.1 Percentage recognition rates (rank-5, rank-10, and rank-15) with respect to age difference groups using local descriptors with no combination with Gabor wavelet images. Bolded are the best values across the ranks.

Age difference	Rank	CSLBP	DLTP	ILBP	LBP	MBLBP	TPLBP	WLD
0-5	5	26	21	25	24	30	29	18
	10	36	31	36	34	40	41	28
	15	44	42	45	40	51	46	40
6-10	5	21	20	22	18	22	24	15
	10	29	30	31	29	31	35	25
	15	37	39	37	38	39	45	34
11-15	5	18	13	15	15	13	15	12
	10	25	21	24	24	21	21	20
	15	32	28	33	31	33	32	27
16-20	5	13	12	16	13	17	12	12
	10	19	16	25	20	24	19	19
	15	25	21	33	30	31	30	22
21-30	5	6	15	6	9	17	13	9
	10	9	21	13	15	28	17	11
	15	19	21	23	26	36	32	21
30+	5	24	20	8	12	4	20	24
	10	28	40	20	32	12	28	24
	15	28	40	32	36	20	32	24



Fig. 2.6 Incorrectly and correctly identified faces. The image in the first row belongs to the training set while in the second and third rows are depicted the examples of correctly and incorrectly identified images (i.e., they were found in the rank 15 list), respectively.

The results enlisted in the tables suggest that the obtained accuracies can be significantly different in dependence on the local descriptors used in the process of classification. The best descriptor is the MBLBP when compared the descriptors not combined with Gabor wavelets. It gives very good results when the age differences between the images coming from the training and testing sets are relatively short (0-5 years) or high (21-30 years). Still good results, but not so well as the produced by the MBLBP, observed were for LBP and its modifications: TPBLBP, CSLB, ILBP. The first modification is very robust to short and middle differences of age.

Table 2.2 Percentage recognition rates (rank-5, rank-10, and rank-15) with respect to age difference groups using local descriptors with combination with Gabor wavelet images. Bolded are the best values across the ranks.

Age difference	Rank	Gabor magnitude	CS LBP	DLTP	ILBP	LBP	MB LBP	TP LBP	WLD	LGPDP	LGXP	LXP
0-5	37	33	36	37	36	38	36	34	26	29	30	37
	45	43	47	46	47	47	44	43	37	36	37	45
	50	54	51	53	54	54	52	49	46	44	45	50
6-10	21	28	28	31	31	29	28	26	17	20	20	21
	32	39	37	42	41	40	37	36	25	30	28	32
	44	49	47	50	47	49	45	45	34	38	36	44
11-15	17	19	21	23	22	21	20	18	15	17	14	17
	24	31	30	33	35	33	28	30	21	23	20	24
	34	38	39	39	41	43	35	36	28	31	28	34
16-20	14	12	10	15	12	12	17	11	13	14	14	14
	22	24	19	25	22	26	25	20	24	21	25	22
	26	30	31	31	31	36	37	31	35	29	33	26
21-30	32	23	19	30	28	26	23	21	17	9	6	32
	43	47	34	47	40	40	32	36	28	30	23	43
	47	55	51	53	45	45	47	40	34	38	32	47
30+	24	20	20	20	20	16	20	16	20	20	12	24
	24	28	36	32	32	28	28	24	40	32	32	24
	40	44	44	44	40	28	40	28	40	32	40	40

Moreover, the combination of local descriptors with Gabor wavelet images significantly improves the accuracy of recognition even up to 16% for age differences in between 0 and 5 years and up to 38% for 21-30 years age differences. The average increase was 13%, 23%, and 23% for rank 5, 10, and 15, respectively. Similarly, as in the previous case, the best recognition rates were obtained for the MBLBP local descriptor and other LBP variations. However, also the Gabor wavelet images treated as local descriptors on their own present very well values of accuracies at similar level as the MBLBP combined with Gabor magnitudes in the group of short age differences. However, the results decrease faster for larger differences of age. As a conclusion one can state that local descriptors, particularly in combination with Gabor filters, are quite robust to facial age changes (differences) and can be a substantial part of face recognition systems oriented on the age-invariance problems. It is worth noting that the local descriptors which are based on the Gabor phase images, i.e., LGPDP, PGXP, and LXP do not exhibit good results (high accuracies) as the local descriptors combined with Gabor magnitudes do.

Now, we repeat the experiments with slightly different settings. Namely, the image dataset is divided into the training set containing 189 images of 69 subjects who are 16-22 years old. The probe sets are the face pictures at the ages

of 23-30, 31-40, 41-50, and 51-69 years and contain 108, 66, 37, and 16 pictures, respectively. The values of rank 5, 10, and 15 recognition rates for the above-defined age groups are listed in Table 2.3 and Table 2.4, with respect to local descriptors combined and not combined with Gabor magnitudes. As in the previous case, the local descriptors are represented by their versions producing the best average recognition rates across the groups.

Table 2.3 Rank-5, rank-10, and rank-15 percentage accuracies for the age groups obtained with local descriptors not combined with Gabor magnitude filters.

Age group	Rank	CSLBP	DLTP	ILBP	LBP	MBLBP	TPLBP	WLD
23-30	5	38	34	43	40	44	42	28
	10	53	47	52	52	50	52	37
	15	59	54	59	58	58	62	46
31-40	5	17	15	17	14	23	24	11
	10	26	30	27	23	35	29	17
	15	36	44	35	38	36	39	26
41-50	5	16	16	14	14	24	16	11
	10	27	24	27	24	30	22	22
	15	32	27	30	38	32	30	24
51-69	5	0	6	0	6	6	6	13
	10	13	6	0	13	13	6	13
	15	13	13	13	19	13	6	19

Table 2.4 The results of rank-5, rank-10, and rank-15 accuracies for the age groups obtained with local descriptors combined with Gabor magnitude filters.

Age group	Rank	Gabor magn.	CSLB P	DLT P	ILB P	LB P	M B LB P	TP LB P	WL D	LGP DP	LGX P	LX P
23-30	5	51	58	42	54	59	56	46	43	32	37	33
	10	59	60	50	58	60	61	58	53	45	52	46
	15	69	63	61	62	64	64	64	60	56	62	56
31-40	5	27	30	33	29	30	30	29	26	26	20	18
	10	45	45	41	48	45	45	44	39	33	27	33
	15	55	52	55	61	61	58	52	52	44	42	44
41-50	5	38	49	49	54	49	49	43	32	41	32	27
	10	57	54	51	54	57	54	54	46	46	35	43
	15	62	57	54	54	57	59	59	51	49	41	46
51-69	5	19	0	13	6	6	13	0	6	0	0	0
	10	19	13	25	13	6	13	13	6	0	6	0
	15	25	13	31	31	19	25	13	6	19	19	13

The experiments with the images of adult subjects only confirm the main conclusions obtained with the experiments with age differences, i.e., that the best local descriptor with no combination with Gabor images is the Multi-scale Block

Local Binary Pattern. LBP and its modifications give quite worse results. Good options to choose are, in particular, ILBP, MBLBP, and LBP when we are able to combine them with Gabor magnitudes. Moreover, the Gabor wavelets are very valuable local descriptors on their own when the age groups are considered. As in the previous case, in combination with local descriptors, the recognition accuracies are essentially improved. Not satisfactory are the results produced by WLD, LGPDP, LXP, and LGXP. One can observe that the recognition rates are better than those obtained with the age differences. It is caused by removing the children images from the dataset. In the considered age groups the changes bones and muscles are not as rapid as in the childhood.

The next set of tests is consisted of the experiments with the leave-one-out approach which is based on excluding one image from the set and the rest serving as a probe set. The results recorded at this stage of experiments are illustrated in Fig. 2.7 and in Fig. 2.8 without and with Gabor filtering, respectively. As before, the MBLBP local descriptor performs better than other descriptors. Local descriptors following the Gabor filters are the most valuable choice. The LBP descriptor and its modifications such as MBLBP, ILBP, and CSLBP produce evidently good results here. Weber local descriptor and DLTP do not seem to be good choices in this case. Again, LXP, LGXP, and LGPDP do not yield satisfying results. It is worth to stress that Gabor filter with no fusion with local descriptors is relatively worse than the rest of local descriptors merged with Gabor wavelets.

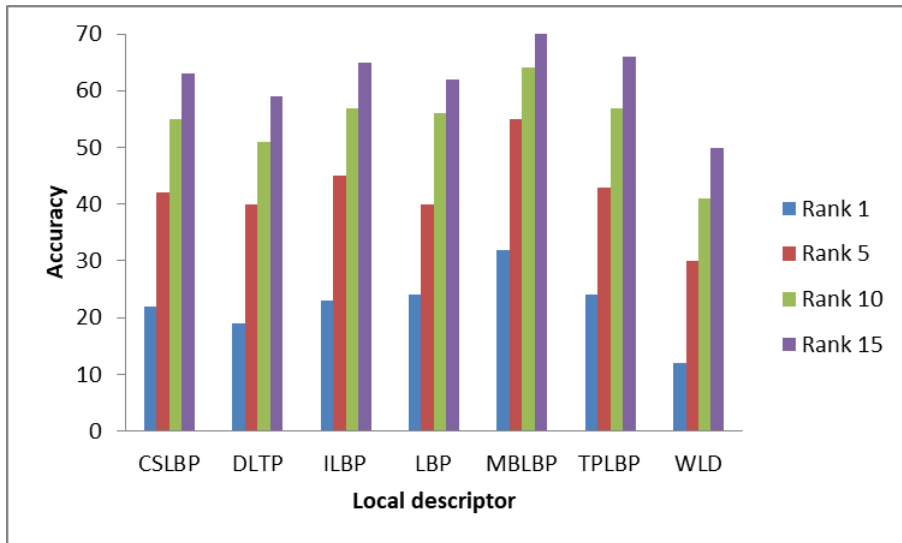


Fig. 2.7 Rank-1, rank-5, rank-10, and rank-15 percentage recognition rates in case of leave-one-out experimental settings for the FG-NET dataset (descriptors with no Gabor filters).

The results of the above discussion are summarized in Table 2.5.

The last series of experiments is devoted to the comparison of the most popular and intuitive similarity/dissimilarity measures and their impact on the recognition rates. The measures such as Bray-Curtis, Canberra, χ^2 statistics, correlation, cosine, Euclidean, histogram intersection, log-likelihood statistics, Manhattan, and Chebyshev are compared. To find the best option of the above measures we analyze the performance of ILBP, LBP, and MBLBP local descriptors fused with Gabor wavelet images. We have chosen the descriptors since they have exhibited the highest average recognition rates in the previous series of experiments. The results for four age groups (rank 15) and for leave-one-out experimental setting (rank-1, -5, -10, and -15) are enlisted in Table 2.6.

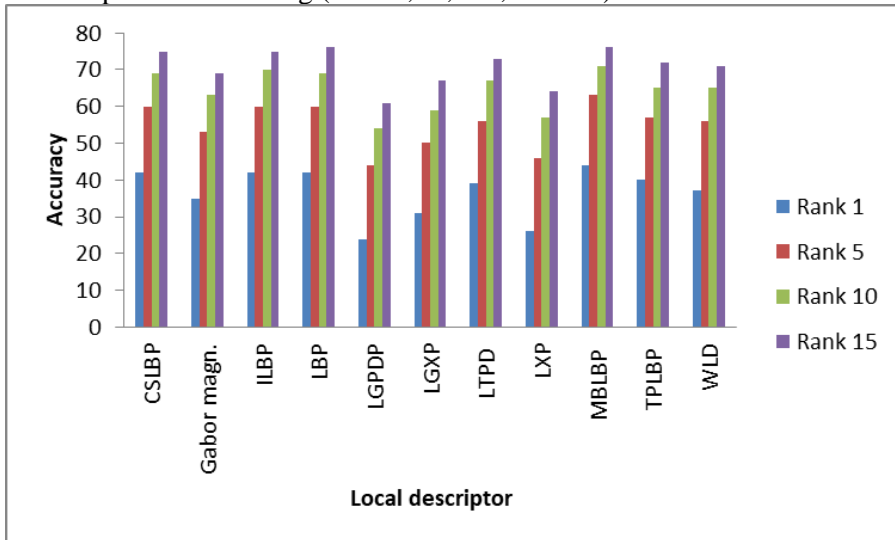


Fig. 2.8 Rank-1, rank-5, rank-10, and rank-15 percentage recognition rates in case of leave-one-out experimental settings for the FG-NET dataset (descriptors merged with Gabor filters).

The results show that the optimal choice of the distance function (i.e., the similarity or dissimilarity measure) can essentially improve the accuracy of the method. Specifically, the Bray-Curtis, χ^2 statistics, histogram intersection, and Manhattan distances exhibit relatively better results than popular Euclidean norm. This is visible especially when the differences in the age between the training and testing images are more than 7 years. In case of the differences less than 7 years, the Canberra measure can also be a proper choice. Leave-one-out experiments have shown that the differences between the results are almost negligible. However, the improvement of the recognition accuracy is possible when the distance measure is properly chosen.

The set of the experiments and their results confirm the intuitive observation that the bigger age difference between the compared faces the lower the

recognition rate is. This trend is, however, not equal in all the discussed cases. Some exceptions can appear. They are mainly caused by the changing counts of images belonging to the compared training and testing sets.

Table 2.5 Summary of the main results obtained for local descriptors in a context of aging problem.

Descriptor	Version with no Gabor filters	Version combined with Gabor filters
CSLBP	Good results for age difference 10-15 and age group 23-30	Relatively high accuracy exhibited for all age differences and age groups <50
DLTP	Good accuracies for age differences 20+	Relatively good option for the age group 50+
ILBP	Very satisfying results for middle age differences and age groups <50 and unsatisfactory results for the age group 50+	High accuracies for all the age groups and differences. Very good performance for the age difference <30
LBP	Average rates for all the age differences	Good option for all the age groups and differences
MBLBP	High recognition rates for almost all the age groups and differences	The highest average recognition rates when all the results are taken into account
TPLBP	Very good results for the age differences <15 and age group 23-40	Good accuracies for all the age differences, in particular 16-20 and age group 23-30
WLD	Relatively robust for high age differences and age groups, for the other collections of data the results are worse	Inefficient in case of all the age groups
Gabor magnitude filters	Efficient for age differences >20 and age group >40	n/a
LGPD	Quite efficient for the age group of 31-50	n/a
LGXP	Good exhibition for age differences <20 and age group 23-40	n/a
LXP	Quite effective for age differences 11-20 and for age group 41-50	n/a

2.7. Chain Code-Based Local Descriptor and Its Extension

Let us explain the main idea behind the Chain Code-Based Local Descriptor. In contrary to other local descriptors, CCBLD is not built on the basis of a given region (dimension and shape). The description of a pixel being under consideration comes from a larger number of pixels or block count. The image local properties influence this description. To be more precise, assume that we

begin at any arbitrary point of the image, i.e., a pixel $I(x,y)$, and assume that in its cross-neighborhood there exists a pixel containing higher gray-scale value, namely that one of the pixels $I_L(x-1,y)$, $I_U(x,y-1)$, $I_R(x+1,y)$, or $I_D(x,y+1)$ is of the highest value. The first letter of the chain code (pixel description) stands one of the corresponding signs, i.e., L , U , R , or D . The process is continued at the pixel of the maximal value and it stops when no pixel in some cross-neighborhood is of higher value. Note that the process must finish since there is a finite number of pixels in an image. Moreover, one can note that the cycles are not possible. In case when more than one maximum is found the special rule must be defined to choose the next step. Of course, the description of a pixel can be an empty string if no maxima is found in its neighborhood. In the same way, using the letters L , U , R , or D , we can build the chain on a basis of the consecutive local minima. The process of creating such a string is depicted in Fig. 2.9.

Table 2.6 A comparison of the results obtained with various similarity/dissimilarity measures for different age groups and leave-one-out experiments (rank-1, -5, -10, and -15).

Descriptor / method	Age group	Bray-Curt.	Canb.	Cheb.	Chi sq.	Corr.	Cos.	Eucl.	Hist.	Log-lik.	Manh.
ILBP	23-30	62	66	32	62	63	62	62	62	61	62
	31-40	58	50	35	58	56	58	61	58	56	58
	41-50	59	59	35	59	57	57	54	59	54	59
	50+	31	19	13	31	38	38	31	31	25	31
LBP	23-30	66	65	33	67	66	65	64	66	62	66
	31-40	61	53	29	59	62	64	61	61	56	61
	41-50	59	57	32	59	57	57	57	59	54	59
	50+	19	6	31	25	31	38	19	19	13	19
MBLBP	23-30	65	61	38	65	63	64	64	65	62	65
	31-40	55	44	29	53	58	58	58	55	52	55
	41-50	65	43	32	65	59	59	59	65	62	65
	50+	13	0	19	13	25	25	25	13	19	13
ILBP, leave-one-out, rank	1	42	38	8	42	42	42	42	42	42	42
	5	60	58	20	61	61	60	60	60	59	60
	10	70	68	28	70	70	70	70	70	69	70
	15	77	74	37	76	75	75	75	77	75	77
LBP, leave-one-out, rank	1	42	36	9	42	42	43	42	42	43	42
	5	60	55	21	61	59	60	60	60	60	60
	10	69	65	32	69	69	69	69	69	68	69
	15	75	73	40	76	75	75	76	75	75	75
MBLBP, leave-one-out, rank	1	43	41	7	43	45	44	44	43	43	43
	5	63	61	19	63	62	62	63	63	62	63
	10	73	71	30	73	71	71	71	73	71	73
	15	78	77	38	79	76	76	76	78	78	78

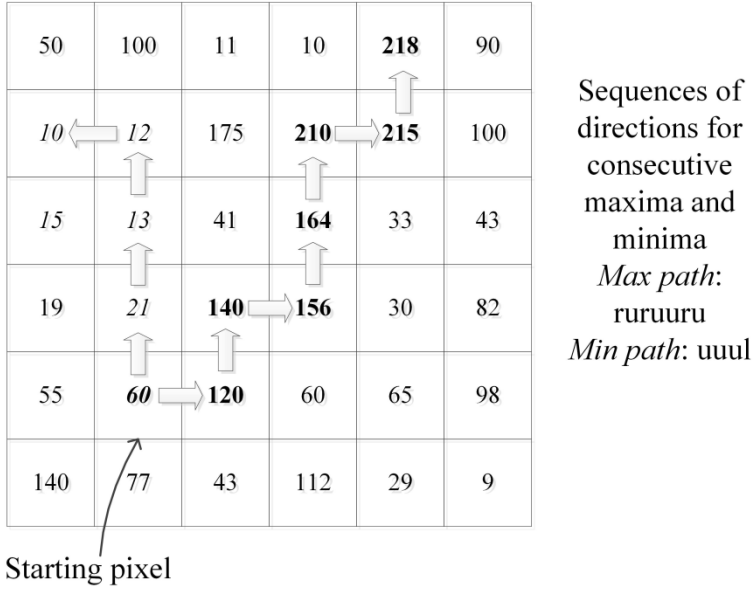


Fig. 2.9 The process of building the chain-codes describing an image.

The number of chain-codes (string descriptions of the pixels) can be large. Therefore, to reduce the dimension of the histogram one can introduce a dictionary of possible words. If a word does not appear in the dictionary it can be matched to the most similar one among the words contained in the dictionary. A comparison can be realized on a basis of Levenshtein distance, see (Levenshtein 1965, Navarro 2001).

The whole image is described in a form of concatenated histograms of words coming from image subareas. In this way both local and global image properties are preserved. In a formal form this is written as

$$B = \left[c_{1,1}^+, \dots, c_{1,\frac{n}{2}}^+, c_{1,1}^-, \dots, c_{1,\frac{n}{2}}^-, \dots, c_{i,j}^+, \dots, c_{i,j}^-, \dots, c_{r,1}^+, \dots, c_{r,\frac{n}{2}}^+, c_{r,1}^-, \dots, c_{r,\frac{n}{2}}^- \right] \quad (2.13)$$

Here $i=1, \dots, r$ and $j=1, \dots, n/2$ are a number of subarea and a number of word in a histogram associated with the subarea, respectively. The words (c) created using the consecutive minima are marked by $-$ while $+$ is used to denote the words created by using the rule of consecutive maxima.

The dictionary of words is designed in a slightly similar way as in Chan et al. (2015). We randomly pick up images and then the pixels from the dataset. The classifier is the Nearest Neighbor method based on any similarity measure to compare two histograms (i.e., the vectors of a form $\mathbf{x} = (x_1, \dots, x_N)$, $\mathbf{y} = (y_1, \dots, y_N) \in \mathbb{R}^N$).

Similarly as in case of the presented MBLBP descriptor one can modify the procedure described above using blocks of pixels instead of applying single

pixels. These blocks of pixels are represented as the means of the $(2p + 1) \times (2p + 1)$ -size blocks, namely

$$\bar{I}(x, y) = (I(x - p, y - p) + \dots + I(x, y) + \dots + I(x + p, y + p)) / (2p + 1)^2 \quad (2.14)$$

As an example of construction of the descriptor let us consider the block of pixels presented in Fig. 2.4. If we analyze the central pixel with value 75 then the block of pixels are created as in the MBLBP example. Next, the chain codes are built as in the procedure depicted in Fig. 2.9. The final result is $c^- = UL$ and $c^+ = R$. Another example is depicted in Fig. 2.10. The presented description is $c^- = UUR$ and $c^+ = R$. An overall processing scheme is shown in Fig. 2.11.

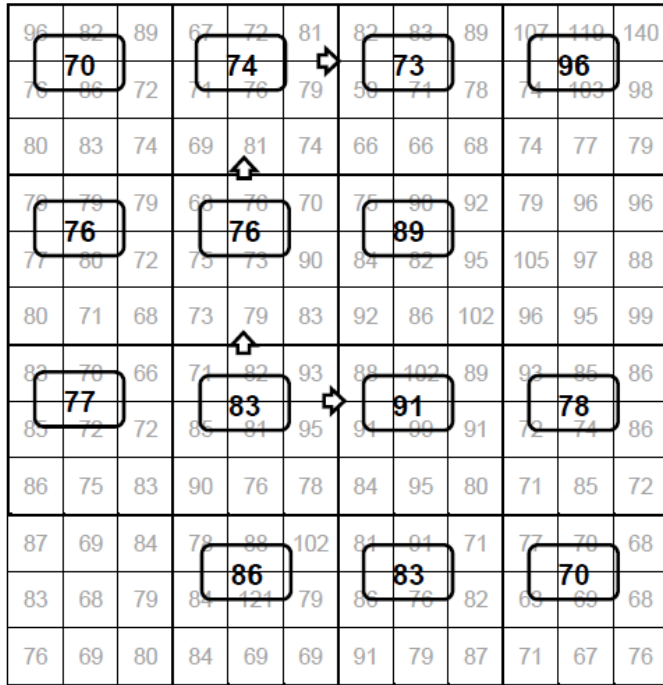


Fig. 2.10 An example of building Chain Code-Based Local Descriptor for blocks of pixels of size 3x3 px.

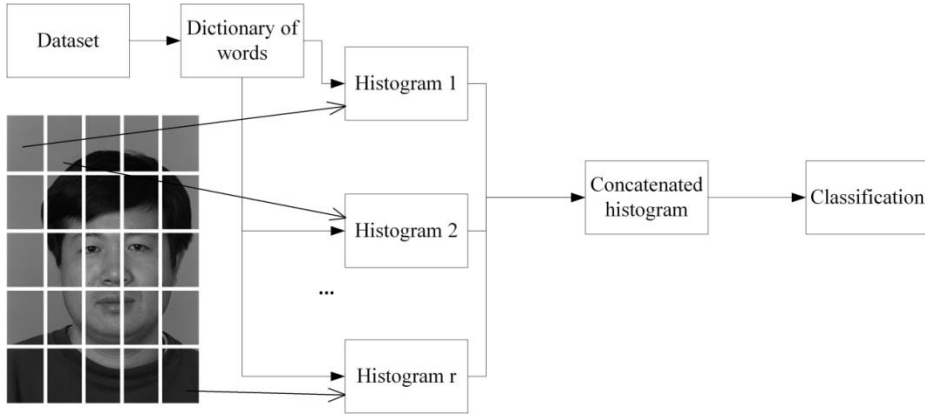


Fig. 2.11 A processing scheme.

2.8. Experimental Results

We expose the results of an application of the CCBLD descriptor in an application to face recognition problems with seven various datasets. We compare the results with LBP and its modifications as well as Full Ranking descriptor presented in Ahonen et al. 2004 and Chan et al. 2015, respectively.

2.8.1. AT&T Database

In the first series of experiments we have compared various similarity/dissimilarity measures introduced in (Ahonen et al. 2004), namely correlation, histogram intersection, χ^2 statistics, and Hellinger (called also Bhattacharyya coefficient) as the methods of comparison of the histograms of words. The length of histograms is 500 and each of the images was divided into 25 regions. The results of 100 repetitions of the tests (200 images, i.e., 5 images per person, randomly chosen to the training set and the rest selected to the testing set) are listed in Table 2.7. They show that the most promising results can be obtained on a basis of Hellinger similarity measure given by the following formula (see OpenCV documentation):

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{1 - \frac{\sum_{i=1}^n \sqrt{x_i y_i}}{\sqrt{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}}} \quad (2.15)$$

where $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$. Therefore, for all the next series of experiments, the results obtained with Hellinger measure will be presented.

The next series of experiments are set as follows: The dataset was divided randomly by choosing 5 images of each subject to the gallery set and 5 photographs to the testing set. We repeated the experiments 20 times. It is worth noting that each time a new dictionary of words was randomly built. Moreover, the experiments were carried with various dictionary lengths and divisions of the

images into the subareas. The division was always in the form $d \times d$, i.e., there is the same number of blocks horizontally and vertically and the subareas have the same proportions as the whole images.

Table 2.7 The initial results of a comparison of similarity/dissimilarity measures.

	Leave-one-out		200 training and 200 testing images	
	Recognition rate	Std. dev.	Recognition rate	Std. dev.
Correlation	97.5	0.46	92	2.07
Histogram intersection	98.43	0.37	93.46	1.94
Chi-square statistics	95.67	0.69	89.03	2.47
Hellinger	98.48	0.33	93.67	2.03

The rank 1 recognition rates for the dimension of block surrounding the starting pixel 1×1 , 3×3 , 5×5 , and 7×7 , respectively, are listed in Table 2.8. The tests were carried with the dictionary lengths in the range 10, 20, ..., 300. The number of subareas were in the range 1 to 100. Additional selected (the best with respect to categories) results for longer dictionaries are shown in Table 2.9. They were specified during an analysis of the progress of the accuracies obtained for CCBLD with lower length dictionaries. The choice of a dictionary containing 100 words for the basic CCBLD and a dictionary with the number of words in the range 400 – 600 for the block version of CCBLD is a sound alternative. The optimal image subregions count for the AT&T dataset is 9 or 16, i.e., there are 9 or 16 subregions of the equal size.

The rank 1 – rank 10 for five versions of CCBLD local descriptor in comparison with LBP and Full Ranking descriptors visualized are in Fig. 2.12. Hereafter the writing $CCBLD(Kpx, L \times M, N)$ means the local descriptor with K -pixel wide neighborhood starting image point, partition of the image on L rows and M columns, and N -element count dictionary. The results depicted at the image show the high efficiency and potential applicability of the proposed CCBLD local descriptor, particularly when referred to other published local descriptors. The notation $FR(Z, L \times M, N)$ means Full Ranking with circular neighborhood built of 17, 25 and square neighborhood built of 17 and 9 pixel values when $Z = A, B, C$, and D , respectively.

The next series of experiments with the AT&T database is conducted after the preprocessing stage, i.e., the faces were cropped to the size 90×94 px, scaled, and the histograms were equalized. CCBLD produces over 98% recognition rate and for all the ranks successfully outperforms the other compared descriptors, see Fig. 2.13.

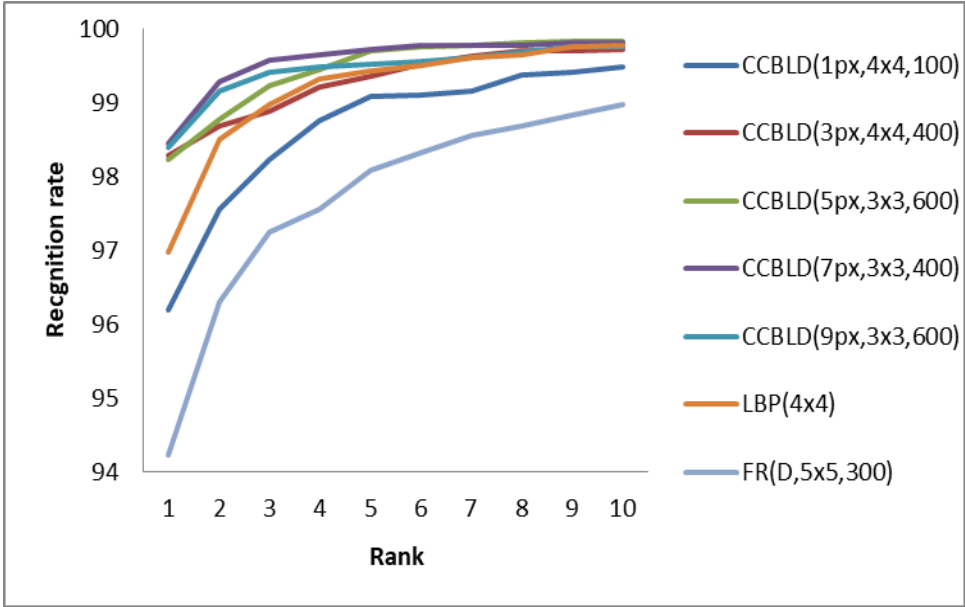


Fig. 2.12 Rank 1, rank 2, ..., rank 10 recognition rates produced by chosen local descriptors.

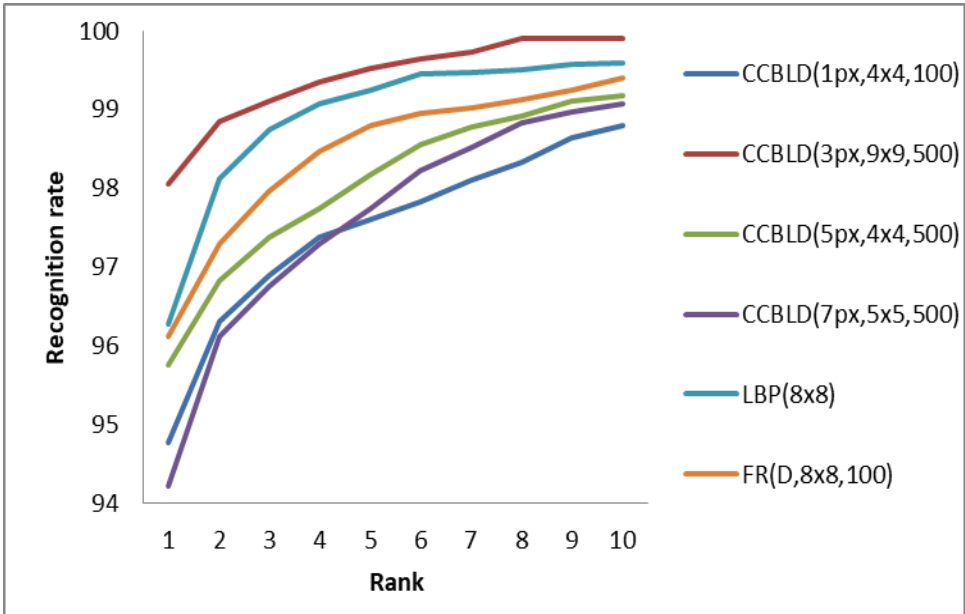


Fig. 2.13 Analogous results obtained for the AT&T images after the preprocessing.

2.8.2. FERET

Here, we use the merged *ba*, *bk*, and *bj* subsets of the set. Similarly as in the case of the previous dataset, we have run twenty iterations of experiments based on the random choice of two images of each individual to the training set and one image to the testing set, respectively. CCBLD descriptor in its simplest form (1 px size neighborhood) produces better results than the other methods. However, the recognition rates are relatively low, i.e., 73.75%. From the other hand, they show the potential hidden in our approach as well as its ability to increase the efficiency on a basis of incorporation of other methods related with image recognition and preprocessing. Fig. 2.14 visualizes the results. The input to the next series of experiments was the same set of images but after the preprocessing based on eye detection, rescaling, face cropping to the size 100×140 px, and histogram equalization. The results depicted at Fig. 2.15 show again the usefulness of our proposal. Moreover, it is seen that the preprocessing stage slightly improves the accuracy by returning over 75% rate of recognition.

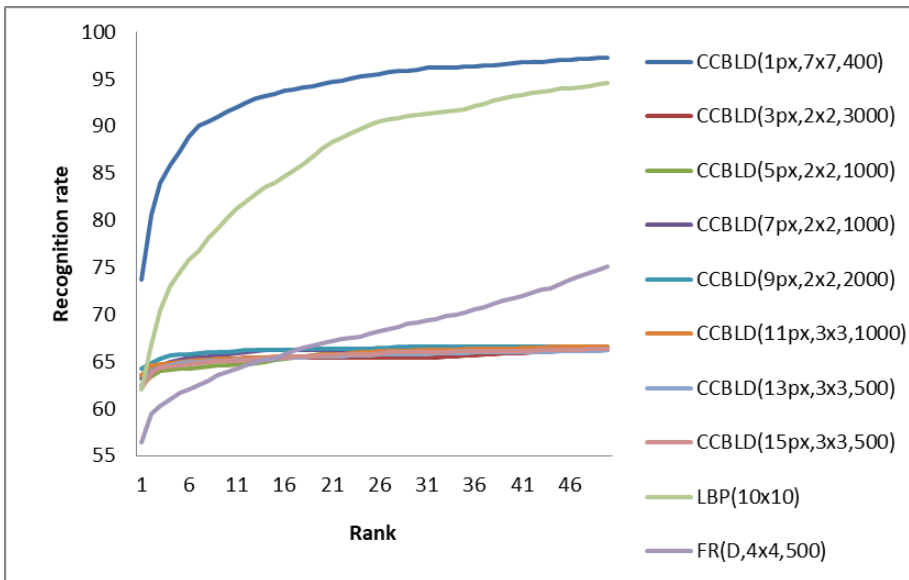


Fig. 2.14 The cumulated recognition rates for rank 1 to rank 50 produced with 600 FERET grayscale images.

Now, let us discuss the set of experiments carried for the ColorFERET dataset. The only preprocessing was conversion to the grayscale. Fig. 2.16 illustrates the results. Similar tests were run for cropped and rescaled to the size 100×140 px images included in the grayscale part of the FERET. The subsets considered are called *fa*, *fb*, *ba*, *bk*, and *bj*. A comparison of CCBLD and other

descriptors is in Fig. 2.17. All the produced results show an evident applicability of the proposed approach.

Table 2.8 Relation between dictionary length, number of subareas, size of blocks, and rank 1 recognition rates.

Number of words in the dictionary in the case of 1x1 pixel blocks														
Sub-region count	10	20	30	40	50	60	70	80	90	100	150	200	250	300
1	52.95	69.33	74.78	77.58	79.68	81.45	82.28	82.80	82.93	83.75	85.23	84.60	85.50	85.53
4	83.00	88.80	89.70	90.63	91.00	91.50	91.45	92.00	92.63	92.28	92.93	92.75	92.80	93.18
9	90.23	92.58	93.38	94.15	94.28	94.65	94.85	94.75	95.43	95.23	95.68	95.75	95.93	95.83
16	92.95	94.70	95.48	95.68	95.83	95.83	95.98	96.03	96.10	96.20	95.93	96.15	96.00	95.60
25	93.68	95.10	95.33	95.30	95.53	95.60	95.55	95.55	95.53	95.58	95.45	95.15	95.10	94.85
36	93.03	94.15	94.80	95.30	95.18	95.03	95.25	95.05	95.35	95.33	95.18	94.73	94.53	93.98
49	93.55	94.30	94.20	94.53	94.28	94.48	94.18	94.28	94.15	94.28	93.93	93.75	93.53	93.10
64	92.58	93.35	93.50	93.60	93.73	93.43	93.73	93.53	93.55	93.45	93.28	92.78	92.80	92.53
81	92.45	93.18	93.38	93.40	93.20	93.40	93.28	93.20	93.13	93.25	92.85	92.63	92.43	92.25
100	92.38	92.70	92.78	92.80	92.80	92.50	92.63	92.68	92.50	92.63	92.08	91.83	91.68	91.03
Number of words in the dictionary in the case of 3x3 pixel blocks														
Subr. count	10	20	30	40	50	60	70	80	90	100	150	200	250	300
1	72.13	85.28	88.93	91.08	92.03	92.88	93.38	93.13	94.08	94.13	94.93	95.30	95.35	95.75
4	92.20	94.80	96.43	96.73	96.60	96.95	97.25	97.10	97.13	97.43	97.60	97.93	98.00	97.93
9	96.05	97.28	97.60	97.88	97.83	97.93	98.00	97.85	97.85	97.95	97.95	98.20	98.15	98.20
16	96.70	97.63	97.75	98.03	98.10	98.23	98.18	98.20	98.08	98.18	98.28	98.18	98.20	98.25
25	96.58	97.10	97.20	97.05	97.35	97.28	97.08	97.28	97.43	97.23	97.20	97.00	97.23	97.20
36	96.28	96.83	96.90	96.83	96.70	96.88	96.63	96.83	96.88	96.93	96.90	96.93	96.68	96.83
49	96.50	96.58	96.43	96.58	96.50	96.58	96.48	96.60	96.53	96.43	96.40	96.55	96.40	96.45
64	95.25	96.03	96.00	95.90	96.00	96.00	95.88	96.03	95.90	95.98	95.98	96.00	95.93	95.93
81	95.38	95.50	95.80	95.70	95.53	95.65	95.70	95.55	95.60	95.55	95.68	95.73	95.70	95.55
100	95.45	95.65	95.35	95.53	95.50	95.60	95.38	95.35	95.45	95.43	95.35	95.35	95.48	95.30
Number of words in the dictionary in the case of 5x5 pixel blocks														
Subr. count	10	20	30	40	50	60	70	80	90	100	150	200	250	300
1	72.90	85.98	89.28	91.18	91.73	92.65	92.88	93.25	93.43	93.78	94.78	95.03	95.75	95.65
4	91.98	95.13	96.10	97.08	96.93	97.28	97.15	97.50	97.45	97.60	97.60	97.95	98.00	97.90
9	95.78	97.38	97.30	97.43	97.68	97.88	97.83	97.75	97.85	98.10	98.00	98.00	98.00	98.00
16	96.90	97.35	97.48	97.30	97.50	97.53	97.43	97.55	97.45	97.50	97.50	97.53	97.50	97.45
25	95.90	96.25	96.33	96.45	96.48	96.45	96.60	96.70	96.53	96.58	96.73	96.73	96.73	96.73
36	95.85	96.18	96.00	96.18	96.23	96.35	96.28	96.28	96.28	96.33	96.18	96.28	96.33	96.35
49	95.33	96.03	95.95	96.13	95.93	95.95	96.00	96.00	95.93	95.95	96.10	95.93	95.90	95.93
64	95.23	95.63	95.65	95.55	95.63	95.65	95.58	95.58	95.48	95.55	95.50	95.35	95.38	95.63
81	95.13	95.33	95.50	95.53	95.30	95.48	95.35	95.30	95.25	95.28	95.15	95.18	95.28	95.08
100	94.73	95.18	95.05	95.05	94.98	95.20	94.93	94.95	94.68	95.00	94.88	94.78	94.63	94.70
Number of words in the dictionary in the case of 7x7 pixel blocks														
Subr. count	10	20	30	40	50	60	70	80	90	100	150	200	250	300
1	70.13	84.13	87.08	90.10	91.50	91.85	92.78	92.55	92.78	93.28	94.03	94.68	95.18	95.28
4	92.18	94.65	95.88	96.45	96.58	96.78	96.85	97.15	97.20	97.05	97.48	97.65	97.53	97.70
9	96.33	97.35	97.73	97.93	98.00	98.15	98.13	98.10	98.33	98.33	98.33	98.33	98.33	98.28
16	96.65	97.70	97.80	97.73	97.83	97.85	97.68	98.05	97.88	98.05	97.88	97.90	97.88	97.95
25	96.18	96.85	96.80	97.03	96.95	97.23	97.03	97.23	97.18	97.20	97.18	97.15	97.05	96.95
36	96.60	96.68	96.75	96.88	96.60	96.65	96.65	96.70	96.68	96.58	96.58	96.33	96.38	96.50
49	95.80	96.13	96.10	96.38	96.30	96.35	96.25	96.55	96.28	96.35	96.33	96.28	96.43	96.40
64	95.65	95.88	95.85	96.13	96.13	96.15	96.28	96.30	95.98	96.05	96.18	96.18	96.23	96.33
81	95.60	95.83	95.70	95.70	95.90	95.78	95.85	95.48	95.65	95.65	95.73	95.60	95.70	95.63
100	95.08	95.55	95.75	95.63	95.58	95.50	95.58	95.40	95.45	95.58	95.35	95.55	95.40	95.40

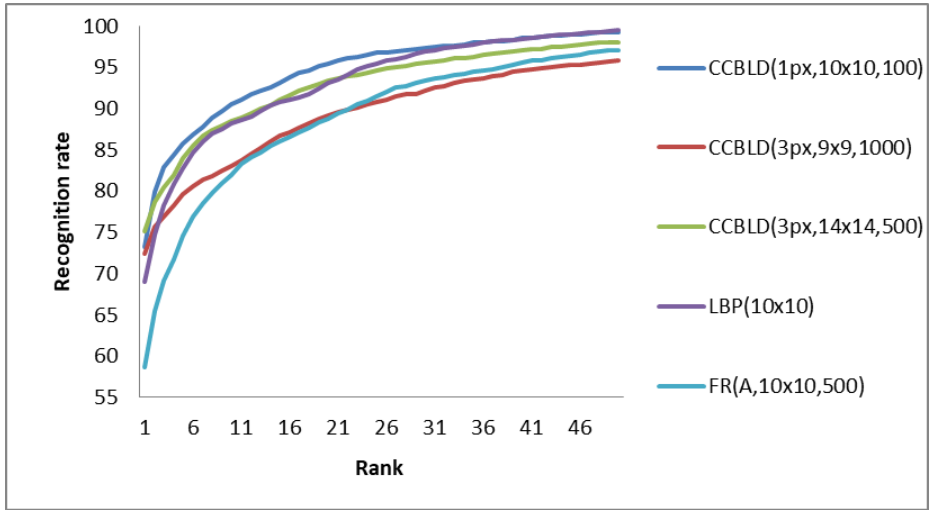


Fig. 2.15 The cumulated recognition rates for rank 1 to rank 50 produced with 600 FERET grayscale preprocessed images.

Table 2.9 Selected rank 1 recognition rates.

Blocks dimension	Subregion count	Dictionary length						
		400	500	600	700	800	900	1000
3	4	97.95	98.2	98.15	98.23	98.08	98	98.08
3	9	98.13	97.98	-	-	-	-	-
3	16	98.28	98.15	98.15	98.23	-	-	-
5	4	98	97.93	-	-	-	-	-
5	9	98.15	98.05	98.23	98.15	-	-	-
5	16	97.6	97.45	-	-	-	-	-
7	4	97.68	97.5	-	-	-	-	-
7	9	98.45	98.3	98.23	98.3	98.23	98.23	98.23
7	16	97.85	97.95	97.75	97.78	-	-	-
9	4	-	96.73	-	-	-	-	-
9	9	98.35	98.38	98.4	98.4	98.28	98.28	98.28
9	16	-	97.83	-	-	-	-	-
11	4	-	96.15	-	-	-	-	-
11	9	-	97.73	-	-	-	-	-
11	16	-	97.53	-	-	-	-	-

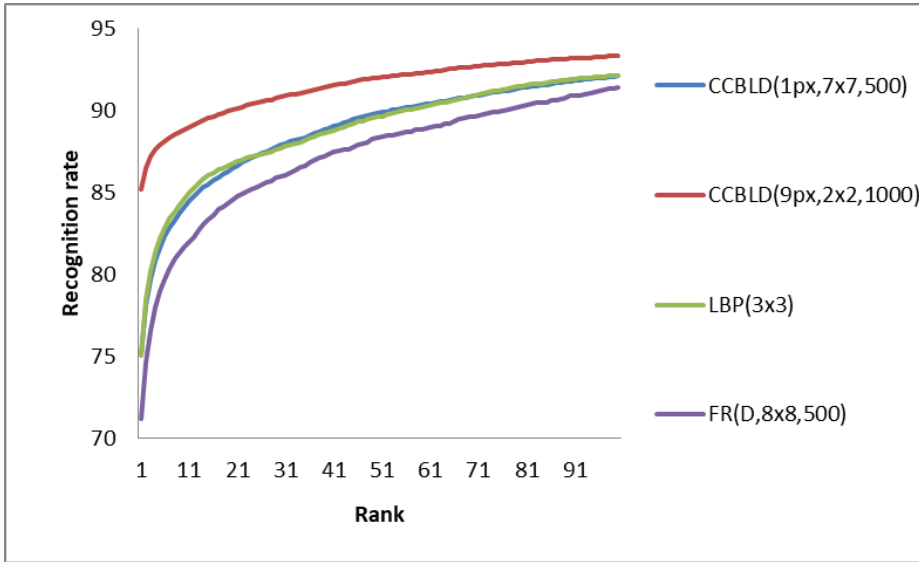


Fig. 2.16 Rank 1 – rank 100 recognition rates produced by the chosen techniques in group of FR, LBP, and CCBLD with the ColorFERET images.

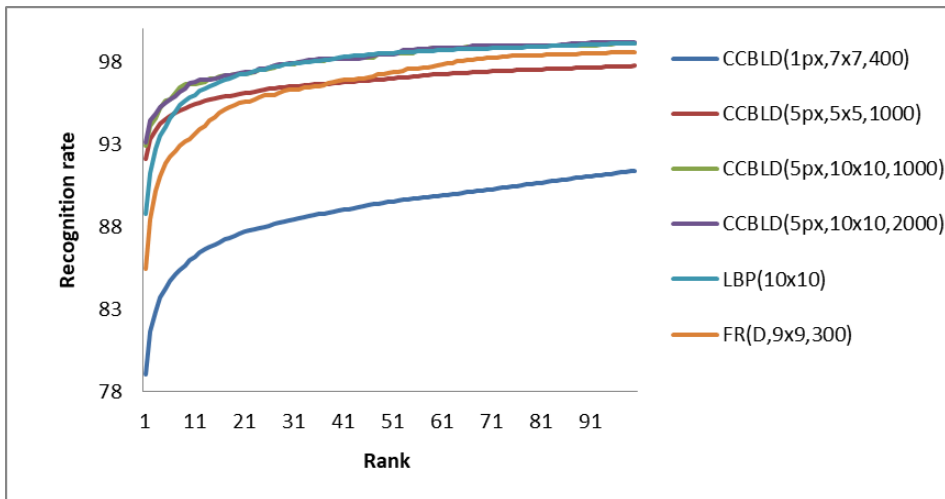


Fig. 2.17 Rank 1 – rank 100 recognition rates produced by the chosen techniques in group of FR, LBP, and CCBLD with the grayscale FERET 3880 images.

2.8.3. Yale

Fig. 2.18 illustrates the recognition rates obtained with various feature extractors. Here, one can observe an essential dependence of the accuracy on the way of partition of the image into the sub-areas. Moreover, one can note a very

huge dependence of the of the accuracy on the dictionary length. Here, the experiment settings are as follows: 5 or 6 images per person were randomly taken to the training set and the rest were taken to the testing set. We have repeated all the experiments 20 times.

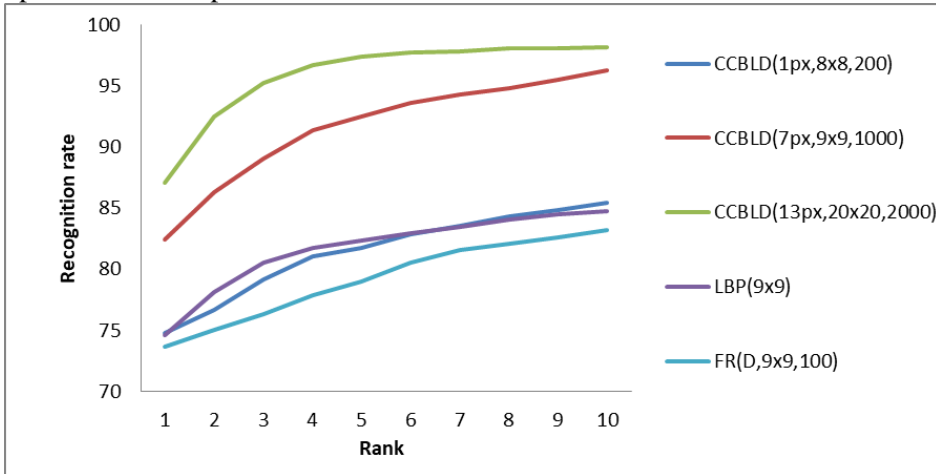


Fig. 2.18 Rank 1 – rank 10 recognition rates for the Yale Face Database.

2.8.4. CAS-PEAL

The results are shown in Fig. 2.19. One can observe that in the case of lower ranks the CCBLD descriptor gives better results than other compared methods. However, CCBLD with one-pixel neighborhood gives very similar results when the ranks over 20 are discussed. Again, 20 iterations of the experiments with half of each person’s images randomly selected to the training set and the rest being in testing set were executed.

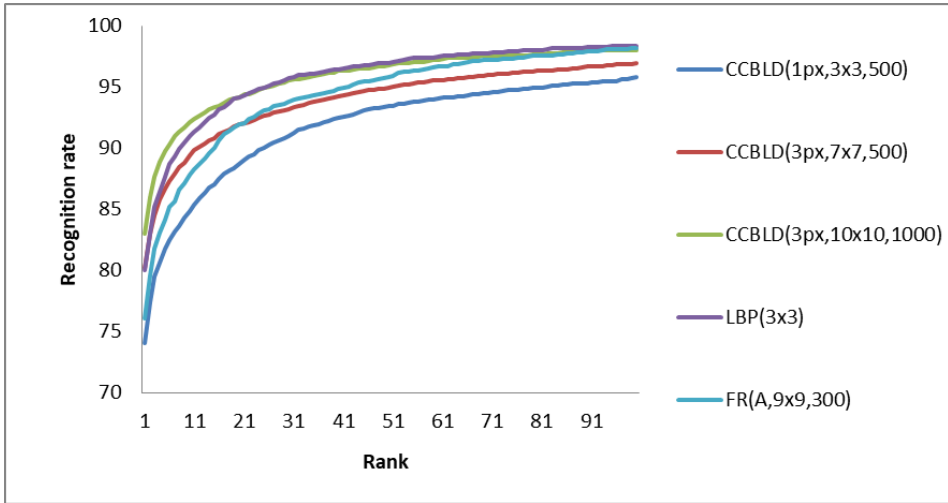


Fig. 2.19 Rank 1 – rank 100 recognition rates yielded with the CAS-PEAL dataset.

2.8.5. Essex

Here, the results are presented in Fig. 2.20. The set is relatively easy. It means that the compared local approaches bring satisfying results. However, similarly to the previous cases, CCBLD produces better recognition rates than the LBP and FR local descriptors. The experimental setting were the same as in the CAS-PEAL set.

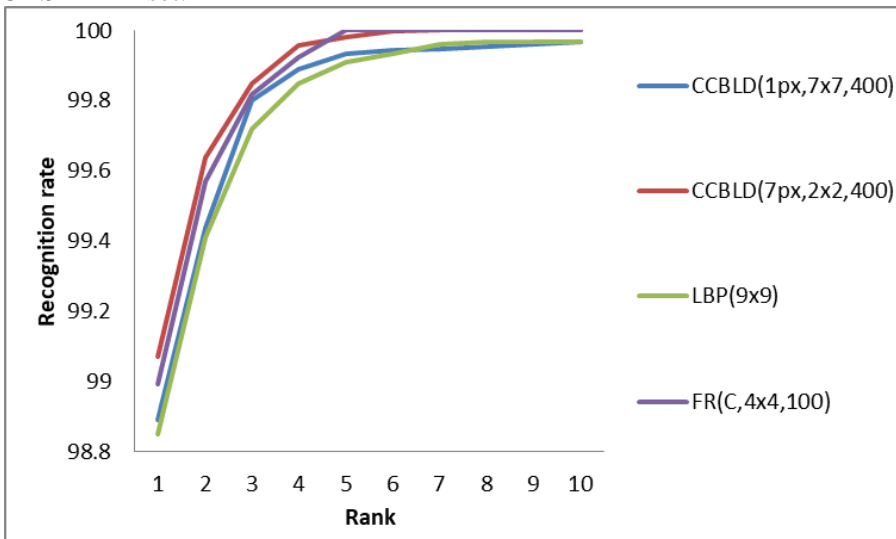


Fig. 2.20 Rank 1 – rank 10 recognition rates for CCBLD, LBP, and FR local descriptors produced for the Essex dataset.

2.8.6. FG-NET

Again, we consider the FG-NET database. The results of comparison are shown in Fig. 2.21. In the sequel, the images were preprocessed (i.e., converted to the grayscale, cropped (128×150 px), rescaled, and the histogram equalization procedure was applied). In this case the results are presented in Fig. 2.22. The first part of this chapter has proven that despite the fact that local approaches are not sufficient to bear the aging progress task in the context of face recognition they have the potential to be combined with other techniques. The results produced by CCBLD suggest that it can be a sound alternative when compared with other local approaches, in particular, in the case of preprocessed images. Note that the parameters of the tests are the same as in the previous cases, i.e., 20 iterations, half of the images contained in the training and the rest in the testing set, respectively.

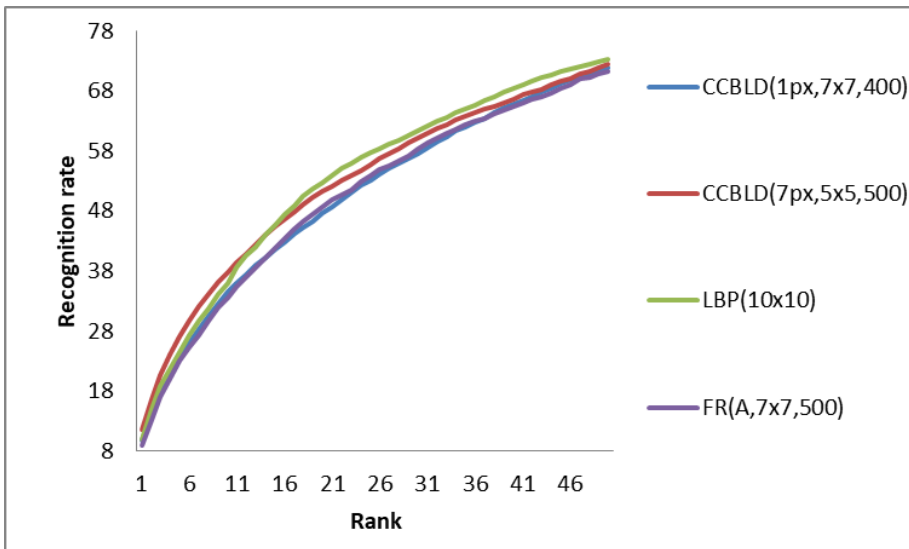


Fig. 2.21 Rank 1 – rank 50 recognition rates in the case of FG-NET dataset.

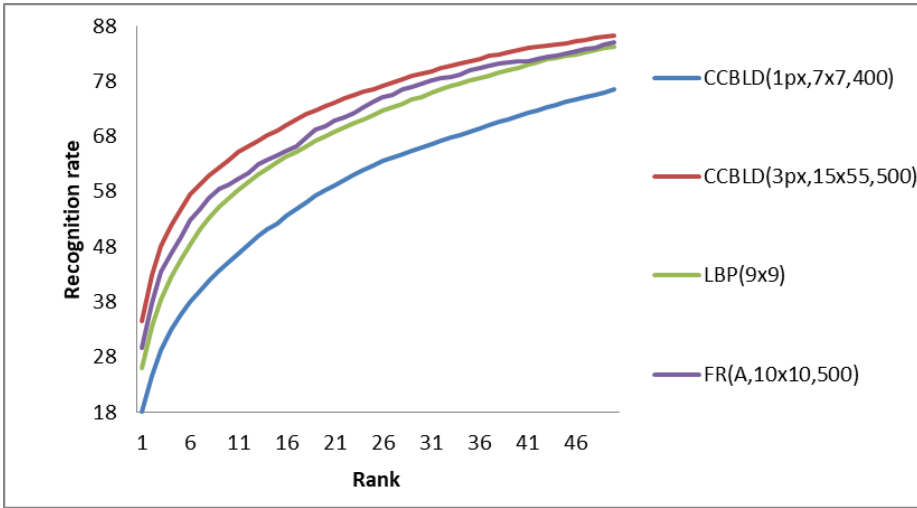


Fig. 2.22 Rank 1 – rank 50 recognition rates in the case of preprocessed images from the FG-NET dataset.

2.8.7. Computational Effort

Here, we discuss the time of performance for chosen procedures and operations related to the execution of the face recognition methods based on CCBLD descriptor. They are the time needed to build a dictionary, processing the files containing the words associated with image pixels, i.e., their changing to the files containing the words best matching to the words contained in the dictionary, and the time needed to compare two histograms of words. We analyze the computational effort on the 600 grayscale images from the FERET image base. The algorithm speed values with respect to its particular steps are listed in Table 2.10, Table 2.11, and Table 2.12. One can observe that the most impact on the application execution time has the dictionary length.

Table 2.10 Computing time (measured in ms) of selected steps of the CCBLD algorithm in its version with 1x1 px block size, no image partition, and the length of dictionary being 100 to 1000 words.

Step	The length of dictionary									
	100	200	300	400	500	600	700	800	900	1000
Dict. build.	561	2636	5163	11529	19923	29001	40544	55474	74823	94489
Chang. text imag.	1950	3532	5229	6855	8540	10445	12548	14017	15819	17714
Hist. build.	193	196	197	198	198	204	200	199	201	200
Two hist. comp.	0.24	0.36	0.49	0.62	0.74	0.86	0.97	1.09	1.21	1.32

Table 2.11 Computing time (measured in ms) of chosen steps of the CCBLD version with no image partition, the length of dictionary set to 100, and the size of starting pixel neighboring block from 1x1 to 15x15.

Action	Size of the starting pixel neighboring block							
	1x1	3x3	5x5	7x7	9x9	11x11	13x13	15x15
Dictionary building	561	187	171	124	140	140	124	124
Changing text images	1950	2033	2119	2075	2071	2008	1972	1982
Histogram building	193	194	194	194	195	193	193	206
Two histograms comparison	0.24	0.24	0.24	0.25	0.24	0.24	0.24	0.24

Table 2.12 Computing time (measured in ms) of chosen steps of CCBLD version with no image partition, the length of dictionary set to 100, the size of starting pixel neighboring block set to 1x1, and the changing partition of an image.

Action	Image partition									
	1x1	2x2	3x3	4x4	5x5	6x6	7x7	8x8	9x9	10x10
Dictionary building	561	452	670	561	702	468	655	483	530	670
Changing text images	1950	1967	1901	1935	1958	1896	1912	1946	2003	1948
Histogram building	193	193	195	196	198	200	203	204	208	211
Two histograms comparison	0.24	0.62	1.23	2.07	3.14	4.42	5.89	7.6	9.44	11.62

2.8.8. Robustness to Noise and Occlusion and Block Version of CCBLD Based on Median

Here, we analyze the problem of noise and occlusion robustness when they are artificially added to the photographs. The facial images coming from the AT&T and grayscale FERET datasets were modified using the well-known salt and pepper and Gaussian noises. The recognition rates are presented in Table 2.13. Similarly, the images were modified by adding white rectangles of random size and placement. Their maximum dimensions were 30×60 px and 90×150 px with respect to AT&T and FERET sets, respectively. Moreover, we tested the CCBLD version for which the method of finding the maxima or minima from the set of averages of the block was replaced by finding the extrema among the block medians. This makes the descriptor significantly more robust to the occurrence of noise.

2.8.9. Similarity/Dissimilarity Measures Performance

The last series of experiments is devoted to statistics exhibiting the performance of four well-known measures of similarity/dissimilarity which are applied to the comparison of word histograms representing the images. They are Hellinger, χ^2 statistics, correlation, Hellinger, and histogram intersection measure. The results are gathered in Table 2.14. As supposed in the first series

of experiments (see above), the best results are produced by the Hellinger measure. A good alternative may be also histogram intersection measure.

Table 2.13 Recognition rates for the images with occlusion, Gaussian noise, and salt and pepper noise.

Set	Method	Occl		Salt & pepper noise (%)					Gaussian noise (%)			
		10	20	30	40	50	10	20	30	40	50	
AT&T	CCBLD (1px, 4x4,100)	36.35	81.25	61.55	37.08	18.30	9.08	83.88	61.50	36.43	19.75	13.95
	CCBLD (7px, 3x3,400) with an average of pixel block	63.30	97.45	95.60	95.83	93.70	89.38	97.43	95.40	93.58	90.90	87.48
	CCBLD (7px, 3x3,400) with a median of pixel block	69.30	97.65	96.83	96.23	95.85	95.48	97.23	96.85	95.93	95.90	95.28
	LBP	32.23	83.53	62.83	39.50	21.58	11.40	83.68	61.20	33.58	16.08	6.88
	FR(D, 5x5,300)	23.53	31.38	20.35	10.98	8.48	5.80	33.63	19.58	12.68	5.33	4.38
FERET	CCBLD (1px, 7x7,400)	7.08	67.38	60.15	53.13	44.30	36.58	20.15	67.93	61.75	54.68	43.78
	LBP	7.53	62.30	56.85	55.18	53.70	44.83	33.45	63.95	61.18	53.60	49.68
	FR(D, 4x4,500)	1.43	53.48	47.85	41.10	32.33	21.25	16.28	54.95	49.4	43.73	30.53

2.9. Conclusions

In this chapter, first, we have discussed an application of local descriptor-based techniques to the problem of aging in face recognition task. The descriptors were considered as standalone entities or in combination with Gabor magnitude images and Gabor phase images. The tests show that the accuracies are the best in case of the fusion of local descriptors and Gabor magnitudes and the descriptor such as MBLBP. The recognition rates obtained with this descriptor seem to be the most stable in comparison to other descriptors. On the other hand, the recognition rates obtained with Gabor phase-based descriptors are essentially low. Finally, various dissimilarity/similarity measures were compared in an application to histograms built of LBP, ILBP, and MBLBP values. The results of the experiments show that the fusion of local descriptors and Gabor filters is a method which can be successfully utilized in the tasks related with age progress. However, the methods have to be supplemented by other reducing the dimension of the data and increasing the accuracy of recognition.

Next, we have thoroughly examined the performance of the Chain Code-Based Local Descriptor and its extensions based on the descriptions of pixels

(blocks of pixels) being the neighbors of a given central image point (or block). Considered were various problems of face recognition (aging, noise, occlusion, pose, illumination, preprocessing) and an application of CCBLD in their context. The carried series of experiments show that CCBLD outperforms other methods such as Local Binary Pattern and its extensions and Full Ranking.

Table 2.14 Performance of the measures used to compare the histograms of words.

Local descriptor	Chi-square	Correlation	Hellinger	Histogram intersection
CCBLD(1px,4x4,100)	93.30	93.80	96.20	95.88
FR(D,5x5,300)	94.90	93.53	94.23	94.93
LBP	97.00	93.20	97.00	96.83

References

Abid, N., Loukil, K., Ouni, T., Ayedi, W., Ammari, A. C., Abidet, M.: An improvement of multi-scale covariance descriptor for embedded system. *J. Real-Time Image Proc.* 1–17 (2018)

Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with Local Binary Patterns. In: Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science **3021**, pp. 469–481 (2004)

Bartyzel, K.: Invariant Levenshtein distance for comparison of Brownian strings. *J. Appl. Comput. Sci.* **18** (1), 7–17 (2010)

Becerra-Riera, F., Méndez-Vázquez, H., Morales-Gonzalez, A., Tistarelli, M.: Age and gender classification using local appearance descriptors from facial components. In: 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, pp. 799–804 (2017)

Belver, C., Arganda-Carreras, I., Dornaika, F.: Comparative study of human age estimation based on hand-crafted and deep face features. In: Nasrollahi, K. et al. (Eds.): Video Analytics. Face and Facial Expression Recognition and Audience Measurement. VAAM 2016, FFER 2016. Lecture Notes in Computer Science **10165**, pp. 98–112 (2017)

Bereta, M., Pedrycz, W., Reformat, M.: Local descriptors and similarity measures for frontal face recognition: A comparative analysis. *J. Vis. Commun. Image. Represent.* **24**, 1213–1231 (2013)

Bendada, A., Akhloufi, M. A.: Multispectral face recognition in texture space. In: 2010 Canadian Conference Computer and Robot Vision, pp. 101–106 (2010)

Best-Rowden, L., Jain, A. K.: Longitudinal study of automatic face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40** (1), 148–162 (2018)

Bharkad, S. D., Kokare, M.: Performance evaluation of distance metrics: Application to fingerprint recognition. *Int. J. Pattern Recognit. Artif. Intell.* **25**, 777–806 (2011)

- Boussaad, L., Benmohammed, M., Benzid, R.: Age invariant face recognition based on DCT feature extraction and kernel Fisher analysis. *J. Inf. Process. Syst.* **12**, 392–409 (2016)
- Bribiesca, E.: A new chain code. *Pattern Recognit.* **32**, 235–251 (1999)
- Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern histograms for face recognition. In: Lee, S.-W., Li, S. Z. (Eds.): *ICB 2007, Lecture Notes in Computer Science* **4642**, pp. 809–818 (2007)
- Chan, C.H., Yan, F., Kittler, J., Mikolajczyk, K.: Full ranking as local descriptor for visual recognition: A comparison of distance metrics on s_n . *Pattern Recognit.* **48**, 1328–1336 (2015)
- Chen, J., Shan, S., He, C., Zhao, G., Pietikäinen, M., Chen, X., Gao, W.: WLD: A robust local image descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **32** (9), 1705–1720 (2010)
- Choi, W.-P., Tse, S.-H., Wong, K.-W., Lam, K.-M.: Simplified Gabor wavelets for human face recognition. *Pattern Recognit.* **41**, 1186–1199 (2008)
- Deb, D., Nain, N., Jain, A. K.: Longitudinal study of child face recognition. In: *The 11th IAPR International Conference on Biometrics (ICB 2018)*, arXiv:1711.03990 (2017)
- Dornaika, F., Bosaghzadeh, A., Salmane, H., Ruichek, Y.: Graph-based semi-supervised learning with Local Binary Patterns for holistic object categorization. *Expert Syst. Appl.* **41** (17), 7744–7753 (2014)
- Du, L., Ling, H.: Cross-age face verification by coordinating with cross-face age verification. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2329–2338 (2015)
- Georgopoulos, M., Panagakis, Y., Pantic, M.: Modelling of facial aging and kinship: A survey. *Image Vis. Comput.* doi: 10.1016/j.imavis.2018.05.003 (2018)
- Girish, G. N., Naika, S. C. L., Das, P. K.: Face recognition using MB-LBP and PCA: A comparative study. In: *2014 International Conference on Computer Communication and Informatics, Coimbatore*, pp. 1–6 (2014)
- Guo, G., Mu, G., Ricanek, K.: Cross-age face recognition on a very large database: The performance versus age intervals and improvement using soft biometric traits. In: *20th International Conference on Pattern Recognition*, pp. 3392–3395 (2010)
- Guo, Y., Xu, Z.: Local Gabor phase difference pattern for face recognition. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4 (2008)
- Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with center-symmetric Local Binary Patterns. In: *Indian Conference on Computer*

- Vision, Graphics and Image Processing (ICVGIP) 2006, Lecture Notes in Computer Science **4338**, pp. 58–69 (2006)
- Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with Local Binary Patterns, *Pattern Recognit.* **42**, 425–436 (2009)
- Jin, H., Liu, Q., Lu, H., Tong, X.: Face detection using improved LBP under Bayesian framework. In: *International Conference on Image and Graphics*, pp. 306–309 (2004)
- Jindal, A., Gupta, S., Kaur, L.: Face recognition techniques with permanent changes: A review. In: *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, Noida, pp. 689–693 (2015)
- Juefei-Xu, F., Luu, K., Savvides, M., Bui, T. D., Suen, C. Y.: Investigating age invariant face recognition based on periocular biometrics. In: *Biometrics (IJCB), 2011 International Joint Conference on*, pp. 1–7 (2011)
- Jun, B., Lee, H. S., Lee, J., Kim, D.: Statistical face image preprocessing and non-statistical face representation for practical face recognition. In: *Signal Processing and Information Technology (ISSPIT), 2009 IEEE International Symposium on*, pp. 392–397 (2009)
- Lanitis, A., Taylor, C. J., Cootes, T. F.: Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (4), 442–455 (2002)
- Lee, T. S.: Image representation using 2D Gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.* **18** (10), 959–971 (1996)
- Levenshtein, V. I.: Binary codes with correction for deletions and insertions of the symbol 1. *Probl. Peredachi Inf.* **1**, 12–25 (1965)
- Li, Y., Wang, G., Nie, L., Wang, Q., Tan, W.: Distance metric optimization driven convolutional neural network for age invariant face recognition. *Pattern Recognit.* **75**, 51–62 (2018)
- Li, Z., Park, U., Jain, A. K.: A discriminative model for age invariant face recognition. *IEEE Trans. Inf. Forensics Secur.* **6** (1), 1028–1037 (2011)
- Liao, W.: Contextual patch feature learning for face recognition. *J. Softw.* **9** (7), 1827–1832 (2014)
- Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S. Z.: Learning multi-scale block Local Binary Patterns for face recognition. In: *Advances in Biometrics, International Conference, ICB 2007, Lecture Notes in Computer Science* **4642**, pp. 828–837 (2007)
- Ling, H., Soatto, S., Ramanathan, N., Jacobs, D. W.: Face verification across age progression using discriminative methods. *IEEE Trans. Inf. Forensics Secur.* **5** (1), 82–91 (2010)

- Luu, K., Seshadri, K., Savvides, M., Buil, T. D., Suen, C. Y.: Contourlet appearance model for facial age estimation. In: Biometrics (IJCB), 2011 International Joint Conference on, pp. 1–8 (2011)
- McKee, J.W., Aggarwal, J. K.: Computer recognition of partial views of curved objects. *IEEE Trans. Comput.* **26**, 790–800 (1977)
- Mehetre, B. M., Kankanhalli, M. S., Lee, W. F.: Shape measures for content based image retrieval: A comparison. *Inf. Process. Manag.* **33**, 319–337 (1997)
- Memiş, A.: Facial feature representation and face recognition with Neighborhood-based Binary Patterns. In: 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, pp. 1–4 (2018)
- Meng, C., Lu, J., Tan, Y.-P.: A comparative study of age-invariant face recognition with different feature representations. In: 11th International Conference on Control, Automation, Robotics and Vision, pp. 890–895 (2010)
- Nagpal, S., Singh, M., Vatsa, M., Singh, R.: Regularizing deep learning architecture for face recognition with weight variations. In: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–6 (2015)
- Navarro, G.: A guided tour to approximate string matching. *ACM Comput. Surv.* **33**, 31–88 (2001)
- Naveena, C., Manjunath Aradhya, V. N., Niranjana, S. K.: 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**, pp. 781–787 (2010)
- Nimbarte, M., Bhojar, K. K.: Distinctive feature representations and classifiers for age invariant face recognition - a survey. *Int. J. Comput. Syst. Eng.* **2**, 222–235 (2016)
- Nimbarte, M., Bhojar, K.K.: Age invariant face recognition using convolutional neural network. *Int. J. Electr. Comput. Eng.* **8**, 2126–2138 (2018a)
- Nimbarte, M., Bhojar, K.K.: Face recognition across aging using GLBP features. In: Satapathy, S., Joshi, A. (Eds.): Information and Communication Technology for Intelligent Systems (ICTIS 2017) - Volume 2. ICTIS 2017. Smart Innovation, Systems and Technologies, vol **84**. Springer, Cham, pp. 275–283 (2018b)
- OpenCV 2.4.13.0 documentation. [online]
http://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/histogram_comparison/histogram_comparison.html (Accessed 22 July 2016)
- Osman, A. M., Viriri, S.: Face verification across age progression: A survey of the state-of-the-art. In: 2018 Conference on Information Communications Technology and Society (ICTAS), Durban, pp. 1–6 (2018)

- Panis, G., Lanitis, A.: An overview of research activities in facial age estimation using the FG-NET Aging Database. In: Agapito, L., Bronstein, M., Rother, C. (Eds.): *Computer Vision – ECCV 2014 Workshops*. ECCV 2014. *Lecture Notes in Computer Science* **8926**, pp. 737–750 (2015)
- Park, U., Tong, M., Jain, A. K.: Age-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **32** (5), 947-954 (2010)
- Perez, C. A., Cament, L. A., Castillo, L. E.: Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count. *Pattern Recognit.* **44**, 951–963 (2011)
- Perlibakas, V.: Distance measures for PCA-based face recognition. *Pattern Recognit. Lett.* **25**, 711–724 (2004)
- Pontes, J. K., Britto, A. S., Fookes, C., Koerich, A. L.: A flexible hierarchical approach for facial age estimation based on multiple features. *Pattern Recognit.* **54**, 34–51 (2016)
- Pontes, J. K., Fookes, C., Britto, A. S., Koerich, A. L.: Two-stage facial age prediction using group-specific features. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, pp. 2577–2581 (2017)
- Punyani, P., Gupta, R., Kumar, A.: A comparison study of face, gait and speech features for age estimation. In: Kalam, A., Das, S., Sharma, K. (Eds.): *Advances in Electronics, Communication and Computing*. *Lecture Notes in Electrical Engineering* **443**. Springer, Singapore, pp. 325–331 (2018)
- Ramanathan, N., Chellappa, R.: Learning facial aging models: A face recognition perspective. In: Boulgouris, N. V., Plataniotis, K. N., Micheli-Tzanakou E. (Eds.): *Biometrics: Theory, Methods, and Applications*. John Wiley & Sons, Hoboken, New Jersey, pp. 271–296 (2010)
- Reddy, P. K. K.: Derivation of shape descriptors on uniform Local Binary Patterns for classification of textures. *IJREAT Int. J. Res. Eng. Adv. Technol.* **3**, 12–22 (2015)
- Ren, X., Guo H., Di, C., Han, Z., Li, S.: Face recognition based on local Gabor binary patterns and convolutional neural network. In: Liang Q., Mu J., Wang W., Zhang B. (Eds.): *Communications, Signal Processing, and Systems*. *CSPS 2016*. *Lecture Notes in Electrical Engineering* **423**. Springer, Singapore, pp. 699-707 (2016)
- Ren, J., Jiang, X., Yuan, J.: Learning LBP structure by maximizing the conditional mutual information. *Pattern Recognit.* **48** (10), 3180–3190 (2015)
- Ricanek, K., Boone, E.: The effect of normal adult aging on standard PCA face recognition accuracy rates. In: *Proceedings of International Joint Conference on Neural Networks*, pp. 2018–2023 (2005)

- Ricanek Jr, K., Tesafaye, T.: MORPH: A longitudinal image database of normal adult age-progression. In: IEEE 7th International Conference on Automatic Face and Gesture Recognition. Southampton, pp. 341–345 (2006)
- Serrano, Á., de Diego, I. M., Conde, C., Cabello, E.: Recent advances in face biometrics with Gabor wavelets: A review. *Pattern Recognit. Lett.* **31**, 372–381 (2010)
- Sethuram, A., Patterson, E., Ricanek, K., Rawls, A.: Improvements and performance evaluation concerning synthetic age progression and face recognition affected by adult aging. In: ICB 2009, Lecture Notes in Computer Science **5558**, pp. 62–71 (2009)
- Shen, L., Bai, L.: A review on Gabor wavelets for face recognition. *Pattern Anal. Appl.* **9** (2–3), 273–292 (2006)
- Shylaja, S. S., Balasubramanya Murthy, K. N., Natarajan, S.: Dimensionality reduction techniques for face recognition. In: Corcoran P. M. (Ed.): Reviews, Refinements and New Ideas in Face Recognition. InTech, pp. 141–166 (2011)
- Sivic, J., Zisserman, A.: Video Google: A text retrieval approach to object matching in videos. In: Proceedings, Ninth IEEE International Conference on Computer Vision, Vol. 2, pp. 1470–1477 (2003)
- Smiatecz, M.: Similarity measures for face images: An Experimental Study. In: Chmielewski, L., Datta, A., Kozera, R., Wojciechowski, K. (Eds.): Computer Vision and Graphics. ICCVG 2016. Lecture Notes in Computer Science **9972**, pp. 341–352 (2016)
- Smiatecz, M., Rumiński, J.: Local texture pattern selection for efficient face recognition and tracking. In: Burduk, R., Jackowski, K., Kurzyński, M., Woźniak, M., Żołnierok, A. (Eds.): Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015, pp. 359–368 (2015)
- Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures, Lecture Notes in Computer Science **4778**, pp. 168–182 (2007)
- Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1991)
- Viola, P., Jones, M.: Robust real-time object detection. *Int. J. Comput. Vis.* **57** (2), 137–154 (2004)
- Wang, J., Shang, Y., Su, G., Lin, X.: Age simulation for face recognition. In: 18th International Conference on Pattern Recognition, pp. 913–916 (2006)
- Wiskott, L., Fellous, J.-M., Krüger, N., von der Malsburg, C.: Face recognition by Elastic Bunch Graph Matching, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 775–779 (1997)

- Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: Faces in Real-Life Images Workshop in ECCV, pp. 1–14 (2008)
- Xie, S., Shan, S., Chen, X., Chen, J.: Fusing local patterns of Gabor magnitude and phase for face recognition. *IEEE Trans. Image Process.* **19** (5), 1349–1361 (2010)
- Xue, Y., Tong, C. S., Zhang, W.: Survey of distance measures for NMF-based face recognition. In: Wang, Y., Cheung, Y., Liu, H. (Eds.): CIS 2006, LNAI **4456**, pp. 1039–1049 (2007)
- Yang, C., Tiebe, O., Shirahama, K., Łukasik, E., Grzegorzek, M.: Evaluating contour segment descriptors. *Mach. Vis. Appl.* **28** (3–4), 373–391 (2017b)
- Yang, Y., Duan, F., Jiang, J., Ma, L., Zheng, H.: Robust method for interest region description based on local intensity binary pattern. *J. Electr. Imaging* **26** (4), 043025 (2017b)
- Yang, Y., Duan, F., Ma, L.: A rotationally invariant descriptor based on mixed intensity feature histograms. *Pattern Recognit* **76**, 162–174 (2018)
- Yang, H., Huang, D., Wang, Y.: Age invariant face recognition based on texture embedded discriminative graph model. In: IEEE International Joint Conference on Biometrics, Clearwater, FL, 2014, pp. 1–8 (2014)
- Yang, J., Zhang, D., Yang, J., Niu, B.: Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **29** (4), 650–664 (2007)
- Zhang, B., Shang, S., Chen, X., Gao, W.: Histogram of Gabor phase pattern (HGPP): A novel object representation approach for face recognition. *IEEE Trans. Image Process.* **16** (1), 57–68 (2007)
- Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Computer Vision, ICCV 2005. Tenth IEEE International Conference on, pp. 786–791 (2005)
- Zhang, W., Shan, S., Qing, L., Chen, X., Gao, W.: Are Gabor phases really useless for face recognition?. *Pattern Anal. Appl.* **12** (3), 301–307 (2009)

3. An Application of Linguistic Descriptors to the Face Recognition Problems

In this chapter, we discuss an application of linguistic descriptors to the problem of facial recognition. The idea behind this concept comes from the strong believe that people are highly efficient in the task of recognizing faces. Of course, it is impossible to retrieve huge datasets of facial images without any computational effort. However, the knowledge coming from the observation of human recognition processes and expert opinions can be an invaluable input to the computational methods of face recognition. Particularly, we propose a method of assessing the facial features saliency in the process of facial recognition and analyze a set of classifiers based on expert knowledge. These two important processes may be effectively supported by Analytic Hierarchy Process.

3.1. Introductory Notes

Capturing a way people recognize and describe other people has been still a challenging research problem. Much research is still focused on how to describe the essence of the recognition process. It is obvious that we are very efficient in recognizing others. This task is even easier if the face to be recognized has been seen previously. Moreover, we are extremely efficient in the processes of comparison of the chosen facial regions or parts, particularly if they belong to one of the most important from the recognition point of view, namely pericocular area (Hollingsworth 2014). Despite of this ability, it is impossible for the people to remember and correctly recognize thousands of faces. Moreover, this task is even more difficult to do in a reasonable amount of time. On the other hand, computer algorithms work relatively well. However, they still cannot fully manage the various illumination, noise, pose, distance to the camera, quality of an image, or age of the depicted person-related problems. Furthermore, the mechanisms of recognition and description of the facial elements and, in general, faces by humans, still remain uncaptured by machines. It is noteworthy that the manner people describe the faces and the facial regions using natural language is relatively simple. Moreover, it is relatively common for the whole population. This fact is used by specialists from the fields of forensics and criminology, at least from the area of one culture. It is related to the so-called phenomenon of the own race bias which means that the faces from one's own ethnicity or culture are better remembered than of another ethnicity (see DeLozier & Rhodes 2015).

O'Toole et al. 2007 reported that the fusion of subjects' answers and the computational algorithms can improve the rate of correct face verification. This observation can be an input to in-depth studies on the linguistic descriptors of the face and the computational face classification methods. Undoubtedly, the

saliency of specific facial parts may be crucial here. There may be two reasons of this. First, we can exclude the non-important facial features from the consideration what can save our time and resources. The second cause is that the saliency of information contained within a concrete face region or feature can be different than the information covered by the other feature. And here, the problem of estimation of the relevance of the features appears.

However, one cannot focus only on the specific facial features treated separately. People process a face in a holistic way (Sinha et al. 2006) with a pivotal role of the second-order relations between features (spacing), see Rotshtein et al. 2007. Particularly, the internal features are proven to be more important than the external (contour of a face, hair, etc.) in the process of familiar faces recognition. Inversely, when unfamiliar faces are considered the external features are more important (Ellis et al. 1979; Young et al. 1985). Classic studies by Davies et al. 1977, Haig 1986, and Matthews 1978 demonstrate that eyebrows/eyes followed by the regions of mouth and nose are the most descriptive areas. Furthermore, the saliency of eyebrows was proved and intensively examined, for instance, by Sadr et al. 2003. An interesting research on various region fusions and their importance were considered by Tome et al. (2015a). Literature surveys on recognition of familiar and unfamiliar (i.e., trained and untrained) faces and cue saliency are covered in Shepherd et al. 1981, Johnston & Edmonds 2009, and Vignolo et al. 2013. Beside these facts, there are the publications presenting the importance of particular facial regions in the computational methods of face recognition. A description of these methods can be found in chapter 6. Finally, noteworthy is the fact that the holistic way of processing faces by humans strictly corresponds to the Gestalt theory which is one of the trends in psychology. In this theory the concept of holism arises as a very important idea (Wagemans et al. 2012). The meaning of the whole is something more than its parts (composites) summing process (Koffka 1935). One of the ideas of the Gestalt theory is the global precedence hypothesis (Navon 1977). It states that the processing of visual objects constituting a hierarchical structure including a set of dependencies among the parts proceeds from the top of hierarchy, namely the global structures, towards the analysis of local properties which are at the bottom positions (Wagemans et al. 2012). When the objects like a face are considered, we can observe the spatial relationships which exist between the parts of the face. Next, there are spatial relations between the subregions or subparts, etc. According to the theory, these relations are more general than the specific attributes and their properties (e.g., length, width, etc.). Capturing these spatial relationships appears as an interesting task here. To fully address this problem, one has to estimate the saliency of particular facial features. This evaluation can be proceeded on a basis of judgments of professional experts in the field of psychology or forensics. Next, the obtained results (weights) can be an input to build an efficient classifier based on many formal tools, e.g., aggregation or fusion operations.

Moreover, the above-mentioned global precedence hypothesis directly corresponds to Granular Computing paradigm, see Pedrycz 2013. That is, the features can be collected into semantically sound entities. The entities refer to internal or external facial features, lower or upper half of a face and other areas which constitute information granules. It is worth noting that the last partition was proposed in Kurach et al. 2014. All these granules of information are consisted of “atomic” features (facial parts). For instance, the lower half of face can be represented (or decomposed) by mouth, chin, cheeks, etc. At the end, the entire face consists of these groups. Having in mind the above assumption about the difference between the whole face and the direct summing of its features, our main task is to find the essence of the process of working with the relationships between the cues at each level of granulation (abstraction). It means that we must find out how to proceed with the atomic facial cues, their groups (regions or areas), and the face positioned at the highest abstraction level. Such approach is intuitively appealing and the feature space being a result of the proceeding is linguistic. Hence, it can be described using the granular information terms.

Hence, the need of improvement and work on the linguistic descriptors-based methods applied to face recognition processes has been constantly discussed in the literature topic. In fact, it is believed that computer can achieve an efficient and realistic interaction between human and machine (see, Iwamoto & Ralescu 1992). However, this process can vary depending on people. Moreover, factors like their culture, age, profession, etc. can be a source of misunderstandings. Fukushima & Ralescu (1995) proposed a so-called adjustment procedure to minimize these differences between descriptions. Moreover, applications of Granular Computing models as well as engaging specialists having an experience in psychology or criminology to assess the facial regions can be pivotal here. For instance, Kumar et al. (2009; 2011) experimented with a large set of “witnesses” who used the Amazon Mechanical Turk service taking into account 65 and 73 facial features, respectively, to get descriptions of a set of facial images. Fukushima & Ralescu (1995), Nakayama et al. (1992a; 1992b) proposed a system based on 19 features which were evaluated by subjects using the terms such as *small*, *rather small*, *medium*, *rather big*, and *big*. Next, these terms were considered as fuzzy sets and the classification process was based on the measure of overlap. The method of obtaining the linguistic values was conducted in the way described by Miyajima et al. (1992). Norita (1994) improved the method with an application of so-called total impressions words, e.g., *cold face*. A neuro-fuzzy algorithm using linguistic descriptors built of triangular and trapezoidal membership functions was applied in (Lee et al. 1998). In Wu & Narasimhalu (1998) it was proposed fuzzy retrieval system where the chosen facial parts were described using linguistic descriptions and fuzzy sets. A latent semantic space built of exact pixel lengths and descriptions of unmeasurable features like person’s character were presented by Ito & Koshimizu (2004). Also, latent semantic space was discussed thoroughly in Ito

& Koshimizu (2006). A metadata set containing automatically generated face descriptions was presented in Sridharan et al. 2005. Conceptual fuzzy sets and linguistic descriptions of parts of a face in an application to draw a face using a so called average face was discussed by Benhidour & Onisawa (2007; 2008). Semantic feature extraction and tensor subspace analysis were discussed by Zhou & Schaefer (2010) where semantic features are three regions of a face. However, the information about facial parts was yielded computationally. In Alattab & Kareem (2012b) 100 participants of a series of experiments gave the semantic descriptions of the faces. However, the number of proceeded faces had to be narrowed. In Conilione & Wang (2012) the labelling of the features was obtained using fuzzy clustering. Again, 100 respondents established (by filling questionnaires) the weights of facial features to be applied in a facial retrieval system. However, the features were detected using the method of segmentation proposed in (Alattab & Kareem 2012a). An interesting approach was proposed by Kurach et al. 2014. A concept of hierarchical granulation of facial features and fuzzy rules used to classify the faces were applied. The fuzzy sets were strictly related to the measures of the most salient facial regions. A so called soft biometrics used in a communication between the robot and its human partner were discussed in (Martinson et al. 2013) including the linguistic descriptors of a face. A very interesting concept of sketching with words was proposed by Rahman & Beg (2015). Fuzzy geometry, fuzzy granules, and computing with words (Zadeh 2001; 2009) were applied to convert the imprecise linguistic descriptors into complete faces. It is worth noting that a quite similar kind of facial description was applied in the processes of recognition of emotions and social interaction. For instance, a very popular method here is FACS (Facial Action Coding System). It is used to quantify the actions of a face using Action Units, see Ekman & Friesen (1978), Pantic & Rothkrantz (2000a; 2000b). Fuzzy inference and clustering were applied by Conilione & Wang (2012) while an automatic conversion of landmarks of the face to a set of features was proposed by Tome et al. (2015b). Finally, neural network-based classifier based on the experts' description, a concept of weights related to fuzzy memberships, and quantitative methods of evaluating the importance of facial features in the recognition processes were discussed in Dolecki et al. 2016, Kiersztyn et al. 2016a; 2016b, respectively. The modern descriptions of facial features which can be utilized in criminal investigations may be found in standardizing documents such as FISWG (2016), police websites (e.g. a Chicago Police Department (2016)) or textbooks by Czerw 1995 and Lindsay et al. 2007. An interested reader can read a survey of the methods contained in (Karczmarek et al. 2015).

The ultimate goal of the study presented in this chapter is to find the saliency of the facial features which are utilized in the processes of human face recognition. Moreover, we are interested in an investigation of the efficiency of the process. The obtained weights of the facial areas (regions, parts, etc.) can be

applied in the processes of computational face recognition methods based on aggregation or fusion schemes which are one of the important trends in the investigation of classification methods, see, e.g., Anderson et al. 2018. Moreover, the results may be applied in the fields of criminology, forensics, psychology of emotion, etc. The method of obtaining their rankings at various levels of abstraction can be of particular interest of the professionals and researchers of the disciplines.

The novelty of our proposal comes from an application of Analytic Hierarchy Process introduced by Saaty (1980). Here, we propose to use a hierarchy built of three levels. It means that the information of the whole face in general is placed at the top. Particular features of a face are positioned at the bottom. Moreover, we use only the linguistic data (information). They are not expressed in a numerical form. To obtain the reliable results we have engaged experts in the field of psychology and criminology having an experience in a practical problems. Finally, the originality of our approach is that we use the entropy measure to evaluate the results of the AHP at the three levels of hierarchy. It can be used to verify the confidence of information produced in the AHP process. When working with experts evaluating particular facial features or estimating the importance of facial areas in general, an application of the mechanism of evaluation of their work, namely the entropy measure and reciprocity property can lead to the reduction of biases occurring in an expert's work, see, e.g., Dror et al. 2012 where an impact of technology is described or Zhang et al. 2016 where the influence of internal feelings on the emotion recognition is discussed.

Noteworthy is the fact that the approach based on an expert opinions cannot be overrated in an applications to forensics. The modern state-of-the-art methods based on, e.g., deep learning models (Sun et al. 2014), sparse recognition (Wright et al. 2009), or local descriptors (Ahonen et al. 2004) can be supplemented by the expert-based knowledge. This fact corresponds to the point of view presented in the literature of forensic science, e.g., Arca et al. 2011. Here, it is important to stress, that always at the end of the recognition process, an expert must give an opinion whether a person is classified or eliminated (Spaun et al. 2001). Inversly, an expert or a group of experts can be effectively supplemented by the algorithms of feature extraction followed by MCDM. In addition, the data collected by the expets can be an invaluable source of information for the research in other domains of science. The method presented here can be a kind of a novel road map to obtain the saliency of newly introduced facial features considered in various problems. Moreover, the experts' presence can be a form of reducing the semantic gap between the low level and high level features (Liu et al. 2007). Even if the experts evaluate the features subjectively or their assessments are dominated by a so-called own race bias the techniques of ajustment of their evaluations can be applied here. Finally, it is worth to stress that the information granularity which is the outcome of the

method can be quantified (or captured) on a basis of the structure of multilevel hierarchy of AHP at various levels of abstraction.

3.2. Facial Features Saliency

Here, we are highly interested in obtaining the saliency of particular facial features and their groups. Our ultimate goal is to determine the order of the most salient, important, and useful from the point of view of a classification task, features. The AHP criterion is the importance of the face parts while the particular facial features constitute the AHP alternatives.

The experts (or expert, AHP user) evaluate the pairwise comparisons (alternatives) of the elements of the AHP hierarchy. Assume that we have a set of n alternatives (features of a face). Then the judgements' results generate the $n \times n$ matrix A and the principal eigenvectors $w_i = [w_{i,1}, \dots, w_{i,n}]$ of the AHP reciprocal matrices are derived. The specific element values correspond to the saliency of particular features according to the judgements of i th expert ($i=1, \dots, p$). Next, they are normalized, $y_i = w_i / \max_j w_{i,j}$ and the coefficients μ_i , denoting consistency ratios, are obtained. The weights related to the experts are $\omega_i = 1 - \mu_i$, and, again, they are normalized, i.e., $u_i = \omega_i / (\omega_1 + \dots + \omega_p)$. Finally, the saliency of the j^{th} feature is given by

$$x_j = y_{1j} u_1 + y_{2j} u_2 + \dots + y_{pj} u_p, \quad j=1, \dots, n \quad (3.1)$$

It is worth noting here, that AHP has been also applied to image retrieval based on semantic representation of an image (Cheng et al. 2005; Chou and Cheng 2006) and to emotion recognition, see Cheng et al. (2007).

The authors of many experimental studies analyzing the processes of recognition divide the face onto obligatory parts. The examples may be the upper half and the lower half (Haig 1986), the areas of forehead, nose, and mouth (Kurach et al. 2014), eyebrows, eyes, nose, mouth, chin, hair (Matthews 1978), eyebrows, eyes, nose, mouth, and cheeks (Karczmarek et al. 2014), etc. Here, we are interested in a departure from this way of face partitioning. The main objective is to find the set of features being the most salient in practical problems.

The details of the method are as follows. First, the Analytic Hierarchy Process is utilized as described above to quantify the saliency of the cues. To get the result three experts from the fields of psychology and criminology are asked to estimate the cues according to their personal experience and knowledge.

The objective is to find the most salient features of a face in a collection of the features. The saliency is understood as the importance of a feature in a process of face recognition by humans. There are the following groups (high level features sets – compositions of features) of alternatives:

- (a) An information deduced from the observation of the whole face, e.g., gender, age, etc.
- (b) Eye area with a forehead;

- (c) The region of a nose and ears;
- (d) Mouth and chin region;
- (e) Hair, neck, and other external facial features.

Note that (a) – (d) represent internal features of a face.

The motivation behind this partition is that the witnesses of crimes or the subjects of psychological examinations describe facial images containing internal or external features of unknown criminals with usage of many appearance details. Note that during an examination of computational applications there are difficulties in comparisons of the images on a basis of hair or ears regions. Of course, it is difficult since the external features can vary or be hidden under hair.

In the series of our experiments three experts are incorporated to evaluate the facial features. Namely a police detective (with more than 30 years of experience) and two psychologists (with over 10 years of experience) are engaged. The pairwise comparisons to be done by them are carried to determine whether the feature a is preferred over b to the value from the range 1 (equal) to 9 (extreme preference). The results of the comparisons are inputs to the corresponding reciprocal matrices. The final result of AHP running is the normalized eigenvector associated with the maximal eigenvalue. Moreover, the inconsistency index and consistency ratio are the results. They give a detailed information (insight) into the features from the points (a) to (e). Next, AHP is runned for the components of the regions (a) – (e). At the end, AHP is used to produce the orders of features in 10 groups of internal parts of the face. The topology of the overall process is presented in Fig. 3.1.

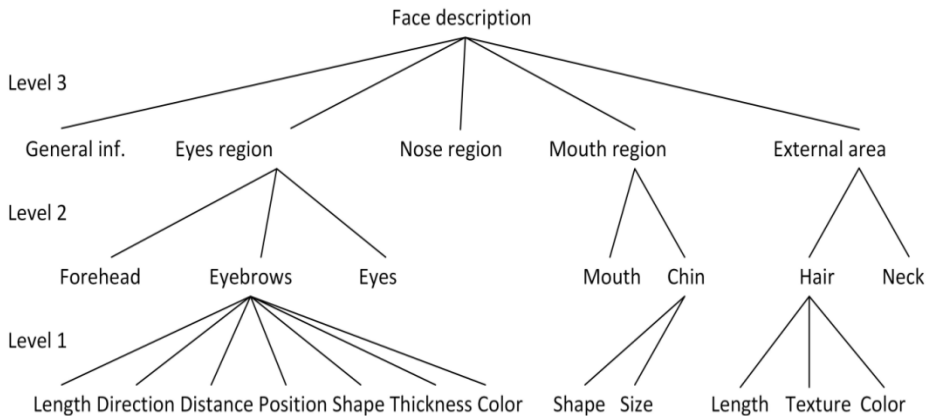


Fig. 3.1. Three levels of AHP hierarchy. Only chosen leaves are presented at the bottom level for the clarity of presentation. Similarly, only a few features at the middle level are represented at the image.

Table 3.1 The selection of features of a face with their linguistic descriptors.

Features	Linguistic descriptors
General information	
Shape of the face	rectangular, pentagonal, oval, round, triangular, ellipsoidal, trapezoidal, rhomboidal
Gender	female, male
Origin	Caucasian, Spanish, Asian, African, etc.
Age (estimated)	child, young adult, middle age adult, older adult
Hair	
Length	short, average, long
Texture	straight, wavy, curly
Color	light blonde, blonde, dark blonde, auburn, chestnut, black, red, turning gray, gray
Forehead area	
Width	low, average, high
Height	narrow, average, wide
Shape	rectangular, square, trapezoidal, inversely trapezoidal
Skin	smooth, creased, wrinkled
Eyebrows	
Length	short, average, long
Direction	horizontal, turned up, turned down
Distance between the eyebrows	merged, narrow, average, wide
Position	low, average, high
Shape	arched, straight, broken-lined, wavy, bushy
Thickness	narrow, average, wide
Color	light, average, dark
Eyes	
Shape of the lower eyelid	normal, thickened, saggy
Distance between eyelids	narrow, average, wide
Fissures length	short, average, long
Direction of the fissures	horizontal, turned up, turned down
Inter-eye distance	narrow, average, wide
Color	hazel, blue, green, gray
Nose	
Length	short, average, long
Width	narrow, average, wide
Width of the nasal bridge	narrow, average, wide
Shape of the nasal bridge	rectangular, trapezoidal, inversely trapezoidal
Shape of the nasal tip	rounded, spiked, blunt, angular
Nostrils	narrow, average, wide
Ears	
Protrusion	flat against head, average, protruding
Length	short, average, long
Cheeks	
Fullness	sunken, normal, filled

Length of the bones	short, average, long
Width of the bones	narrow, average, wide
Mouth	
Shape of the opening between lips	straight, concave, convex, wavy
Fullness	low, average, high
Width	short, average, long
Width of the philtrum	narrow, average, wide
Chin	
Shape	round, oval, angular, triangular, concave
Size	small, average, big

The data contained in Table 3.1 show that the atomic facial features have the properties more easily described in linguistic terms than the regions with a single feature and its neighborhood or grouped (collected) features. One can note that the table contains specific details, e.g., the eyebrows shape as well as the regions containing the groups of features which cover the importance of the possibility to affect the perception of a user in more general way. It cannot be seen in terms of typically physical measures such as length, width, and others.

It is worth noting that the facial description details are thoroughly presented in (FISWG 2016) or description sheets provided by the Chicago Police Department (2016). In the table contained are the attributes being, in our opinion, relatively easy to get from the two-dimensional pictures of the faces and, as well as, the most descriptive.

3.3. Observer’s Classification Process Confidence

A witness of a crime or an observer describes a suspect (a specific individual). The hierarchy developed in a previous section can be useful in an evaluation of the process of identification (classification). The idea comes from the entropy concept. Assume that the suspect evaluates the length of eyebrows using Analytic Hierarchy Process applied to the following attribute quantification: *short, average, long*. The question can be formulated as follows: To which extent do you prefer long eyebrows in describing someone’s (in this case, the suspect) eyebrows as opposed to long eyebrows? Moreover, we assume that the following values were obtained as the AHP process result: z_1, z_2, z_3 . They, obviously, correspond to the values of this specific subject’s length of eyebrows. On the basis of these values one can find the entropy of the attribute length of eyebrows, namely $H(\text{Length})$. The remaining entropies are $H(\text{Color})$, $H(\text{Position})$, $H(\text{Direction})$, $H(\text{Shape})$, $H(\text{Distance between the eyebrows})$, and $H(\text{Thickness})$. According to the Fig. 3.2, the eyebrows feature entropy can be yielded as

$$H(\text{Eyebrows}) = v_{color} H(\text{Color}) + v_{length} H(\text{Length}) + v_{direction} H(\text{Direction}) + v_{distance} H(\text{Distance between the eyebrows}) + v_{position} H(\text{Position}) + v_{shape} H(\text{Shape}) + v_{thickness} H(\text{Thickness}) \quad (3.2)$$

where v_{color} , v_{length} , $v_{direction}$, ... denote the weights of the attributes color, length, direction ..., respectively. For example, $H(\text{Length})$ can be found as

$$H(\text{Length}) = -(z_1 \log z_1 + z_2 \log z_2 + z_3 \log z_3) \quad (3.3)$$

Similar considerations can be repeated for the other facial features. Note that the values v_{color} , v_{length} , $v_{direction}$, $v_{distance}$, $v_{position}$, $v_{thickness}$, and v_{shape} stand for the assessments of the facial features which are not related to any specific individual.

Similarly, the entropies for specific areas of the face can be found at the second level of Analytic Hierarchy Process hierarchy. For instance, the entropy for the eyes area can be obtained as

$$H(\text{Eyes region}) = v_{eyebrows} H(\text{Eyebrows}) + v_{eyes} H(\text{Eyes}) + v_{forehead} H(\text{Forehead}) \quad (3.4)$$

Here, $v_{eyebrows}$, v_{eyes} , and $v_{forehead}$ stand for the weights found using Analytic Hierarchy Process for eyebrows, eyes, and forehead, respectively.

A total entropy for the whole face can be found yielded from the formula

$$H(\text{Face}) = v_{general} H(\text{General info}) + v_{eyes} H(\text{Eyes region}) + v_{nose} H(\text{Nose region}) + v_{mouth} H(\text{Mouth region}) + v_{external} H(\text{External area}) \quad (3.5)$$

Here $v_{general}$, v_{eyes} , v_{nose} , v_{mouth} , and $v_{external}$ are, as previously, the weights produced for the five features groups at the highest AHP level, see points (a) – (e). Note that the lowest confidence of the observer is obtained for the highest H_{total} value. If the value is too high (i.e., if it extends the assumed value) this fact may call for repeating the identification process or engaging another expert (observer).

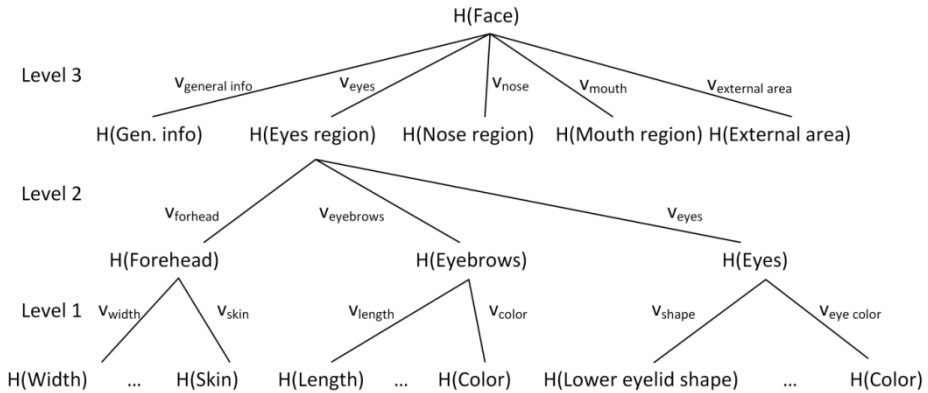


Fig. 3.2. The hierarchy of entropies related to the particular facial features and their weights.

3.4. Experimental Studies

We have asked each of the above mentioned $p=3$ experts to proceed the pairwise comparisons between the features of face. The experts are assumed to take into account their independent and subjective opinion about the saliency of the facial features in the process of classification (recognition). They evaluated these relations using Analytic Hierarchy Process method in the 10 sets of facial features. The experts filled the special questionnaires in a spreadsheet program. One may claim that the number of experts should be increased. However, the obtained results are intuitively appealing. Moreover, the level of consistency is satisfying. Similarly, the factors such as availability, position in a group, the heterogeneity level, experience, and others can play a pivotal role in a process of evaluation. The number of engaged experts (three) seems to be a well-established compromise between the agility of the pairwise comparisons process (their total number) and their preferences representativeness. Finally, if the number of experts would be too large, the opinions might be averaged and lead to the disappearance of differences between the experts' preferences obtained in the opinion aggregation process.

Let us describe the process of obtaining the features saliency. First, AHP is used at the first level. We consider the question to which extent texture is suitable in evaluating hair over length. Analogously, the experts have to compare hair texture and color as well as hair color and length. Example answers are detailed in Table 3.2. Similar questionnaires related to mouth region are listed in Table 3.3. Finally, a comparison of the main groups of facial features in a form of reciprocal matrix is presented in Table 3.4. The consistency ratio values regarding to the experts and the estimated features are depicted in Fig. 3.3.

The collected set of results is presented in Fig. 3.4. They are the average values of reciprocal matrices eigenvectors being the outputs of AHP. It is easy to observe that the most common and intuitive opinions are reflected in this setting, for example the most descriptive and salient facial areas are eyes, nose, and mouth regions. However, one of the surprises can be that the experts do not confirm the saliency of the eyebrows excluding eyes. The explanation can be that despite the studies on the computational facial recognition discuss it, in real life the people are focused on the eye region rather than on the eyebrows. Another interesting fact may be a relatively low position of an origin of a subject. It may be caused by the fact that the experts work and live in relatively homogeneous society. The experts establish the color as a very dominating feature in many cases (for instance, hair and eye colors). Unfortunately, this information may be difficult to utilize in the computational processes in the cases when the images are grayscale. However, if the model of color is RGB this fact may be very useful. Finally, note that the experts do not see the importance of the details such as the philtrum width. Nevertheless, in our opinion, the

presence of this kind of features in such rankings can be practically helpful when the set of images is of a good quality, for instance, when the resolution of images is high.

Next, we realize the AHP at the second level of hierarchy. The question to the experts is: To which extent a group of features *A* is regarded more salient than a group of features *B* in the face classification processes? Features *A* and *B* are general information, hair, and the areas of cheeks, chin, ears, eyebrows, eyes, mouth, and nose. Fig. 3.5 and Fig. 3.6 depict the results and associated consistency ratios, respectively. Hair and the area of eyes are, according to the experts' opinions, the most salient.

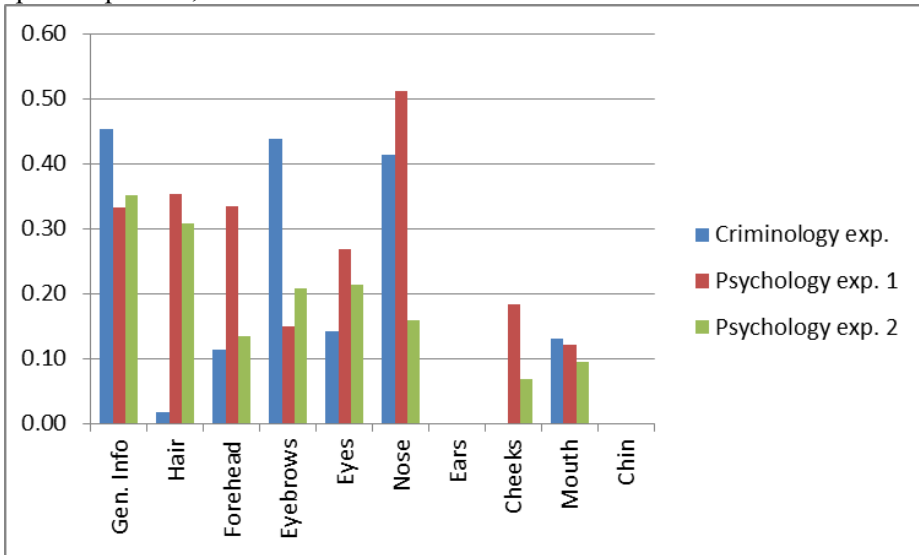


Fig. 3.3. Consistency ratios related to expert evaluations.

Table 3.2 Experts' reciprocal matrices regarding to hair feature.

Criminology expert			
Hair feature	Length	Texture	Color
Length	1	1	0.25
Texture	1	1	0.1(6)
Color	4	6	1
Psychology expert no. 1			
Hair feature	Length	Texture	Color
Length	1	9	9
Texture	0.(1)	1	0.1(6)
Color	0.(1)	6	1
Psychology expert no. 2			
Hair feature	Length	Texture	Color
Length	1	4	0.25
Texture	0.25	1	0.(3)
Color	4	3	1

Table 3.3 Experts' reciprocal matrices related to mouth area.

Criminology expert					
Mouth feature	Opening between lips shape	Fullness	Width	Philtrum width	
Opening between lips shape	1	6	4	6	
Fullness	0.1(6)	1	0.1(6)	0.3(3)	
Width	0.25	6	1	5	
Philtrum width	0.1(6)	3	0.2	1	
Psychology expert no. 1					
Mouth feature	Opening between lips shape	Fullness	Width	Philtrum width	
Opening between lips shape	1	0.2	0.(3)	3	
Fullness	5	1	4	4	
Width	3	0.25	1	3	
Philtrum width	0.(3)	0.25	0.(3)	1	
Psychology expert no. 2					
Mouth feature	Opening between lips shape	Fullness	Width	Philtrum width	
Opening between lips shape	1	0.(3)	0.25	3	
Fullness	3	1	0.25	4	
Width	4	4	1	5	
Philtrum width	0.(3)	0.25	0.2	1	

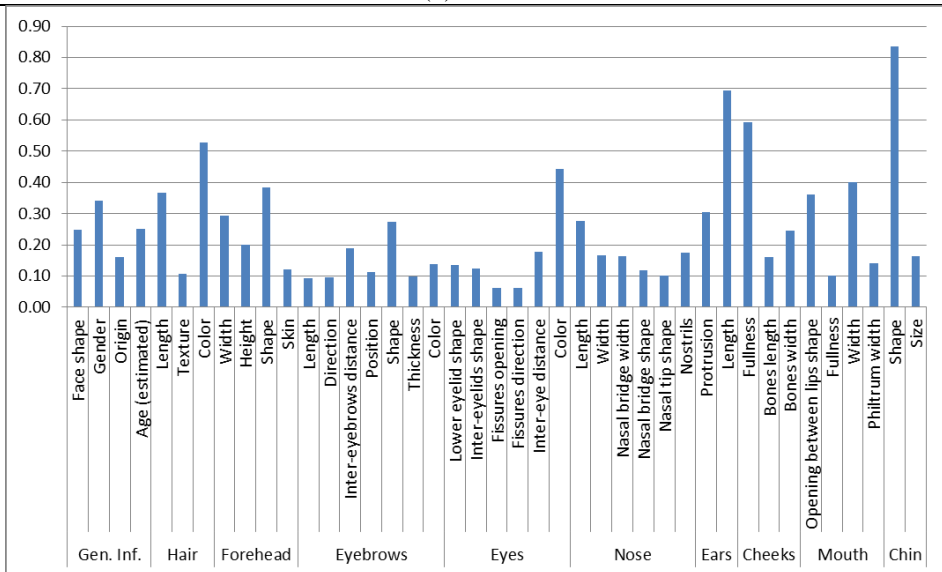


Fig. 3.4. Facial features saliency (averaged experts opinions) normalized to the sum per group (equal to 1).

Table 3.4 Experts' reciprocal matrices regarding to the general groups of features.

Criminology expert					
Main group	General information	Eyes & forehead	Nose & ears	Mouth & chin	External features
General inf.	1	0.25	0.1(6)	0.2	0.5
Eyes & forehead	4	1	3	3	7
Nose & ears	6	0.(3)	1	0.25	8
Mouth & chin	5	0.(3)	4	1	8
External features	2	0.14	0.125	0.125	1
Psychology expert no. 1					
Main group	General information	Eyes & forehead	Nose & ears	Mouth & chin	External features
General inf.	1	6	6	6	6
Eyes & forehead	0.1(6)	1	6	4	4
Nose & ears	0.1(6)	0.1(6)	1	0.25	3
Mouth & chin	0.1(6)	0.25	4	1	4
External features	0.1(6)	0.25	0.(3)	0.25	1
Psychology expert no. 2					
Main group	General information	Eyes & forehead	Nose & ears	Mouth & chin	External features
General inf.	1	0.25	0.25	0.5	1
Eyes & forehead	4	1	0.(3)	0.5	4
Nose & ears	4	3	1	5	4
Mouth & chin	2	2	0.2	1	4
External features	1	0.25	0.25	0.25	1

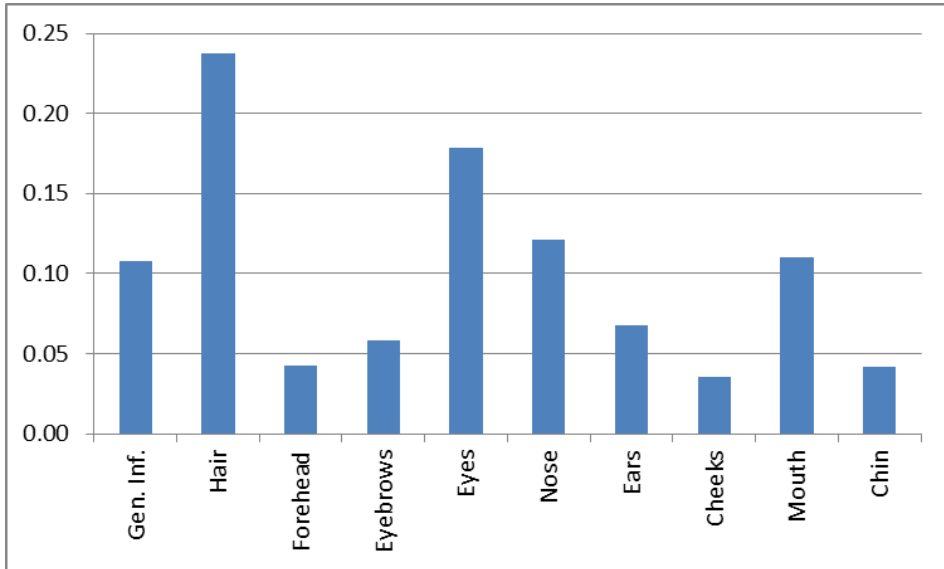


Fig. 3.5. The AHP second level facial features saliency.

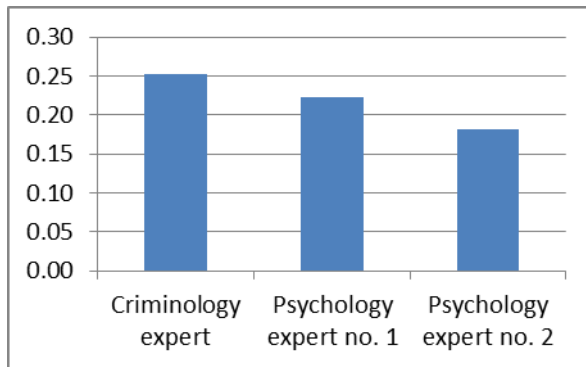


Fig. 3.6. Consistency ratios related to the second level of AHP.

Additionally, we have conducted another set of experiments and considered the more general group of facial features: General information, eyes and forehead, nose and ears, mouth and chin areas, and external features which constitute the third AHP level. Taking into account only the face, one can obtain the following dependencies: The highest importance is associated with the highest features, see Fig. 3.7 and Fig. 3.8 for the detailed results.

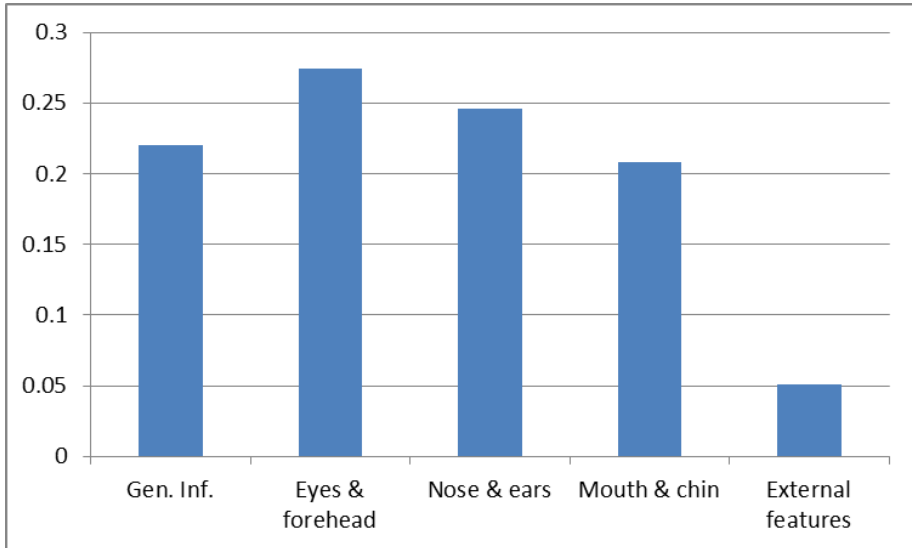


Fig. 3.7. AHP results for the most general facial features (third level of AHP).

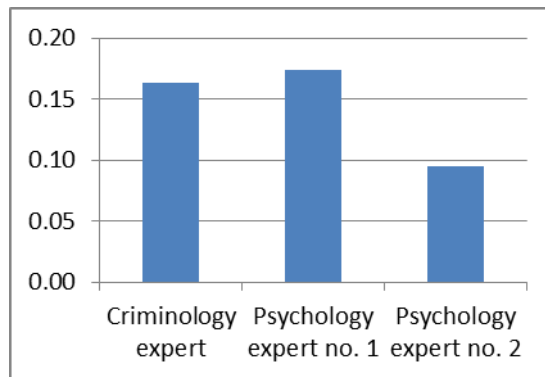


Fig. 3.8. Consistency ratios related to the most general facial features.

Let us now consider the confidence of the process of identification realized by a witness (observer) as described in the previous section. The experts evaluated six images of persons contained in the ColorFERET database. The example face images are presented in Fig. 3.9. These are the first six images of the database which were RGB and containing the persons with no glasses. The experts assessed the features coming from the forehead, eyebrows, and eyes regions. The questions were in the following form: To which extent the eyebrows length is regarded *long* versus *short*? The sum

$$H(\text{Eyes region}) = v_{\text{eyebrows}} H(\text{Eyebrows}) + v_{\text{eyes}} H(\text{Eyes}) + v_{\text{forehead}} H(\text{Forehead}) \quad (3.6)$$

returns the entropy for the eyes area and serves as an observation on expert confidence as the witness. Here, we consider the eye and forehead areas only to narrow the number of pairwise comparisons. However, this fact does not lead to the loss of generalization.



Fig. 3.9. Example images from the ColorFERET Dataset. In a text, we relate to the faces as person 1 (the first image in the top row), person 2, person 3, person 4, person 5, and person 6 (the last picture in the bottom row), respectively.

The experts' evaluations are listed in Table 3.5. Here, there are presented the average weights for the width and height of the forehead, length and thickness of the eyebrows, length and color of eyes. The winning values of the facial features are bolded. Table 3.6 consists of the average sums (i.e., the sums of entropies of particular experts divided by their number) of all considered entropies of the features for all the images. The conclusion from the data is that the features of the persons no. 1 and no. 2 are very difficult to estimate. On the other hand,

persons no. 5 and no. 6 are relatively easy to assess. A partial explanation of this fact may be that the forehead covered by hair and pose of the subject can lead to the difficulties in estimation. However, the data included in Table 3.6 can be helpful here. In case of 3 last persons the experts' opinions formed in AHP resulting vectors gave more than 50% certainty to have concrete linguistic values. The entropies determined when all the features were used are listed in Table 3.7. One can observe that the shape of face, between eyebrows distance, the shape of the eyebrows, and the eyes color are the most difficult to evaluate. Table 3.8 shows the average entropies yielded from the results of the descriptions of all the faces by each of the experts taking part in the series of experiments. There were two ways of calculations of the values. In the first case, all the entropies for all the features were summed and averaged. The second manner was based on the accumulation using the hierarchy introduced above and the weights obtained by the same experts. The explanation of the results is that evaluating the specific and concrete features of the face the most confidence appears in case of criminology expert (denoted as expert no. 1) and the first expert in the field of psychology (expert no. 2). From the other hand, considering the weights applied to abstract facial features we can note that the most confident is second expert from the field of psychology (expert no. 3). It leads to the conclusion that this expert found the best possible relationship between the abstract facial features.

Table 3.5 AHP results regarding to the chosen images of the face and features.

Feature	Value/Person number	1	2	3	4	5	6
Width of a forehead	Low	0.30	0.10	0.07	0.29	0.28	0.05
	Average	0.40	0.48	0.40	0.35	0.39	0.16
	High	0.30	0.42	0.52	0.36	0.33	0.79
Height of a forehead	Narrow	0.30	0.10	0.38	0.08	0.29	0.32
	Average	0.49	0.31	0.26	0.40	0.35	0.18
	Wide	0.13	0.61	0.45	0.28	0.57	0.76
Length of the eyebrows	Short	0.06	0.31	0.53	0.06	0.05	0.07
	Average	0.21	0.36	0.35	0.43	0.18	0.52
	Long	0.73	0.33	0.12	0.51	0.77	0.41
Thickness of the eyebrows	Narrow	0.17	0.10	0.11	0.07	0.05	0.08
	Average	0.56	0.58	0.32	0.42	0.19	0.54
	Wide	0.28	0.31	0.57	0.51	0.76	0.38
Length of the eyes	Short	0.07	0.13	0.33	0.10	0.09	0.33
	Average	0.55	0.35	0.27	0.54	0.61	0.52
	Long	0.38	0.52	0.39	0.37	0.30	0.14
Color of the eyes	Hazel	0.69	0.66	0.67	0.62	0.08	0.50
	Blue	0.06	0.06	0.04	0.05	0.66	0.05
	Green	0.09	0.09	0.09	0.10	0.08	0.20
	Gray	0.16	0.19	0.19	0.22	0.17	0.25

Table 3.6 The entropies related to the individuals. They are accumulated through all the facial features. The second row presents the number of AHP vectors containing the entries higher than 0.5.

Person number	1	2	3	4	5	6
Average sums of all features' entropies (no weights)	21.27	21.51	20.82	21.00	19.85	20.26
Number of AHP vectors' elements > 0.5	45	45	46	50	51	50

Table 3.7 The average entropies of all the features considered in the experiments.

	Facial feature	Entropy value
Forehead	Height	1.13
	Width	1.14
	Shape	1.51
	Skin	1.11
Eyebrows	Length	1.06
	Direction	1.06
	Between eyebrows distance	1.57
	Position	1.15
	Shape	1.84
	Thickness	1.13
	Color	1.10
Eyes	Lower eyelid shape	1.16
	Between eyelids distance	1.08
	Length of fissures	1.20
	Fissures direction	1.03
	Distance between eyes	1.10
	Color	1.40

Table 3.8 The values of entropies associated with the particular experts.

	Police expert	Psychology exp. 1	Psychology exp. 2
Average sum of all features' entropies (no weights)	20.51	20.58	21.25
Average entropies, see formula (3)	1.33	1.28	1.27

3.5. Classification Based on the Linguistic Descriptors

In the previous section we have proposed an application of Analytic Hierarchy Process to the estimation of importance of the facial features in the processes of face recognition realized by humans. Here, proposed is an approach to the face recognition problem realized both by humans as well as computers which is based on linguistic descriptors. Specifically, we apply Analytic Hierarchy Process to obtain linguistic values of the parts of face, particularly of the face features.

As mentioned above, we apply the Analytic Hierarchy Process to find the weights associated with specific features of the face and the membership degrees

of concrete facial features to the linguistic attributes. Therefore, we assume that we are able to extract the most important and descriptive features of the face and that they can be relatively easily assessed by people analyzing two-dimensional image of the face. The general idea of the method is that the experts taking part in an assessment of the features have to answer the questions of a form: To which extent the feature X is preferred over the feature Y ? We can extend this form of a question as: To which extent an attribute a is preferred over an attribute b of this feature? The result of this process is a normalized vector $\mathbf{w} = [w_1, \dots, w_N]$. It is consisted of N weights related with N facial features under consideration. Now, let us assume that a concrete face and its specific features are considered. All the features are described using linguistic descriptors. Hence, the vectors of the form $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N$ are produced. They describe each feature of the considered face. They can be easily concatenated into one vector. It describes the whole face: $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_N]$. Denote the whole set of descriptions of the images by Σ . The goal is to classify a face coming from this set as belonging or not belonging to one of the classes of faces in Σ . Let us assume that this new face is described by a vector \mathbf{f}_{new} . The classification process can be realized by the NN-classifier by finding a minimal distance between \mathbf{f}_{new} and the vectors \mathbf{f} coming from Σ . For example, let us analyze the feature length of the eyebrows and assume that the faces were evaluated by an expert and the expert's answers with regard to the eyebrow length were: *short-middle*: 1 – 5, *short-long*: 1 – 9, and *middle-long*: 1 – 7. Then

the AHP reciprocal matrix is
$$\begin{bmatrix} 1 & 0.2 & 0.14 \\ 5 & 1 & 1 \\ 9 & 7 & 1 \end{bmatrix}$$
. Then the eigenvector associated

to its maximal eigenvalue is $[0.055, 0.173, 0.772]$. The entries correspond to the linguistic values *short*, *middle*, and *long*, respectively. This example shows that the example eyebrow is rather long than middle or short. In this manner we can build the vector describing the whole face on a basis of concatenation of the vectors corresponding to particular facial features. This method corresponds to the psychological studies (LaVergne et al. 2016) suggesting that people have difficulties with evaluation of physical attributes of humans and can be a sound alternative to the approaches based on the direct evaluations.

3.6. Linguistic and Numeric Information Fusion

In the previous section, we have discussed a manner of including the linguistic terms being the outputs of the expert's opinion. Such kind of vectors can be supplemented by numerical values. These values can come directly from the geometrical relations appearing between the specific parts of a face. Similarly to the linguistic descriptions of the faces contained in an image set, one can determine the membership degrees of the measurable features' numeric values such as eye length to the linguistic attributes *short*, *middle*, *long*, etc. One

of the techniques appearing here can be the well-known K -means method (Hartigan & Wong 1979). Now, we will discuss two particular methods based on K -means and membership functions. Note that other methods can be applied here as well.

First, let us discuss the first of the proposed methods, namely the one based on K -means and normalization of the lengths of features. Consider all the N possible (available) facial features. One can divide them into two subsets of features: measurable and non-measurable. The first class can be characterized by the fact that they can be measured in terms of width or height. Also, they can be linearly ordered, e.g., *short*, *medium*, and *long*. The second class is built of the features such as the shape of the face and others. The specific values of the measurable facial features can be considered as input data which can be clustered by Fuzzy C -means (Bezdek et al. 1984) or K -means methods. Hence, a sound alternative here can be an application of K -means for a clustering of the analyzed dataset into three groups, which correspond to the descriptors *short*, *medium*, *long* or *small*, *average*, *long* with respect to the M measurable features separately.

This clustering conducted with respect to each separate feature allows us to more deeply analyze the crucial differences between the faces being under consideration and the result is N clusters, not three general and multidimensional ones. The clustering is based on data being the results of measuring the distances between the chosen points positioned on the facial images contained in an image database. The localization of example landmarks is presented in Fig. 3.10. For example, let us consider the feature forehead width. It can be found by the formula $(d(P_1, P_2) + d(P_3, P_4))/2$. Here, $P_k, k = 1, 2, \dots, 55$, are the coordinates of the k th point and, of course, they do not exhaust the set of possible landmarks to be used in the method. For instance, for the feature nose length the centers of clusters associated to the linguistic descriptors such as *short*, *medium*, and *long* can be designed. Next, the membership degree of each person to each cluster can be found. These clusters serve to describe the values of the nose length feature.



Fig. 3.10 Landmarks and their positions. The facial image comes from FERET dataset.

Once the lengths of M measurable features α_i^k ($i = 1, 2, \dots, M, k = 1, 2, \dots, m$) (m is the considered faces number), the results have to be normalized. Next, the obtained distances are scaled. It is done by setting the constant distance for all the pupils of the faces: $\alpha_i^{k*} = const \alpha_i^k$ where $const$ is a coefficient of scaling. They are clustered using K -means method. After normalization, the length of specific feature can be the starting point for examining the membership degree of every person to a cluster separately built for each feature. K -means returns the set of cluster centers. Precisely, for each of the measurable parts of faces, one can receive three numerical values which

correspond to the descriptors of the form: *small*, *average*, and *long*. The descriptions are contained in the vectors \mathbf{c}_k ($k = 1, 2, \dots, M$). Now, the membership degrees to the respective cluster centers can be found as follows.

Denote the vector containing the measurable features of the k th person as \mathbf{d}_k ($k = 1, 2, \dots, m$). The distances between each of the features to the centers \mathbf{c}_k can be found. It means that for each person $k = 1, 2, \dots, m$ the new vectors \mathbf{z}_i^k ($i = 1, 2, \dots, M$) elements can be obtained as $z_{i,j}^k = |d_{k,i} - c_{i,j}|$. Here, the index $j = 1, 2, 3$ corresponds to the descriptor *short*, *middle*, or *long*. To visualize the method consider the following example. If a specific length of the eyebrows is $d = 35$ and the centers vector obtained in clustering process is $\mathbf{z} = [25, 37, 47]$ then the distance vector is $\mathbf{z} = [10, 2, 12]$. Next, one can standardize the vectors \mathbf{z}_i^k , namely $z_{i,j}^{k*} = (r_i - z_{i,j}^k)/z_{i,j}^k$, where $r_i = \max_{1 \leq k \leq m} d_{k,i} - \min_{1 \leq k \leq m} d_{k,i}$ and $d_{k,i}$ means a dispersion of measurable feature no. i , see Fig. 3.11. In the example, if the spread for the considered feature is 48 then $\mathbf{z}^* = [3.8, 23, 3]$. The final result is the set of normalized to the value of 1 vectors \mathbf{z}_i^{k*} . They contain the membership degrees to the clusters related to specific linguistic values.

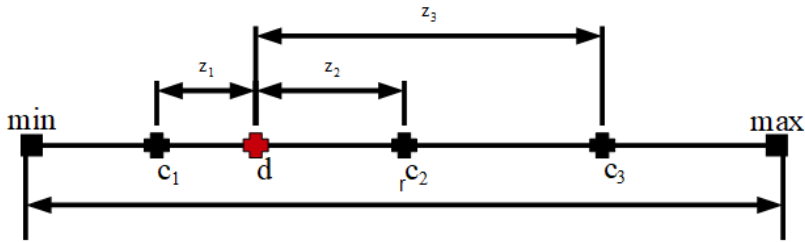


Fig. 3.11 The relations between feature lengths and cluster centers. For details, see the description in text.

Now, let us discuss the method of membership functions formation. Here, we focus triangular membership functions as an efficient way of describing (quantifying) the variables in linguistic form, namely *long*, *medium*, and *short*. The functions we consider are

$$A_1(x) = \begin{cases} 1 & \text{for } x \leq x_{min} \\ \frac{c-x}{c-x_{min}} & \text{for } x_{min} < x \leq c \\ 0 & \text{for } c < x \end{cases} \quad (3.7)$$

$$A_2(x) = \begin{cases} \frac{x-x_{min}}{c-x_{min}} & \text{for } x_{min} < x \leq c \\ \frac{x_{max}-x}{x_{max}-c} & \text{for } c < x \leq x_{max} \\ 0 & \text{for } x_{max} < x \end{cases} \quad (3.8)$$

$$A_3(x) = \begin{cases} 0 & \text{for } x \leq c \\ \frac{x-c}{x_{max}-c} & \text{for } c < x \leq x_{max} \\ 1 & \text{for } x_{max} < x \end{cases} \quad (3.9)$$

Here, $x_{min} = \min_{1 \leq k \leq m} x_k$, $x_{max} = \max_{1 \leq k \leq m} x_k$. This manner of feature description suggests its realization by means of three membership functions which overlaps. This overlap is 0.5. One can adjust the modal value c of the function $A_2(x)$ to get the flexibility at a certain level. This adjustment is realized on a basis of values of vectors f_j . The example of membership functions model is depicted in Fig. 3.12. For each of M facial features which are measurable we have to minimize the sum

$$\sum_{j=1}^n \sum_{k=1}^m \sum_{l=1}^3 \left(A_l(x_k) - f_{k,l}^{(j)} \right)^2 \quad (3.10)$$

Here, n is the experts count or more precisely, it denotes the number of Analytic Hierarchy Processes conducted for this particular feature, m denotes a number of facial images being under examination, while $f_{k,l}^{(j)}$ stands for the entries of vectors related to the k th face. Finally, the feature's length is x_k . Note that other types of memberships such as, for instance, Gaussian membership functions can be also used here.

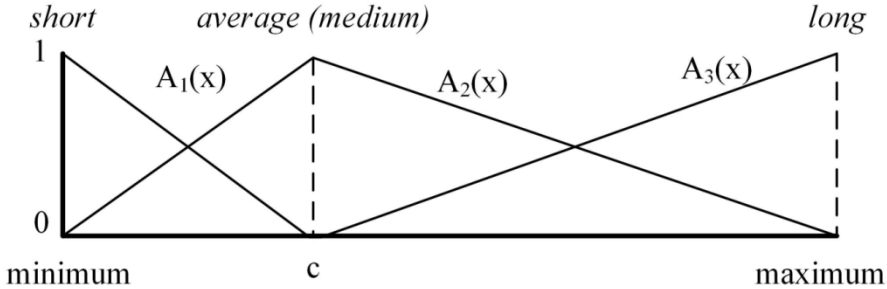


Fig. 3.12 Examples of fuzzy membership functions.

3.7. The Process of Classification

The main processing flow is depicted in Fig. 3.13. One expert or a group of experts describe a face by evaluating the features using Analytic Hierarchy Process. The results of particular specific features evaluations are then concatenated into the vectors which represent activation levels of facial descriptors. These descriptors form a linguistic space. They can be easily averaged on a basis of arithmetic mean. Parallely, the measures of facial features are kept in a form of the vectors of membership values of linguistic terms *short*, *medium*, and *long*. These vectors are the input to the classification process realized by an intuitively appealing classifier which is Nearest Neighbor with a weighted Euclidean distance function in the following form:

$$d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \sqrt{w_i} (x_i - y_i)^2 \quad (3.11)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_n]$, $\mathbf{x} = [x_1, x_2, \dots, x_n]$, $\mathbf{y} = [y_1, y_2, \dots, y_n]$. Note that the form of a square root of the i th weight was determined in a series of experiments by trial and error.

Let us consider that p experts take part in the AHP process. Then p weight vectors are obtained, namely $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$. They describe the saliency (importance) of face features. It means that $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \dots, w_{i,N}]$, $i = 1, 2, \dots, p$. Now, one can reform the vectors of weights and rebuild the vectors of the form $\mathbf{v}_i = [v_{i,1}, v_{i,2}, \dots, v_{i,Q}]$, $i = 1, 2, \dots, p$. Q denotes the sum of all the linguistic values which correspond to the N features of a face. Note that this kind of building weight vectors means that, for example, the eyebrow length weight vector associates with vector elements \mathbf{v}_i which corresponds to *short*, *medium*, and *long* linguistic values. Moreover, note that $v = \frac{(v_1 + v_2 + \dots + v_p)}{p}$, i.e., they are averaged.

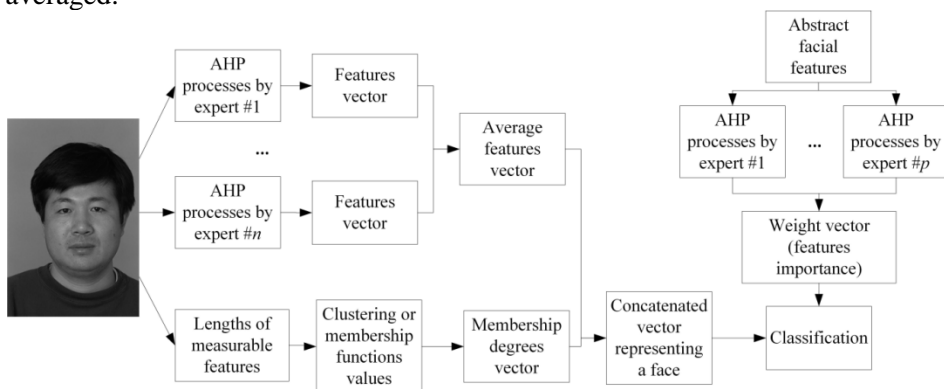


Fig. 3.13 An overall scheme of the process of classification.

3.8. Experimental Studies

The series of experiments was carried with using the well-known FERET dataset. To show the efficiency of our proposal we work with the first 50 facial images coming from the subset called *ba* and the first 50 images from the subset called *bk*. We denote the first set as A and the second one as B (i.e., training and testing set, respectively). Three experts (our friends or lab members) were to analyze the facial images coming from the set A , while three experts were asked to work with the images being in the set B . Two experts filled the questionnaires related to the set A and set B . To be precise, they had to play a role of a witness of a crime who describes the images using AHP method in a way proposed above. The images were evaluated using specially prepared forms in a spreadsheet program. As a result, three questionnaires for sets A and B were

obtained. Additionally, an experienced face recognition expert (our lab member) described $N = 52$ of the most descriptive, in our opinion, features which come with linguistic descriptors. To choose the features we used the standards explained in Chicago Police Department (2016), Czerw (1995), FISWG (2016). The features which are non-measurable were chosen as follows (alphabetically): cheeks fullness, chin details, chin prominence, chin shape, depth of the philtrum, direction of the fissures, eyebrows direction, eyebrows shape, earlobe shape, ears protrusion, ears shape, forehead profiling, forehead shape, forehead skin, gender, hair length, hair texture, hairstyle, mouth fullness, origin, placement of eyes shallow, position of the earlobe, shape of the face, shape of the lower eyelid, shape of the nasal bridge, shape of the nasal tip, shape of the opening between lips, size of the earlobe while the measurable features are: chin size, distance between eyelids, distance between the eyebrows, ears length, eyebrows length, eyebrows position, eyebrows thickness, fissures length, forehead height, forehead width, height of the nasal bridge, inter-eye distance, length of the cheeks, lower lip height, mouth width, nose length, nose width, nostrils, size of the nose holes, upper lip height, width of the cheek bones, width of the nasal bridge, width of the philtrum. Table 3.9 contains the experiments' results. They show that AHP is a useful tool here. In particular, if more than one expert evaluates the images, a good recognition rate level is observed, i.e., more than 90%. A detailed analysis of the results shows that the methods of assessing the membership degrees to related linguistic values represented by clusters or membership functions may successfully supply the process of description proceeded by the experts. The fusion of information (in the form of vectors) coming from the measurements of the lengths of features and linguistic descriptors obtained from the experts strongly improves the classification algorithm accuracy. Particularly, in the case when the images are assessed by a single expert, 94% of subjects are correctly classified. When more experts are involved the efficiency of our proposal improves. Two experts participation in the process of face evaluation seems to be relatively inexpensive option. In the case when *K*-means is utilized to construct an augmented vector of features, good identification accuracy is reported for only two experts involved in the evaluation of the training set and one expert describing the testing images (more than 97% of recognition rate). Moreover, an application of weights generated in the AHP method improves the performance of the method at the level of 6 percentage points. In addition, an application of the well-known PSO optimization method (see Kennedy et al. 2001, Kacprzyk & Pedrycz 2015, and introductory chapter) to the experts' answers regarding to the specific facial features as well as to the abstract relations between the features (weight generation of the features) leads to an improvement of the classification results up to 1.5% recognition rate level. Note that here, the termination criterion of the PSO was that the inconsistency index should be less or equal 0.1. The results are denoted in Table 3.9 by AHP & PSO and AHP & PSO & *K*-means. The numbers

of PSO particles and generations were 30 and 300, respectively. Finally, we compare the proposed method with other approaches, particularly with local descriptors-based methods (LBP, Local Binary Patterns, see Ahonen et al. 2004, and MBLBP, Multi-scale Block Local Binary Pattern, see Chan et al. 2007 and Liao et al. 2007). The setup for the series of numerical experiments was as follows: Each image in the dataset was divided into $n \times n$ subrectangles. The best recognition rate was obtained for $n = 6$. The sizes of MBLBP blocks were 3, 5, and 7 pixels, respectively. Our method outperforms the LBP-based approaches. Moreover, we compare our proposal with other linguistic descriptors-based approaches such as AHP with no distances information and the classifier based on neural networks (Dolecki et al. 2016), voting on the chosen lengths of features (see the next part of this chapter), and fuzzy sets obtained directly from the weights given by users (Kiersztyn et al. 2016).

3.9. Voting and Linguistic Descriptors for Face Recognition

Here, the linguistic descriptors obtained directly from an expert or a group experts treated as the votes when evaluating face are considered as an input to face recognition processes.

We develop a feature space which is linguistic and we assess its capabilities of discrimination. Assume that there are considered n features and that we have their quantification denoted by linguistic labels $c_i, i = 1, 2, \dots, n$. Then, we dispose a total number $\sum_{i=1}^n c_i$ of 0-1 element Boolean vectors. In contrast, the feature space with numeric entries is built of n -dimensional vectors. If more than one expert evaluates the faces, the Boolean vector can be substituted by probabilities of occurrences of some linguistic labels. Consider a face and selected n particular features (descriptors) f_1, f_2, \dots, f_n such as length of eye. Each of descriptors can be quantified using small number of fuzzy sets (or, more generally, granular values) such as *small*, *medium*, *large* and others. Such descriptors can be concatenated into a single vector describing the face, say $f = [f_1, f_2, \dots, f_n]$. Denote the whole collection of facial images by Σ . The goal is to identify any facial image as belonging or not belonging to the set of faces contained in Σ . As in previous section, the face image can be described as g . The identification can be realized using nearest neighbor classifier by finding a minimal distance between all the face images $f \in \Sigma$ and g . For instance, consider the feature length of nose and assume that ten experts assess this feature as follows: Three of them evaluated someone's nose as *short*, five of them described it as *middle*, and two experts said it is *long*. In this way, the membership values vector $f = [0.3, 0.5, 0.2]$. Moreover, if all the features are evaluated, the face contained in the image is described in a form of the vector of a dimension $\sum_{i=1}^n c_i$. This vector is a result of concatenation of n vectors built similarly to f . As in the previous section, the weights of these features can be

obtained in the pairwise comparison process. The main flow of the whole process of classification is presented in Fig. 3.14. The k experts assess abstract facial cues using AHP. The results of their consideration, after an aggregation process, constitute a weight vector to be used in a classification. The compositions of the vectors coming from abstract and specific features of the faces are compared using NN classifier.

Table 3.9 Recognition rates.

Experts count in sets A (train.) and B (test.)	1/1	1/2	2/2	3/1	3/2	3/3
AHP	74	90	96.67	96	100	100
AHP & weights	72. 89	91.33	96.44	96	100	100
AHP & K -means	94.22	97.33	99.33	98.67	100	100
AHP & K -means & weights	94.67	97.11	99.11	99.33	100	100
AHP & Particle Swarm Optimization	75.33	90	96.44	96.67	100	100
AHP & Particle Swarm Optimization & weights	74.44	90.44	95.78	96	100	100
AHP & PSO & K -means	94.44	97.56	99.33	99.33	100	100
AHP & PSO & K -means & weights	94.67	97.78	99.11	99.33	100	100
AHP & triangular membership functions	87.56	93.11	96.22	95.33	98	100
AHP & triangular MF & weights	87.56	93.56	96.67	97.33	99.33	100
AHP & K -means & triangular MF	94	97.78	99.33	98.67	100	98
AHP & K -means & triangular MF & weights	94.22	96.67	98.44	98.67	100	100
AHP & normalized lengths	77.33	91.78	96.89	96.67	100	100
AHP & normalized lengths & weights	78.44	92.44	96.67	96.67	100	100
Other methods based on linguistic descriptors						
AHP & neural networks (Dolecki et al. 2016)	96		95.3			98
Voting	38.94	48.13	62.75	53.79	69.41	78.18
Fuzzy measures (Kiersztyn et al. 2016)	20.8	29.3	43.6	35.3	53.5	67.2
Example approaches with no attendance of experts						
Lengths				66		
Lengths & weights				68		
K -means				56		
K -means & weights				54		
Triangular membership functions				58		
Triangular membership functions & weights				64		
LBP (6×6)				88		
MB-LBP (7 px square block, 6×6)				88		
MB-LBP (3 px, 6×6)				84		
MB-LBP ($5, 6 \times 6$)				84		

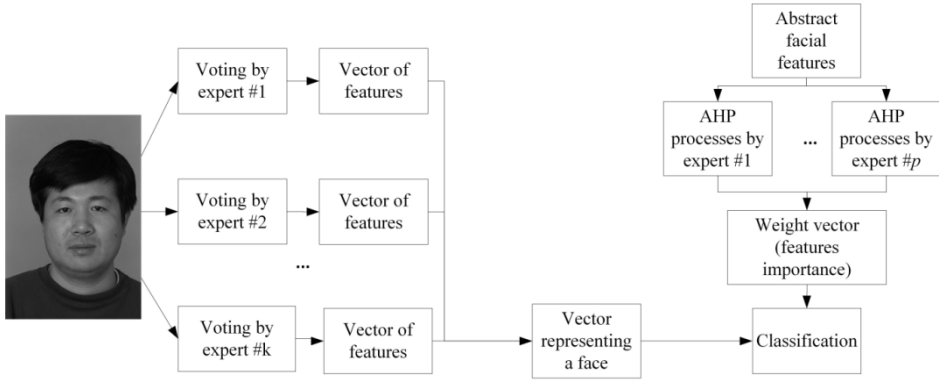


Fig. 3.14 An overall processing scheme.

If we assume that p experts are involved in the AHP process of evaluation of n facial cues we obtain p weight vectors: $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$ associated with the saliency of these facial cues having the following form $\mathbf{f} = [w_{i,1}, w_{i,2}, \dots, w_{i,n}]$, $i = 1, 2, \dots, p$. Analogously, the concatenated weight vectors produce $\sum_{i=1}^n c_i$ -element vectors denoted as $\bar{\mathbf{w}}_i$. These weights can be used (in dependence of needs) in the present form or in their inverted form:

$$\bar{\mathbf{v}}_i = \max[\bar{w}_1, \bar{w}_2, \dots, \bar{w}_n] - \bar{w}_i \quad (3.12)$$

NN classifier is realized using the distance or similarity functions such as Euclidean, Manhattan, cosine, correlation, modified Euclidean of the form

$$d(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{j=1}^n x_j^2 \sum_{k=1}^n y_k^2} \quad (3.13)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_n]$, $\mathbf{y} = [y_1, y_2, \dots, y_n]$ are non-zero vectors. Moreover, the classification can be realized using weighted versions of the above distances (i.e., Euclidean, Manhattan, squared Euclidean, and modified Euclidean).

3.10. Experimental Results

To show the efficiency of the method we used the same set of images A (training set) and B (testing set) as in the previous section. Next, seventeen people being our lab members and friends were asked to assess the images in the above presented feature space. Each of them described either 50 images contained in the set A or the same number of images included in the set B . The experts used special application to fill the questionnaires consisted of Boolean values associated to the facial linguistic descriptors. The number of the features was narrowed to $n = 27$ which are, in our opinion, quite easy to be estimated by experts. The features and they possible values are collected in Table 3.10.

Table 3.10 Linguistic descriptors of facial features.

Cue	Linguistic descriptors
Chin shape	round, oval, angular, triangular, concave
Chin size	small, average, big
Distance between eyelids	narrow, average, wide
Distance between eyes	narrow, average, wide
Ears protrusion	fitting, average, protruding
Eye length	short, average, long
Eyebrows direction	horizontal, turned up, turned down
Eyebrows length	short, average, long
Eyebrows position	low, average, high
Eyebrows shape	arched, straight, broken-lined, wavy, bushy
Eyebrows thickness	narrow, average, wide
Facial shape	rectangular, pentagonal, oval, round, triangular, ellipsoidal, trapezoidal, rhomboidal
Fissures direction	horizontal, turned up, turned down
Forehead height	low, average, high
Forehead shape	rectangular, square, trapezoidal, inversely trapezoidal
Forehead width	narrow, average, wide
Gender	female, male
Inter-eyebrows distance	merged, narrow, average, wide
Lower lip height	low, average, high
Lower eyelid shape	normal, average, saggy
Mouth width	short, average, long
Nasal bridge length	narrow, average, wide
Nasal tip shape	rounded, spiked, blunt, angular
Nose length	short, average, long
Nose width	narrow, average, wide
Origin	Caucasian, Spanish, Asian, African
Upper lip height	low, average, high

In addition, 4 experts (our lab members or friends) estimated the above-mentioned 27 features (abstract, not concrete) using the Analytic Hierarchy Process to obtain the weights being the averages of their individual estimations. The values are depicted in Fig. 3.15. Note that the weights are relatively intuitively appealing because of the fact that the confidence of the experts' evaluations is at a high level. The exception is the case of head and forehead shapes, origin, and gender. The source of this fact can be that some experts put the weight on the importance of the features in the automatic face recognition processes while other experts put the weight on the features related with face recognition realized by humans, particularly related with feelings.

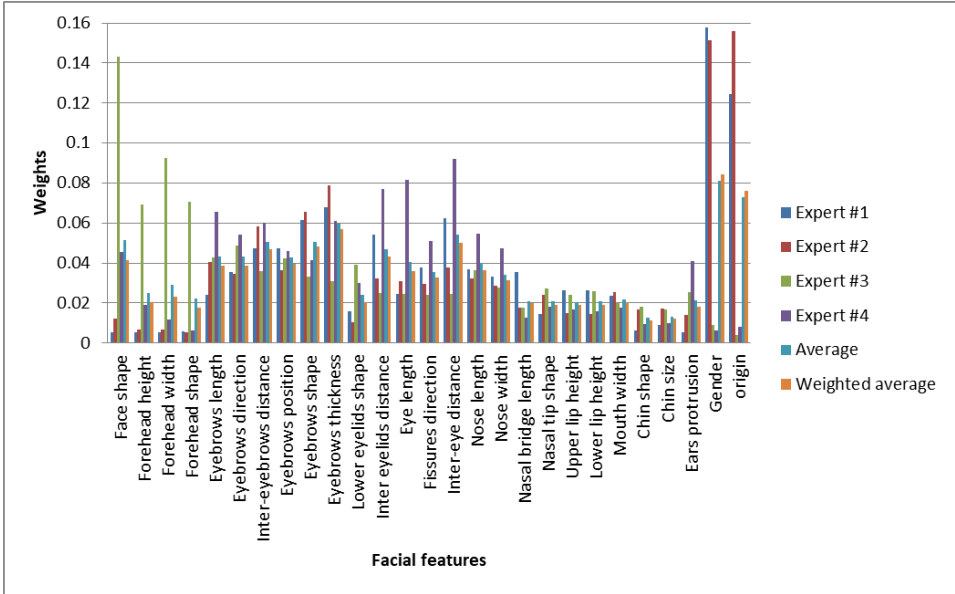


Fig. 3.15 Average values of facial features weights obtained by experts.

In the series of experiments the tests for all the combinations of three, five, and seven experts estimated the training and testing set *A* and *B*, respectively, were run. The results presenting the accuracy of classification based on majority voting were presented in Table 3.11. One can observe that the best distance function is the Manhattan distance. Moreover, the weights obtained on the basis of experts' knowledge with regards to the importance of facial features can improve the recognition rates, specifically for the lower ranks they reach 100% level. Rank-1 recognition rates for the chosen classifiers such as voting, Manhattan distance-based Nearest Neighbor, modified Euclidean distance-based NN, and weighted modified Euclidean distance-based NN are enlisted in Table 3.12. Presented are the average values for all the combinations of experts. An intuitively appealing fact is that the more experts is involved in the estimation of the facial features process, the recognition rate is closer to 100%. In addition, as before, one can observe that the weights improve the accuracy of the classification methods, particularly when the lower number of experts is engaged in the experiments. The justification is that the weights fill the lack of the data coming from the estimation proceeded by the experts. Adding the weights to the modified Euclidean measure can improve the average recognition result by 1%. If more than one expert estimates training set and only one expert estimates testing set, the results are better by even 5%. In case of six or more experts, the results obtained with weighted modified distance are worse by ca. 1%. Interesting is fact that the improvement of classification rate for one expert

estimating the training set increases when the number of experts estimating the testing set grows up to 3.55%.

It is noteworthy that considering only the space of features which are numeric, namely such that they can be sorted in a linear order (i.e., *short*, *medium*, and *long*) and if the weighted squared Euclidean distance is applied to their lengths after normalization, i.e., 1 is equal to the longest cue in the set, the accuracy reaches 54%, see Fig. 3.16. Here, the experts' presence can effectively increase the recognition rate.

Table 3.11 The recognition rates obtained with majority voting and the method based on comparison of concatenated vectors using various distance functions.

Three experts estimating training and testing sets										
Rank	Voting	Eucl.	Manh.	Cosine	Correl.	Mod. Euclid.	W. Euclid.	W. Manh.	W. sq. Euclid.	W. mod. Eucl.
1	56.03	77.96	78.18	77.85	77.70	77.61	76.90	76.64	78.24	78.87
5	85.82	96.97	97.00	97.01	96.90	96.81	97.66	97.20	97.99	97.84
10	93.93	99.21	99.23	99.25	99.20	99.16	99.63	99.46	99.74	99.65
Rank number for which 100% rec. rate is reached	48	39	36	36	38	42	32	28	25	27
Five experts estimating training and testing sets										
1	71.40	92.36	92.65	92.77	92.64	92.85	91.60	91.28	92.42	92.99
5	93.13	99.86	99.84	99.87	99.86	99.86	99.88	99.85	99.94	99.91
10	97.68	99.99	99.99	99.99	99.99	99.99	100	99.99	100	100
Rank number for which 100% rec. rate is reached	47	17	15	17	18	19	11	20	11	11
Seven experts estimating training and testing sets										
1	79.38	97.40	97.80	97.21	97.20	97.40	96.72	97.42	97.28	97.18
5	96.43	100	100	100	100	100	100	100	100	100
10	99.33	100	100	100	100	100	100	100	100	100
Rank number for which 100% rec. rate is reached	21	3	4	3	3	4	4	4	4	3

Table 3.12 Rank 1 recognition rates for various methods of classification.

Number of experts estimating training (vertically) and testing (horizontally) set								
Voting								
	1	2	3	4	5	6	7	8
1	38.94	36.52	46.16	45.46	50.62	49.10	53.14	52.67
2	42.23	47.55	50.58	55.57	56.36	57.99	58.45	58.94
3	46.26	44.18	56.03	55.77	61.79	60.13	64.65	62.95
4	51.38	53.95	61.60	65.63	68.60	69.62	71.85	71.95
5	52.92	51.25	64.05	65.02	71.40	70.60	74.55	74.21
6	55.07	56.56	66.25	69.95	74.10	75.08	77.64	77.88
7	55.93	54.92	68.33	69.54	75.87	75.55	79.38	77.50
8	58.06	58.46	70.05	73.24	77.87	78.95	81.81	80.44
9	56.00	55.86	70.46	71.26	78.68	78.21	82.75	78.00
Manhattan distance								
	1	2	3	4	5	6	7	8
1	38.94	48.13	53.79	57.47	59.97	61.90	63.00	62.67
2	47.21	62.75	69.41	74.42	76.93	79.23	80.39	82.56
3	52.31	69.49	78.18	82.38	85.33	87.70	88.61	89.69
4	55.78	75.06	82.51	87.41	89.83	91.81	92.76	93.79
5	58.25	77.58	85.78	89.90	92.65	94.03	95.13	95.94
6	59.99	80.52	88.23	92.05	94.10	95.76	96.50	97.14
7	61.40	81.59	89.28	93.02	95.32	96.61	97.80	98.06
8	62.58	83.24	90.40	94.16	96.19	97.77	98.31	99.11
9	63.50	83.57	91.43	94.51	96.71	98.64	99.75	100.00
Modified Euclidean distance								
	1	2	3	4	5	6	7	8
1	38.94	48.13	53.79	57.47	59.97	61.90	63.00	62.67
2	44.38	60.38	69.06	74.35	77.88	80.20	81.94	83.83
3	50.13	68.48	77.61	82.87	86.11	88.39	89.99	91.29
4	54.34	73.36	82.45	87.44	90.44	92.42	93.81	94.71
5	57.08	76.52	85.48	90.11	92.85	94.60	95.73	96.35
6	59.12	78.65	87.53	91.83	94.29	95.86	96.82	97.55
7	60.67	80.21	88.94	92.96	95.23	96.69	97.40	97.89
8	62.39	81.24	89.92	93.83	95.77	96.95	97.47	97.78
9	62.25	82.43	90.50	94.51	96.21	97.07	97.50	98.00
Weighted modified Euclidean distance								
	1	2	3	4	5	6	7	8
1	38.25	49.26	55.06	59.01	61.66	63.35	64.69	66.22
2	49.30	63.70	71.06	75.60	78.43	80.35	81.70	83.00
3	55.54	71.08	78.87	83.41	86.36	88.30	89.70	90.88
4	59.16	75.57	83.41	87.82	90.50	92.39	93.62	94.57
5	61.61	78.45	86.33	90.44	92.99	94.62	95.77	96.51
6	63.25	80.52	88.30	92.19	94.44	95.89	96.69	97.31
7	64.67	81.96	89.78	93.40	95.36	96.57	97.18	97.44
8	65.83	82.95	90.75	94.25	95.98	96.93	97.33	97.56
9	67.50	83.93	91.46	94.80	96.57	97.36	97.00	98.00

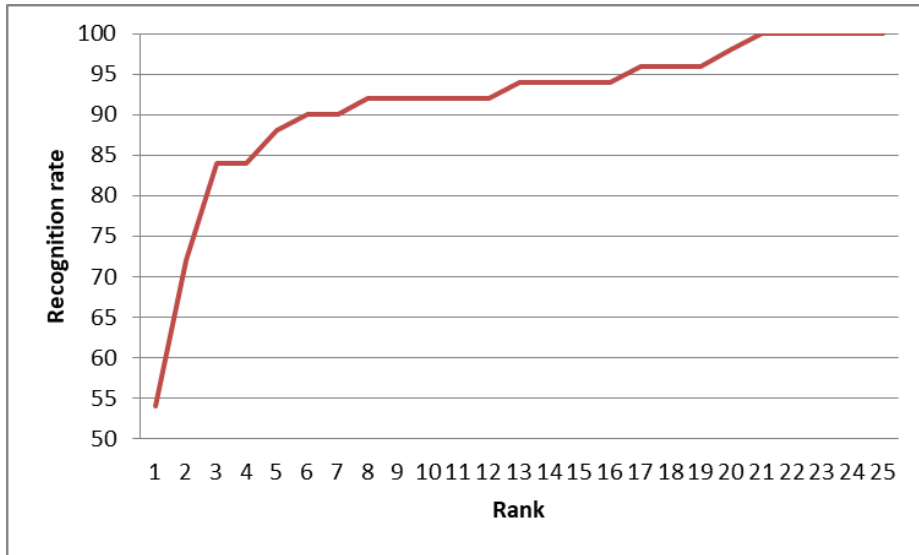


Fig. 3.16 Recognition rates regarding to rank 1 – rank 25 (a method based on the normalized lengths of features).

3.11. Conclusions

In this chapter, we have introduced a method of evaluation of the importance of face features in the process of face recognition realized by humans. Our approach is based on the knowledge and opinions of the experts. Moreover, the method can be easily adjusted in the computational face recognition methods depending upon the practical needs. We have proposed the Analytic Hierarchy Process method as a generic approach of obtaining the weights of the essential facial features to be used by humans describing faces. The AHP structure is built of three levels of hierarchy. It has been developed along with the method of finding the relevance and confidence of the assessment of subjects which is based on the entropy measure. The novelty and originality of the approach lays in the fact that linguistic values (and not numerical) were used to generate intuitively appealing and interesting results. Moreover, the set of results of the pairwise comparisons conducted by experts can be an invaluable and interesting example to other studies.

Furthermore, we have discussed a novel and original approach to the facial recognition realized by humans with an application of the Analytic Hierarchy Process. This method can be an efficient tool to realize the process of evaluation of faces by an expert or a group of experts. Very good results were obtained in a series of experiments, namely the recognition rates varied from 94% to 100%. It means that the AHP-based method can be an important vehicle to improve the

processes where the presence of experts is very important (or necessary, e.g., in forensic science).

Finally, we have thoroughly studied an original combination of AHP and human face images linguistic descriptors. These descriptions were received from the experts' evaluations. Such kind of approach can work as a successful complement to other classifying methods.

References

Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with Local Binary Patterns. In: Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science 3021, pp. 469–481 (2004)

Alattab, A. A., Kareem, S. A.: Efficient method of visual feature extraction for facial image detection and retrieval. In: Computational Intelligence, Modeling and Simulation (CIMSIM), 2012 Fourth International Conference on, pp. 220–225 (2012a)

Alattab, A. A., Kareem, S. A.: Facial image retrieval based on Eigenfaces and semantic features. *UACEE Int. J. Artif. Intell. Neural Netw.* **2** (3), 15–19 (2012b)

Anderson, D. T., Scott, G. J., Islam, M., Murray, B., Marcum, R.: Fuzzy Choquet integration of deep convolutional neural networks for remote sensing. In: Pedrycz, W., Chen, S. M. (Eds.): Computational Intelligence for Pattern Recognition. Studies in Computational Intelligence, vol. 777. Springer, Cham, pp. 1–28 (2018)

Arca, S., Campadelli, P., Lanzarotti, R., Lipori, G., Cervelli, F., Mattei, A.: Improving automatic face recognition with user interaction. *J. Forensic Sci.* **57** (3), 765–771 (2011)

Benhidour H., Onisawa, T.: Drawing human faces using words and conceptual fuzzy sets. *Systems, Man and Cybernetics*, 2007. ISIC. IEEE International Conference on, pp. 2068–2073 (2007)

Benhidour, H., Onisawa, T.: Interactive face generation from verbal description using conceptual fuzzy sets. *J. Multimed.* **3** (2), 52–59 (2008)

Bezdek, J. C., Ehrlich, R., Full, W.: FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **10** (2–3), 191–203 (1984)

Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern histograms for face recognition. In: Lee, S.-W., Li, S. Z. (Eds.): ICB 2007, Lecture Notes in Computer Science **4642**, pp. 809–818 (2007)

Cheng, S. C., Chen, M. Y., Chang, H.-Y., Chou, T. C.: Semantic-based facial expression recognition using analytical hierarchy process. *Expert Syst. Appl.* **33**, 86–95 (2007)

- Cheng, S. C., Chou, T. C., Yang, C. L., Chang, H. Y.: A semantic learning for content-based image retrieval using analytical hierarchy process. *Expert Syst. Appl.* **28**, 495–505 (2005)
- Chicago Police Department: How to describe a suspect. <https://portal.chicagopolice.org/portal/page/portal/ClearPath/Get%20Involved/Hotlines%20and%20CPD%20Contacts/How%20to%20Describe%20a%20Suspect>. Accessed 7 January 2016
- Chou, T. C., Cheng, S. C.: Design and implementation of a semantic image classification and retrieval of organizational memory information systems using analytical hierarchy process. *Omega* **34**, 125–134 (2006)
- Conilione, P., Wang, D.: Fuzzy approach for semantic face image retrieval. *Comput. J.* **55**, 1130–1145 (2012)
- Czerw, Z.: Human identification using appearance. In: Kędziński, W. (Ed.): *Forensic technique*, Vol. II. WSPol, Szczytno, pp. 141–171 (1995)
- Davies, G., Ellis, H., Shepherd, J.: Cue saliency in faces as assessed by the ‘Photofit’ technique. *Percept.* **6**, 263–269 (1977)
- DeLozier, S., Rhodes, M. G.: The impact of value-directed remembering on the own-race bias. *Acta Psychol.* **154**, 62–68 (2015)
- Dolecki, M., Karczmarek, P., Kiersztyn, A., Pedrycz, W.: Face recognition by humans performed on basis of linguistic descriptors and neural networks. In: *Proceedings of 2016 International Joint Conference on Neural Networks (IJCNN 2016)*, pp. 5135–5140 (2016)
- Dror, I. E., Wertheim, K., Fraser-Mackenzie, P., Walajtys, J.: The impact of human–technology cooperation and distributed cognition in forensic science: Biasing effects of AFIS contextual information on human experts. *J. Forensic Sci.* **57** (2), 343–352 (2012)
- Ekman, P., Friesen, W. V.: *Facial action coding system (FACS)*. Consulting Psychologists Press, Palo Alto (1978)
- Ellis, H. D., Shepherd, J. W., Davies, G. M.: Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Percept.* **8**, 431–439 (1979)
- FISWG (Facial Identification Scientific Working Group): Facial image comparison feature list for morphological analysis. Version 1.0 2013.08.12. <http://www.fiswg.org>. Accessed 7 January 2016
- Fukushima, S., Ralescu, A. L.: Improved retrieval in a fuzzy database from adjusted user input. *J. Intell. Inf. Syst.* **5**, 249–274 (1995)
- Haig, N. D.: Exploring recognition with interchanged facial features. *Percept.* **15**, 235–247 (1986)

- Hartigan, J. A., Wong, M. A.: A k-means clustering algorithm. *J. Roy. Statist. Soc. Ser. C* **28**, 100–108 (1979)
- Hollingsworth, K. P.: Eye-catching eyebrows: Training and evaluating humans for periocular image verification. *J. Forensic Sci.* **59** (3), 648–658 (2014)
- Ito, H., Koshimizu, H.: Keyword and face image retrieval based on latent semantic indexing. In: Proc. 2004 IEEE International Conference on Systems, Man and Cybernetics, pp. 358–363 (2004)
- Ito, H., Koshimizu, H.: Some experiments of face annotation based on latent semantic indexing in FIARS. In: Gabrys, B., Howlett, R. J., Jain, L. C. (Eds.): KES 2006, Part II, LNAI **4252**, pp. 1208–1215 (2006)
- Iwamoto, H., Ralescu, A.: Towards a multimedia model-based image retrieval system using fuzzy logic. In: Proc. SPIE 1827, Model-Based Vision, pp 177–185 (1992)
- Johnston, R. A., Edmonds, A. J.: Familiar and unfamiliar face recognition: A review. *Mem.* **17**, 577–596 (2009)
- Kacprzyk, J., Pedrycz, W.: Springer handbook of Computational Intelligence. Springer-Verlag, Berlin Heidelberg (2015)
- Karczmarek, P., Kiersztyn, A., Rutka, P., Pedrycz, W.: Linguistic descriptors in face recognition: A literature survey and the perspectives of future development. In: Proceedings of the Conference Signal Processing Algorithms, Architectures, Arrangements, and Applications (SPA 2015), pp. 98–103 (2015)
- Karczmarek, P., Pedrycz, W., Reformat, M., Akhoundi, E.: A study in facial regions saliency: A fuzzy measure approach. *Soft Comput.* **18**, 379–391 (2014)
- Kiersztyn, A., Karczmarek, P., Dolecki, M., Pedrycz, W.: Linguistic descriptors and fuzzy sets in face recognition realized by humans. In: Proc. 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1120–1126 (2016a)
- Kiersztyn, A., Karczmarek, P., Rutka, P., Pedrycz, W.: Quantitative methods for linguistic descriptors in face recognition. In: Zapła, A. (Ed.): Recent Developments in Mathematics and Informatics, Contemporary Mathematics and Computer Science, Vol. 1. The John Paul II Catholic University of Lublin Press, Lublin, pp. 123–138 (2016b)
- Koffka, K.: Principles of Gestalt psychology. Harcourt Brace Jovanovich, New York (1935)
- Kumar, N., Berg, A. C., Belhumeur, P. N., Nayar, S. K.: Attribute and simile classifiers for face verification. Proc. Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 365–372 (2009)

- Kumar, N., Berg, A. C., Belhumeur, P. N., Nayar, S. K.: Describable visual attributes for face verification and image search. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 1962–1977 (2011)
- Kurach, D., Rutkowska, D., Rakus-Andersson, E.: Face classification based on linguistic description of facial features. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., Zurada, J. M. (Eds.): *Artificial Intelligence and Soft Computing 2014, Part II. LNAI 8468*, pp. 155–166 (2014)
- LaVergne, D., Tiferes, J., Jenkins, M., Gross, G., Bisantz, A.: Linguistic estimations of human attributes. In: *Proc. HFESAM'16*, vol. **60**, pp. 318–322 (2016)
- Lee, S. Y., Ham, Y. K., Park, R. H.: Recognition of human front faces using knowledge-based feature extraction and neuro-fuzzy algorithm. *Pattern Recognit.* **29**, 1863–1876 (1998)
- Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S. Z.: Learning multi-scale block Local Binary Patterns for face recognition: In: *Advances in Biometrics, International Conference, ICB 2007, Lecture Notes in Computer Science 4642*, pp. 828–837 (2007)
- Lindsay, R. C. L., Ross, D. F., Read, J. D., Togliola, M. P.: *The handbook of eyewitness psychology: Volume II: Memory for People*. Psychology Press, Mahwah (2007)
- Liu, Y., Zhang, D., Lu, G., Ma, W.-Y.: A survey of content-based image retrieval with high-level semantics. *Pattern Recogn.* **20**, 262–282 (2007)
- Martinson, E., Lawson, W., Trafton, J. G.: Identifying people with soft-biometrics at Fleet Week. In: *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, pp. 49–56 (2013)
- Matthews, M. L.: Discrimination of Identikit constructions of faces: Evidence for a dual processing strategy. *Percept. Psychophys.* **23**, 153–161 (1978)
- Miyajima, K., Nakayama, M., Iwamoto, H., Norita, T.: Top-down image processing using fuzzy reasoning. In: Terano, S., Sugeno, M., Mukaidono, M., Shigemasa K. (Eds.): *Fuzzy engineering toward human friendly systems*. Ohmsha: IOS Press, pp. 983–994 (1992)
- Nakayama, M., Miyajima, K., Iwamoto, H., Norita, T.: Interactive human face retrieval system based on linguistic expression. In: *Proc. 2nd International Conference on Fuzzy Logic and Neural Networks, IIZUKA'92*, **2**, pp. 683–686 (1992a)
- Nakayama, M., Norita, T., Ralescu, A.: A fuzzy logic based qualitative modeling of image data. *Proc. IPMU'92*, pp. 615–618 (1992b)
- Navon, D.: Forest before trees: The precedence of global features in visual perception. *Cogn. Psychol.* **9**, 353–383 (1977)

- Norita, T.: Fuzzy theory in an image understanding retrieval system. In: Relascu, A. L. (Ed.): *Applied Research in Fuzzy Technology*. Springer Science+Business Media, New York, pp. 215–251 (1994)
- O’Toole, A. J., Abdi, H., Jiang, F., Phillips, P. J.: Fusing face-verification algorithms and humans. *IEEE Trans. Syst. Man Cybern. B* **37**, 1149–1155 (2007)
- Pantic, M., Rothkrantz, L. J. M.: An expert system for recognition of facial actions and their intensity. *IAAI-00 Proc.*, pp. 1026–1033 (2000a)
- Pantic, M., Rothkrantz, L. J. M.: Expert system for automatic analysis of facial expressions. *Image Vis. Comput.* **18**, 881–905 (2000b)
- Pedrycz, W.: *Granular Computing. Analysis and design of intelligent systems*. CRC Press, Boca Raton (2013)
- Rahman, A., Beg, M. M. S.: Face sketch recognition using sketching with words. *Int. J. Mach. Learn. Cyber.* **6**, 597–605 (2015)
- Rotshtein, P., Geng, J. J., Driver, J., Dolan, R. J.: Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *J. Cogn. Neurosci.* **19**, 1435–1452 (2007)
- Saaty, T. L.: *The Analytic Hierarchy Process*. McGraw-Hill, New York (1980)
- Sadr, J., Jarudi, I., Sinha, P.: The role of eyebrows in face recognition. *Percept.* **32**, 285–293 (2003)
- Shepherd, J., Davies, G., Ellis, H.: Studies of cue saliency. In: Davies, G., Ellis, H. D., Shepherd, J. W. (Eds.): *Perceiving and remembering faces*. Academic Press, New York, pp. 105–131 (1981)
- Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE* **94**, 1948–1962 (2006)
- Spaun, N. A.: Face recognition in forensic science. In: Li, S. Z., Jain, A. K. (Eds.): *Handbook of Face Recognition*. Springer-Verlag, London, pp. 655–670 (2001)
- Sridharan, K., Nayak, S., Chikkerur, S., Govindaraju, V.: A probabilistic approach to semantic face retrieval system. *Lect. Notes Comput. Sci.* **3546**, 977–986 (2005)
- Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 1891–1898 (2014)
- Tome, P., Fierrez, J., Vera-Rodriguez, R., Ortega-Garcia, J.: Combination of face regions in forensic scenarios. *J. Forensic Sci.* **60**, 1046–1051 (2015a)

- Tome, P., Vera-Rodriguez, R., Fierrez, J., Ortega-Garcia, J.: Facial soft biometric features for forensic face recognition. *Forensic Sci. Int.* **257**, 271–284 (2015b)
- Vignolo, L. D., Milone, D. H., Scharcanski, J.: Feature selection for face recognition based on multi-objective evolutionary wrappers. *Expert Syst. Appl.* **40**, 5077–5084 (2013)
- Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J., van der Helm, P. A.: A century of Gestalt psychology in visual perception II. Conceptual and theoretical foundations. *Psychol. Bull.* **138**, 1218–1252 (2012)
- 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)
- Wu, J. K., Narasimhalu, A. D.: Fuzzy content-based retrieval in image databases. *Inf. Process. Manag.* **34**, 513–534 (1998)
- Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., Ellis, A. W.: Matching familiar and unfamiliar faces on internal and external features. *Percept.* **14**, 737–746 (1985)
- Zadeh, L. A.: From Computing with Numbers to Computing with Words. *Ann. N. Y. Acad. Sci.* **929**, 221–252 (2001)
- Zadeh, L. A.: Toward extended fuzzy logic: A first step. *Fuzzy Set. Syst.* **160**, 3175–3181 (2009)
- Zhang, L., Song, Y., Liu, L., Liu, J.: Dissociable roles of internal feelings and face recognition ability in facial expression decoding. *Neuroim.* **132**, 283–292 (2016)
- Zhou H., Schaefer, G.: Semantic features for face recognition. In: *Proc. 52nd International Symposium ELMAR-2010*, pp. 33–36 (2010)

4. User Centric Graphic Enhancements of Methods of Decision-Making

In this chapter, we present a novel and innovative approach to the enhancement of Analytic Hierarchy Process which is one of the most important techniques used in decision-making theory. Typically, the matrices of pairwise comparisons are built on the basis of numerical inputs given by the experts evaluating the set of alternatives. These numerical values are often natural numbers from the range 1-9 or fuzzy memberships built on a basis of this scale. Here, we introduce the graphic approach to the AHP method which lets the experts to be independent on the constraints which the numerical scale brings with itself. To obtain such effect, we propose to use well-known GUI components such as slider and dial arc. The series of experiments show that the application of the graphical elements significantly improves the AHP inconsistency indices. Moreover, using Particle Swarm Optimization, we find the parameters of non-linear mappings which let to preserve the low inconsistency of the experts' opinions and, of course, the obtained values of preferences related to particular features (alternatives).

4.1. Enhancements of Analytic Hierarchy Process

Despite the classical approach to Analytic Hierarchy Process (Saaty 1980) is relatively intuitive and simple, in a number of situations, the consistency is difficult to be maintained at high (satisfactory) level. The reasons may be of various kind, namely misunderstanding of the AHP method by people or simply their lack of experience, difficulties with matching the numerical (or even linguistic) values of the AHP scales with their meaning which can be really important problem when two or more experts are engaged in the process of evaluations. Even if their general preferences are relatively similar, the aggregated results may lead to higher inconsistency. Finally, many people do not feel comfortable with somehow restricting scales. The process of choice of the numbers may be exhausting and lead to perfunctory treating of the process, particularly in the situations when higher number, say 10 or more, of alternatives is considered.

In the literature there were a number of proposals generalizing or improving the generic AHP approach, e.g., fuzzy weights describing the criteria of decision and alternatives (van Laarhoven & Pedrycz 1983), aggregation mechanism when a group of experts make the pairwise comparisons (Forman & Peniwati 1988, Dong et al. 2015), different scales, e.g., linear, power, and others, see Saaty (1977), Harker & Vargas (1987), or Ishizaka & Labib (2011). However, in Tavana et al. (1997) it was presented that people prefer to use verbal evaluations than plain numeric values. Linguistic modeling or Granular Computing-based techniques can be an important improvement here, see Pedrycz & Vasilakos

1999, Cabrerizo et al. 2018, Pedrycz 2013, Pedrycz & Song 2014, Liu et al. 2018, and many others. However, there is still a need of a significant improvement of decision tools by involving graphical input by user (Kabassi & Virvou 2015, Kersten 1987). The experts could choose among tools based on numeric, verbal or graphical inputs in experiments presented by Weistroffer et al. (1999). The advantages of an application of the graphical approaches were presented in details in Power & Sharda (2007), Larichev et al. (2003), von Winterfeldt & Edwards (1986). Mustajoki & Hämäläinen (2000) proposed an application of a slider. However, despite of promising concept no discussion of the impact of a slider on the decision-making process performance was presented. Moreover, in (Ito & Shintani 1997) it was presented a graphical user interface for persuasion mechanism for negotiation among agents based on integer values. Bhargava et al. 1999 presented a comparison of graphical decision support systems including interfaces to AHP based on numerical values and linguistic option panels. Drop down menus as well as sliders (but with linguistic descriptions of particular positions) were discussed by, for instance, Thirumalaivasan & Karmegam, 2001 and Thirumalaivasan et al. 2003. Perini et al. (2009) proposed radio buttons with a descriptions of the form $A \gggg B$, $A \ggg B$, etc., where A and B are compared alternatives. An et al. (2011) described an interface to fuzzy AHP (based on numerical values). Cay & Uyan described an interface with typical radio buttons and their linguistic descriptors. Wang et al. (2014) proposed numerical interface for a hybrid method based on fuzzy AHP and GRA originally developed by Deng (1989). Numerical drop down menus built in reciprocal matrices were discussed in Hanine et al. (2016). Other improvements of AHP techniques such as aggregation techniques and optimization approaches to the process were discussed in the previous chapters. Kiersztyn et al. (2018) proposed a method of determination of a matrix of dependencies between experts' opinions based on a slider.

Here, we propose a method which helps to depart from the user forcing method of quantifying the preferences. It is based on two graphical tools which are components available in almost all of the graphical programming environments. They are slider, sometimes called track bar, and dial arc. Sliding the position of the slider or rotating the dial arc can free the user from the numeric-oriented choice, particularly if the differences between the succeeding slider's values are negligibly small and imperceptible for the experts. This is easy to obtain when the property of maximal value of the slider is sufficiently large, say 100. In the case of the dial arc the positions are related with the angle rotation in-between -180 to 180 degrees. The examples of a slider and a dial arc are presented in Fig. 4.1 and Fig. 4.2, respectively.

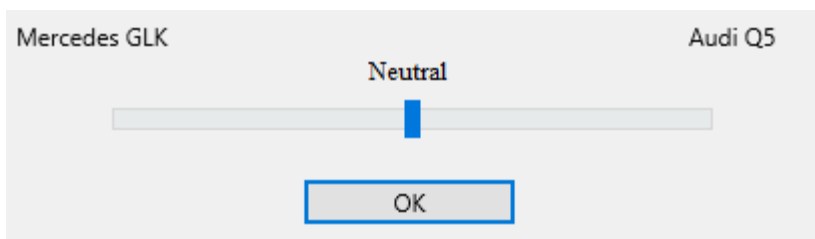


Fig. 4.1 An example of the graphic interface based on a slider. All the values of the component including neutral 0 value and its maximal and minimal positions are hidden to not suggest any decision to the user. The highest preference of Mercedes is at the maximal left position while the highest preference of Audi is pointed by the slider's maximal right position.

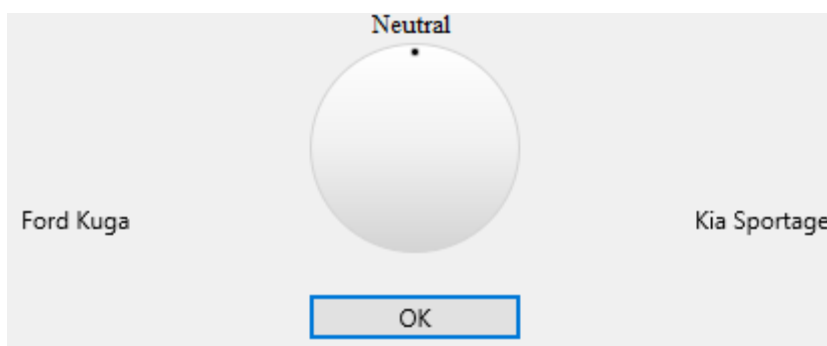


Fig. 4.2 An example of a realization of graphic interface with the use of a dial arc. Similarly to the previous case, the central position is neutral. The left minimal (-180°) and the right maximal (180°) mean the preference of Ford over Kia and an opposite situation, respectively.

To obtain the matrix of pairwise comparisons preserving the reciprocity property, the slider's positions should be transformed to the values from the scale $[1/9, 9]$ or another, say $[1/7, 7]$. It is worth to stress, that the expert is not able to take part in this step of the process. The conversion is realized with using the following formula:

$$t(q) = \begin{cases} \frac{8}{9r}q + 1 & \text{for } q \in [-r, 0) \\ \frac{8}{r}q + 1 & \text{for } q \in [0, r] \end{cases} \quad (4.1)$$

where q denotes the geometric position of the slider or dial arc, while r is a maximal value of the graphical component.

4.2. Non-linear Transformation of the Reciprocal Matrix

The use of graphic interfaces alone does not completely eliminate the inconsistencies of the pairwise comparisons process. They should be reduced to minimum. The role of a transformation (1) is to simplify the scale to be $[1/9, 9]$. However, some inevitable inconsistencies might still occur. Therefore, to minimize the inconsistency index we introduce a non-linear transformation which is non-decreasing and defined with a set (vector) of parameters to be established. The procedure of their finding has to include their adjustment to minimize the inconsistencies. One of such models of non-linear and non-decreasing functions is a piecewise linear function given by the formula

$$f(t) = \frac{(b_i - b_{i-1})(t - a_{i-1})}{a_i - a_{i-1}} + b_{i-1} \text{ for } t \in [a_{i-1}, a_i) \quad (4.2)$$

where $t \geq 1$ and the coefficients $a_i, b_i \in [1, 9]$ ($i = 2, \dots, p + 1$). If $t < 1$ the function value is obtained as $1/f(1/t)$.

Obviously, the mapping parameters are the coefficients a_2, a_3, \dots, a_p which are positioned on the x -axis and the coefficients b_2, b_3, \dots, b_p lay on the y -axis. It means that p linear segments is formed. Observe that the boundary coefficients are $a_1=1, a_{p+1}=9, b_1=1, \text{ and } b_{p+1}=9$. An example of such kind of the transformation is shown in. Fig. 4.3.

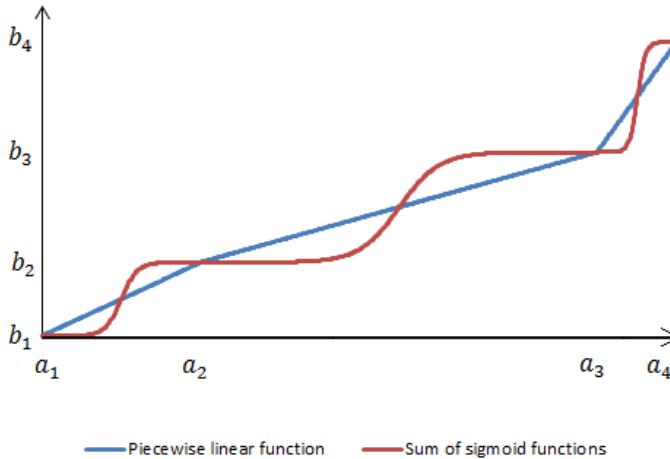


Fig. 4.3 Examples of a piecewise linear function f and a combination of sigmoid functions g for $p=3$ segments.

As a conclusion we can stress that (1) maps the slider values to the cells of the reciprocal matrix while (2) and the next expression (3) are established to result in more consistent matrix.

Of course, there is no need to limit to only one non-linear mapping. Another function of this kind may be a combination of sigmoidal functions. Its value in the segment $[a_{i-1}, a_i)$ is read as follows

$$g(t) = \frac{L}{1+e^{-k\left(t-\frac{a_i+a_{i-1}}{2}\right)}} \text{ for } t \in [a_{i-1}, a_i] \quad (4.3)$$

where L can be obtained from the formula

$$L = b_{i-1} \left(1 + e^{k\frac{a_i-a_{i-1}}{2}}\right) \quad (4.4)$$

while

$$k = 2 \frac{\ln \frac{b_i}{b_{i-1}}}{a_i - a_{i-1}} \quad (4.5)$$

for $t \geq 1$ and the coefficients $a_i, b_i \in [1, 9]$. Again, note that for the value $t < 1$ we take $1/g(1/t)$.

4.3. Minimization of Inconsistency

So far, we have established a collection of graphic data written in the scale $[1/9, 9]$. However, the AHP user is interested in minimizing the inconsistencies of the experts' reciprocal matrices. This conviction is due to the fact that the experts are not necessarily rational. A properly chosen non-linear transformation of the results can significantly improve reaching this aim. Of course, the piecewise linear function and combination of sigmoidal functions seem to be a good choices to serve as this mapping. This implies that their coefficients, i.e., a_2, a_3, \dots, a_p and b_2, b_3, \dots, b_p have to be found. A good strategy here is to use Particle Swarm Optimization (see Kennedy et al. 2001, Kacprzyk & Pedrycz 2015, and chapter 1). The results of evaluations of m (in particular $m=1$) experts are contained in the form of m reciprocal matrices, namely A_1, A_2, \dots, A_m . The goal here is to establish the parameters of the transformations f or g for which the sum of inconsistency indices v_k is minimal:

$$\arg \min_{a_2 < \dots < a_p, b_2 < \dots < b_p} \sum_{k=1}^m v_k \quad (4.6)$$

All the particles (in particular, their initial positions and velocities) of the swarm are initialized using the random method. Next, in the iterations of the method, they are reset using the following formulae:

$$\mathbf{v}_i = \mathbf{v}_i + 2[\mathbf{R}_1 \otimes (\mathbf{p}_i - \mathbf{x}_i) + \mathbf{R}_2 \otimes (\mathbf{p}_g - \mathbf{x}_i)] \quad (4.7)$$

$$\mathbf{x}_i = \mathbf{x}_i + \mathbf{v}_i \quad (4.8)$$

Here, \mathbf{x}_i denotes a vector particle which means that it is an i -th vector of values $a_2, a_3, \dots, a_p, b_1, b_2, \dots, b_p$ of f or g , respectively, the rest of symbols are described in the similar way as in chapter 1, see, formulae (1.1) and (1.2). It is worth to stress that the conditions $a_{i-1} < a_i$ and $b_{i-1} < b_i$ are to be preserved. It is easy to fulfill it by an application of the increases $\Delta_i = a_{i+1} - a_i$ and $\delta_i = b_{i+1} - b_i, i = 1, \dots, p$, in the optimization space instead of the origin values a_i and b_i , respectively. After the optimization process the yielded transformations f and g are applied to all the reciprocal matrices. The average eigenvector of all the reciprocal matrices is a result of the process. Of course, one could apply a geometric mean to the initial reciprocal matrices entries and optimize just one

matrix, see (Dong et al. 2010). Fig. 4.4 depicts an overall processing scheme of the process.

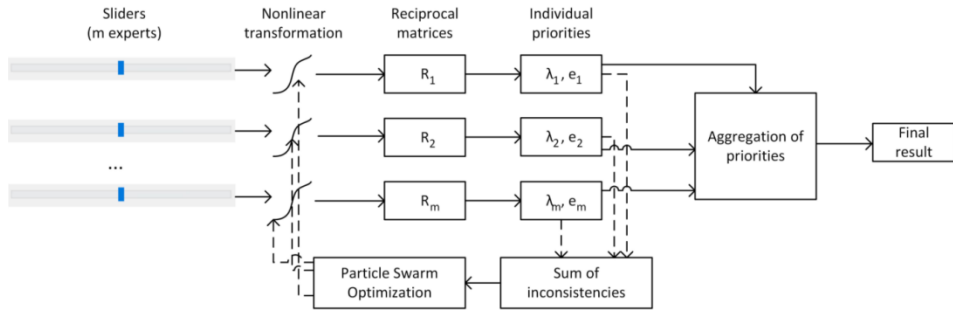


Fig. 4.4 Main processing phases of the method.

4.4. Experimental Results

In the series of experiments, we consider four typical problems (which model of 4-5 year old crossover or SUV is less emergency, where would you like to go on vacation, who is a better footballer, and which TV series would you like to watch?).

4.4.1. Crossovers and SUVs' Reliability

We have decided to compare the list of fifteen car models, namely Audi Q5, BMW X1, Chevrolet Captiva, Dacia Duster, Ford Kuga, Honda CR-V, Hyundai ix35, Kia Sportage, Mercedes GLK, Mitsubishi ASX, Nissan Juke, Suzuki Grand Vitara, Toyota RAV4, Volvo XC60, and Volkswagen Tiguan, see the report of reliability TÜV-Report 2017. Three experts (consumers with general but not specialized knowledge) were to fill the scoresheet questionnaires in three manners, namely the questionnaires with integral values corresponding to the descriptions given in the previous sections and simple dedicated applications with graphical interfaces containing slider and dial arc, see print screens presented in Fig. 4.1 and Fig. 4.2. Note that the order of pairwise comparisons presented to the experts was random. The comparison between SUV and crossover models on a basis of normalized experts priorities values along with inconsistency indices are listed in Table 4.1. The results listed in the table show that the lowest inconsistency level can be reached when experts use a very intuitive graphical component, namely slider. However, a dial arc is not as intuitive as a slider. The participants of our tests stressed it during the experiment series. It may be a reason why the inconsistencies obtained when using this tool are higher. No matter of inconsistency analysis, it is worth to stress that the general preference is always the same, i.e., it is the same model of the car and the relations between the second, third, and fourth place are almost negligible. Let us consider the results of an application of non-linear

transformations which were found by the PSO method in a series of 20 repetitions. Note that the first and the fifth SUV are the same (Mercedes and Volvo), while the cars being at places no. 2, 3, and 4 change and their results are dependent on the method. The reason of this may be as follows: After 200 PSO iterations, the inconsistency indices are very close to 0, in particular in the case when the experts use dial arc. To focus on the properties of the method from now on, the default number of PSO particles is 40 while the number of segments (i.e., p) is 10. Additionally, an interesting result is presented in Fig. 4.5. The transformation plots giving the lowest inconsistencies lay under the rest of the plots. As one can see, this rule is satisfied in other kinds of experiments. Furthermore, the segments count (p) and the particles number do not influence the final results of AHP, see Fig. 4.6, Fig. 4.7, and Fig. 4.8. Note that the last three figures show the results obtained with slider. However, similar results can be obtained using other variants of our approach (dial arc). Interesting information is presented in Fig. 4.9. It depicts all the experts answers after transformation through the non-linear piecewise linear function being the result of the PSO process. Observe that here, the values laying almost at the right and left slider's end, but not being the end itself (which is $s = 100$ and $-s = -100$) are used relatively rarely. The figure shows that almost the whole range of the floating point scale values is used provided that a relatively high features number is under comparison.

Table 4.1 The preferences (normalized to 1) for SUVs and crossovers.

Compared cars	Num.	Sliders	Dial arcs	Num. +piec. lin. f.	Num. +comb. sigm. f.	Sliders +piec. linear f.	Sliders +comb. sigm. f.	Dial arcs +piec. lin. f.	Dial arcs +comb. sigm. f.
Mercedes GLK	1	1	1	1	1	1	1	1	1
BMW X1	0.78	0.79	0.8	0.79	0.81	0.83	0.86	0.89	0.95
Honda CR-V	0.75	0.84	0.79	0.79	0.84	0.83	0.82	0.96	0.96
Audi Q5	0.68	0.85	0.75	0.72	0.76	0.88	0.82	0.96	0.98
Volvo XC60	0.63	0.67	0.66	0.65	0.68	0.72	0.76	0.86	0.92
Mitsubishi ASX	0.37	0.44	0.28	0.42	0.5	0.53	0.63	0.83	0.88
Toyota RAV4	0.31	0.42	0.4	0.37	0.47	0.51	0.59	0.79	0.86
VW Tiguan	0.28	0.35	0.36	0.34	0.45	0.45	0.55	0.74	0.83
Ford Kuga	0.22	0.35	0.29	0.29	0.40	0.43	0.51	0.77	0.85
Hyundai ix35	0.17	0.25	0.2	0.24	0.36	0.36	0.47	0.76	0.83
Kia Sportage	0.16	0.2	0.19	0.23	0.35	0.31	0.43	0.74	0.81
Nissan Juke	0.14	0.2	0.17	0.21	0.33	0.31	0.43	0.72	0.8
Suzuki Grand Vitara	0.12	0.15	0.13	0.18	0.29	0.25	0.38	0.73	0.79
Chevrolet Captiva	0.1	0.14	0.15	0.16	0.27	0.23	0.35	0.67	0.76
Dacia Duster	0.05	0.07	0.09	0.09	0.17	0.13	0.22	0.6	0.7
Inconsistency index	0.17	0.09	0.2	0.02	0.02	0.02	0.02	0	0

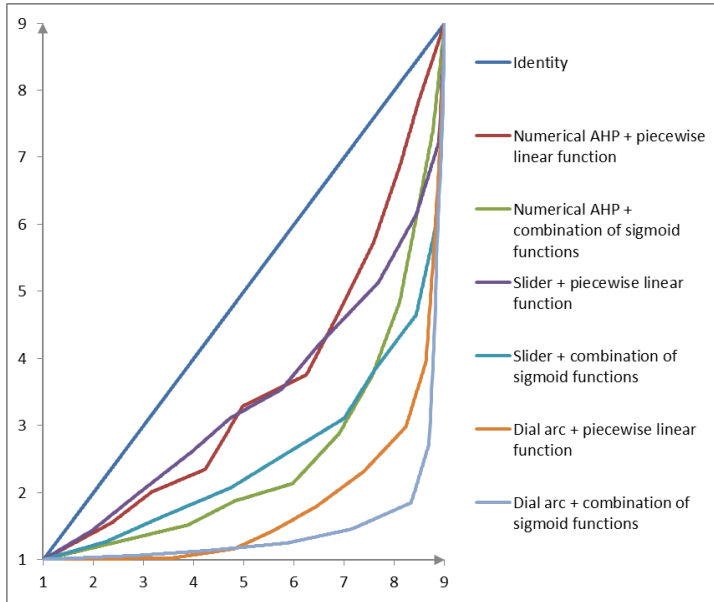


Fig. 4.5 The average non-linear transformations (piecewise linear function and combination of sigmoid functions) after 20 repetitions of experiments. Numerical AHP refers to the scoresheet-based questionnaires.

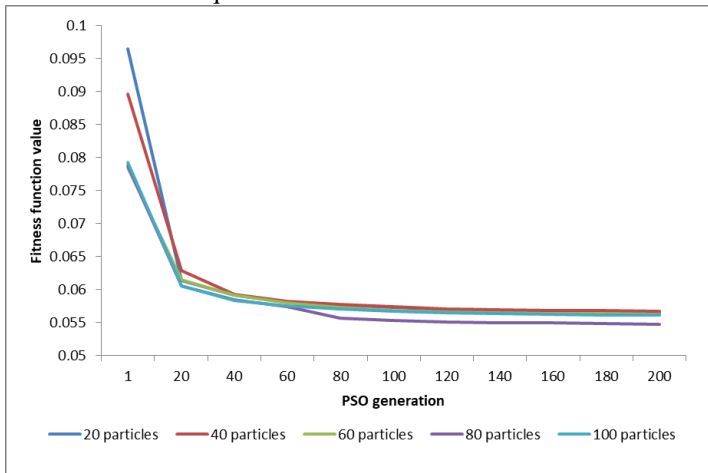


Fig. 4.6 A convergence of PSO fitness function values in consecutive iterations for 20, 40, 60, 80, and 100 particle populations, respectively (a case of slider and piecewise linear transformation function).

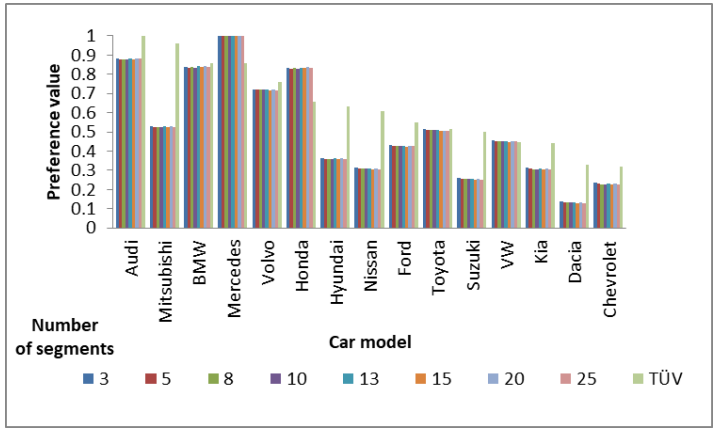


Fig. 4.7 The results in dependency on the number of segments of piecewise linear function (3, 5, 8, 10, 13, 15, 20, and 25 segments) in comparison with an original ranking (TÜV).

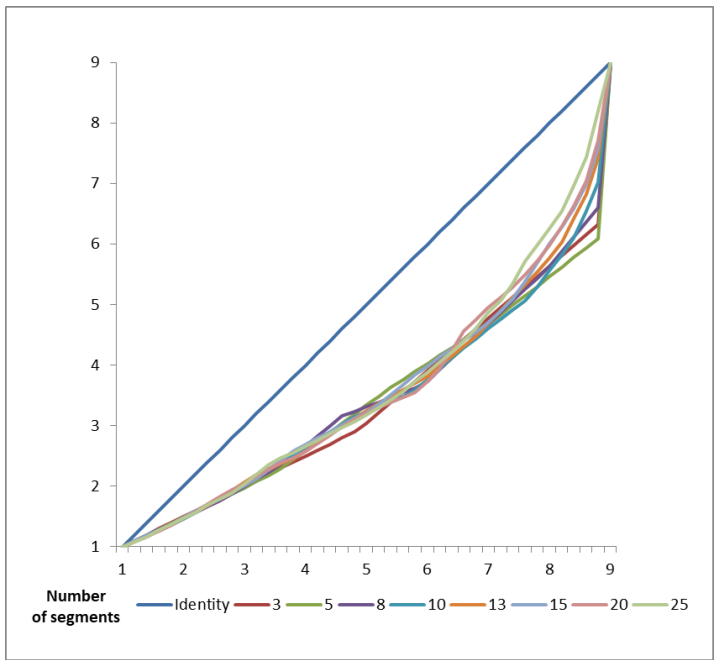


Fig. 4.8 The average plots of transformations when the function is built from 3, 5, 8, 10, 13, 15, 20, and 25 segments, respectively.

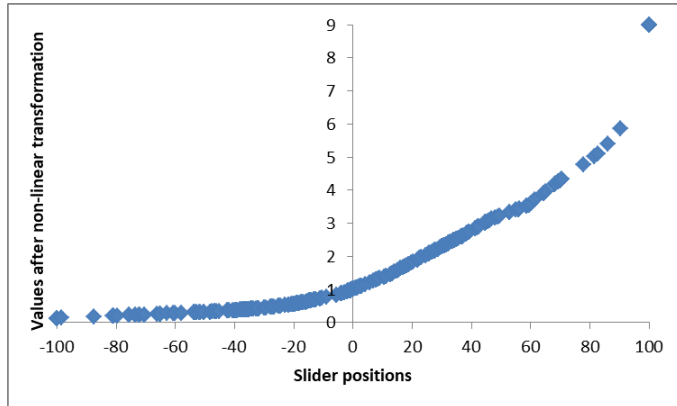


Fig. 4.9 The experts answers before (slider positions) and after transformation by piecewise linear function.

4.4.2. The Choice of the Holidays

Here, the set of possible choices was built of the countries or islands in the Mediterranean Sea area (Bulgaria, Crete, Cyprus, Egypt, Madeira, Mallorca, Sicilia, and Turkey). Again, similarly to another considered cases, the inconsistency index was the smallest when the sliders were in use.

Table 4.2 shows the results. The order of preferences of the experts is as follows: 1. Mallorca, 2. Madeira, 3. Crete. Relatively small changes start at the position no. four. It depends on the method's choice. It is worth noting that the combination of the classic Analytic Hierarchy Process, non-linear transform and Particle Swarm Optimization leads to the swap between two last holiday choices, namely Egypt and Turkey. Moreover, when slider is used, no changes after PSO are seen. Dial arc followed by the PSO swaps the places no. 4 and 5 (Bulgaria and Cyprus). As in the case of SUVs, the smallest inconsistency indices were yielded by the AHP based on the dial arc angular values inputs. Fig. 4.10 contains the plots of average transformations after 20 series of experiments.

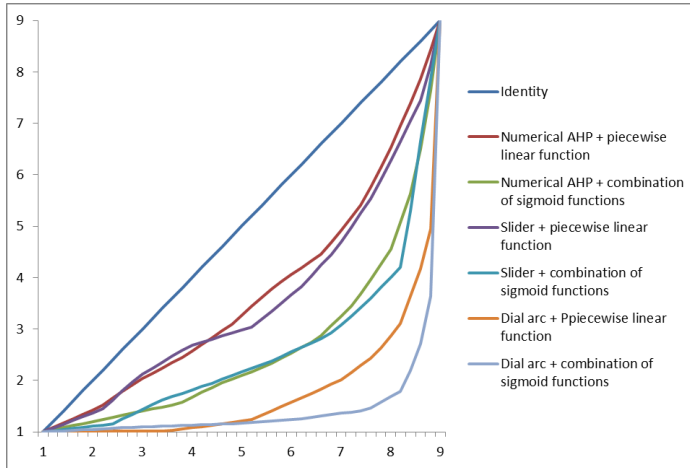


Fig. 4.10 Holidays choice: The average plots of transformations sought in the PSO method.

Table 4.2 Holidays choice.

Place	Numbers	Sliders	Dial arcs	Num. + piec. lin. f.	Num. + comb. sigm. f.	Sliders+ piec. lin. f.	Sliders+ comb. sigm. f.	Dial arc+ piec. lin. f.	Dial arc+ comb. sigm. f.
Bulgaria	0.23	0.26	0.24	0.27	0.30	0.32	0.39	0.56	0.75
Crete	0.43	0.36	0.4	0.46	0.5	0.41	0.48	0.68	0.84
Cyprus	0.16	0.18	0.21	0.20	0.25	0.25	0.34	0.59	0.75
Egypt	0.05	0.04	0.05	0.07	0.1	0.06	0.1	0.36	0.56
Madeira	0.57	0.59	0.55	0.6	0.63	0.62	0.68	0.73	0.88
Mallorca	1	1	1	1	1	1	1	1	1
Sicilia	0.19	0.18	0.2	0.23	0.29	0.23	0.31	0.54	0.73
Turkey	0.05	0.05	0.06	0.07	0.09	0.07	0.11	0.36	0.58
Inconsistency index	0.2	0.16	0.21	0.05	0.05	0.03	0.03	0	0

4.4.3. Top Football Players

In the third series of experiments the experts were to choose top football (soccer) players among the following footballers: Gareth Bale, Antoine Griezmann, Robert Lewandowski, Cristiano Ronaldo, Lionel Messi, Neymar, and Luis Suárez. In this case an application of non-linear transformation being the result of PSO slightly changed the order, namely Messi followed Ronaldo. However, the reciprocal matrices being the results of slider and dial arc applications were relatively stable and the orders have not been changed after using the PSO-based non-linear transformation. Again, the transformation plots and inconsistency values are very similar to the corresponding statistics obtained in the previous topics (car models and holidays choice), see Table 4.3 and Fig. 4.11.

Table 4.3 The choice of the best football player.

Player	Num.	Slid.	Dial arc	Num.+ piec. lin. f.	Num.+comb. sigm. f.	Slid.+ piec. lin. f.	Slid.+ comb. sigm. f.	Dial arc+piec. lin f.	Dial arc+ comb. sigm. f.
Ronaldo	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00
Messi	0.96	0.87	0.81	1.00	1.00	0.90	0.92	0.90	0.96
Griezmann	0.10	0.08	0.11	0.12	0.15	0.11	0.14	0.47	0.72
Suárez	0.19	0.20	0.20	0.22	0.26	0.25	0.29	0.53	0.76
Neymar	0.40	0.39	0.43	0.43	0.46	0.45	0.47	0.74	0.88
Bale	0.15	0.14	0.26	0.19	0.22	0.18	0.22	0.60	0.78
Lewandowski	0.34	0.39	0.35	0.39	0.44	0.41	0.44	0.71	0.84
Inconsistency index	0.2	0.18	0.27	0.03	0.03	0.02	0.02	0	0

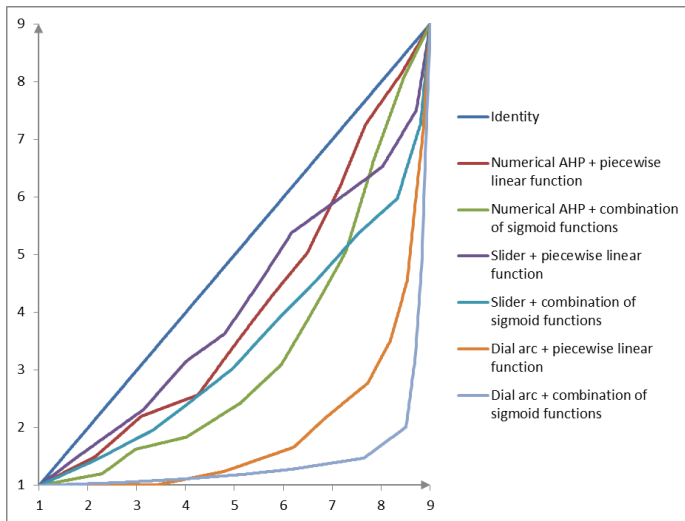


Fig. 4.11 PSO-optimized non-linear transformations obtained for the problem of the choice of the best footballer.

4.4.4. Preferred TV Series

In this series of experiments our experts were to choose the TV series which is most preferred among the following: Game of Thrones, House of Cards, Narcos, Vikings, and The Walking Dead. Game of Thrones and The Walking Dead have won in the ranking of preferences. The results show that the rest of places in ranking depend on the method of pairwise comparisons. One of the reasons may be an emotional relation to the particular series which implied that the experts have often used maximal values of the AHP scale. Moreover, the observed inconsistencies were different in a comparison to the previous experimental series. The best result was noted when a slider approach was in use. Table 4.4 consists of the full set of results while Fig. 4.12 depicts the average plots of obtained non-linear transformations.

Table 4.4 TV series' preferences.

Place	Num.	Slid.	Dial arc	Num. + piec. lin. f.	Num.+ comb. sigm. f.	Slid.+ piec. lin. f.	Slid.+ comb. sigm. f.	Dial arc+ piec. lin. f.	Dial arc+ comb. sigm. f.
Game of Thrones	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
House of Cards	0.37	0.40	0.40	0.44	0.59	0.72	0.90	0.63	0.71
Narcos	0.19	0.18	0.18	0.29	0.45	0.60	0.81	0.53	0.64
Vikings	0.46	0.52	0.47	0.35	0.28	0.36	0.37	0.40	0.54
The Walking Dead	0.70	0.67	0.89	0.61	0.59	0.81	0.94	0.80	0.92
Inconsistency index	0.11	0.06	0.31	0.06	0.06	0	0	0	0.02

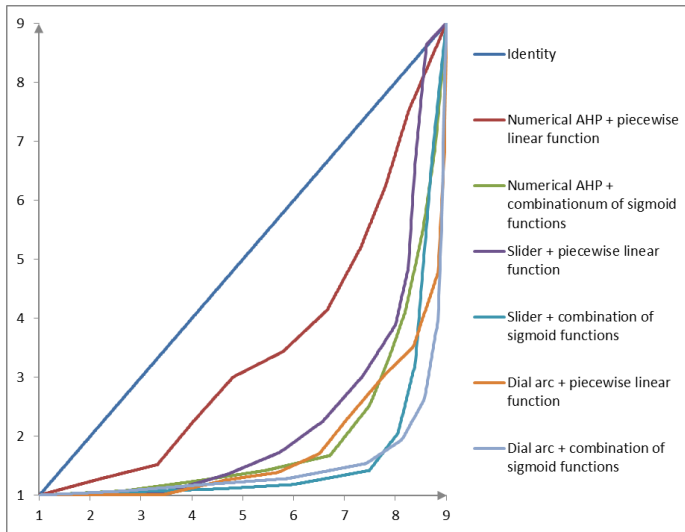


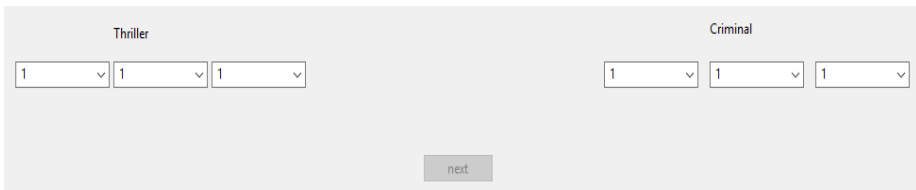
Fig. 4.12 The plots of the average non-linear transformation found in the Particle Swarm Optimization process for the TV series choice.

A conclusion after the analysis of the considered cases is that the graphical tools, particularly a slider followed by an optimization method, are able to improve the consistency of the experts pairwise comparisons. Moreover, their application can lead to the effective determination of the importance of particular preferences.

4.5. Fuzzy Extensions of the Analytic Hierarchy Process

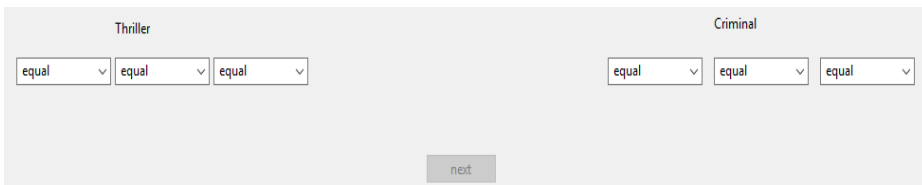
In this section, we consider an extension of Analytic Hierarchy Process based on fuzzy numbers, namely interval, triangular, and trapezoidal. In case of the interval membership function the expert sets the range of the preferred values and they are equivalently important. Using triangular fuzzy membership functions, the user chooses the most important preference. However, two boundary values are the points where the preference vanishes. Finally, these two approaches are merged in the case of trapezoidal fuzzy number, where the expert has to select the interval of the most preference and the points where the preference is vanishing. The examples of numeric interfaces for triangular numbers are depicted in Fig. 4.13, Fig. 4.14, and Fig. 4.15. It is worth to note that implementing such kind of interface the developer has to remember about the monotonicity of the preference values to avoid the unnecessary errors when transforming to the AHP fuzzy values. Moreover, it is worth stressing that the intervals are enabled to be in a degenerated form (a point value).

The slider values are transformed onto the fuzzy AHP scale using formula (1). The final crisp value can be chosen by, for instance, the method of inverse distribution, where the membership functions serve as distributions.



The interface shows two columns: 'Thriller' and 'Criminal'. Each column contains three dropdown menus, all of which are set to the value '1'. Below the dropdowns is a 'next' button.

Fig. 4.13 Numerical interface for triangular fuzzy number-based Analytic Hierarchy Process.



The interface shows two columns: 'Thriller' and 'Criminal'. Each column contains three dropdown menus, all of which are set to the value 'equal'. Below the dropdowns is a 'next' button.

Fig. 4.14 Linguistic interface for triangular fuzzy number-based Analytic Hierarchy Process.



Fig. 4.15 Graphical interface for three slider-based Analytic Hierarchy Process.

4.6. Experimental Results

The two following series of experiments with the following problems were conducted to show the efficiency of our proposal:

- Problem #1 Which national football team will get the title of world champion in football (soccer) in 2018?
- Problem #2 Which movie genre do you prefer?

The experiments were carried in the similar way as in the crisp version of AHP with one difference: There is no optimization process following the decision by experts. Three football fans and three TV series fans were to decide which team will win the football world championship and a favorite TV series, respectively. The results are listed in Table 4.5. It is worth noting that the slider-based graphical interfaces give the best results in the sense of lowest inconsistency independently onto the way of aggregation of the results of a group of experts, i.e., arithmetic mean of three experts' evaluations or geometric mean of AHP reciprocal matrices' entries. Such situation is observed in 9 cases of 16.

Table 4.5 A comparison of inconsistency indices for the fuzzy extensions of AHP interface.

Problem	1 num.	1 ling.	1 slid.	2 num.	2 ling.	2 slid.	3 num.	3 ling.	3 slid.	4 num.	4 ling.	4 slid.
World champion-arithm. approach	0.071	0.086	0.111	0.174	0.101	0.087	0.081	0.126	0.064	0.108	0.124	0.106
World champion-geom. approach	0.028	0.024	0.054	0.082	0.023	0.02	0.028	0.048	0.022	0.037	0.039	0.039
Movie-arithm. approach	0.118	0.122	0.197	0.152	0.154	0.128	0.128	0.146	0.155	0.127	0.182	0.175
Movie-geom. approach	0.011	0.016	0.036	0.037	0.047	0.025	0.02	0.056	0.018	0.024	0.047	0.015

4.7. An Application of Graphical Interface to the Biometric Features Description

In this section, we give a thorough in-depth analysis of the problem of face description by the experts. A situation when the witness of a crime or a specialist has difficulties in a correct description of the facial parts is quite often. One of the promising ways of improving this process may be by an application of AHP. The pairwise comparison method could help experts to use more intuitive approach than typical numerical or linguistic values. Here, we show chosen interesting results of an application of the above-mentioned method which helps to improve the judgements of the experts regarding the description of facial features as well as their rationality. The experiments were carried out using the PUT Face Database. An example of a face image is presented in Fig. 4.16. We asked one experienced face recognition expert to assess a set of 20 images. These photos are of high resolution and a few facial features including eye position are precisely found. It will help to thoroughly check the dependencies between the real values of facial parts' lengths and experts' evaluations. An example of an interface to the AHP face estimation is presented in Fig. 4.16.

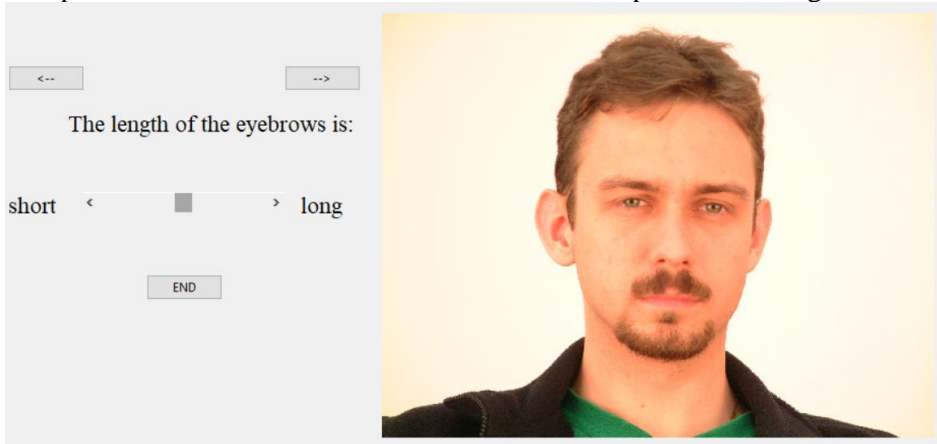


Fig. 4.16 A print screen of the executed application. The facial image comes from PUT Face Database.

The first of the features under consideration is the eyebrow width. Our expert described 20 images of people presenting their head central position using the AHP pairwise comparison method applied to three linguistic values: *short*, *average*, and *long*. The optimization process based on Particle Swarm Optimization was repeated ten times to get the 10-segment piecewise linear function f . Its plot is presented in Fig. 4.17. The values of coefficients of maximal eigenvectors of AHP reciprocal matrices before the PSO process are shown in Fig. 4.18. Fig. 4.19 presents analogical values after the transformation

by the function f . One can observe an improvement of the results. It is manifested in the form of trend lines corresponding to *short*, *average*, and *long* linguistic variables. An important fact is that the average inconsistency index before and after optimization was 0.072 and 0.007, respectively. Similar considerations were made for the facial width feature. The list of detailed values is presented in Table 4.6. Fig. 4.20 presents the mapping f while Fig. 4.21 and Fig. 4.22 depict the trends of the results before and after the PSO process of optimization, respectively. It is worth noting, that again the inconsistency index was improved and changed from 0.048 to 0.007. These two results present the effectiveness and potential applicability of the proposed method to the description of biometric features.

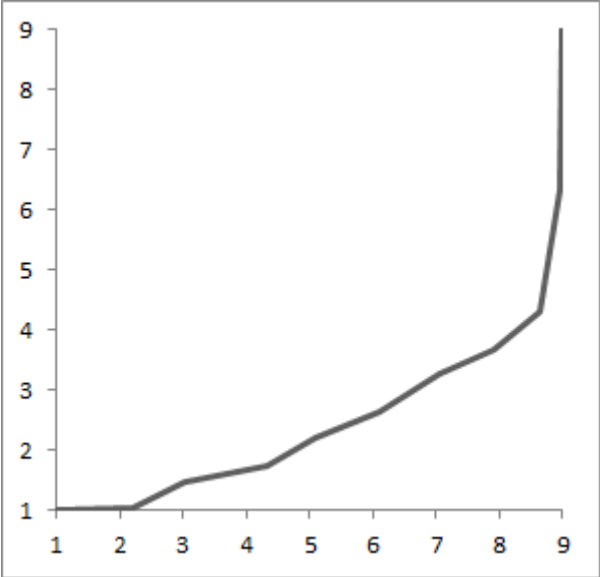


Fig. 4.17 The transformation function f (eyebrows width).

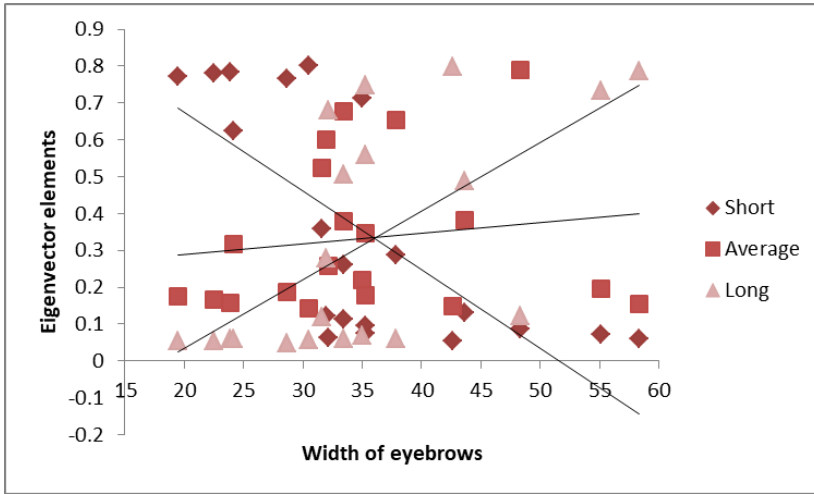


Fig. 4.18 The set of AHP results including trend lines (before the optimization process).

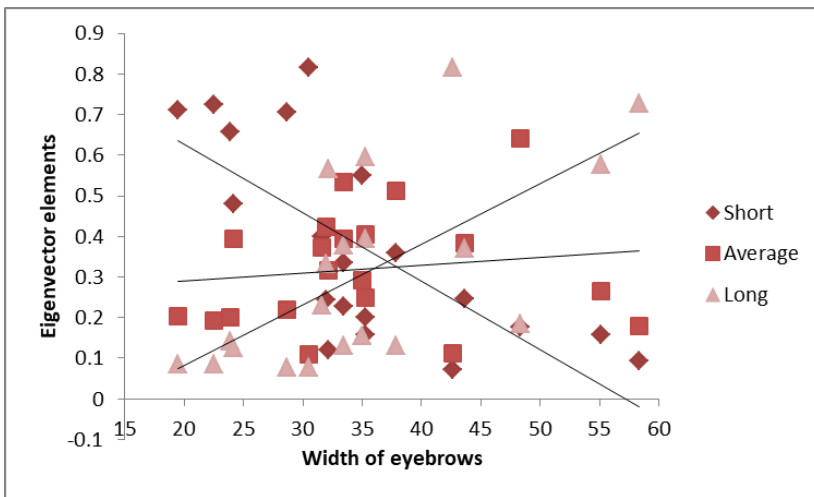


Fig. 4.19 The AHP results after PSO-based optimization with respect to the sum of maximal eigenvalues.

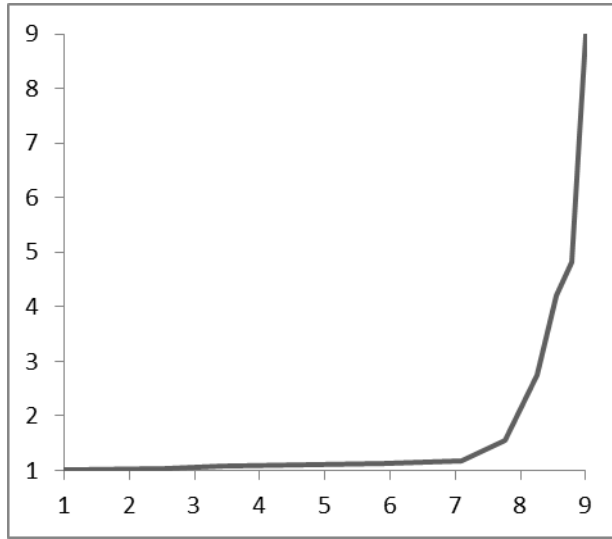


Fig. 4.20 The transformation function f (a case of face width). Here the Ox -axis corresponds to the a_i coefficients while Oy -axis represents the values of b_i .

Table 4.6 The detailed results corresponding to Fig. 4.19.

Pixel length	Short	Average	Long
19.45	0.71	0.20	0.08
22.50	0.72	0.19	0.08
23.92	0.66	0.20	0.14
24.16	0.48	0.39	0.13
28.69	0.71	0.22	0.08
30.51	0.82	0.11	0.08
31.65	0.40	0.37	0.23
31.95	0.24	0.42	0.33
32.16	0.12	0.32	0.57
33.45	0.23	0.39	0.38
33.48	0.34	0.53	0.13
34.97	0.55	0.29	0.16
35.26	0.16	0.25	0.59
35.27	0.20	0.41	0.39
37.82	0.36	0.51	0.13
42.67	0.07	0.11	0.81
43.62	0.25	0.38	0.37
48.31	0.18	0.64	0.18
55.12	0.16	0.26	0.58
58.32	0.09	0.18	0.73

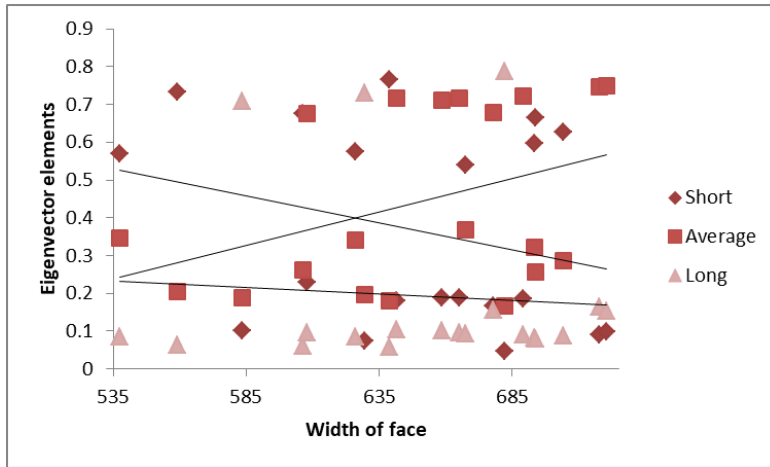


Fig. 4.21 Results related to the face width feature (before the optimization procedure).

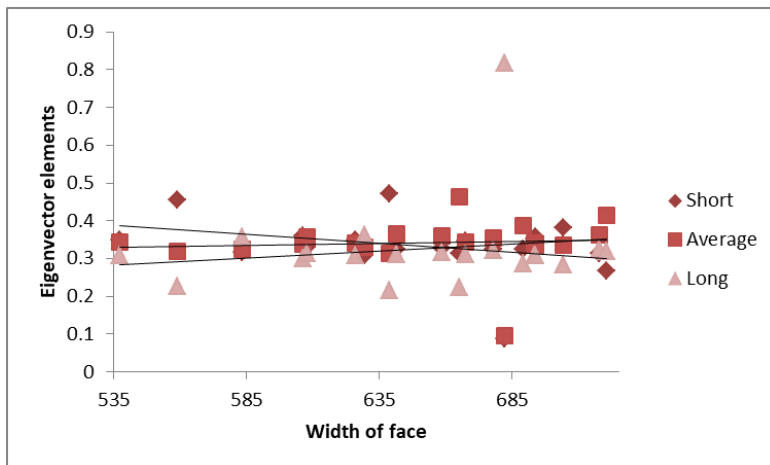


Fig. 4.22 Results related to the face width feature (after the optimization procedure).

4.8. Conclusions

In this chapter, we have proposed a novel and in-depth approach to the Analytic Hierarchy Process and fuzzy Analytic Hierarchy Process decision-making techniques based on the graphic interfaces such as slider or dial arc. This way of obtaining the experts' opinions gives very consistent results. A series of experiments related to real life problems such as the prediction of sport results, preferred TV series, etc. show the applicability of the method. Moreover, we have shown that the method is worth considering in tasks related to biometric

features description. Moreover, an application of PSO optimization helps to decrease the level of inconsistency of pairwise comparison reciprocal matrix.

References

- An, M., Chen, Y., Baker, C. J.: A fuzzy reasoning and fuzzy-analytical hierarchy process based approach to the process of railway risk information: A railway risk management system. *Inf. Sci.* **181**, 3946–3966 (2011)
- Bhargava, H. K., Sridhar, S., Herrick, C.: Beyond spreadsheets: Tools for building decision support systems. *Comput.* **32** (3), 31–39 (1999)
- Cabrerizo, F. J., Morente-Molinera, J. A., Pedrycz, W., Taghavi, A., Herrera-Viedma, E.: Granulating linguistic information in decision-making under consensus and consistency. *Expert Syst. Appl.* **99**, 83–92 (2018)
- Cay, T., Uyan, M.: Evaluation of reallocation criteria in land consolidation studies using the Analytic Hierarchy Process (AHP). *Land Use Policy* **30**, 541–548 (2013)
- Deng, J. L.: Introduction to grey system theory. *J. Grey Syst.* **1** (1), 1–24 (1989)
- Dong, Y., Fan, Z. P., Yu, S.: Consensus building in a local context for the AHP-GDM with the individual numerical scale and prioritization method. *IEEE Trans. Fuzzy Syst.* **23** (2), 354–368 (2015)
- Dong, Y., Zhang, G., Hong, W.-C., Xu, Y.: Consensus models for AHP group decision-making under row geometric mean prioritization method. *Decis. Support Syst.* **49** (3), 281–289 (2010)
- Forman E., Peniwati, K.: Aggregating individual judgments and priorities with the analytic hierarchy process. *Eur. J. Oper. Res.* **108**, 165–169 (1988)
- Hanine, M., Boutkhoul, O., Tikniouine, A., Agouti, T.: Application of an integrated multi-criteria decision-making AHP-TOPSIS methodology for ETL software selection. *SpringerPlus* **5** (263) (2016)
- Harker, P. T., Vargas, L. G.: The theory of ratio scale estimation: Saaty's analytic hierarchy process. *Manag. Sci.* **33** (11), 1383–1403 (1987)
- Ishizaka, A., Labib, A.: Review of the main developments in the analytic hierarchy process. *Expert Syst. Appl.* **38** (11), 14336–14345 (2011)
- Ito, T., Shintani, T.: Persuasion among agents: An approach to implementing a group decision support system based on multi-agent negotiation. In: *Proceedings of the 5th International Joint Conference on Artificial Intelligence (IJCAI'97)*. Morgan Kaufmann, pp. 592–597 (1997)
- Kabassi, K., Virvou, M.: Combining decision-making theories with a cognitive theory for intelligent help: A comparison. *IEEE Trans. Hum. Mach. Syst.* **45** (2), 176–186 (2015)

- Kacprzyk, J., Pedrycz, W.: Springer handbook of Computational Intelligence. Springer–Verlag, Berlin Heidelberg (2015)
- Kennedy, J. F., Eberhart, R. C., Shi, Y.: Swarm intelligence. Academic Press, San Diego (2001)
- Kersten, G. E.: Two aspects of group decision support system design. In: Sawaragi, Y., Inoue, K., Nakayama, H. (eds): Toward interactive and intelligent decision support systems. Lect. Notes Econ. Math. Syst. **286**, 373–382, Springer, Berlin, Heidelberg (1987)
- Kiersztyn, A., Karczmarek, P., Zhadkovska, K., Pedrycz, W.: Determination of a matrix of the dependencies between features based on the expert knowledge. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J. (Eds.): Artificial Intelligence and Soft Computing. ICAISC 2018. Lecture Notes in Computer Science **10842**, 570–578 (2018)
- Larichev, O., Kochin, D., Ustinovičius, L.: Multicriteria method for choosing the best alternative for investments. Int. J. Strateg. Prop. Manag. **7**, 33–43 (2003)
- Liu, F., Wu, Y. H., Pedrycz, W.: A modified consensus model in group decision-making with an allocation of information granularity. IEEE Trans. Fuzzy Syst. doi: 10.1109/TFUZZ.2018.2793885 (2018)
- Mustajoki, J., Hämäläinen, R. P.: Web-HIPRE: Global decision support by value tree and AHP analysis. INFOR Inf. Syst. Oper. Res. **38** (3), 208–220 (2000)
- Pedrycz, W.: Granular Computing. Analysis and design of intelligent systems. CRC Press, Boca Raton (2013)
- Pedrycz W., Song, M.: A granulation of linguistic information in AHP decision-making problems. Inf. Fusion **17**, 93–101 (2014)
- Pedrycz, W., Vasilakos, A. V.: Linguistic models and linguistic modeling. IEEE Trans. Syst. Man Cybern. B Cybern. **29** (6), 745–757 (1999)
- Perini, A., Ricca, F., Susi, A.: Tool-supported requirements prioritization: Comparing the AHP and CBRank methods. Inf. Softw. Technol. **51**, 1021–1032 (2009)
- Power, D. J., Sharda, R.: Model-driven decision support systems: Concepts and research directions. Decis. Support Syst. **43** (3), 1044–1061 (2007)
- Saaty, T. L.: A scaling method for priorities in hierarchical structures. J. Math. Psychol. **15** (3), 234–281 (1977)
- Saaty, T. L.: The Analytic Hierarchy Process. McGraw-Hill, New York (1980)
- Tavana, M., Kennedy, D. T., Mohebbi, B.: An applied study using the analytic hierarchy process to translate common verbal phrases to numerical probabilities. J. Behav. Decis. Mak. **10** (2), 133–150 (1997)

- Thirumalaivasan, D., Karmegam, M.: Aquifer vulnerability assessment using analytic hierarchy process and GIS for upper palar watershed. In: 22nd Asian Conference on Remote Sensing, pp. 1–6 (2001)
- Thirumalaivasan, D., Karmegam M., Venugopal, K.: AHP-DRASTIC: software for specific aquifer vulnerability assessment using DRASTIC model and GIS. *Env. Model. Softw.* **18**, 645–656 (2003)
- TÜV-Report 2017. Auto Bild Special, pp. 36–39 (2017)
- van Laarhoven P. J. M., Pedrycz, W.: A fuzzy extension of Saaty's priority theory. *Fuzzy Set. Syst.* **11** (1–3), 199–227 (1983)
- von Winterfeldt, D., Edwards, W.: Decision analysis and behavioral research. Cambridge University Press (1986)
- Wang, Y., Xi, C., Zhang, S., Yu, D., Zhang, W., Li, Y.: A combination of extended fuzzy AHP and fuzzy GRA for government e-tendering in hybrid fuzzy environment. *Sci. World J. Art. ID: 123675* (2014)
- Weistroffer, H. R., Wooldridge, B. E., Singh, R.: A multi-criteria approach to local tax planning. *Socio Econ. Plan. Sci.* **33** (4), 301–315 (1999)

5. Fuzzy Measures and Facial Features Saliency

In this chapter, we are focused on the design of fuzzy measures (specifically, λ -fuzzy measure) related to the importance of particular facial features. Moreover, we are interested in finding its optimal set of parameters which lead to the best possible performance of classifiers built on the basis of Choquet integral.

5.1. Introduction

One of the most challenging research problems has been still an understanding of the mechanisms of recognition and perception of faces. Despite the fact that humans recognize others in relatively different way, some properties of the mechanisms are common. They are, for instance, important features perception, the mutual relationships between them, and finally, an identification of responsible brain areas. These facts can significantly supply the computational methods of facial recognition alleviating the limitations of humans.

One of the divisions of face recognition methods is that there exist holistic matching and feature-based matching approaches (see, Zhao et al. 2003 for details). The first of them is, for instance, a well-known Fisherfaces (Belhumeur et al. 1997), a representative of the latter group is, for instance, LBP (Ahonen et al. 2004, Heikkilä et al. 2009). However, there are many approaches combining these two main trends. They are called hybrid. Examples are, among others, the work by Pentland et al. 1994 (a combination of the eigenfeatures, Eigenfaces by Turk & Pentland 1991, and the combined modular representation) or component (parts of face)-based approaches (Heisele et al. 2003, Huang et al. 2003, Bonnen et al. 2013). Among all these methods, fuzzy information fusion produces promising results (Kwak & Pedrycz 2005).

From the point of view of this kind of methods an in-depth studies and understanding the way of faces perception by people is a key task. For instance, people process the faces holistically, see Sinha et al. 2006. Here, the spacing between face elements (i.e., second-order spatial relations) play a pivotal role (Rotshtein et al. 2007). However, the familiar (trained) faces are recognized on a basis of internal parts (e.g., eyes, nose, or mouth). The external features (e.g., face contour or hair properties) are very salient in the process of untrained (unfamiliar) face recognition (Ellis et al. 1979, Young et al. 1985). Davies et al. (1977) proved that the forehead, eyes, or mouths change causes the lowest error rates in a series of experiments with subjects. The well-known Photofit Kit was used in this work. Haig (1986) and Matthews (1978) presented similar results. The observers claimed that eyebrows/eye followed by mouse and nose areas are the most dominant parts if the internal features are compared. Moreover,

O'Donnell & Bruce (2001) presented the results showing that people are very sensitive to changes in the region of eyes of familiarized faces. Upper face of face and eyebrows were, again, confirmed as very salient by Haig (1986). Sadr et al. (2003) revealed the results of experiments where subjects recognized the celebrities' faces with no eyebrows worse than the images without eyes. The mean difference was 9.5%. Literature surveys on human face recognition and cue importance are presented in (Johnston & Edmonds 2009) and (Shepherd et al. 1981). The psychological mechanisms and cue saliency in the processes of face recognition were also discussed by Venkat et al. 2013, Da et al. 2010, Pujol & García 2012, Fang et al. 2011, Choi et al. 2012, or Robins et al. 2018.

Computational face recognition methods in relation to cue saliency were examined in a series of works. Here, it is noteworthy that they are important from another point of view. In many forensic applications the face can be covered and only a small region can be visible (e.g., when a subject wears sunglasses, a helmet, a balaclava, a veil, or a mask). We describe a few works with details. Template matching strategy was applied by Brunelli & Poggio (1993) with the ranking of saliency as follows: eyes, mouth, nose and the template of a whole face. Similar order of features was yielded by Lam & Yan 1998 on a basis of correlation values treated as a similarity measure and by Kwak & Pedrycz 2005 with an application of Fisherfaces. The techniques based on Radial Basic Functions Networks were applied to determine the importance of facial features or facial parts by Sato et al. 1998, Gutta et al. 2002, and Gutta & Wechsler 2003. A region-based partitioning and generic approach were compared by Ekenel & Stiefelhagen (2009) on a basis of discrete cosine transform-based feature extraction (Ekenel & Stiefelhagen 2005). For example, the following order of overlapping regions was obtained: Forehead, left eye and right eye, left cheek and right cheek. Noteworthy is that an increased number of features (14) did not produce satisfying results. In (Yan & Osadciw 2004) discussed was the combination of Eigenfaces carried for eyes, mouth, nose and forehead region. An interesting result was that adding an individual eigenfeature can improve the recognition rate except for the nose area. Fisherfaces for chosen regions of face was discussed in (Dargham et al. 2012). 3D morphable model approach was considered in an application to 14 facial regions by Heisele & Blanz (2005). Almost all of the studies produce one common result that the region of eyes exhibits the highest value of discrimination. Other results on an importance of particular facial features were published by Neo et al. 2007, Neo et al. 2010, Park et al. 2011, Savvides et al. 2004a, Savvides et al. 2004b, Savvides et al. 2006, Teo et al. 2007, Woodard et al. 2010, and Wright et al. 2009.

The advantage of information about the importance of facial features in the process of face classification can be utilized particularly in the techniques based on an aggregation of different classifiers. Such classifiers can be built using specific facial features, regions, or whole faces, transformations of the images,

methods, and multi-modal biometrics. A proper choice of the aggregation operator (function) leads to the increment of accuracy of the method. However, the optimal choice of this aggregation operator is difficult and has not been studied intensively in the literature so far.

Let us recall a few significant results related to the aggregation of face recognition classifiers. Brunelli & Poggio 1993 proposed a technique of scoring for the regions of eyes, nose, mouth, and the whole face. Pentland et al. 1994 presented proposals for aggregation of Eigenfaces. Haddadnia & Ahmadi 2004 used the majority rule for RBF neural networks. Jarillo et al. 2008 discussed Bayesian product and majority voting to the aggregation process of methods based on dimensionality reduction. Liu & Liu (2010) discussed a weighted sum rule and similarity matrices. Dolecki et al. (2016) applied utility functions. The classifiers were the parts of face. Hu et al. (2015) analyzed t-norms as the aggregation functions. Al-Hmouz et al. 2017 proposed fuzzy set and three-valued logic-based decision mechanism. Embedding of the colors fusion with deep learning framework was proposed by Alrjebi et al. 2016.

A special role in the aggregation methods is played by the fuzzy measure. Its application may help in a determination of weights related to the criteria corresponding to particular classifiers. On a basis of this fact the procedure of final classification can be realized by fuzzy integral. Such approach was proposed in (Kwak & Pedrycz 2005), where the Fisherfaces-based classifiers applied to the regions of the whole face, eyes, nose, and mouth were aggregated. Similarly, Melin et al. (2005) applied modular neural networks to eyes, nose, and mouths areas. Wavelet-based classifiers were aggregated by Kwak & Pedrycz 2004. Yan et al. 2006 aggregated the separate component SVMs outputs using their importance. Lee & Marshall (2008) applied similar techniques to the three-dimensional case. Gender recognition on a basis of fuzzy measure was proposed in Li et al. 2012. The particular classifiers were SVMs applied to chin, mouth, nose, eyes, forehead, hair, and clothing. Other fuzzy measure applications to the problems of pattern recognition can be found in Pedrycz (1990), Graves & Nagarajah (2007), Keller et al. (1994), Yan & Keller (1991), Grabisch (1995), Mirhosseini et al. (1998), Martínez et al. (2014; 2015).

In this study, we build the fuzzy measure on a basis of the results of computational and psychological experiments and relate it to the importance of face areas for the human processes of recognizing faces. Note that the fuzzy measure is able to potentially capture the important information associated with the significance of special face areas and their connections or combinations. To the recognition purpose, the information contained in the merged areas and in the specific areas is used by people. An intuitively appealing fact is that the more features are included in the process of face recognition, the better accuracy can be obtained. Here, the property of monotonicity appears as a key factor. Formally, this fact can be transformed using fuzzy measure which is based on the pivotal concept of monotonicity, and then it can be experimentally evidenced

on the basis of a series of experimental experiments in order to evaluate the performance of fuzzy measure. In the experimental section we consider the areas of eyes, nose, mouth, and left and right cheeks. They can be seen as the most descriptive facial features. The main goals of the study discussed in this work are:

- Investigating the fuzzy measure abilities to reflect the saliency of the information contained in facial areas and their groups.
- Evaluation and numeric quantification of the facial regions and their combinations roles in the face recognition processes.
- Finding the relations between the importance (quantified by fuzzy measure) of merged (combined) facial regions and the accuracy produced using Eigenfaces and Fisherfaces as well-known representatives of the face recognition algorithms.
- A comparison between the results of experiments obtained with the presence of subjects and realized by computational face recognition methods.
- Design of the Sugeno fuzzy measure on a basis of the psychological studies results related to cue saliency and a proposition of a novel model of the face identification mechanism.

Moreover, we are interested in a comprehensive studying of the fuzzy measure abilities to cover and evaluate the significance of the information of chosen parts of a face. Similarly, the concatenations of the regions of a face are worth investigating in this context. Particularly, an interesting is to examine the way on how the appearance of these regions impacts the process of recognition. Finally, one of the main objectives of the study is to determine the optimal parameters (densities) of the Sugeno λ -fuzzy measure by a thorough comparison of the results obtained for the above six facial parts as well as for different combinations and the whole area of the face. To find the relations we will examine different classification techniques like PCA (Principal Component Analysis, Turk & Pentland 1991), LDA (Linear Discriminant Analysis, Belhumeur et al. 1997), LBP (Local Binary Patterns, Ahonen et al. 2004), MBLBP (Multi-scale Block LBP, Chan et al. 2007, Liao et al. 2007), FR (Full Ranking, Chan et al. 2015), CCBLD (Chain Code-based Local Descriptor, Karczmarek et al. 2016; 2017). The results are obtained with utilizing the image datasets as follows: AT&T, FERET, Yale (Yale Face Database), and Labeled Faces in the Wild cropped version.

5.2. A General Scheme of the Aggregation Process

Here, we discuss the general scheme of processing. It highlights a sequence of tasks in a classification based on an aggregation method, see Fig. 5.1. The general flow is relatively typical to all the processes which utilize the aggregation mechanisms. The first stage is preprocessing of a face including

cropping, scaling, an eventual histogram equalization. Next, the positions of important regions (eyes, nose, mouth, left and right cheeks) are determined manually. Then, the accuracies for the selected segments of the face and their combinations on a basis of PCA, and PCA followed by LDA, are determined. Next, the λ -fuzzy measure is constructed on a basis of atomic facial regions-related recognition rates. Moreover, the recognition rates associated with chosen combinations of the areas are yielded.

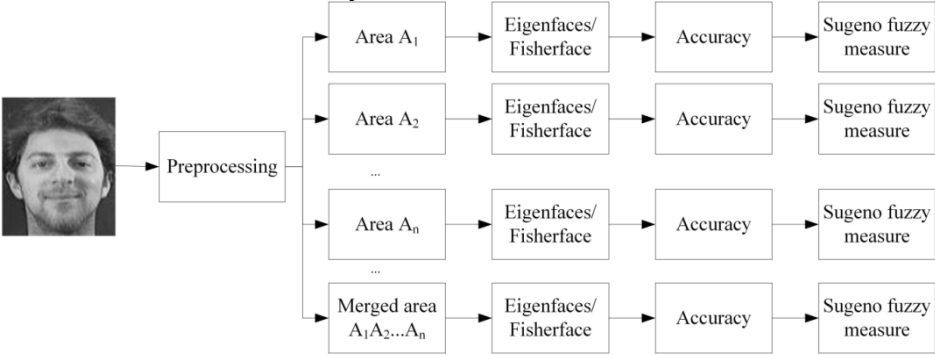


Fig. 5.1 An overall processing scheme.

5.3. Fuzzy Measure and Choquet Integral Interpretation

To identify an image on a basis of a given number of regions (classifiers), one should take into account the weights associated with criteria (individual and related to the groups, i.e., facial regions concatenations). The weights present recognition qualities of particular classifiers and affect the final decision in a proper way, i.e., the aggregation result. It is important that the fuzzy measure can express the relations between the regions on a basis of particular classifiers (see Grabisch 1995). In a formal way, denote by $X = \{x_1, \dots, x_n\}$ the overall area of a face. The entities x_1, \dots, x_n are non-overlapping segments (eyes, nose, etc.). Then, we can define a fuzzy measure as a set function fulfilling the relations (for a general theory, see the preliminary chapter) (1.4), (1.5), and (1.6). The first two conditions say that the whole face carries the complete information associated with the face. The last monotonicity property quantifies the observation from the area of psychology that the possibility of a proper classification of a face increases when the information (knowledge) about the facial region is augmented by parts of knowledge about other regions of face. Recall that the parametrized version of the measure has the form (1.8). In this formula A and B are disjoint sets. The λ parameter shows the dependency between the regions of the face. If $\lambda < 0$ (super-additivity), the measure is sub-additive. It means that the satisfaction which arises from one evidence source (face area) entails the satisfaction of the second one. It leads to the conclusion that they are in redundancy (competition) and that a combination of regions is

not efficient. If $\lambda > 0$ then the synergy effect is present and the evidence sources efficiently support each other, see Grabisch 1995, Pedrycz & Gomide 1998. The value of λ can be yielded in a unique form ($\lambda > -1, \lambda \neq 0$) from the equation (Sugeno 1974) (1.9), where x_1, \dots, x_n represent the regions of face which do not overlap and g_i are fuzzy measure densities. If we denote $A_i = \{x_1, \dots, x_i\}, A_{i+1} = \{x_1, \dots, x_i, x_{i+1}\}$, the fuzzy measure over the area being the combination regions is determined recursively as (1.10).

The fuzzy measure g is linked with the classification concept proposed in (Kwak & Pedrycz, 2005). Denote by d_{ij} the distance between a test image and the j th image in i th classifier (e.g., a region of a face). Moreover, let $d^{(i)}$ denote average distance between all the vectors of images in the i th classifier and let C_k be the k th class. This leads to the formula

$$h(x_{ik}) = \frac{1}{N_k} \sum_{v_{ij} \in C_k} v_{ij} \quad (5.1)$$

for the k th class values with the maximal membership grade in the i th classifier. Here,

$$v_{ij} = \frac{1}{1 + \frac{d_{ij}}{d^{(i)}}} \quad (5.2)$$

Then, Choquet integral of a function h is defined in the following way

$$\int g \circ h = \sum_{1 \leq i \leq n} \left(\left(h(x_{ik}) - h(x_{i+1,k}) \right) g(A_i) \right) \quad (5.3)$$

with an assumption that

$$h(x_{n+1,k}) = 0 \quad (5.4)$$

and that the values $h(\cdot)$ are non-increasingly sorted. Note that this is a direct incorporation of the formulas (1.12) and (1.11), respectively with the assumption that here is the double indexing used. Choquet integral is required to return as high as possible values when the patterns (elements) are contained in the same class and low values when they come from different classes.

5.4. Experimental studies

In the experiments' series the main goal is to find the recognition rates for important face regions and their combinations. Moreover, we are interested in finding the corresponding fuzzy measure values for these parts. In parallel, we find these values on a basis of the results of psychological experiments described in the literature. The fuzzy measure properties are one of our interests here. The experimental results in this study were carried for the AT&T facial images and FERET Database.

In the preliminary series of tests, we obtained the recognition rates for the following six subregions of the face separately: Eyebrows, eyes, nose, mouth, and cheeks (left and right). Our choice of the regions was motivated by the fact that they cover almost the whole area of a face, have very descriptive value, and are important in the processes of classification by humans. These regions (for

which the preliminary tests gave the best results) are depicted in Fig. 5.2 while Table 5.1 shows their size details.



Fig. 5.2 The whole face and its subregions.

Table 5.1 Characteristics of particular facial segments.

Region	AT&T		FERET	
	Width (px)	Height (px)	Width (px)	Height (px)
Original image	92	112	256	384
Face after cropping	90	94	100	140
Eyebrows area	88	14	91	14
Eye area	82	14	84	15
Nose area	35	28	37	31
Mouth area	51	28	54	29
Left/right cheek region	22	55	24	72

The first series of experiments is devoted to obtaining the fuzzy measures for atomic facial regions and their merges (combinations) on a basis of Principal Component Analysis technique. The AT&T image dataset was randomly divided into two equal sets (training and testing) containing five people of each subject. Similar experiments were done for the images from FERET set with two images per person in the training set and one image in the testing set. In parallel, Fisherfaces (PCA followed by LDA) was carried for the same sets. Each experimental setting was repeated 100 times. The final reported accuracies are the averages of all the results.

Table 5.2 presents the recognition rates obtained for the atomic salient regions while Fig. 5.3 depicts the values obtained from the fuzzy measure calculations with respect to the connected (combined) regions as well as the accuracies yielded for the concatenated facial images. The concatenation is understood as follows: It is the result of merging two pictures treated as vectors with pixel values entries. Similar results are illustrated in Fig. 5.4, Fig. 5.5, Fig. 5.6 for the respective combinations of 3, 4, and 5 areas. Next, the scatter plot of the accuracies and accuracies obtained with an application of fuzzy measure are visualized in Fig. 5.7, Fig. 5.8, Fig. 5.9, and Fig. 5.10. Moreover, a linear regression expressing the relation between the fuzzy measure-based results and

the classification schemes are shown. Here, it is important to note that the accuracy values are rescaled to consist with the fuzzy measure boundary condition. Analogical results were obtained for the results produced on a basis of probability measure. It represents the additive class of measures, see Fig. 5.11, Fig. 5.12, Fig. 5.13, and Fig. 5.14. It is constructed using the basic areas of the face (e.g., eyes) and applying the condition of additivity for the concatenation of areas. Again, the values are normalized to satisfy the condition (1). λ -values, maximal and minimal differences between Sugeno fuzzy measure values along with the recognition rates are enlisted in Table 5.3. A visualization of the Fisherfaces recognition rates and fuzzy measures corresponding to combination of eyebrows, eyes, nose, and mouth areas as well as nose, mouth, left and right cheeks is presented in Fig. 5.15. We distinguish these parts since they represent the upper and lower part of a face, respectively (this partition is widely discussed in the literature). Finally, Fig. 5.16 includes the values of the accuracies and the fuzzy measure values for the regions of mouth and eyes which are gradually augmented by areas in their neighborhoods.

Table 5.2 Atomic facial regions and corresponding percentage recognition rates.

Area	AT&T set		FERET set	
	Eigenfaces	Fisherfaces	Eigenfaces	Fisherfaces
Eyebrows	62.16	81.75	28.81	72.93
Eyes	67.03	79.86	15.42	52.4
Nose	59.77	66.23	10.29	31.28
Mouth	49.31	60.89	4.08	18.75
Left cheek	36.55	68.06	9.4	30.68
Right cheek	39.42	67.96	10	36.17

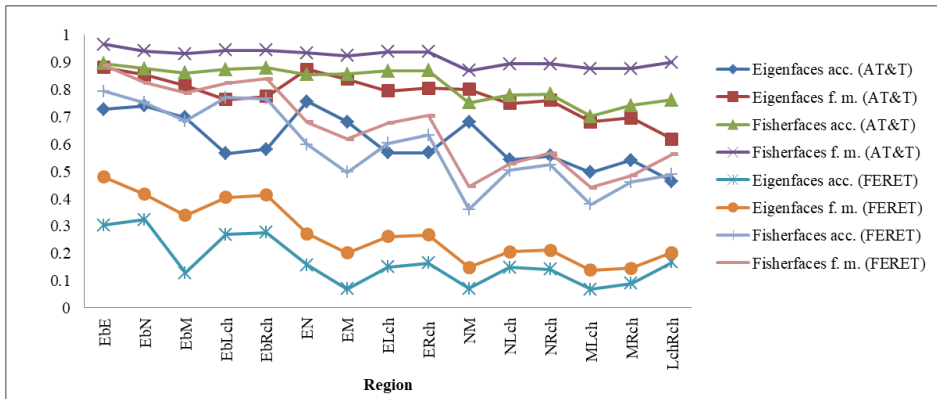


Fig. 5.3 Accuracies and fuzzy measure obtained for concatenations of (a) two facial areas using Eigenfaces and Fisherfaces. Here Eb, E, N, M, Lch, and Rch denote eyebrows, eyes without eyebrows, nose, mouth, left cheek, and right cheek region, respectively.

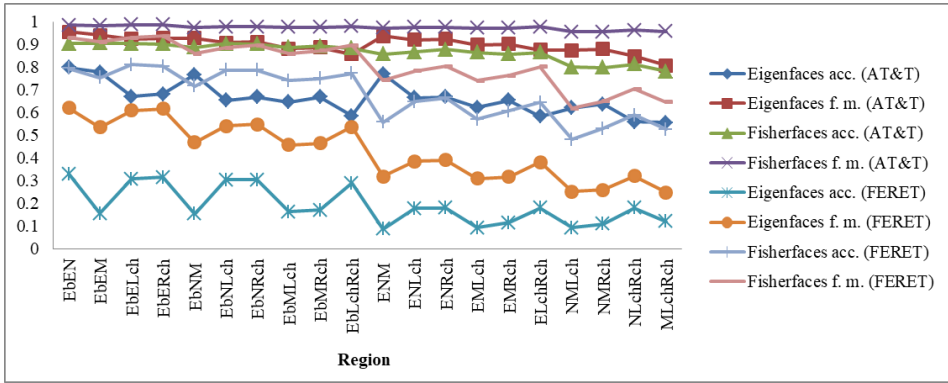


Fig. 5.4 Accuracies and fuzzy measure obtained for concatenations of three facial areas using Eigenfaces and Fisherfaces.

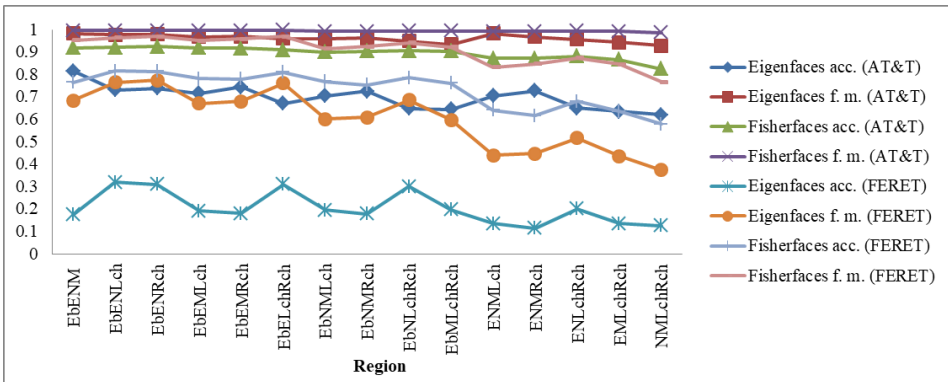


Fig. 5.5 Accuracies and fuzzy measure obtained for concatenations of four facial areas using Eigenfaces and Fisherfaces.

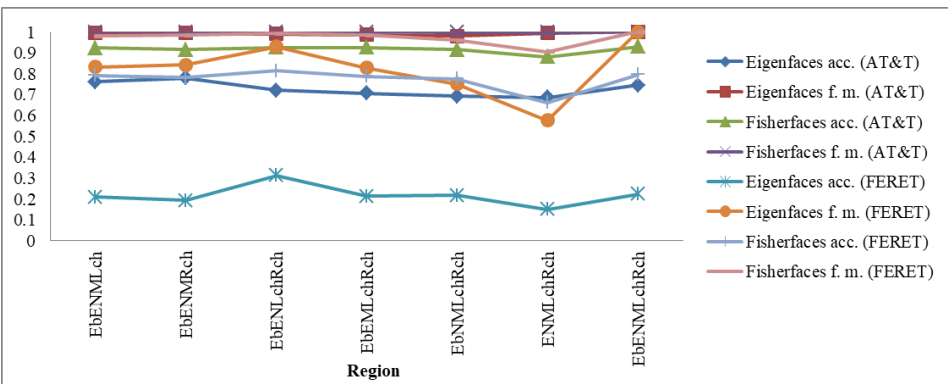


Fig. 5.6 Accuracies and fuzzy measure obtained for concatenations six facial areas using Eigenfaces and Fisherfaces.

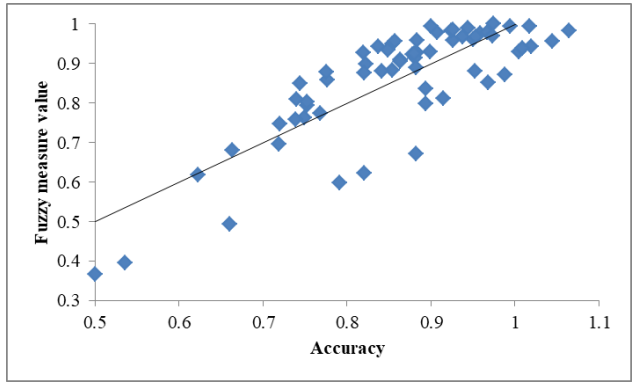


Fig. 5.7 Accuracies versus fuzzy measure values for Eigenfaces and AT&T database.

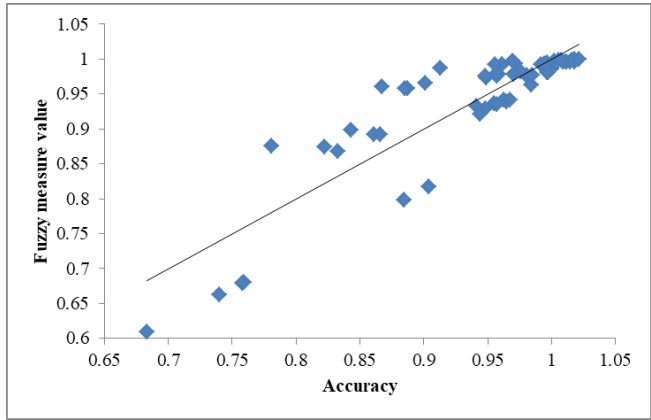


Fig. 5.8 Accuracies versus fuzzy measure values for Fisherfaces and AT&T database.

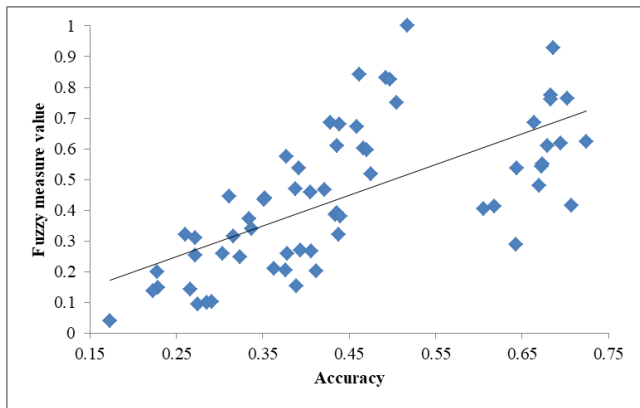


Fig. 5.9 Accuracies versus fuzzy measure values for Eigenfaces and FERET database.

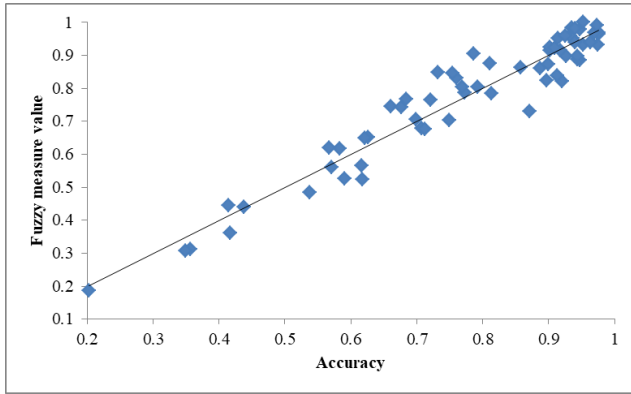


Fig. 5.10 Accuracies versus fuzzy measure values for Fisherfaces and FERET database.

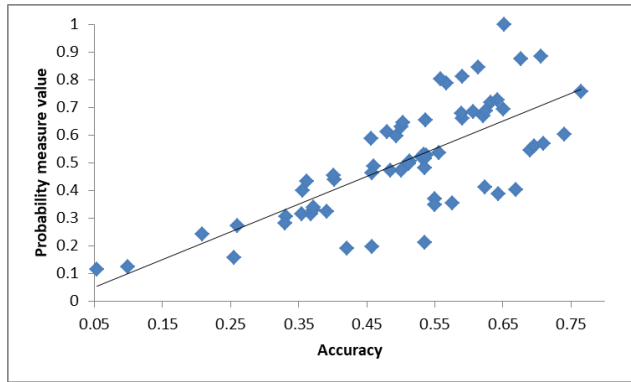


Fig. 5.11 Accuracies versus probability measure values for Eigenfaces and AT&T database.

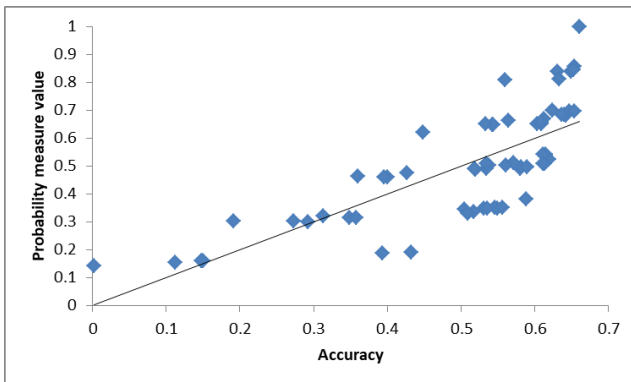


Fig. 5.12 Accuracies versus probability measure values for Fisherfaces and AT&T database.

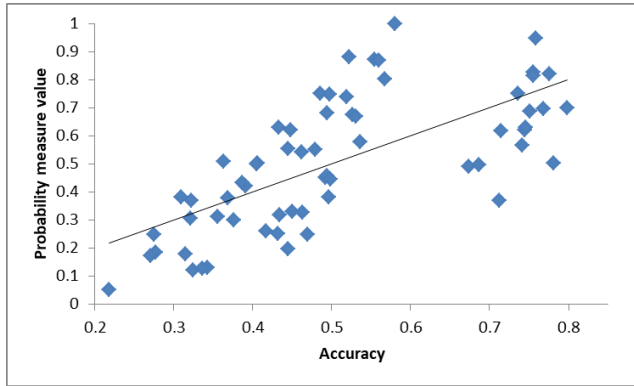


Fig. 5.13 Accuracies versus probability measure values for Eigenfaces and FERET database.

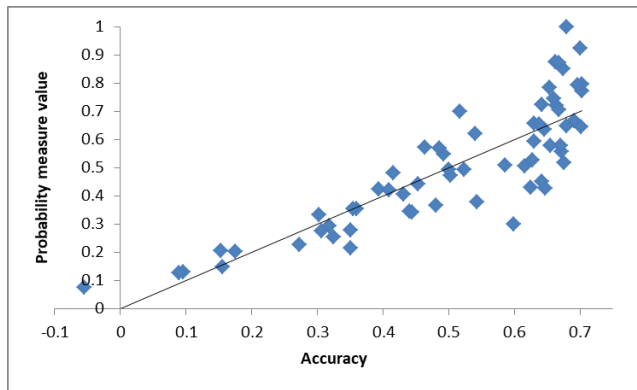


Fig. 5.14 Accuracies versus probability measure values for Fisherfaces and FERET database.

Table 5.3 A Sugeno fuzzy measure parameter, minimal and maximal differences between the obtained fuzzy measure and recognition rates on the concatenated facial regions.

Technique	λ value	Min. difference	Max. difference
Eigenfaces (AT&T)	-0.98944	0.11	0.31
Fisherfaces (AT&T)	-0.9995	0.06	0.17
Eigenfaces (FERET)	0.82529	0.04	0.78
Fisherfaces (FERET)	-0.9608	0.02	0.24

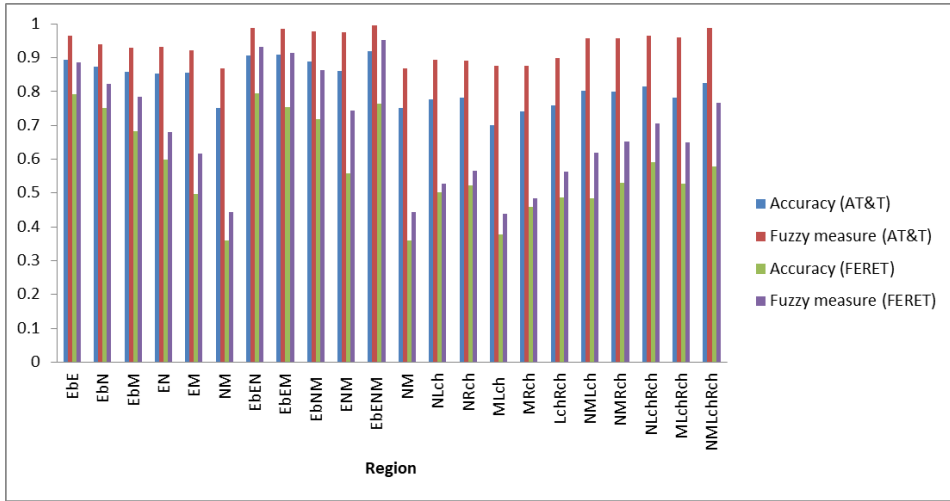


Fig. 5.15 Fisherfaces accuracy and corresponding fuzzy measures for merged regions in case of the upper part (bins on the left) and lower part (bins on the right) of face.

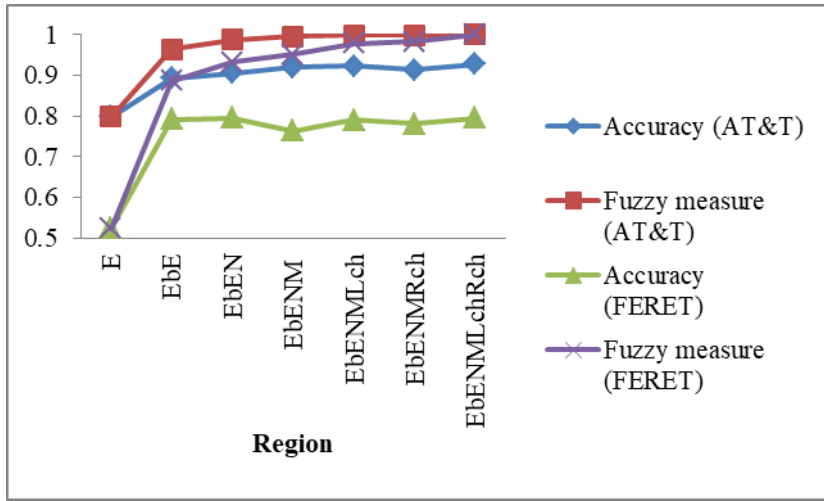


Fig. 5.16 Fisherfaces accuracy and the fuzzy measure values for the eyes area gradually augmented by other areas.

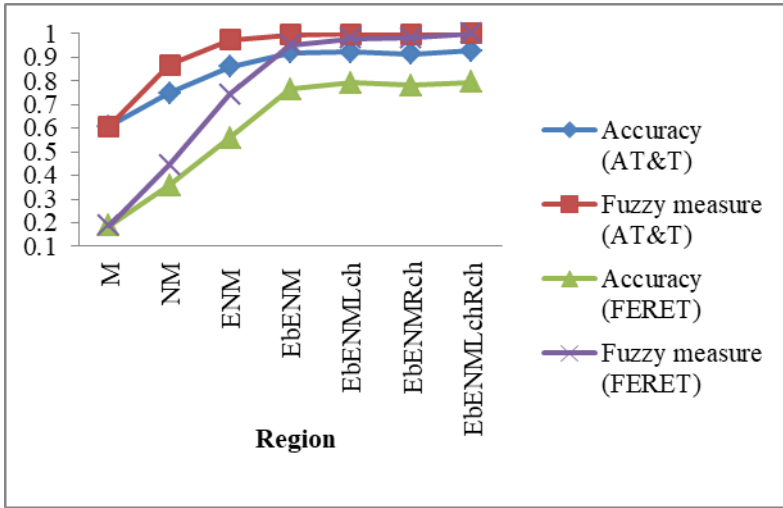


Fig. 5.17 Fisherfaces accuracy and the fuzzy measure values for the mouth area gradually augmented by other areas.

The following conclusions can be deduced from the results. First, for any face region, particularly, eyes, the accuracy of the recognition is better when the region of interest is getting bigger. In addition, the rate of recognition increases when considered is the combination of important facial parts. Second, the most important and descriptive area is the eyebrows and eyes region (more generally, the upper part of the face). In particular, the eyebrows are very significant in the computational face recognition process with over 81% accuracy in case of the AT&T database. Their occurrence in the area of consideration can significantly increase the rate of recognition. It is depicted in Fig. 5.24. Note that augmenting the region of eyes by any other areas of face does not increase the performance of the method in a meaningful fashion, see Fig. 5.23.

Another important observation is that the λ -fuzzy measure seems to be strongly sub-additive with $\lambda \leq -0.9608$. This property occurs in all considered cases. The exception is the Eigenfaces method for the FERET dataset. The explanation is that the accuracies are low and the method is ineffective. Hence, the λ is positive. This fact comes from the boundary conditions (1) and (2). However, fuzzy measure may be treated as a very good source of evidence for important regions of the face since it corresponds to the recognition rates yielded for the combined regions. This observation means that the interactions between the salient areas are reflected by Sugeno λ -fuzzy measure. This fact is easily seen from the plots depicting the scatter between the values of the measure and real recognition rates. Note that the points far away from the regression line correspond to atomic areas and the merges of areas underrating the accuracies, for instance left and right cheeks or nose segments.

In addition, the correlation between the Sugeno fuzzy measure and accuracy in the case of six atomic facial areas is less than the correlation in the case of four areas, see Fig. 5.15. It leads to the conclusion that if the chosen parts of face are occluded one can take into account only the most important parts, e.g., eyebrows and eyes with combination with the segments being available. However, the general tendency is that both Sugeno fuzzy measure and accuracy increase when the greater number of face segments is concatenated (merged). The greatest differences between accuracy and fuzzy measure are observed for lower regions of the face, e.g., mouth. These segments are considered as less useful classifiers than, for example, the area of eyes. One can note that the Sugeno fuzzy measure slightly overvalues the potential importance (weight) of information covered in these areas.

Let us compare the scatter between the accuracies of classifiers, fuzzy measure, and probability measure presented in Fig. 5.9 and Fig. 5.13. It shows that the fuzzy measure is flexible and its values are closer to the scaled accuracies than the probability measures. Again, in the case of Eigenfaces carried on the FERET dataset the scatters are similar. Recall that this is the above-discussed case of low efficiency of the method.

Finally, consider the experiments presented in the work by Matthews (1978). The subjects answered a question about dissimilarity and similarity of images being modified: one or more of the facial features such as eyes, eyebrows, nose, mouth, chin, or hair were changed. We removed the last feature (hair) from our considerations since it is an external face area. The face images were built with a police "Identikit" from the transparent overlays of face areas. A part of results along with Sugeno fuzzy measure values are included in Fig. 5.18. Accuracies observed in the psychological experiments and in our series of experiments with the PCA followed by LDA technique for chosen facial regions are presented in Fig. 5.19. A similar comparison with the Sugeno fuzzy measure is depicted in Fig. 5.20.

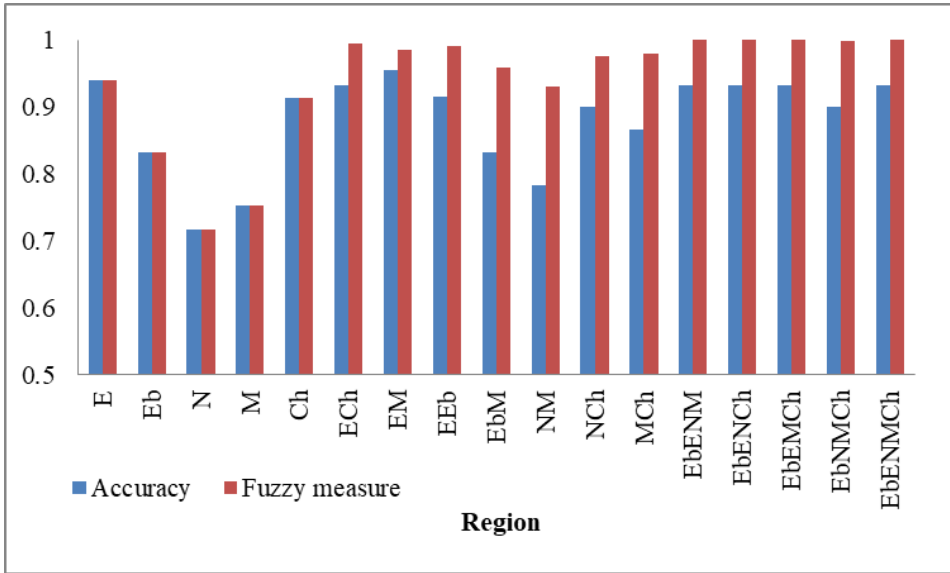


Fig. 5.18 Fuzzy measure built from the psychological experiments' results in comparison to recognition accuracies. Abbreviations are similar as in the previous case, Ch denotes a chin area.

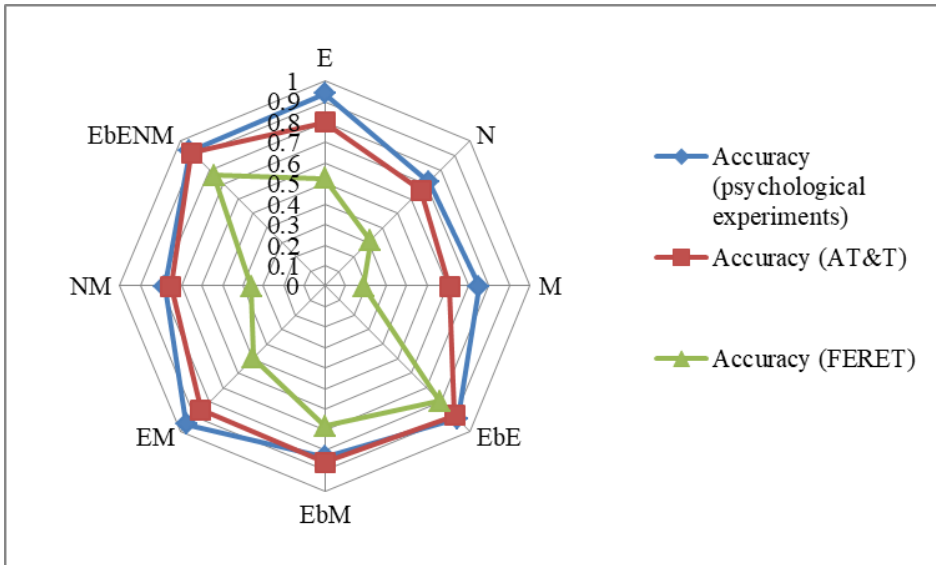


Fig. 5.19 Accuracies obtained in experiments by Matthews (1978) and in computational experiments using Fisherfaces.

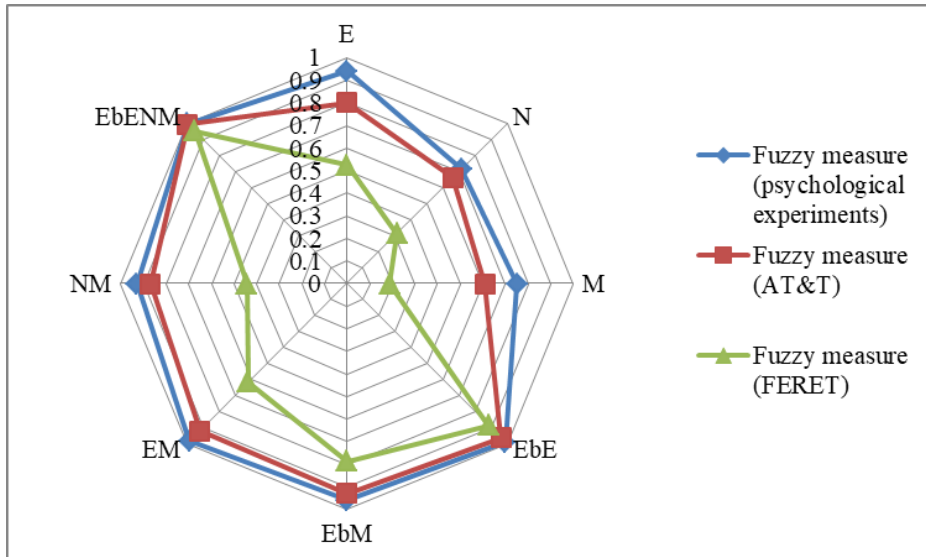


Fig. 5.20 A comparison of the values of fuzzy measure obtained on a basis of results of psychological and computational experiments.

On the basis of human recognition accuracies the value of parameter λ is -0.99994 while the correlation between accuracies and computed Sugeno fuzzy measures for chosen elements of face is 0.846 . This leads to the conclusion that the λ -fuzzy measure reflects the manner humans recognize others. Moreover, it can be applied to the process of modeling the interactions between the features of face. In the Fig. 5.19 it is presented that in the process of human face recognition on a basis of particular facial features and their concatenations (combinations) have very similar meaning as well as their importance is comparable with the importance in the process of computational identification and that the relationships between their values are kept. Similar situation, as a consequence of this fact, takes place in the fuzzy measure values case.

The last series of experiments is based on the division of the individuals into 8 sets A_1, \dots, A_8 in such a way that $A_1 \subset A_2 \subset \dots \subset A_8$. The subsets of AT&T dataset are built as follows: The first is contained of five subjects, second contains ten individuals, etc. Similarly, FERET subsets are built of 25, 50, ..., 200 individuals, respectively. As in the previous experiments, we observe the recognition accuracies with Eigenfaces and Fisherfaces methods. Next, the fuzzy measure is constructed with respect to the densities being the accuracies for the atomic face areas (eyebrows, eyes, nose, mouth, left cheek, right cheek). These values of λ are shown in Fig. 5.21 and Fig. 5.22. Note that λ tends to -1 with decreasing number of people in the dataset and with the efficiency of the method (i.e., if it reaches the highest accuracies). The explanation lays in boundary conditions on the fuzzy measure. It tends to satisfy them by the results

overvaluation. Moreover, note that the value of λ is linearly dependent on the number of people in the set in case of Eigenfaces method.

The fuzzy measure values and their dependence on the number of classes from each dataset are depicted in Fig. 5.23 and Fig. 5.24. In this case four lower and four upper combinations of facial segments were considered. The measure is stable in the case of effective methods, e.g., Fisherfaces with five training images per person obtained for AT&T or if the important facial part is considered, e.g., the eye region with its surrounding. But in the cases of less effective methods such as Eigenfaces, the fuzzy measure decreases with increasing number of classes. It is strictly associated with the real accuracies which decrease very similarly.

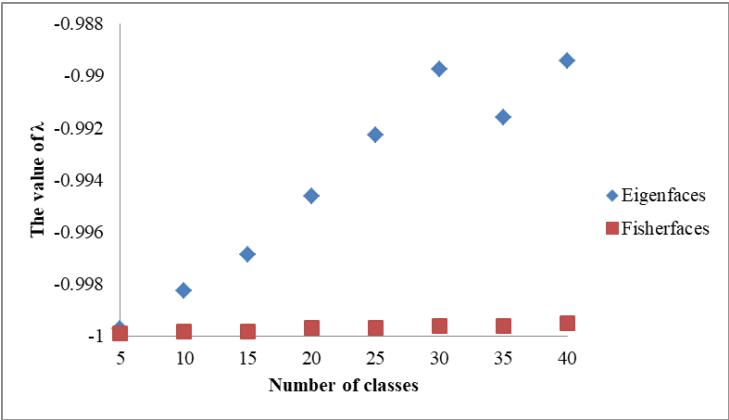


Fig. 5.21 λ values for AT&T.

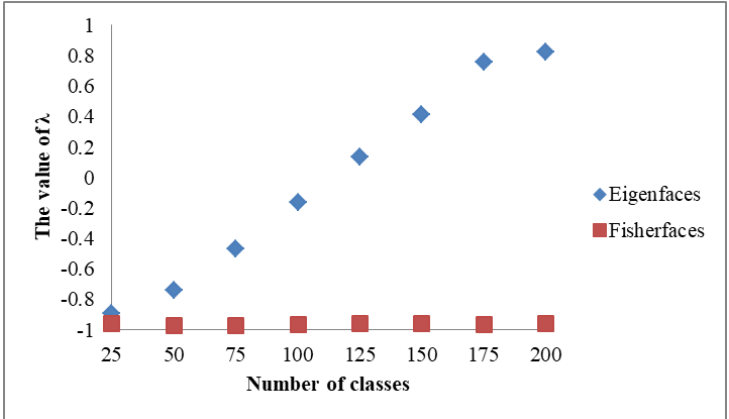


Fig. 5.22 λ values for FERET.

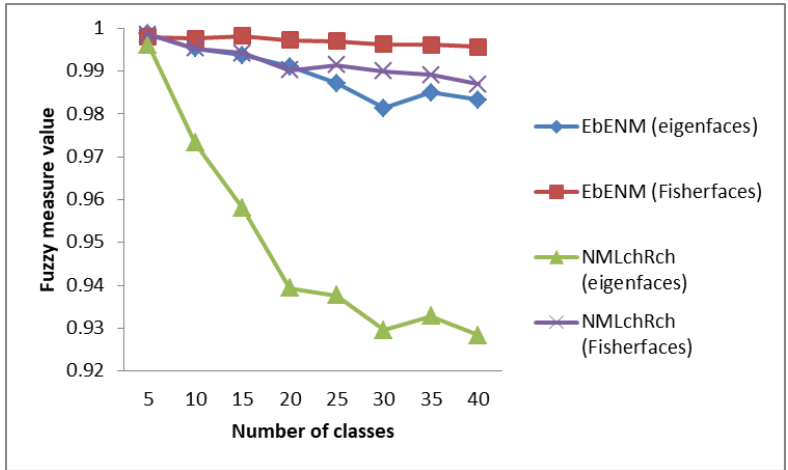


Fig. 5.23 Fuzzy measure values obtained for combinations of areas for AT&T dataset in dependence on the number of classes.

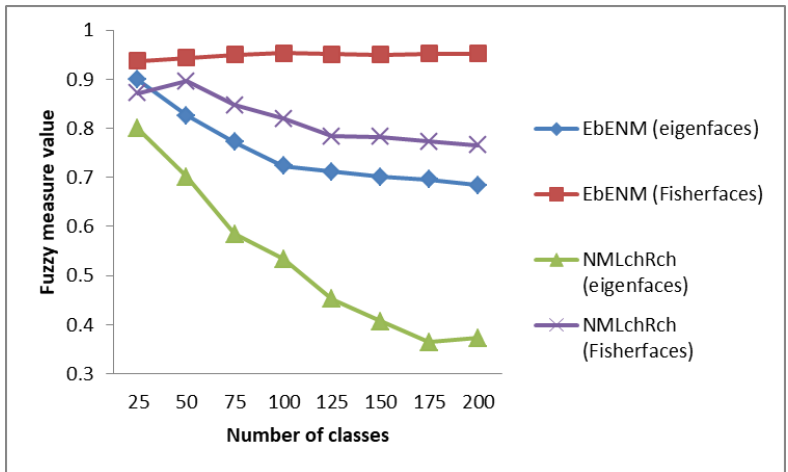


Fig. 5.24 Fuzzy measure values obtained for combinations of areas for FERET dataset in dependence on the number of classes.

5.5. Sugeno Fuzzy Measure Densities Evaluation

In the previous part of this chapter, we it was stated that the Choquet integral is required to return as high as possible values when the patterns (elements) belong to the same class and low values when they belong to different classes. Therefore, we propose an optimization method which maximizes the Choquet integral values for intra-class relations and minimizes them for inter-class relations, i.e., one has to find the values of fuzzy measure densities $g(x_i)$ ($i = 1,$

..., n). They are associated with importance (weights) of classifiers related to the methods which they represent or to the particular facial regions.

The above situations will be referred as *positive* and *negative* optimization, respectively. Therefore, the task can be simply defined as finding the facial regions saliency or the importance of classifiers in these two classification aspects. Here, a suitable optimization vehicle can be Particle Swarm Optimization. In our situation the parameters to be optimized are $g(x_i)$ ($i = 1, \dots, n$). The particles and their velocities are initially randomly set. Next, their positions (\mathbf{x}_i , $i = 1, \dots, n$) are updated according to the following manner (see the preliminary chapter for details):

$$\mathbf{v}_i = \frac{1}{2}(1 + r)\mathbf{v}_i + 2\mathbf{r}_1 \otimes (\mathbf{p}_i - \mathbf{x}_i) + 2\mathbf{r}_2 \otimes (\mathbf{p}_g - \mathbf{x}_i) \quad (5.5)$$

$$\mathbf{x}_i = \mathbf{x}_i + \mathbf{v}_i \quad (5.6)$$

see formulae (1.1) and (1.2) for a detailed description. Note that this version of PSO incorporates the random inertia weight weight concept reducing the time of execution to reach the satisfying level of convergence after realization of a series of generation.

It is worth noting that the process of optimization is relatively sensitive on random numbers r , \mathbf{r}_1 , and \mathbf{r}_2 . Therefore, it should be repeated many times to obtain the final average results. It means that this approach can be difficult to use in real time systems. Thus, one can find the densities of the fuzzy measure once. Next, they can be updated successively not necessarily during the face recognition system execution. In addition, the *negative* and *positive* results of optimization can differ significantly since the first of them is responsible for verification that two given images are not of the same individual while the second is more helpful in statement that the two images are of the same person. Moreover, the most important question is: How to connect these two approaches? The most intuitive answer is to find the weighted average in the most optimal way. It can be found during the daily practice and on the basis of trial-error-based method in dependence on the image dataset structure. Fig. 5.25 depicts an overall processing scheme of our proposal.

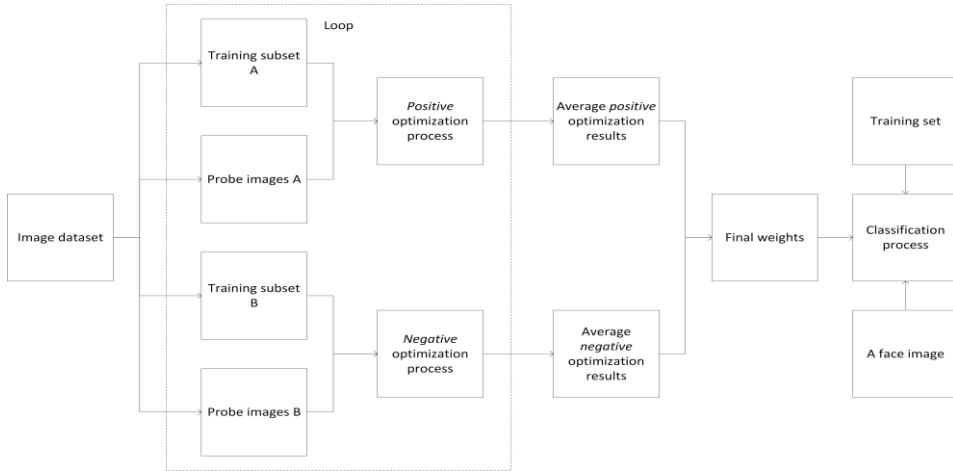


Fig. 5.25 An overall scheme of processing.

5.6. Experimental Results

Here, we analyse the series of experiments with four face image databases. We examine the efficiency of the method of determining the classifiers importance (saliency) on a basis of introduced concept of *negative* and *positive* optimization and compare them with the weights obtained from votes of the experts and the recognition rates of particular classifiers treated as the weights.

5.6.1. AT&T

First, we analyse the series of experiments with AT&T database. The images from the set were cropped and scaled. Moreover, the images were partitioned. An example of such partition is shown in Fig. 5.2 (F – whole (cropped) face, Eb – eyebrows, Eo – eyes only, N – nose, M – mouth, Lch – left cheek, Rch – right cheek) and in Fig. 5.26 (E – eye region, Ee, En, Em – extended areas of eyes, nose, and mouth, respectively). We asked over 30 subjects (our students and lab members) to give the weights of particular facial features in the human or computational face recognition. The results (average) are shown in Table 5.4.

Next, we carry the series of numerical experiments. We run the process of optimization to find the *negative* and *positive* results for each of the eight sets of face regions. The average values after 100 series of experiments in two cases: (a) two randomly chosen images to the training set per person, (b) five randomly selected photos of one person to the training set, respectively, are enlisted in Table 5.5. The considered facial parts were eyebrows, eyes only, nose, mouth, cropped face, left and right cheek areas. Table 5.5 presents also similar results for the collection of extended eyes, extended nose, extended mouth, and cropped face regions. The reference methods were the simple forms of Local Binary Patterns, Multi-scale Block LBP, Full Ranking, and Chain Code-Based Local

Descriptor, Principal Component Analysis with Canberra distance between the vectors representing features, and Linear Discriminant Analysis with cosine distance measure.

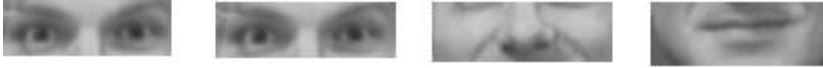


Fig. 5.26 Chosen parts of face (eye area and extended parts of eyes, nose, and mouth).

Table 5.4 Saliences of face areas yielded from the experts (questionnaires).

Set of features	Weights
Eyebrows, eyes only, nose, mouth	0.12; 0.43; 0.23; 0.21
Eyebrows, eyes only, nose, mouth, cropped face	0.08; 0.2; 0.12; 0.11; 0.49
Eyebrows, eyes only, nose, mouth, cropped face, left cheek, right cheek	0.08; 0.2; 0.11; 0.11; 0.39; 0.05; 0.05
Eyebrows, eyes only, nose, mouth, left cheek, right cheek	0.12; 0.34; 0.2; 0.18; 0.08; 0.08
Eyes, nose, mouth	0.52; 0.25; 0.23
Eyes, nose, mouth, cropped face	0.28; 0.13; 0.12; 0.46
Extended eyes, extended nose, extended mouth	0.46; 0.27; 0.26
Extended eyes, extended nose, extended mouth, cropped face	0.26; 0.16; 0.15; 0.43

The data gathered in Table 5.5 show the differences between the saliency of various facial regions in the processes of aggregation of classifiers. In addition, noteworthy is the fact that the abilities of *positive* classification and *negative* classification are at different levels. However, in the case for Eigenfaces and Fisherfaces this general rule is not confirmed.

Next, the results of optimization (both *negative* and *positive*) were used in the role of an input to the process of identification where the λ -fuzzy measure densities are built in the following form:

$$g_i = \alpha w_{i,\text{positive}} + (1 - \alpha) w_{i,\text{negative}} \quad (5.7)$$

where i is associated with the number of feature (or classifier), $w_{i,\text{negative}}$ and $w_{i,\text{positive}}$ are a *negative* and *positive* weight values, respectively, α changes from 0.05, 0.1, ... to 1. Here, the most important question arises: Which (if any) g_i does produce the best recognition rates? To answer this question one has to find a suitable α value. Fig. 5.27 depicts the example results obtained for various classifiers after the dataset division into two parts with equal number of images. Note that, for instance, the choice of α in case of LDA with cosine distance is not important while in case of the CCBLD descriptor the parameter $\alpha=0.35$ gives the best results of classification. Moreover, it is easy to observe that it is possible to find the classifiers importance with no presence of experts.

Now, we are interested in a comprehensive examination for which techniques or parts of a face one should ask experts, seek the optimal weight values, or conduct the initial process of identification. In Table 5.6 there are presented the winning forms of obtaining the fuzzy densities applied in the aggregation

processes with respect to specific set of facial parts. Cl. denotes the situation where the best aggregation result was obtained on a basis of the initial results. Ex. means that the experts found the best densities while the specific value is related with some parameter α . One can conclude that in case of more than four classifiers both initial results and experts cannot establish the densities. A possible explanation is that the experts may have difficulties in estimation of the weights for such divisions of face and if the regions' overlapping is possible. If the experts see the regions as relatively vast and are able to notice more details then the weights providing is easier. Moreover, if one analyzes the methods resulting in high recognition rates, it may be observed that they can provide relatively optimal weights during the initial pre-tests. The case of local descriptors it better to proceed with optimization methods since they are chosen in their simplest forms.

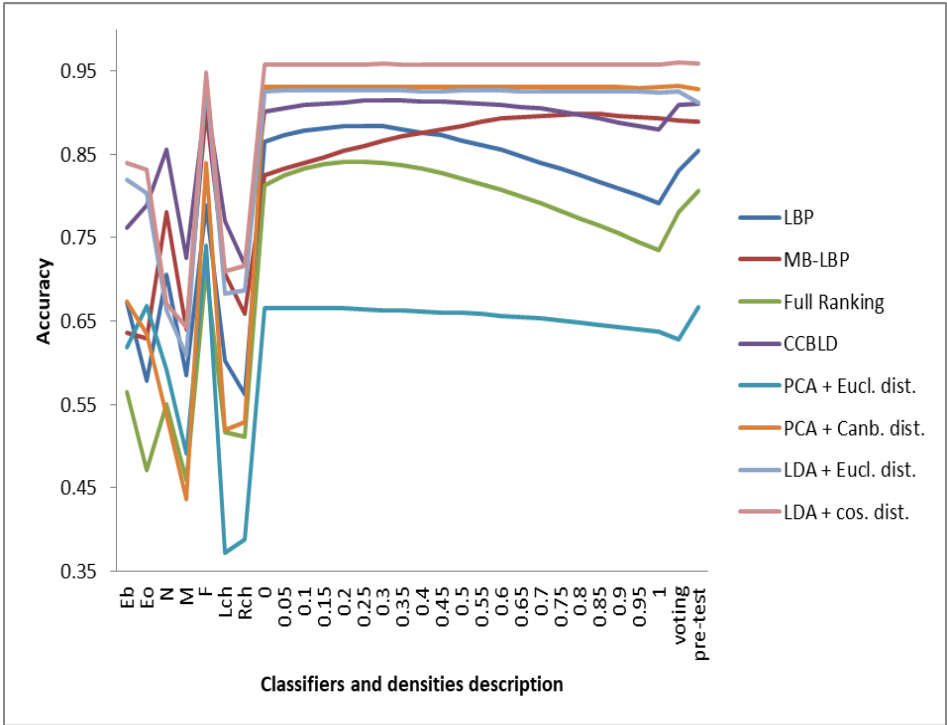


Fig. 5.27 The results of aggregation of classifiers. First 7 bars are for the particular classifiers based on facial parts, next 21 are for α which connects the *negative* and *positive* weights. Two last bars describe weights obtained from initial pre-tests and questionnaires. Initial pre-tests are the normalized results of classification with respect to particular classifiers.

Table 5.5 *Negative* and *positive* weight values found in the optimization process when two/five images per person were taken to the training set.

Two images per person in the training set (Eb, Eo, N, M, Lch, Rch)												
	LBP		MB-LBP		Full Ranking		CCBLD		Eigenfaces		Fisherfaces	
Optimization:	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>
Eyebrows	0.03	0.35	0.01	0.43	0.03	0.24	0.02	0.43	0.11	0.3	0.07	0.3
Eyes only	0.26	0.02	0.07	0.03	0.28	0.03	0.11	0.04	0.2	0.04	0.24	0.06
Nose	0.32	0.04	0.28	0.12	0.32	0.03	0.4	0.06	0.05	0.03	0.1	0.04
Mouth	0.27	0.03	0.17	0.02	0.26	0.02	0.17	0.02	0.07	0.02	0.16	0.01
Cropped face	0.04	0.37	0.34	0.26	0.04	0.41	0.24	0.3	0.51	0.1	0.38	0.15
Left cheek	0.03	0.12	0.07	0.09	0.03	0.17	0.03	0.1	0.03	0.3	0.02	0.14
Right cheek	0.05	0.07	0.06	0.05	0.03	0.1	0.02	0.05	0.03	0.21	0.02	0.29
Five images per person in the training set (Eb, Eo, N, M, Lch, Rch)												
	LBP		MB-LBP		Full Ranking		CCBLD		PCA		LDA	
Optimization:	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>
Eyebrows	0.03	0.31	0.01	0.41	0.03	0.26	0.02	0.37	0.17	0.2	0.16	0.14
Eyes only	0.29	0.03	0.06	0.04	0.28	0.02	0.11	0.05	0.16	0.14	0.16	0.15
Nose	0.32	0.04	0.29	0.1	0.32	0.03	0.38	0.07	0.04	0.03	0.04	0.05
Mouth	0.25	0.03	0.18	0.01	0.27	0.02	0.19	0.01	0.03	0.03	0.04	0.02
Cropped face	0.04	0.39	0.33	0.29	0.04	0.41	0.24	0.35	0.51	0.5	0.5	0.53
Left cheek	0.03	0.12	0.06	0.1	0.03	0.16	0.03	0.1	0.05	0.05	0.05	0.05
Right cheek	0.03	0.07	0.07	0.04	0.03	0.1	0.03	0.04	0.05	0.04	0.04	0.05
Two images per person in the training set (Ee, En, Em, F)												
	LBP		MB-LBP		Full Ranking		CCBLD		PCA		LDA	
Optimization:	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>
Ee	0.28	0.04	0.05	0.08	0.25	0.04	0.01	0.04	0.18	0.49	0.31	0.47
En	0.07	0.1	0.02	0.29	0.04	0.08	0.07	0.08	0.03	0.11	0.04	0.1
Em	0.05	0.04	0.77	0.01	0.7	0.04	0.48	0.04	0.02	0.01	0.07	0.01
F	0.6	0.83	0.16	0.62	0.01	0.85	0.44	0.85	0.77	0.39	0.58	0.42
Five images per person in the training set (Ee, En, Em, F)												
	LBP		MB-LBP		Full Ranking		CCBLD		PCA		LDA	
Optimization:	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>
Ee	0.61	0.04	0.05	0.06	0.26	0.03	0.01	0.19	0.27	0.27	0.18	0.25
En	0.03	0.09	0.02	0.3	0.04	0.06	0.03	0.04	0.03	0.03	0.04	0.03
Em	0.34	0.04	0.79	0.01	0.7	0.05	0.53	0.02	0.01	0.02	0.02	0.02
F	0.01	0.83	0.14	0.63	0.01	0.86	0.43	0.75	0.69	0.69	0.76	0.7

5.6.2. FERET Database

Now, we consider the subset of FERET dataset consisted of 200 grayscale images (3 images per person). The experiments are organized similarly as in the previous case with the exception that the set of methods is narrowed to Fisherfaces with Euclidean and cosine norms. Local descriptor-based classifiers did not produce satisfying results. Again, the face images were preprocessed. First, they were cropped and scaled. Next, the histograms were equalized and, finally, they were splitted onto subregions.

The parts of face were chosen in the same way as in the AT&T case. We do not carry new experiments with experts and work with the weights obtained for the AT&T set. The average values (after 100 series of experiments) of *negative* and *positive* optimization methods for two sets of regions (eyebrows, eyes only, nose, mouth, cropped face, left cheek, right cheek and extended eyes, extended nose, extended mouth, and cropped face areas) are listed in Table 5.7.

Note that in case of relatively small seven regions the proportions between *negative* and *positive* optimization results relatively vary. In the case of wider parts of face (e.g., extended parts) *negative* and *positive* weights are quite similar. Fig. 5.28 depicts the example aggregation results for the collection of classifiers based on eyebrows, eyes only, nose, mouth, cropped face, left cheek, and right cheek areas for Fisherfaces with Euclidean and cosine norms and Eigenfaces with Canberra norm. Finally, Table 5.7 contains the results of finding the best possible weights to set the fuzzy densities. From Fig. 5.28 and Table 5.7 it is possible to observe that the classifiers' initial values are relatively easier to obtain for the small parts of the face but in case of the whole images included into the process of aggregation, the optimization method seems to be a good choice to find the densities of fuzzy measure. Moreover, for the FERET dataset the weights evaluated by experts are not as well as the weights found from the computations.

Table 5.6 Winning forms of obtaining the densities of fuzzy measure.

Classifier	No. Train. imag.	LBP	MB-LBP	Full Ran.	CCB LD	PCA +Eucl. dist.	PCA +Canb. dist.	LDA +Eucl. dist.	LDA +cos. dist.
Eb + Eo + N	2	Cl.	Cl.	Cl.	Cl.	Cl.	Cl.	Cl.	Cl.
+ M	5	Cl.	Ex.	Cl.	Cl.	Cl.	Cl.	Cl.	Cl.
Eb + Eo + N	2	0.35	0.1	0.35	Cl.	Cl.	0.25	Cl.	Cl.
+ M + F	5	0.3	0.15	0.25	Cl.	Cl.	Ex.	Ex.	Ex.
Eb + Eo + N	2	0.25	0.1	0.35	0.3	0.05	0.25	0.4	0.4
+ M + F + Lch + Rch	5	0.25	0.8	0.25	0.3	Cl.	Ex.	0.2	Ex.
Eb + Eo + N	2	0.3	Cl.	0.35	0.3	0.05	Cl.	0.55	0.45
+ M + Lch + Rch	5	0.25	0.2	0.25	Cl.	Cl.	Cl.	Cl.	Cl.
E + N + M	2	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	Ex.	Ex.
	5	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	Ex.	Ex.
E + N + M + F	2	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	0.7	Ex.
	5	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	Ex.	Ex.
Ee + En + Em	2	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	Ex.	Ex.
	5	Cl.	Cl.	Cl.	Cl.	Ex.	Ex.	Ex.	Ex.
Ee + En + Em + F	2	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	Cl.	Cl.
	5	Cl.	Cl.	Cl.	Cl.	Cl.	Ex.	Ex.	Ex.

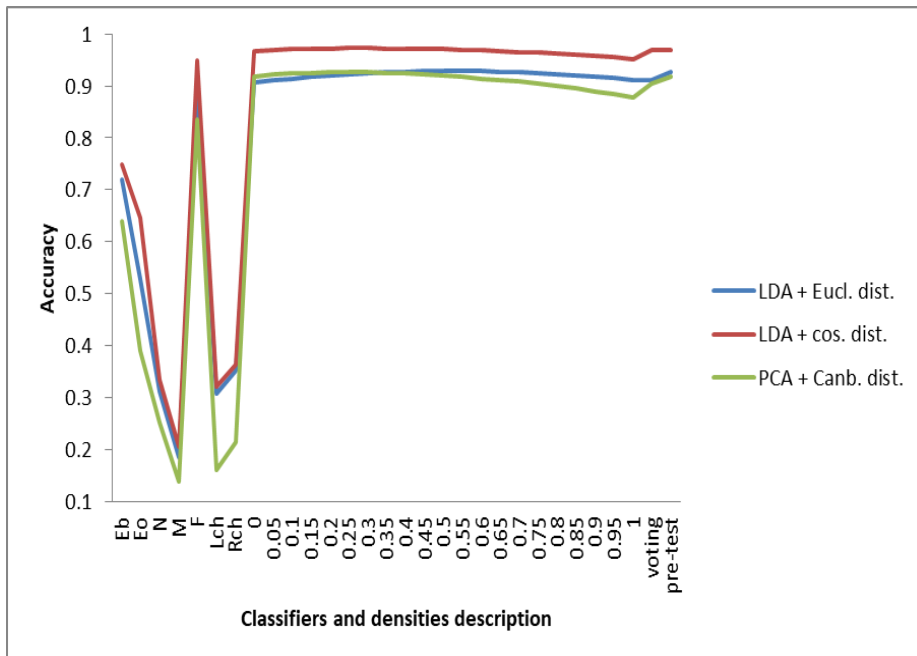


Fig. 5.28 The results of aggregation of classifiers.

Table 5.7 *Negative* and *positive* weights obtained in the optimization process.

Optimization:	Eigenfaces		Fisherfaces (Eucl. dist.)		Fisherfaces (cos. dist.)	
	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>
Eyebrows	0.11	0.45	0.09	0.46	0.05	0.42
Eyes only	0.21	0.08	0.12	0.06	0.27	0.06
Nose	0.06	0.07	0.17	0.04	0.08	0.05
Mouth	0.06	0.02	0.13	0.02	0.08	0.02
Cropped face	0.5	0.29	0.45	0.26	0.44	0.29
Left cheek	0.03	0.04	0.02	0.09	0.03	0.07
Right cheek	0.03	0.05	0.02	0.07	0.03	0.09
Optimization:	Eigenfaces		Fisherfaces (Eucl. dist.)		Fisherfaces (cos. dist.)	
	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>	<i>neg.</i>	<i>pos.</i>
Extended eyes	0.29	0.4	0.27	0.38	0.24	0.36
Extended nose	0.02	0.02	0.03	0.02	0.03	0.02
Extended mouse	0.02	0.01	0.01	0.01	0.02	0.02
Cropped face	0.67	0.57	0.68	0.58	0.71	0.6

Finally, we conduct the experiments with various methods of face recognition applied to the whole facial images. The chosen methods are: Eigenfaces with Canberra norm, Fisherfaces with cosine norm, Chain Code-Based Local Descriptor, Local Binary Pattern, and Multi-Scale Block Local Binary Pattern with three pixel blocks. The last methods (local descriptors) are implemented with their best settings, i.e., giving best recognition rates. The

results are gathered in Table 5.9. Moreover, there are contained the weights obtained in the set of initial experiments. It is easy to see that there is a slight correlation between the *positive* weights and the normalized classification results of the methods. The winning option (the one which produces the highest accuracy) is the combination of *negative* and *positive* densities obtained on the basis of optimization with the balance parameter $\alpha = 0.1, 0.15, \text{ or } 0.2$, see Fig. 5.29.

Table 5.8 Winning forms of finding the fuzzy measure densities (FERET).

Classifiers	Eigenfaces	Fisherfaces (Eucl. dist.)	Fisherfaces (cos. dist.)
Eb + Eo + N + M	Cl.	Cl.	Cl.
Eb + Eo + N + M + F	0.4	0.25	0.25
Eb + Eo + N + M + F + Lch + Rch	0.55	0.3	0.25
Eb + Eo + N + M + Lch + Rch	Cl.	Cl.	Cl.
E + N + M	Cl.	Cl.	Cl.
E + N + M + F	0.65	0.5	0.85
Ee + En + Em	Cl.	Cl.	Cl.
Ee + En + Em + F	0.95	1	0.55

Table 5.9 Weights related to the methods.

	<i>Negative</i> optimization	<i>Positive</i> optimization	Normalized initial results
Principal Component Analysis	0.47	0.06	0.2
Linear Discriminant Analysis	0	0.77	0.23
Chain Code-Based Local Descriptor	0.17	0.05	0.19
Local Binary Pattern	0.27	0.06	0.19
Multi-Scale Block Local Binary Pattern	0.1	0.06	0.19



Fig. 5.29 The values of α in relation to the accuracy of the aggregation of various classifiers for the whole face FERET images.

5.6.3. Yale Dataset

In a series of 100 repetitions of experiments we used four local descriptors, namely Multi-Scale Block Local Binary Pattern and Chain Code-Based Local Descriptor with 5 pixel width and 7 pixel width block versions. The training and testing sets in each series of experiments was built as follows: Five images from each class were randomly chosen to fill this set while five other random images built the testing one. Similarly to the FERET case, the weights obtained in the experiments are listed in Table 5.10. One can see here that the combination of MBLBP and CCBLD methods leads to the conclusion that *negative* and *positive* optimization can slightly differ and that they stand in an opposition to each other. It is a relative kind of problem. Therefore, the normalized results of initial experiments appear here as the best weights, i.e., the best aggregation accuracy was obtained for the fuzzy measure built on their basis.

Table 5.10 Weights assigned to the CCBLD and MBLBP methods.

	<i>Negative</i> optimization	<i>Positive</i> optimization	Normalized initial results
Chain Code-Based Local Descriptor	0.94	0	0.33
Multi-Scale Block Local Binary Pattern (5 px)	0.02	0.66	0.33
Multi-Scale Block Local Binary Pattern (7 px)	0.04	0.34	0.34

5.6.4. LFW Dataset

The last series of experiments was conducted with the usage of Labelled Faces in the Wild images, specifically on a basis of the version of this dataset containing cropped images. Two groups of images were selected to obtain the *negative* and *positive* optimization results, respectively, while the third group was built to verify the results. Note that in contrast to the other considered sets of images, here one image group is used only to verify results.

The group of images used to obtain the *negative* weights were built of images of 187 individuals having exactly four images and of 187 other individuals (one image per person), i.e., training and testing set, respectively. The second group was created of people with exactly five pictures in the LFW. Four images per person were randomly selected to the training set. The rest images constituted the testing set. The verifier set was consisted of the pictures of 311 individuals with exactly six images. Four images per person were randomly chosen as training and one as testing set. The weights found in the numerical experiments using Chain Code-Based Local Descriptor, Local Binary Pattern, and Multi-Scale Block Local Binary Pattern with 5 px block size are presented in Table 5.11. It is worth to add that the classification results for the combination of the obtained *negative* and *positive* weights were better than the results obtained with weights being the normalized initial classification results. The optimal value of α is 0.05.

Table 5.11 Weights associated with the methods for LFW.

	<i>Negative</i> optimization	<i>Positive</i> optimization	Normalized initial results
Chain Code-Based Local Descriptor	0.01	0.95	0.35
Local Binary Pattern	0.93	0.02	0.35
Multi-Scale Block Local Binary Pattern	0.06	0.04	0.29

5.6.5. General Results

Here, we present a summary of the above experiments. The average normalized (to the maximal values set as 1) recognition rates for all the 155 sets of experiments with classification based on aggregation of classifiers are detailed in Fig. 5.30. The plot suggests that the factor related to *positive* optimization slightly outweighs. However, *a priori* evaluation of a proper value of α highly depends on the type of classifier.

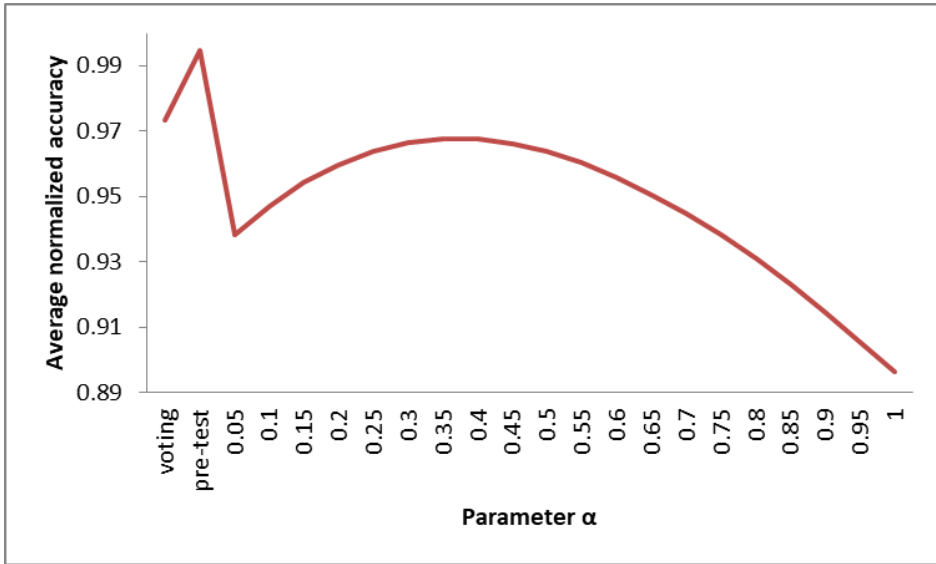


Fig. 5.30 Average (normalized) values of 155 experiments carried in the study in relation to the values of the parameter α as well as the densities obtained in the initial tests (pre-tests) and with experts' presence (voting).

5.7. Conclusions

In the first part of this chapter, we have considered an application of the Sugeno fuzzy measure as a vehicle to quantify a manner of aggregation of discriminatory information covered in facial segments. We have thoroughly analyzed the monotonicity and additivity properties in the context of face classification with relation to the salient facial areas. The series of experiments show that the fuzzy measure is an important vehicle to aggregate the pieces of knowledge which resides within the areas of the face. On the basis of the most cases one can conclude that the fuzzy measure appears as an effective classification model. This fact is strictly related to its monotonicity property.

Next, we have proposed a novel and original approach to obtain the optimal values of the densities of the fuzzy measure being the base of the important concept of classification on a basis of aggregation with the Sugeno λ -fuzzy measure and Choquet integral. The densities of the measure relate to the saliency of facial features and the importance of particular classifiers, i.e., methods not related with particular facial features. Our approach has been thoroughly compared with the expert-based method and has shown its high efficiency in the series of experiments carried for the classical face recognition methods such as Eigenfaces, Fisherfaces, and local descriptors.

References

- Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with Local Binary Patterns. In: Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science 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.* **20**, 128–140 (2018)
- Alrjebi, M., Liu, W., Li, L.: Face recognition against pose variations using multi-resolution multiple colour fusion. *Int. J. Mach. Intell. Sens. Signal Process.* **1** (4), 304–320 (2016)
- Bonnen, K., Klare, B. F., Jain, A. K.: Component-based representation in automated face recognition. *IEEE Trans. Inf. Forensics Secur.* **8**, 239–253 (2013)
- Brunelli, R., Poggio, T.: Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1042–1052 (1993)
- Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern histograms for face recognition. In: Lee, S.-W., Li, S.Z. (Eds.): *ICB 2007*, Lecture Notes in Computer Science **4642**, pp. 809–818 (2007)
- Chan, C. H., Yan, F., Kittler, J., Mikolajczyk, K.: Full ranking as local descriptor for visual recognition: A comparison of distance metrics on s_n . *Pattern Recognit.* **48**, 1328–1336 (2015)
- Choi, E., Lee, S.-W., Wallraven, C.: Face recognition with enhanced local Gabor binary pattern from human fixations. In: *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, pp. 863–867 (2012)
- Da, B., Sang, N., Li, C.: Face recognition by estimating facial distinctive information distribution. In: Zha, H., Taniguchi, R., Maybank, S. (Eds.) *Computer Vision – ACCV 2009, Part III*, LNCS **5996**, pp. 570–580 (2010)
- Dargham, J. A., Chekima, A., Hamdan, M.: Hybrid component-based face recognition system. In: Omatu, S. et al. (Eds.): *Distributed Computing and Artificial Intelligence*, *Adv. Intell. Soft. Comput.* **151**, 573–580 (2012)
- Davies, G., Ellis, H., Shepherd, J.: Cue saliency in faces as assessed by the ‘Photofit’ technique. *Percept.* **6**, 263–269 (1977)
- Dolecki, M., Karczmarek, P., Kiersztyn, A., Pedrycz, W.: Utility functions as aggregation functions in face recognition. In: *Proc. 2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, Athens, pp. 1–6 (2016)
- Ekenel, H. K., Stiefelhagen, R.: Local appearance based face recognition using discrete cosine transform. In: *Proceedings of the 13th European Signal Processing Conference (EUSIPCO)*. Antalya, pp. 2484–2488 (2005)

- Ekenel, H. K., Stiefelhagen, R.: Generic versus salient region-based partitioning for local appearance face recognition. In: Tistarelli, M., Nixon, M. S. (Eds.): *Advances in Biometrics*, LNCS **5558**, pp. 367–375 (2009)
- Ellis, H. D., Shepherd, J. W., Davies, G. M.: Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Percept.* **8**, 431–439 (1979)
- Fang, F., Qing, L., Wang, C., Miao, J., Chen, X., Gao, W.: Attention driven face recognition, learning from human vision system. *Int. J. Comput. Sci. Issues* **8**, 8–13 (2011)
- Grabisch, M.: Fuzzy integral in multicriteria decision-making. *Fuzzy Set. Syst.* **69**, 279–298 (1995)
- Graves, K. E., Nagarajah, R.: Uncertainty estimation using fuzzy measures for multiclass classification. *IEEE Trans. Neural Netw.* **18**, 128–140 (2007)
- Gutta, S, Philomin, V., Trajković, M.: An investigation into the use of partial faces for face recognition. In: *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 28–33 (2002)
- Gutta, S, Wechsler, H.: Partial faces for face recognition: Left vs right half. In: Petkov, N., Westenberg, M. A. (Eds.): *Computer Analysis of Images and Patterns*, LNCS **2756**, pp. 630–637 (2003)
- Haddadnia, J., Ahmadi, M.: N-feature neural network human face recognition. *Image Vis. Comput.* **22**, 1071–1082 (2004)
- Haig, N. D.: Exploring recognition with interchanged facial features. *Percept.* **15**, 235–247 (1986)
- Heikkilä, M., Pietikäinen, M., Schmid, C.: Description of interest regions with Local Binary Patterns. *Pattern Recognit.* **42**, 425–436 (2009)
- Heisele, B., Blanz, V.: Morphable models for training a component-based face recognition system. In: Zhao, W., Chellapa, R. (Eds.): *Face Processing, Advanced Modeling and Methods*, Elsevier, pp. 439–462 (2005)
- Heisele, B., Ho, P., Wu, J., Poggio, T.: Face recognition: component-based versus global approaches. *Comput. Vis. Image Underst.* **91**, 6–21 (2003)
- 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)
- Huang, J., Heisele, B., Blanz, V.: Component-based face recognition with 3D morphable models. In: *Audio- and Video-Based Biometric Person Authentication*, LNCS **2688**, pp. 27–34 (2003)
- Jarillo, G., Pedrycz, W., Reformat, M.: Aggregation of classifiers based on image transformations in biometric face recognition. *Mach. Vis. Appl.* **19**, 125–140 (2008)

- Johnston, R. A., Edmonds, A. J.: Familiar and unfamiliar face recognition: A review. *Mem.* **17**, 577–596 (2009)
- 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)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W., Rutka, P.: Chain code-based local descriptor for face recognition. In: Burduk, R., Jackowski, K., Kurzyński, M., Woźniak, M., Żołnierek A. (Eds.): *Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015*, pp. 307–316 (2016)
- Keller, J. M., Gader, P., Tahani, H., Chiang, J.-H., Mohamed, M.: Advances in fuzzy integration for pattern recognition. *Fuzzy Set Syst.* **65**, 273–283 (1994)
- Kennedy, J. F., Eberhart, R. C., Shi, Y.: *Swarm intelligence*. Academic Press, San Diego (2001)
- Kwak, K.-C., Pedrycz, W.: Face recognition using fuzzy integral and wavelet decomposition method. *IEEE Trans. Syst. Man Cybern. B Cybern.* **34**, 1666–1675 (2004)
- Kwak, K.-C., Pedrycz, W.: Face recognition: A study in information fusion using fuzzy integral. *Pattern Recognit. Lett.* **26**, 719–733 (2005)
- Lam, K.-M., Yan, H.: An analytic-to-holistic approach for face recognition based on a single frontal view. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 673–686 (1998)
- Lee, Y., Marshall, D.: Curvature based normalized 3D component facial image recognition using fuzzy integral. *Appl. Math. Comput.* **205**, 815–823 (2008)
- Li, B., Lian, X.-C., Lu, B.-L.: Gender classification by combining clothing, hair and facial component classifiers. *Neurocomput.* **76**, 18–27 (2012)
- Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S. Z.: Learning multi-scale block Local Binary Patterns for face recognition: In: *Advances in Biometrics, International Conference, ICB 2007, Lecture Notes in Computer Science* **4642**, pp. 828–837 (2007)
- Liu, Z., Liu, C.: Fusion of color, local spatial and global frequency information for face recognition. *Pattern Recognit.* **43**, 2882–2890 (2010)
- 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*, Springer, pp. 437–449 (2014)
- 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, Springer, pp. 59–70 (2015)

Matthews, M. L.: Discrimination of Identikit constructions of faces: Evidence for a dual processing strategy. *Percept. Psychophys.* **23**, 153–161 (1978)

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)

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)

Neo, H. F., Teo, C. C., Teoh, A. B. J.: A study on optimal face ratio for recognition using part-based feature extractor. In: *Signal-Image Technologies and Internet-Based System, 2007. SITIS '07. Third International IEEE Conference on*, pp. 735–741 (2007)

Neo, H. F., Teo, C. C., Teoh, A. B. J.: Development of partial face recognition framework. In: *Computer Graphics, Imaging and Visualization (CGIV), 2010 Seventh International Conference on*, pp. 142–146 (2010)

O'Donnell, C., Bruce, V.: Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Percept.* **30**, 755–764 (2001)

Park, U., Jillela, R. R., Ross, A., Jain, A. K.: Periocular biometrics in the visible spectrum. *IEEE Trans. Inf. Forensics Secur.* **6**, 96–106 (2011)

Pedrycz, W.: Fuzzy sets in pattern recognition: Methodology and methods. *Pattern Recognit.* **23**, 121–146 (1990)

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

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

Pujol, F. A., García, J. C.: Computing the principal Local Binary Patterns for face recognition using data mining tools. *Expert Syst. Appl.* **39**, 7165–7172 (2012)

Robins, E., Susilo, T., Ritchie, K. L., Devue, C.: Within-person variability promotes learning of internal facial features and facilitates perceptual discrimination and memory. doi: 10.31219/osf.io/5scnm (2018)

Rotshtein, P., Geng, J. J., Driver, J., Dolan, R. J.: Role of features and second-order spatial relations in face discrimination, face recognition, and individual face skills: Behavioral and functional magnetic resonance imaging data. *J. Cogn. Neurosci.* **19**, 1435–1452 (2007)

- Sadr, J., Jarudi, I., Sinha, P.: The role of eyebrows in face recognition. *Percept.* **32**, 285–293 (2003)
- Sato, K., Shah, S., Aggarwal, J. K.: Partial face recognition using radial basis function networks. In: *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 288–293 (1998)
- Savvides, M., Abiantun, R., Heo, J., Park, S., Xie, C., Vijayakumar, B. V. K.: Partial & holistic face recognition on FRGC-II data using support vector machine kernel correlation feature analysis. In: *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, doi:10.1109/CVPRW.2006.153 (2006)
- Savvides, M., Kumar, B. V. K. V., Khosla, P. K.: “Corefaces” – robust shift invariant PCA based correlation filter for illumination tolerant face recognition. In: *Computer Vision and Pattern Recognition, CVPR 2004, Proceedings of the 2004 IEEE Computer Society Conference on*, pp. 834–841 (2004a)
- Savvides, M., Kumar, B. V. K. V., Khosla, P. K.: Eigenphases vs. Eigenfaces. In: *Pattern Recognition 2004, ICPR 2004, Proceedings of the 17th International Conference on*, pp. 810–813 (2004b)
- Shepherd, J., Davies, G., Ellis, H.: Studies of cue saliency. In: Davies, G., Ellis, H. D., Shepherd, J. W. (Eds.): *Perceiving and Remembering Faces*. Academic Press, New York, pp. 105–131 (1981)
- Sinha, P., Balas, B., Ostrovsky, Y., Russell, R.: Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proc. IEEE* **94**, 1948–1962 (2006)
- Sugeno, M.: *Theory of fuzzy integral and its applications*. Dissertation. Tokyo Institute of Technology, Tokyo (1974)
- Teo, C. C., Neo, H. F., Teoh, A. B. J.: A study on partial face recognition of eye region. In: *Machine Vision, 2007, ICMV 2007, International Conference on*, pp. 46–49 (2007)
- Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**, 71–86 (1991)
- Venkat, I., Khader, A. T., Subramanian, K. G., De Wilde P.: Recognizing occluded faces by exploiting psychophysically inspired similarity maps. *Pattern Recognit. Lett.* **34**, 903–911 (2013)
- Woodard, D. L., Pundlik, S. J., Lyle, J. R., Miller, P. E.: Periocular region appearance cues for biometric identification. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 162–169 (2010)
- 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)

- Yan, B., Keller, J.: Conditional fuzzy measures and image segmentation. In: NAFIPS 1981-1991: A Decade of Growth in Uncertainty Modeling, Proceedings of NAFIPS-91. University of Missouri-Columbia, Missouri, pp. 32–36 (1991)
- Yan, G., Ma, G., Zhu, L.: Support vector machines ensemble based on fuzzy integral for classification. In: Advances in Neural Networks - ISNN 2006, LNCS **3971**, pp. 974–980 (2006)
- Yan, Y., Osadciw, L. A.: Intra-difference based segmentation and face identification. In: Jain, A. K., Ratha, N. K. (Eds.): Biometric Technology for Human Identification. Proceedings of SPIE **5404**, pp. 502–510 (2004)
- Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., Ellis, A. W.: Matching familiar and unfamiliar faces on internal and external features. *Percept.* **14**, 737–746 (1985)
- Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A.: Face recognition: A literature survey. *ACM Comput. Surv.* **35**, 399–458 (2003)

6. Aggregation Mechanism in Face Recognition

In this chapter, we consider an important technique of aggregation of classifiers' results which can be used in many problems of image recognition. A proper choice of an aggregation operator can essentially improve the classification rate. In the first part of the chapter, we discuss various techniques of aggregation and find the best ones in dependence of the applied measure of similarity between the vectors representing features. The second part is devoted to the efficient and useful method based on Choquet integral. We compare the performance of its various modifications and extensions.

6.1. Introduction

The techniques of information fusion, in general, and the aggregation functions (operators) in particular, realizing them, are applied to many fields of science and research, for instance, computer science, biology, economics, etc. (Torra and Narukawa 2007). The same situation is present in the field of face recognition where many works addressed the tasks of classification based on two or more classifiers. Such classifiers may be constructed on a basis of many regions (parts) of face, many 2D, 3D, or infrared facial images.

The face partitioning may be caused by many factors. The most important are the following: First, in real life problems, not always the whole face is visible but only its small parts. The face can be occluded, for instance, by a helmet, a mask, a veil, or sunglasses. Second, by choosing a few parts of a face we can reduce the size of data proceeded, and in some cases, the computational time and effort. Finally, an intuitively appealing fact is that even few methods of classification applied to the whole face images can significantly improve the recognition rate. However, one should have in mind that the time needed to execute the method can be drastically longer. Here, the question appears: How to find the best possible operator of aggregation of such classifiers. This type of operator should efficiently maximize the classification and minimize the possibility of incorrect identification of a face.

The main goal of this work is to find the best aggregation operators applied to the processes of the information integration. The information comes from different classification methods (classifiers). If we use the most intuitive methods based on the nearest neighbor classifier, they are essentially dependent on the distance (similarity) measures used in the process of vector of features comparison. Therefore, our aim is to find the best possible aggregation techniques in dependence on the distance measures used for one of the most popular (classical) classifiers, namely Fisherfaces (Belhumeur et al. 1997) applied to the well-known facial image datasets (AT&T and FERET). The parts of face used in the experimental series are eyebrows, eyes, nose, and mouth

areas. We carry a quantitative and comprehensive analysis of over 1000 aggregation functions published in the literature along with the guidelines of their application. The work is important not because of finding of the best possible aggregation function but also because it sheds the light on the ways of work of the most important functions considered as potential aggregators.

Our work is highly experimental and we implement methods appearing in the literature. Therefore, let us discuss the most important works appearing in the topic of aggregation operators and face recognition based on aggregation functions. The aggregation operators theory can be found in many monographs such as Beliakov et al. 2007; 2016, Torra & Narukawa 2007, Grabisch et al. 2009, Pedrycz et al. 2011, Calvo et al. 2014, Baczyński et al. 2017a; 2017b, etc. and in the articles or books on T-norms and co-norms (Alsina et al. 2006, Klement et al. 2000, Klement & Mesiar 2005, fuzzy sets and their applications (Grabisch 1995, Liu et al. 2012, Pedrycz & Gomide 1998), ordered weighted averaging operators (OWA, Yager & Kacprzyk 2012), and multi-criteria decision-making theory, e.g. Das & Guha 2015.

It is worth to stress that one of the main properties of aggregation functions are monotonicity and specific boundary conditions. One of the goals of the study presented in this chapter is to examine the generalizations of Choquet integral based on a kind of modification related with a structural flexibility by an incorporating, among others, t-norms instead of product operation under the sign of the integral. Bustince et al. (2016) and Lucca et al. (2016b) introduced such class of functions which is weaker if compared to the conditions of aggregation. The authors called this class using the term pre-aggregation functions. Their main property is a so-called directional monotonicity, i.e., monotonicity along a fixed direction. Therefore, our next main objective is to thoroughly examine this wide class of functions with respect to its effectiveness as the aggregation tool. We are interested in comparing their potential with the original fuzzy Choquet integral which is one of the most effective aggregation operators. We are going to design 25 classes of such integrals which are built using the t-norms instead of the product operation. The accuracies are tested using 4 sets of images: AT&T, FERET, Yale (Yale Face Database), and cropped LFW. The classifiers used in a comparison are Eigenfaces (PCA, Turk & Pentland 1991), Fisherfaces (PCA followed by PCA, Belhumeur et al. 1997), Full Ranking (Chan et al. 2015), Local Binary Pattern (Ahonen et al. 2004), Multi-Scale Block Local Binary Pattern (Chan et al. 2007, Liao et al. 2007), and Chain Code-Based Local Descriptor in an application to the whole facial images and their parts.

It is worth noting that in the image recognition literature there are present two approaches to the fusion of information. The first is based on the combination of information coming from the images (a data-level approach). The second one is based on the classification utilizing the inputs being the results of particular classifiers. This is the score-level approach.

The most known applications of information aggregation in the literature are as follows. Brunelli & Poggio 1993 utilized template matching strategy to yield an improvement of recognition result adding the scores for eyes, nose, mouth, and the whole face. Pentland et al. 1994 used the eigenfeatures approach to the most salient facial cues with a few strategies of aggregation. Similar approaches were proposed by Gottumukkal & Asari 2004 and Kim et al. 2005. Majority rule for RBF neural networks-based classifiers were used by Haddadnia & Ahmadi 2004. Kwak & Pedrycz (2005) proposed an integration of classification results based on the Fisherfaces approach utilizing the images of chosen facial features (eyes, nose, mouth, whole face). The aggregation operator was Choquet integral and fuzzy measure. A similar method was proposed by Kwak & Pedrycz (2004) to aggregate the results of classification based on an original facial images and their three kinds of wavelet decompositions. It is worth noting that the Choquet fuzzy integral as an aggregation operator was also used by Mirhosseini et al. 1998, Melin et al. 2005, Martínez et al. (2014; 2015). Newer results are obtained, among others, by Ekenel and Stiefelwagen 2009 (fourteen facial parts included), Jarillo et al. 2008 (majority voting and Bayesian product for various classifiers), Oh et al. 2013 (polynomial-based RBF neural networks aggregation), Radtke et al. 2014 (a selection and fusion of ensembles yielded through combination of Boolean classifiers), Tome et al. 2013 (fifteen facial features studied), Dolecki et al. 2016 (utility functions as aggregation operators), Kurach et al. 2014 (Granular Computing techniques), Campomanes-Alvarez et al. 2016 (modeling of the craniofacial correspondence in craniofacial superimposition by aggregation functions), Al-Hmouz et al. 2017 (a three-valued logic with decision based on fuzzy sets for multimodal facial images), Karczmarek et al. 2018 (an aggregation of linguistic descriptors-based techniques and computational methods), Kiersztyn et al. 2018 (a concept of a so-called multi-level aggregation), and many others. A comparison of fuzzy measure-based ensemble classifiers was presented by Agrawal et al. 2018.

Finally, it is worth to note that in the works by Ahonen et al. 2004, Bereta et al. 2013, Bharkad & Kokare 2011, Naveena et al. 2012, Perlibakas 2004, Xue et al. 2007, and Smiatacz 2016 the authors analyzed the influence of various chosen distance measures on the classifier performance. A general conclusion appears that the proper choice of the measure can essentially improve the accuracy of the method.

6.2. Aggregation Functions for Face Recognition

Consider the following problem. There is a system of face recognition based on comparing the regions of face. Such kind of system allows the situation that someone is classified as a person A when the eye area region is taken into account. However, when a region of mouth is utilized, he/she is identified as a person B. Finally, on a basis of the nose areas, he/she is supposed to be a person

C. This is a typical misclassification example. Here, a question appears: How to identify this individual properly? The answer is to find an aggregation function which will return a correct class on a basis of the specific classifiers' results. The input to the procedure of aggregation will be the distances between the vectors representing features or the rankings with assigned points. Of course, this is only the wish and it can be impossible to find such an *ideal* function. But an intuitively appealing fact is that the more number of classifiers the better final classification result.

Recall (see chapter 1) that an aggregation function can be defined as the function $f: [0,1]^n \rightarrow [0,1]$ having the properties (1.22) – (1.24).

The example problem discussed above is a typical task of the information fusion and multi-criteria decision-making theory. The authors such as Grabisch et al. (2009) suggest to use, for instance, weighted arithmetic mean or the aggregator operators based on the integral concept like Choquet integral. Similarly, at the level of data fusion the authors propose to use, again, Choquet integral or other concepts.

Here, we revise this problem. First, we conduct the processing stages as follows:

1. Preprocessing (internal face cropping, scaling, histogram equalization).
2. Determining the positions of the eye, eyebrows, nose, and mouth areas. The parameters of the regions were chosen in the initial sets of experiments.
3. Execution of the dimensionality reduction method, i.e., Fisherfaces for the determined regions.
4. A comparison of the feature vectors being the result of the method of dimensionality reduction coming from the training and testing sets with using 16 various distance measures.
5. Normalization of the distances to the interval $[0, 1]$.
6. Aggregation of the results using one of over 1000 aggregation operators.

The set of the aggregation functions was built on a basis of the well-known books describing aggregation operators with their applications such as Alsina et al. 2006, Beliakov et al. 2007; Grabisch et al. 2009, Pedrycz & Gomide 1998. Because of the size constraints we do not show all the formulas here. Only the functions producing the best classification accuracies are reported.

The special kind of the aggregation operators is the Choquet integral. Let us recall its concept in an application to the face recognition since it frequently appears in the further text. Assume that $X = \{x_1, \dots, x_n\}$ represents the whole facial area. Here x_1, \dots, x_n represent the particular facial parts like eyes, nose, etc.

Definitions of fuzzy measure and λ -fuzzy measure were recalled in chapter 1, see formulas (1.1) – (1.3) and (1.8), respectively. Moreover, a definition of the Choquet integral and the way it is constructed in an application to the face recognition problem is presented by the formulas (5.1) – (5.4).

6.3. Similarity Measures

In mathematics considered are many dissimilarity or similarity measures used to find the degree of matching of two vectors of the form $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$. In our experimental settings we utilize the functions as follows:

- Bray-Curtis

$$m(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |x_i + y_i|} \quad (6.1)$$

- Canberra

$$m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{|a_i - b_i|}{|a_i| + |b_i|} \quad (6.2)$$

- Chebyshev

$$m(\mathbf{x}, \mathbf{y}) = \max_i |x_i - y_i| \quad (6.3)$$

- χ^2 -statistics

$$m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{x_i + y_i} \quad (6.4)$$

- Correlation

$$m(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_{i=1}^n \left(\left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right) \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right) \right)}{\left(\sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{j=1}^n x_j \right)^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n \left(y_i - \frac{1}{n} \sum_{j=1}^n y_j \right)^2 \right)^{\frac{1}{2}}} \quad (6.5)$$

- Cosine

$$m(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_{i=1}^n (x_i y_i)}{\left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^n y_i^2 \right)^{\frac{1}{2}}} \quad (6.6)$$

- Euclidean

$$m(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n (x_i - y_i)^2 \right)^{\frac{1}{2}} \quad (6.7)$$

- Manhattan

$$m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n |x_i - y_i| \quad (6.8)$$

- Median of absolute differences

$$m(\mathbf{x}, \mathbf{y}) = \text{med}_i |x_i - y_i| \quad (6.9)$$

- Median of square differences

$$m(\mathbf{x}, \mathbf{y}) = \text{med}_i (x_i - y_i)^2 \quad (6.9)$$

- Modified Euclidean

$$m(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2} \quad (6.10)$$

- Modified Manhattan

$$m(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |x_i| \sum_{i=1}^n |y_i|} \quad (6.11)$$

- Squared Euclidean

$$m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n (x_i - y_i)^2 \quad (6.12)$$

- Weighted cosine

$$m(\mathbf{x}, \mathbf{y}) = 1 - \frac{\sum_{i=1}^n \frac{x_i y_i}{\sqrt{w_i}}}{(\sum_{i=1}^n x_i^2)^{\frac{1}{2}} (\sum_{i=1}^n y_i^2)^{\frac{1}{2}}} \quad (6.13)$$

- Weighted Manhattan

$$m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{|x_i - y_i|}{\sqrt{w_i}} \quad (6.14)$$

- Weighted squared Euclidean

$$m(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \frac{(x_i - y_i)^2}{\sqrt{w_i}} \quad (6.15)$$

In all the above formulas $\mathbf{w} = [w_1, w_2, \dots, w_n]$ is a vector of weights associated with the eigenvalues found during execution of the Fisherfaces procedure. The above functions are commonly used in related studies on face recognition.

6.4. Experimental Studies

Here, we discuss various techniques of aggregation in an application to the results returned by the basic classifiers built on four facial parts: Eyebrows area, eyes with no eyebrows region, nose segment, and mouth area, i.e., the most salient facial features. We test the efficiency of the most common aggregation functions in an application to the face recognition problem solved with the Fisherfaces method for the AT&T dataset and FERET database (its subsets called *ba*, *bk*, and *bj*). Taking on account the results of the previous chapter (i.e., the accuracies of Fisherfaces method for the specified parts of face) one can build the weights for all the aggregation functions appearing in our experiments and which need weights. One of such kind of functions is, for instance, the above-mentioned fuzzy Choquet integral. The weights with respect to eyebrows, eyes, nose, and mouth segments are: 0.28, 0.28, 0.23, and 0.21 in the case of AT&T and 0.42, 0.3, 0.18, and 0.1 in the case of FERET dataset, respectively.

Recall that in the settings presented here, we show the average results of the best accuracies obtained for each of the above discussed measures in an application to the Principal Component Analysis method followed by Linear Discriminant Analysis in an application to the image sets randomly divided into the training and testing sets as follows: 5 images of each person from the AT&T set were placed in the training set and the rest in the testing set. In the case of FERET database 2 images were training set while one image per person was treated as a probe. For each measure the experiments were repeated 200 times to get reliable results. In almost all cases we present maximally the best five aggregation functions provided that they gave satisfactory results, see Table 6.1. For chosen aggregation operators the tests were repeated to find the parameter of the function giving potentially the best results.

6.5. General Results

Table 6.1 Chosen results of classification with respect to various aggregation operator and distance measures.

Image set	Aggregation operator	Accuracy
<i>Euclidean distance</i>		
	$f(x, y) = \min(x, y) \min(1, x^p + y^q)$ for $p = 1, q = 2$	91.52
	$f(x, y) = \min(x\sqrt{y}, y)$ (EV-copula, extreme value copula)	91.5
	$C_\lambda^{FGM}(x, y) = xy + \lambda xy(1-x)(1-y)$ for $\lambda = 0.75$ (Farlie-Gumbel-Morgenstern copula)	91.46
AT&T	$C(x, y) = pxy + (1-p) \min(x, y)$ for $p = 0.75$ (general nonassociative symmetric copula)	91.46
	$RM_{w,\gamma}(x_1, \dots, x_n) = \left(\log_\gamma \sum_{i=1}^n w_i \gamma^{x_i} \right)^{-1}$ for $\gamma = 0.75$ (weighted radical mean)	91.45
	$f(x, y) = \min(x, y) \min(1, x^p + y^q)$ for $p = 2, q = 1$	91.45
	Fuzzy Choquet integral with mem. grades of the form $\mu_{ij} = \frac{1}{1 + \frac{d_{ij}}{\bar{d}_i}}$, i - classifier number, j - training image index, \bar{d}_i - average distance within classifier no. i , d_{ij} - distance between an unknown image and training image no. j within classifier no. i	84.59
	$\text{med}(x_1, \dots, x_n)$ (median)	84.54
FERET	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1 + \frac{d_{ij}}{\bar{d}_i}} \right)$	84.52
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1 + d_{ij}} \right)$	84.5
	Voting (i.e., one vote for any minimum in one of classifiers)	83.98
	$m_w(x_1, \dots, x_n) = \frac{1}{n} \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$ (weighted average)	82.61
<i>Squared Euclidean distance</i>		
	$f(x, y) = \min(x, y) \min(1, x^p + y^q)$ for $p = 0.9, q = 0.9$	91.54
	$T_{\lambda,w}^Y = \max\left(0, \left(1 - \sum_{i=1}^n w_i (1 - x_i)^\lambda\right)^{1/\lambda}\right)$ for $\lambda = 10$ (weighted Yager t-norm)	91.51
	$f(x, y) = \min(x\sqrt{y}, y)$ (EV-copula)	91.5
AT&T	$f(x, y) = \min(x, y)(x^p + y^q - x^p y^q)$ for $p = 0.9, q = 0.9$	91.5
	$C_\alpha(x, y) = \frac{\alpha^2(1-\alpha)}{2} \max(x + y - 1, 0) + (1 - \alpha^2)xy + \frac{\alpha^2(1+\alpha)}{2} \min(x, y)$ for $\alpha = -0.25$ (commutative, non-associative copula)	91.41
FERET	The order of operators as in the case of Euclidean distance with the accuracies 84.59, 84.29, 84.2, 84.03, and 82.49, respectively	

<i>Modified Euclidean distance</i>		
	$OWA_p(x_1, \dots, x_n) = (\sum_{i=1}^n (w_i x_{(i)}^r))^{1/r}$ for $r = 0.1$, $w_i = \frac{2i}{n(n+1)}$	90.7
	and decreasing $(x_{(1)}, \dots, x_{(n)})$ (power-based OWA)	
	$OWGM(x_1, \dots, x_n) = \prod_{i=1}^n x_{(i)}^{w_i}$ for $w_i = 1 - \frac{i-1}{n}$ and decreasing	90.66
	$(x_{(1)}, \dots, x_{(n)})$ (ordered weighted geometric mean)	
AT&T	$T_\alpha(x, y) = \left(1 + \frac{[(1+x)^{-\alpha-1}][(1+y)^{-\alpha-1}]}{2^{-\alpha-1}}\right)^{-1/\alpha} - 1$ for $\alpha = 10$	90.6
	$OWGM(x_1, \dots, x_n)$ for $w_1 = \frac{1}{2}, w_2 = \frac{1}{2}, w_3 = \frac{1}{4}, w_4 = \frac{1}{8}$ (the rest parameters are given as above)	90.58
	$T_\lambda^F(x, y) = \log_\lambda \left(1 + \frac{(\lambda^x - 1)(\lambda^y - 1)}{\lambda - 1}\right)$ for $\lambda = 0.0001$ (Frank t-norm)	90.58
<hr/>		
	$G_\alpha(x, y) = \log(e^{\alpha x} + e^{\alpha y})$ for $\alpha = 0.99$	86.48
	$EM(x_1, \dots, x_n) = \frac{1}{\alpha} \log \left(\frac{1}{n} \sum_{i=1}^n e^{\alpha x_i}\right)$ for $\alpha = 10$ (exp. mean)	86.36
	$T_{\lambda, w}^F(x_1, \dots, x_n) = \log_\lambda \left(1 + \frac{\prod_{i=1}^n (\lambda^{x_i} - 1)^{w_i}}{(1 - \lambda)^{\sum_{i=1}^n w_i - 1}}\right)$ for $\lambda = 0.25$	86.27
FERET	(weighted Frank t-norm)	
	$M(x_1, \dots, x_n) = \frac{GM(x_1, \dots, x_n)}{GM(x_1, \dots, x_n) - GM(1 - x_1, \dots, 1 - x_n)}$ for geom. mean	86.26
	$GM(x_1, \dots, x_n)$ (Kolesárová function)	
	$RM_{w, \gamma}(x_1, \dots, x_n) = (\log_\gamma (\sum_{i=1}^n w_i \gamma^{1/x_i}))^{-1}$ for $\gamma = 5$ (weighted radical mean)	86.26
<hr/>		
<i>Weighted squared Euclidean distance</i>		
	$T_{\lambda, w}^Y$ for $\lambda = 10$ (weighted Yager t-norm, see weighted Euclidean dist.)	85.32
AT&T	$f(x, y) = \min(x\sqrt{y}, y)$ (EV-copula)	85.29
	$OWGM(x_1, \dots, x_n)$ for $w_i = 1 - \frac{i-1}{n}$ (see mod. Eucl. dist.)	85.25
	Power-based OWA (see mod. Eucl. dist.)	85.24
	$T_\lambda^F(x, y)$ for $\lambda = 0.0001$ (Frank t-norm)	85.21
<hr/>		
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{1 + \frac{d_{ij}}{d_i}}$ (see Eucl. dist.)	80.31
	$\text{med}(x_1, \dots, x_n)$ (median)	79.94
FERET	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{\frac{d_{ij}}{d_i}}{1 + \frac{d_{ij}}{d_i}}\right)$	79.75
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1 + d_{ij}}\right)$	79.42
	$T_{\lambda, w}^Y$ for $\lambda = 0.25$ (weighted Yager t-norm, see weighted Eucl. dist.)	78.08
<hr/>		
<i>Manhattan distance</i>		
AT&T	$T_\lambda(x, y) = xy / (\lambda + (1 + \lambda)(x + y - xy))$ for $\lambda = 0.6$ (mod.	90.91

	Hamacher t-norm)	
	Mod. Hamacher t-norm for $\lambda = 0.5$	90.91
	$RM_{w,\gamma}(x_1, \dots, x_n)$ for $\gamma = 1.01$ (weighted radical mean, see Euclid. dist)	90.9
	$\sum_{i=1}^n (w_i/x_i)$ (weighted harmonic mean)	90.9
	$f(x, y) = \min(x\sqrt{y}, y)$	90.89
	Fuzzy Choquet integral (see Euclid. dist.)	82.96
	$\text{med}(x_1, \dots, x_n)$	82.9
FERET	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{\frac{d_{ij}}{d_i}}{1 + \frac{d_{ij}}{d_i}} \right)$	82.88
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1 + d_{ij}} \right)$	82.79
	Voting (as above)	82.18
	<i>Modified Manhattan distance</i>	
	$RM_{w,\gamma}(x_1, \dots, x_n)$ for $\gamma = 1.01$ (weighted radical mean)	90.61
	Weighted harmonic mean	90.59
AT&T	$T_{\alpha,\beta}(x, y) = \frac{1}{\beta} \left(\left(1 + \frac{((1+\beta x)^{-\alpha} - 1)((1+\beta y)^{-\alpha} - 1)}{(1+\beta)^{-\alpha} - 1} \right)^{-1/\alpha} - 1 \right)$ for $\alpha = -10, \beta = -0.5$	90.5
	$\sum_{i=1}^n (w_i/x_{(i)})$ for $w_1 = 1, w_2 = 0.9, w_3 = 0.8, w_4 = 0.7$ (ordered weight. harm. mean)	90.48
	$F_\alpha(x, y) = \alpha / \log(e^{\alpha/x} + e^{\alpha/y} - e^\alpha)$ for $\alpha = 0.1$	90.48
	$F_\alpha(x, y) = -\frac{1}{\alpha} \log(e^{-\alpha x} + e^{-\alpha y})$ for $\alpha = 0.99$ (Wiener-Shannon law extension)	84.99
FERET	$M_g(x_1, \dots, x_n) = \begin{cases} \frac{\sqrt[n]{\prod_{i=1}^n x_i}}{\sqrt[n]{\prod_{i=1}^n x_i + \sqrt[n]{\prod_{i=1}^n (1-x_i)}}} & \text{for } \{0, 1\} \not\subseteq \{x_1, \dots, x_n\} \\ 0 & \text{otherwise} \end{cases}$ (quasi-arithm. mean)	84.7
	$RM_{w,\gamma}(x_1, \dots, x_n)$ for $\gamma = 7$ (weighted radical mean)	84.7
	$M(x_1, \dots, x_n)$ (Kolesárová function, see mod. Euclid. dist.)	84.7
	Weighted Frank t-norm for $\lambda = 0.25$ (see mod. Euclid. dist.)	84.68
	<i>Weighted Manhattan distance</i>	
	Weighted harmonic mean	82.94
	$RM_{w,\gamma}(x_1, \dots, x_n)$ for $\gamma = 1.01$ (weighted radical mean)	82.93
	OWGM for $w_1 = 1, w_2 = \frac{3}{4}, w_3 = \frac{1}{2}, w_4 = \frac{1}{4}$	82.6
AT&T	Power-based OWA (see mod. Euclid. dist.) for $w_1 = \frac{1}{2}, w_2 = 1, w_3 = \frac{3}{2}, w_4 = 2$	82.6
	$T_\lambda(x, y) = xy / (\lambda + (1 + \lambda)(x + y - xy))$ for $\lambda = 0.5$ (Mod. Hamacher t-norm)	82.56
	<i>Chebyshev distance</i>	
AT&T	Mod. Hamacher t-norm for $\lambda = 0.5$ or $\lambda = 0.6$	87.67

	Mod. Hamacher t-norm for $\lambda = 1$	87.6
	$f(x_1, \dots, x_n) = \prod_{i=1}^n x_i^{w_i}$, weighted geom. mean, weighted product	87.58
	$G(x_1, \dots, x_n) = \left(\prod_{i=1}^n x_i^{w_i x_i^p} \right)^{1/\sum_{i=1}^n w_i x_i^p}$ for $p = 0.1$ (weighted Gini mean)	87.57
	Mod. Hamacher t-norm for $\lambda = 2$	87.56
<i>Cosine distance</i>		
	Mod. Hamacher t-norm for $\lambda = 0.5$	93.81
	Mod. Hamacher t-norm for $\lambda = 0.6$	93.81
	Mod. Hamacher t-norm for $\lambda = 1$	93.8
AT&T	$A(x, y) = \begin{cases} x^\alpha y^{1-\alpha} & \text{for } x \leq y \\ x^{1-\beta} y^\beta & \text{otherwise} \end{cases}$ for $\alpha = 0.75, \beta = 0.25$	93.8
	$f(x, y) = \min(x\sqrt{y}, y)$	93.79
	Voting	90.96
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1+d_{ij}} \right)$	90.79
FERET	Fuzzy Choquet integral with mem. 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
	Fuzzy Choquet integral (see Euclidean dist.)	90.35
<i>Weighted cosine distance</i>		
	OWA for $w_1 = \frac{1}{4}, w_2 = \frac{1}{8}, w_3 = \frac{1}{12}, w_4 = \frac{1}{16}$	84.79
	$OWGM(x_1, \dots, x_n) = (\min(x_1, \dots, x_n))^{1-\alpha} (\max(x_1, \dots, x_n))^\alpha$ for $\alpha=0.1$ (special OWGM)	84.65
AT&T	OWA for $w_1 = 1, w_2 = \frac{3}{4}, w_3 = \frac{1}{2}, w_4 = \frac{1}{4}$	84.61
	OWA for $w_1 = 1, w_2 = \frac{1}{2}, w_3 = \frac{1}{4}, w_4 = \frac{1}{8}$	84.56
	$E(x_1, \dots, x_n) = \log \sum_{i=1}^n (w_i \log \alpha^{x_i}) / \log \alpha$ for $\alpha = 10$ (weighted exp. mean)	84.31
FERET	The first four aggregation operators as in the case of cosine distance, the fifth operator is weighted average. The results are 87.77, 87.71, 87.7, 87.67, and 87.67, respectively	
<i>Correlation distance</i>		
	$f(x, y) = \min(x\sqrt{y}, y)$	93.8
	$A(x, y) = \begin{cases} x^\alpha y^{1-\alpha} & \text{for } x \leq y \\ x^{1-\beta} y^\beta & \text{otherwise} \end{cases}$ for $\alpha = 0.75, \beta = 0.25$	93.79
AT&T	$T_\lambda(x, y) = xy / (\lambda + (1 + \lambda)(x + y - xy))$ for $\lambda = 1$ (mod. Hamacher t-norm)	93.76
	Weighted geom. mean	93.76
	$T_\lambda(x, y)$ for $\lambda = 0.5$	93.75
	$G(x_1, \dots, x_n) = \left(\prod_{i=1}^n x_i^{w_i x_i^p} \right)^{1/\sum_{i=1}^n w_i x_i^p}$ for $p = 0.01$ (weighted	93.75

Gini mean	
FERET	The results are the same as in the cosine distance case with accuracies 90.73, 90.59, 90.51, 90.49, and 90.14, respectively
<i>Bray-Curtis distance</i>	
	$T(x, y) = xy/(x + y - xy)$ 93.83
	$f(x, y) = \min(x\sqrt{y}, y)$ 93.69
AT&T	Weighted harmonic mean 93.63
	$RM_{w,\gamma}(x_1, \dots, x_n)$ for $\gamma = 1.01$ (weighted radical mean) 93.63
	$A(x, y)$ for $\alpha = 0.75, \beta = 0.25$ (see correlation distance) 93.61
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1+d_{ij}} \right)$ 90.2
FERET	Fuzzy Choquet integral with mem. 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
	Fuzzy Choquet integral (see Euclidean distance) 90.05
<i>Canberra distance</i>	
	$T_\lambda(x, y) = xy/(\lambda + (1 + \lambda)(x + y - xy))$ for $\lambda = 20$ (mod. Hamacher t-norm) 91.43
	$T_\lambda(x, y)$ for $\lambda = e$ and $\lambda = 4$ (mod. Hamacher t-norm) 91.41
AT&T	$T_{\lambda,w}^Y$ for $\lambda = 2$ (weighted Yager t-norm, see weighted Euclid. dist.) 91.4
	$T_\lambda(x, y)$ for $\lambda = 2$ 91.4
	$T_\lambda(x, y)$ for $\lambda = 1$ 91.37
	Voting 83.92
	Weighted average 83.76
	Fuzzy Choquet integral with mem. grades $\mu_{ij} = \frac{1}{2} \left(1 + \frac{d_{ij}}{1+d_{ij}} \right)$ 83.44
FERET	Fuzzy Choquet integral with mem. 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
<i>χ^2-statistic measure</i>	
	$T_\lambda(x, y)$ for $\lambda = 0.0001$ (mod. Hamacher t-norm) 82.06
	Power-based OWA (see mod. Euclid. dist.) 81.7
	$f(x, y) = \min(x, y) \min(1, x^p + y^q)$ for $p = 0.5, q = 0.5$ 81.63
	$f(x, y) = \min(x, y)(x^p + y^q - x^p y^q)$ for $p = 0.5, q = 0.5$ 81.63
AT&T	$C_\alpha(x, y) = \max\left((x^{-\alpha} + y^{-\alpha} - 1)^{-\frac{1}{\alpha}}, 0\right)$ for $\alpha = 0.25$ (Clayton copula) 81.59
	$F_\alpha(x, y) = (x^{-\alpha} + y^{-\alpha} - 1)^{-\frac{1}{\alpha}}$ for $\alpha = 0.25$ 81.59
	$F_\alpha(x, y) = 1 - e^{-[(-\log(1-x))^{-\alpha} + (-\log(1-y))^{-\alpha}]^{-1/\alpha}}$ for $\alpha = 0.25$ 81.59
	T-norm generated by $g_\lambda^T(t) = \left(\frac{1-t}{t}\right)^\lambda$ for $\lambda = 0.25$ (Dombi) 81.59

	t-norm)	
	$F_\alpha(x, y) = \left(1 + \left(\left(\frac{1}{x} - 1\right)^\alpha + \left(\frac{1}{y} - 1\right)^\alpha\right)^{\frac{1}{\alpha}}\right)^{-1}$ for $\alpha = 0.25$	81.59
<i>Median of absolute differences</i>		
	$T_\lambda(x, y)$ for $\lambda = 0.5$ (mod. Hamacher t-norm)	83.12
	Weighted harmonic mean	83.09
	Weighted radical mean	83.08
AT&T	$RM_{w,\gamma}(x_1, \dots, x_n)$ for $\gamma = 1.01$ (weighted radical mean)	83.07
	$T_\lambda(x, y)$ for $\lambda = 1$ (mod. Hamacher t-norm)	83.03
	T-norm generated by $g_\lambda^{AA}(t) = (-\log t)^\lambda$ for $\lambda = 2$ Aczél- Alsina t-norm (Gumbel-Hougaard copula)	83.03
<i>Median of square differences</i>		
	T-norm generated by $g_\lambda^{AA}(t) = (-\log t)^\lambda$ for $\lambda = 2$ (see above)	83.03
	T-norm generated by $\log((\lambda + (1 - \lambda)t)/t)$ for $\lambda = 0.1$ (Hamacher t-norm)	83.00
AT&T	$T(x, y) = 1 / \left(1 + \left(\left(\frac{1-x}{x}\right)^p + \left(\frac{1-y}{y}\right)^p\right)^{\frac{1}{p}}\right)$ for $p = 0.5$ (Dombi t- norm)	82.98
	$F_\alpha(x, y) = \left(1 + \left(\left(\frac{1}{x} - 1\right)^\alpha + \left(\frac{1}{y} - 1\right)^\alpha\right)^{\frac{1}{\alpha}}\right)^{-1}$ for $\alpha = 0.5$	82.98
	$F_\alpha(x, y) = \exp(1 - ((1 - \log x)^\alpha + (1 - \log y)^\alpha - 1)^{1/\alpha})$ for $\alpha = 2$	82.96

Table 6.1 shows that in the case of Euclidean measure the best results for AT&T set are obtained with aggregation operator $f(x, y) = \min(x, y) \min(1, x^p + y^q)$ and its parameters $p = 1, q = 2$, or $p = 2, q = 1$. Since not all the possible parameters were used in the first series of experiments, we have arranged additional test series to find their optimal values. Fig. 6.1 shows that the best accuracies are yielded when the parameters p and q are of similar values (are placed close to the main diagonal) and are in the range 1 – 3. When the FERET set is considered Choquet integral produces the best results. Two of these results are obtained when a so-called compensation mechanism is used to construct the integral. In addition, median and weighted average can be a good choice here.

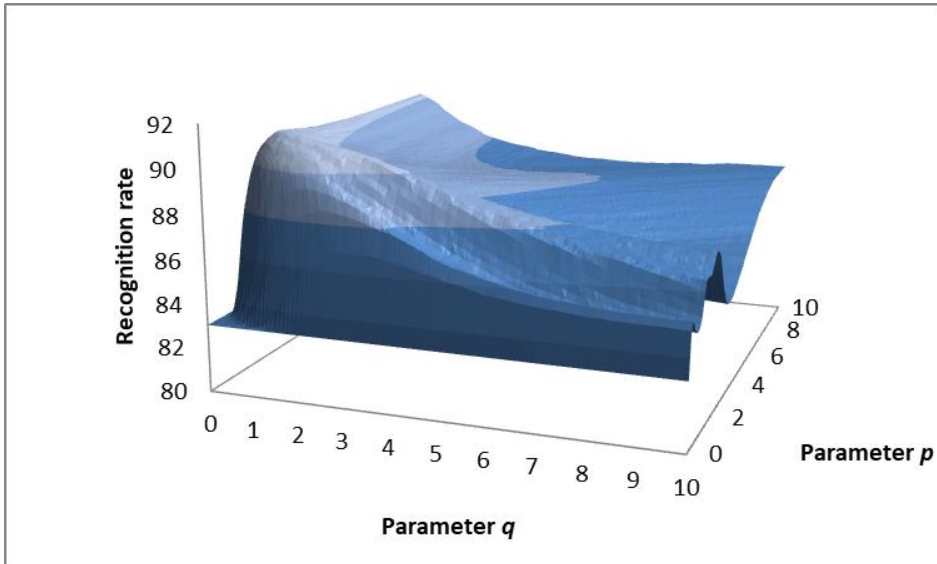


Fig. 6.1 The results of repeated experiments for the winning function in the case of Euclidean distance.

Relatively similar standings were generated for squared Euclidean measure. However, in this case the product $f(x, y) = \min(x, y) \min(1, x^p + y^q)$ gives the best accuracies for AT&T. The additional experiments carried for this operator for wider range of the parameters p and q have shown that the structure of the results is close to the plot depicted in the previous case, see Fig. 6.2.

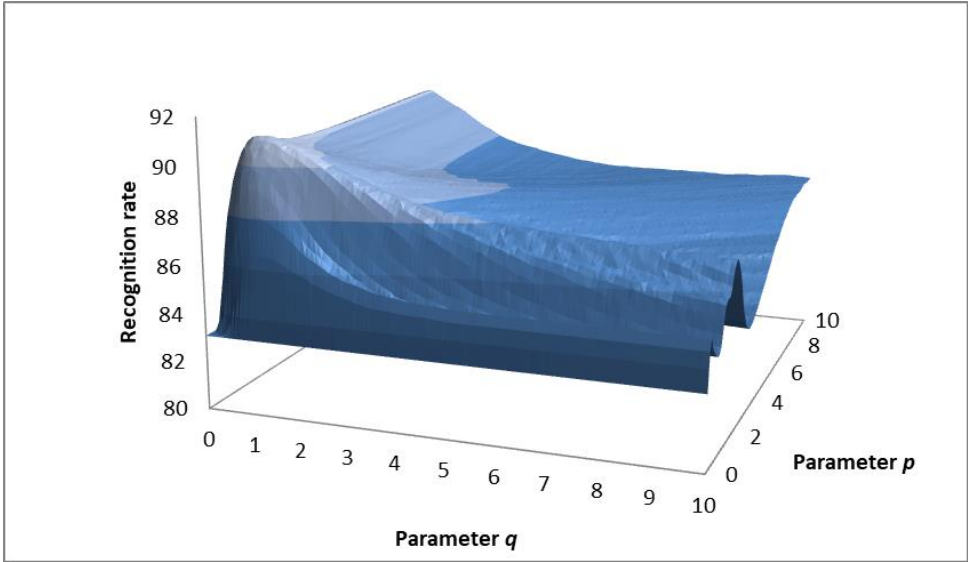


Fig. 6.2 The relation between the accuracies and winning function's parameters in the case of weighted squared distance.

In the case of modified Euclidean distance the families of Frank operators and OWAs dominate. The plot presented in Fig. 6.3 shows the results of repeated experiments for power-based ordered weighted average in relation to the value of the parameter r . To get the best accuracy one should choose relatively small values of r (less than 0.4). When FERET set is considered, the best results are found with the operator $G_\alpha(x, y)$ when $\alpha = 0.99$. But only the values of α being close to 1 guarantee good results here.

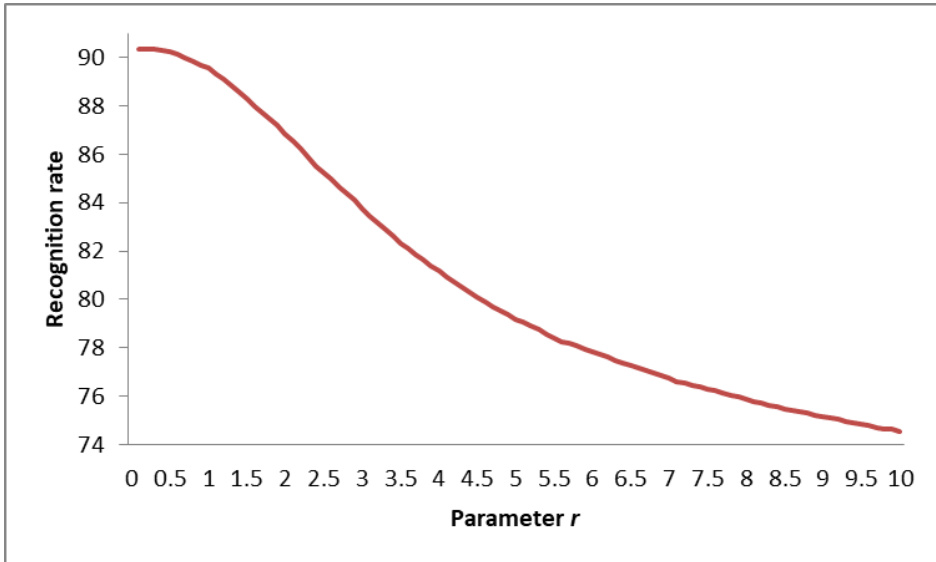


Fig. 6.3 The dependence of the results produced by power-based OWA operator on the values of its parameter.

The analysis of the results obtained with the weighted squared Euclidean distance and AT&T dataset shows that many functions can be applied here with a success (t-norms, multiplication and power functions' modifications, or OWA and OWGM). In the case of FERET dataset the fuzzy Choquet integral produces high accuracies.

Now, let us consider other than Euclidean distance-based measures. When Manhattan measure is under consideration, one can note that the Hamacher t-norm modifications and various weighted means can be sound alternatives. The repeated tests for the modified Hamacher function are illustrated in Fig. 6.4. This function can be a very good operator of aggregation for almost all of the measures of dissimilarity/similarity with relatively wide range of its parameter λ .

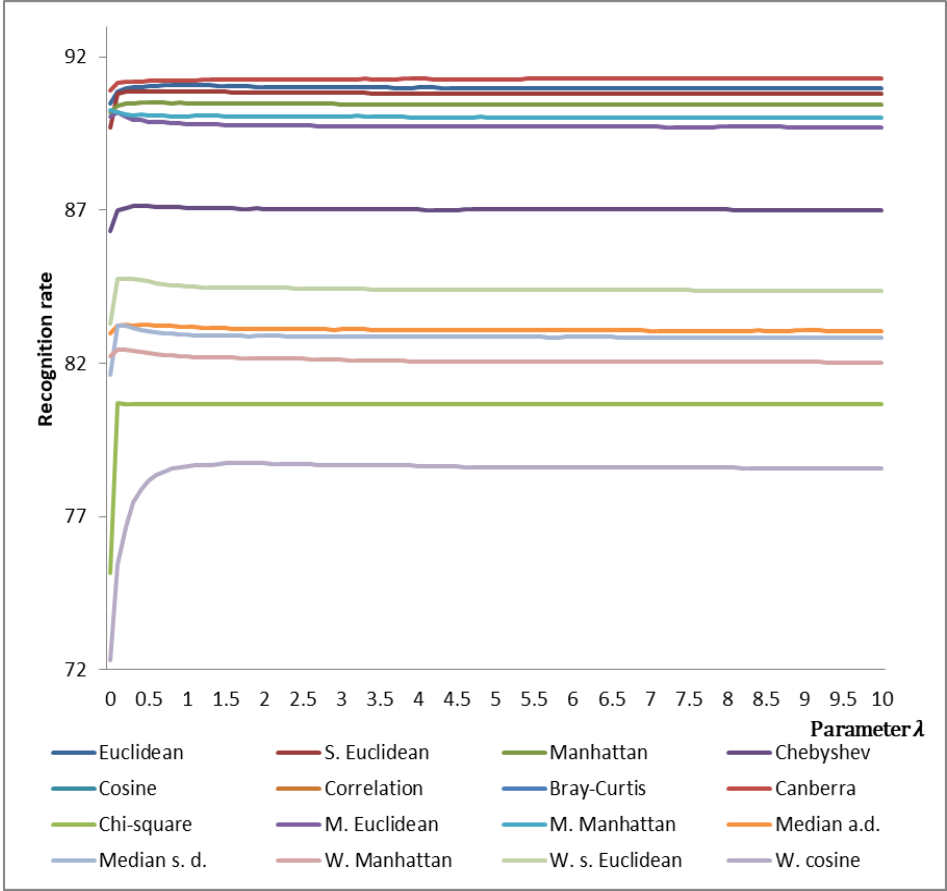


Fig. 6.4 The results produced by the modification of the Hamacher function for various similarity measures.

An interesting fact is that the Modified Manhattan distance produces other results. One can observe that in this case many different families of mean functions may be utilized. Fisherfaces method works well with weighted radical and harmonic means and Wiener–Shannon low extension. From the other hand, a comparison of results for weighted Manhattan measure shows that they are relatively close to the results yielded with the Manhattan measure with no

weights. In the series of experiments for Chebyshev distance, Hamacher t-norm modification, Gini mean, and different averages with weights presented the highest accuracies. In the case of cosine distance measure, fuzzy Choquet integrals, median, and voting work well with the FERET set of images. The experiments with AT&T show, again, that the modification of Hamacher function and weighted geometric mean work well. Similar results can be obtained for weighted cosine measure. EV-copula and geometric mean with weights produce good accuracies for AT&T and correlation distance. Experiments with FERET set have shown that fuzzy Choquet integral, median, and voting are well choices. Bray-Curtis distance applied to Fisherfaces produces well results when fuzzy Choquet integral is used as an aggregation operator for FERET dataset. For the AT&T dataset the functions known from the previously considered distances, namely $A(x,y)$ and $T(x,y)$, dominate. The effectiveness of $A(x,y)$ is illustrated in Fig. 6.5.

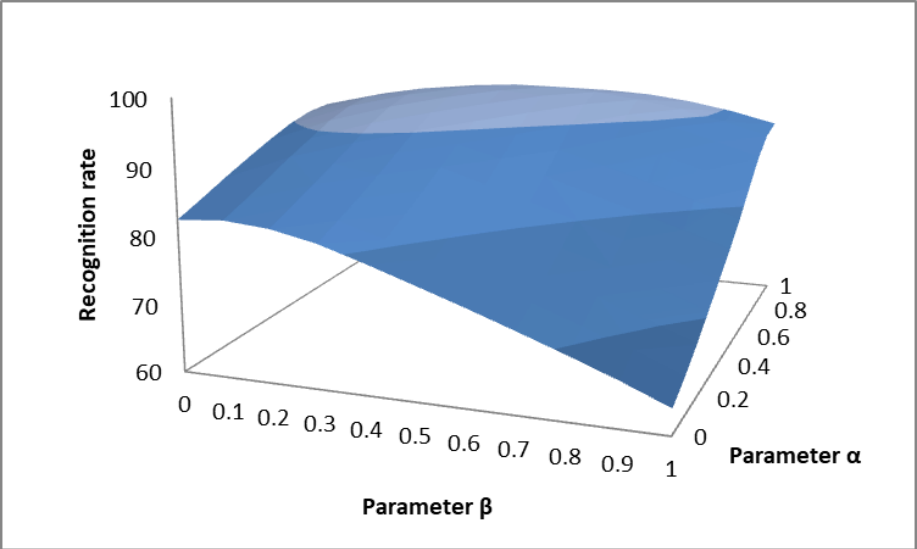


Fig. 6.5 Fisherfaces with Bray-Curtis measure and the function $A(x,y)$ as the aggregator operator.

The next considered distance is Canberra measure. Here, voting, fuzzy Choquet integral, and various modifications of Hamacher functions are the most efficient. In the case of χ^2 -statistic measure again, the Hamacher function modification is a sound alternative. Moreover, a function $f(x,y) = \min(x\sqrt{y}, y)$ can be a good choice. But the highest results are about 80-81% recognition rate which is unsatisfying value. Finally, we discuss the median of absolute differences and median of square differences. The first of them can be used effectively only in the case of AT&T dataset. Here, again Hamacher

function modifications and two means (weighted harmonic and radical) are the most efficient aggregation operators. The t-norm called Aczél-Alsina appears in case of the two median-based distances as one of the best five aggregation functions. However, it is not efficient. The accuracy is, in general, lower than 90%. Fig. 6.6 shows the dependency of the aggregation efficiency on the value of parameter λ . The best is found when $\lambda=1.6$.

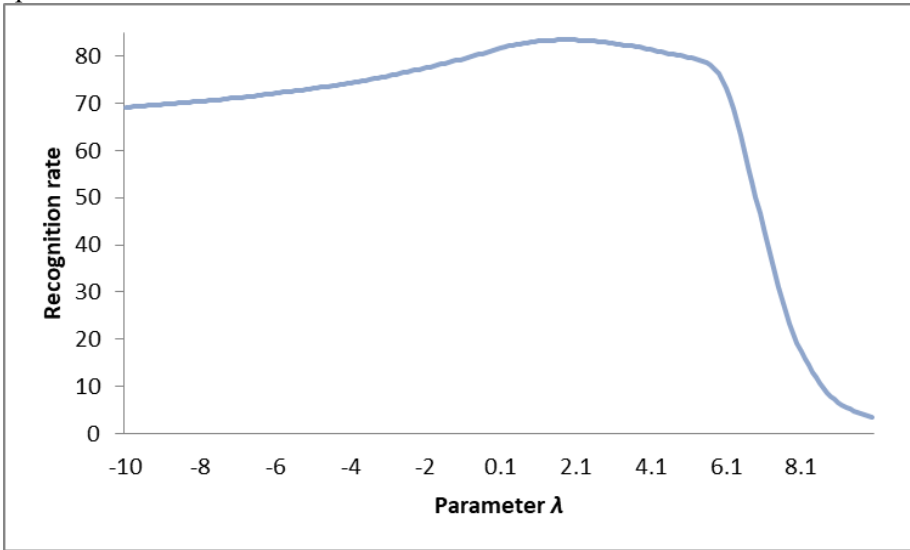


Fig. 6.6 The values obtained with various values of Aczél–Alsina t-norm parameter.

The next comparison is presented for weighted Manhattan function but for PCA only instead of PCA+LDA method with AT&T dataset. The weighted Manhattan function gave the best recognition rates in the case of Eigenfaces. The results are presented in Table 6.2.

Table 6.2 Aggregation results for Eigenfaces and weighted Manhattan function.

Aggregation operator	Accuracy
$A(x, y) = \begin{cases} x^\alpha y^{1-\alpha} & \text{if } x \leq y \\ x^{1-\beta} y^\beta & \text{otherwise} \end{cases}$ for $\alpha = 0.75, \beta = 0.25$	89.18
EV-copula $f(x, y) = \min(x\sqrt{y}, y)$	89
Function $C_\alpha(x, y)$ (see results for Fisherfaces) for $\alpha = 0.5$	88.99
Farlie-Gumbel-Morgenstern copula $C_\lambda^{FGM}(x, y)$ for $\lambda=0.5$	88.98
General nonassociative symmetric copula $C(x, y)$ (see results for Fisherfaces)	88.98

The next standing is the set of standard deviations with respect to the considered similarity/dissimilarity measures which were obtained during our experiments, see Table 6.3. Only the measures giving more than 70% average accuracy were taken into account. The values of standard deviations suggest that

the winning aggregation operators could slightly differ if the experimental series would be repeated. However, the appearance in the top of the rankings of the functions such as fuzzy Choquet integral or modified Hamacher function is rather not accidental.

Table 6.3 Standard deviations yielded in the experimental series with the Fisherfaces method.

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

The last comparison of the results discussed in this part is the set of accuracies reported for the most efficient (in our opinion) aggregation functions, namely fuzzy Choquet integral, median, and voting. The former is considered in the version described in details for Euclidean distance realized for FERET set of images. Note that in the FERET case one can aggregate the recognition results also using the median operator. However, the fuzzy Choquet integral was the best option for ten similarity measures.

One can see that there is a slight difference between the results obtained for the two considered databases. There are at least two reasons of this fact. The first is that the AT&T dataset is relatively *easy* dataset, i.e., there are 10 images of each subject and the number of classes is only 40 while the number of classes in case of FERET is 200 with only 3 images per person. The second fact is that the expression of faces in the FERET dataset varies more than in the case of AT&T. Finally, the preprocessing of FERET images also included histogram equalization procedure.

Table 6.4 Accuracies obtained for fuzzy Choquet integral (c), median (m), and voting (v) with the AT&T and FERET dataset.

Name	Bray-Curt	Canberra	Chebyshev	Chi-square	Correlation	Cosine	Euclidean	Manhattan	Med. abs. dif.	Med. sq. diff.	Mod. Eucl.	Mod.Manh.	Weight.cos.	Sq. Euclidean	W. Euclidean	W. Manh.
AT&T database																
m	91.9	89.0	83.9	73.9	92.2	92.3	88.9	87.8	78.0	77.3	86.2	86.7	78.1	88.4	79.7	77.1
v	88.4	83.8	81.6	77.5	88.4	88.6	87.0	86.1	72.2	72.2	86.0	85.5	75.1	87.0	79.8	75.7
c	92.6	91.3	83.3	67.0	92.6	92.7	87.4	86.5	80.2	78.9	81.6	84.8	81.7	86.5	76.7	75.3
FERET database																
m	90.1	83.3	79.0	51.3	90.5	90.7	84.5	82.9	68.4	68.1	83.1	82.9	87.7	84.3	79.9	77.5
v	90.1	83.9	77.8	40.9	90.7	91.0	84.0	82.2	67.5	64.8	79.3	81.3	87.8	82.5	77.0	76.3
c	90.0	83.3	79.1	63.2	90.1	90.3	84.6	83.0	68.3	68.4	84.3	83.1	87.6	84.6	80.3	77.6

6.6. Generalizations of Aggregation Functions

Lucca et al. (2016b) introduced a concept of a so-called r -increasing function. It is a function $f: [0, 1]^n \rightarrow [0, 1]$ that for all points $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and $p > 0$ satisfies the relation

$$f(x_1, x_2, \dots, x_n) \leq f(x_1 + pr_1, x_2 + pr_2, \dots, x_n + pr_n) \quad (6.16)$$

where $\mathbf{r} = (x_1, x_2, \dots, x_n)$.

Next, the same authors proved that if $T: [0, 1]^2 \rightarrow [0, 1]$ satisfies $T(x, y) \leq x, T(x, 1) = x, T(0, y) = 0$, T is $[0, 1]$ -increasing, and if for any fuzzy measure g the generalized Choquet integral is a function of the form

$$Ch' \int h \circ g(T) = \sum_{1 \leq i \leq n} T(h(x_i) - h(x_{i+1}), g(A_i)) \quad (6.17)$$

with an assumption that

$$h(x_{n+1}) = 0 \quad (6.18)$$

then one can find such a non-zero vector \mathbf{r} that Ch' is \mathbf{r} -increasing, it satisfies the boundary aggregation conditions, and

$$\min(x_1, x_2, \dots, x_n) \leq Ch' \leq \max(x_1, x_2, \dots, x_n) \quad (6.19)$$

This property may suggest an application of the constructions of the form (6.17) and many more generalizations as an efficient aggregation operator. Therefore, in the experimental part of this study we analyse 25 classes of functions based on the formula (6.17) where the role of the function T is played by the functions enlisted in the monograph (Alsina et al. 2006, p. 72, Table 2.6) and article (Lucca et al. 2016b). The former work contains the list of the most important families of t-norms appearing in the literature while the latter study discusses simple functions such as minimum, product, Łukasiewicz t-norm $T_L(a, b) = \max(0, a + b - 1)$, drastic product $T_{DP}(a, b) = b, a$, or 0 for $a = 1, b = 1$, or $x, y \neq 1$, respectively, nilpotent minimum $T_{NP}(a, b) = \min(a, b)$ for $a + b > 1$ and 0 otherwise, and Hamacher product $T_{DP}(a, b) = \frac{ab}{a+b-ab}$ for $a, b \neq 0$ and 0 otherwise.

Note that the general processing scheme is the same as in the previous part. Specific classifiers results are aggregated and the results return the final classification effect, i.e., whether the face is correctly classified or not.

6.7. Experimental Results

In the first part of tests, we use the AT&T image set. The faces were initially cropped and scaled (as in the previous chapter). Next, 18 subjects being our friends or lab members evaluated the saliency of 6 facial features, namely eyebrows, eye, nose, mouth, and left and right cheeks areas, respectively. The normalized weights with respect to these parts of face are as follows: 0.13, 0.32, 0.2, 0.21, 0.07, and 0.07.

In the first series of computational experiments, we conducted 100 iterations of the above-mentioned classification processes for the Eigenfaces method. We used two norms, namely Euclidean and Canberra, which serve as the measures of distances in the set of vectors representing the features of a face after the PCA-based dimensionality reduction of the images. During each of the iterations 5 face images per person were randomly chosen to be placed in the training set. The rest part of the image dataset served as the testing set. Table 6.5 lists the results as follows. In the first column placed are the numbers of t-norm families for which the average accuracies were better than the accuracy obtained with classical Choquet integral (i.e., the product under the integral sign), in the next columns given are the range of the family parameter for which the values are obtained, an argument for which the maximal value is produced, and the difference between the accuracy for the specific t-norm and the accuracy for the product operator. In addition, listed are the median and voting aggregation operators as the methods yielding relatively stable results in the aggregation processes (see the previous part of this chapter). Note that the actual classification results are not essential in the comparison. It is because the fuzzy Choquet integral has proven to be well-established aggregation function utilized in the facial recognition problems (Kwak & Pedrycz 2005). Therefore, we present only the values of positive differences between the analyzed generalization of Choquet integral and the classical Choquet integral, i.e., its version with product t-norm. A similar comparison was done for the classifiers based on LDA and three norms to compare its resulting vectors, i.e., Canberra, cosine, and Euclidean distances, see Table 6.6.

Table 6.7 enlists the detailed results obtained with using six local descriptors (Chain Code-Based Local Descriptor, Full Ranking, Local Binary Pattern, and three versions of Multi-Scale Block Local Binary Pattern, i.e., with 3, 5, and 7 pixel width square blocks, respectively). The descriptors were used in their simplest forms with no partition of the face images into the subareas. Again, 100 iterations of the tests were carried for FERET dataset (with the same division into the training and testing sets as in the previous part of this chapter). The

method applied for the six facial regions was the Fisherfaces technique and the analyzed similarity measures were correlation, cosine, and Euclidean distances, see Table 6.8. The next series of tests were carried not for the parts of the face but for various methods applied for the whole images of face. Eight techniques, i.e., Eigenfaces with Euclidean distance, Fisherfaces with Euclidean measure, CCBLD, Full Ranking, LBP, and three variants of MBLBP were utilized for FERET dataset. The same protocol of testing was used. The weights associated with particular classifiers were their normalized average accuracies. Labeled Faces in the Wild images were analyzed using CCBLD, Full Ranking, LBP, and MBLBP with block of width 5 px. We have chosen the images of the people who have exactly six images in the dataset. Four images of each person were randomly chosen to the training set. One image was selected to the testing set. In the case of Yale database, the six local descriptors discussed above were used in the experiments repeated 100 times for 5 images of each person put in the training set and the rest in the testing one. The summary of the tests with Yale, FERET, and LFW is presented in Table 6.9.

Table 6.5 The results obtained with the AT&T dataset and the Eigenfaces method applied to the chosen feature regions.

T-norm family num.	Canberra measure			Euclidean measure		
	Range of the parameter	Value of the best argument	Corresp. maximal difference	Range of the parameter	Value of the best argument	Corresp. maximal difference
1	[0.1,0.2]	0.1	0.23	[-0.7,-0.1]	-0.6	2.23
2	-	-	-	[1.2,1.6]	1.4	0.85
3	[0.1,0.3]	0.2	0.23	[-10.0,-0.1]	-4.2	4.25
4	1.1	1.1	0.15	[0.6,0.9]	0.7	3.23
5	[0.2,0.8]	0.4	0.25	[-10.0,-0.1]	-5.7	3.18
6	[1.1,1.3]	1.2	0.15	[0.2,0.9]	0.4	2.98
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	[1.1,1.3]	1.2	0.23	[0.1,0.9]	0.1	2.58
17	-	-	-	[1.1,1.4]	1.2	1.65
20	[-0.8,0.1]	-0.3	0.25	[-10.0,-1.1]	-9.8	2.98
24	-	-	-	[1.1,1.5]	1.2	1.55
25	-	-	-	[0.1,1.4]	0.8	2.6
median	-	-	-	-	-	14.25
voting	-	-	-	-	-	8.95

Table 6.6 The results obtained with the AT&T dataset and the Fisherfaces method applied to six face segments.

T-norm fam. num.	Canberra measure			Cosine measure			Euclidean measure		
	Range of the param.	Value of the best arg.	Corresp. max. differ.	Range of the param.	Value of the best arg.	Corresp. max. differ.	Range of the param.	Value of the best arg.	Corresp. max. differ.
1	[0.1,0.2]	0.1	0.48	[0.1,0.3]	0.2	0.4	[0.1,0.2]	0.1	0.23
3	[0.1,0.5]	0.4	0.28	[0.3,0.7]	0.4	0.15	[0.1,0.4]	0.2	0.28
4	[1.1,1.2]	1.1	0.15	[1.1,1.3]	1.2	0.2	1.1	1.1	0.28
5	[0.2,1.3]	0.6	0.18	[0.5,2.4]	0.8,	0.13	[0.1,1.3]	0.4	0.28
6	[1.3,1.4]	1.3	0.03	[1.2,1.6]	1.4	0.1	[1.1,1.5]	1.3	0.3
10	-	-	-	-	-	-	[5.8,10.0]	8.5	0.13
12	0.4	0.4	0.45	[0.5,0.6]	0.5	0.28	-	-	-
14	-0.1	-0.1	0.03	-0.1	-0.1	0.03	-0.1	-0.1	0.3
15	[1.1,1.4]	1.3	0.35	[1.2,2.0]	1.4	0.28	[1.1,1.3]	1.2	0.28
20	[-0.7,1.0]	0.1	0.18	[-0.3,2.4]	0.6	0.18	[-0.8,0.8]	-0.1	0.38
23	0.1	0.1	0.13	0.1	0.1	0.38	0.1	0.1	0.05
median	-	-	-	-	-	-	-	-	1.25
voting	-	-	-	-	-	-	-	-	1.65

Noteworthy is that 25 various families of t-norms were evaluated as the substitutes of the product operator under the Choquet integral sign with their parameter $\alpha = -10, -9.9, \dots, -0.1, 0, 0.1, \dots, 9.9, 10$ (of course, if the formula enables specific values). We have assumed that the choice of such division of arguments represents relatively wide and satisfies covering of their possible values. The results show that the t-norms no. 1, 3, 4, 5, 6, 9, 10, 11, 14, 15, 20, and 25 have a great potential to serve as valuable aggregation operators. Note that their numbers correspond to the enumeration proposed in the monograph by Alsina et al. (2006) at page 72, Table 2.6. Two functions (voting and median) can also serve as aggregation operators in the face recognition problems. It is interesting that only the function (with the index no. 10), namely

$$T_{\alpha}(x, y) = \frac{xy}{(1+(1-x^{\alpha})(1-y^{\alpha}))^{\frac{1}{\alpha}}} \quad (6.20)$$

for $\alpha \geq 3.1$ and median produce the average accuracies being better than the fuzzy Choquet integral with the product under integral sign with respective average differences between their results and the results generated by Choquet integral being 0.15 and 0.77. The best choice of parameter α for the function $T_{\alpha}(x, y)$ is 9.8. However, the range of parameter for which other families of t-norms return satisfying results varies. Hence, in general, it is difficult to predict its optimal value. There are also the functions that are totally impractical from the aggregation point of view, namely the functions with indices 16, 19, and 22. Note that they do not appear in the tables containing the results.

Table 6.7 The results obtained with the AT&T dataset and various local descriptors for the six facial parts.

Chain Code-Based Local Descriptor				Full Ranking			Local Binary Pattern		
T-norm fam. num.	Range of the param.	Val. of the best arg.	Cor-resp. max. differ.	Range of the param.	Val. of the best arg.	Cor-resp. max. differ.	Range of the param.	Val. of the best arg.	Cor-resp. max. differ.
1	0.1	0.1	0.2	-	-	-	-	-	-
3	[-0.2,0.3]	0.3	0.3	[-0.4,-0.3]	-0.3	0.05	[-0.3,-0.2]	-0.2	0.08
4	1.1	1.1	0.23	-	-	-	0.9	0.9	0.05
5	[-0.5,1.0]	0.7	0.23	[-0.6,-0.5]	-0.6	0.1	[-0.6,-0.3]	-0.5	0.1
6	[0.8,1.3]	1.1	1.13	[0.7,0.9]	0.7	0.05	0.8	0.8	0.13
9	-	-	-	-	-	-	-	-	-
10	[3.3,10.0]	9.7	0.18	[1.9,10.0]	8.7	0.1	[4.2,10.0]	7.3	0.15
11	-	-	-	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-
14	-0.1	-0.1	0.03	-	-	-	-	-	-
15	[0.9,1.3]	1.2	0.25	0.9	0.9	0.03	0.9	0.9	0.13
20	[-1.7,0.4]	0.2	0.25	-	-	-	[-1.9,-1.4]	-1.5	0.13
25	-	-	-	-	-	-	-	-	-
median voting	-	-	3.1	-	-	4.48	-	-	1.68
	-	-	1.85	-	-	-	-	-	-
Multi-scale Block Local Binary Pattern (3 px)				Multi-scale Block Local Binary Pattern (5 px)			Multi-scale Block Local Binary Pattern (7 px)		
T-norm fam. num.	Range of the param.	Val. of the best arg.	Cor-resp. max. differ.	Range of the param.	Val. of the best arg.	Cor-resp. max. differ.	Range of the param.	Val. of the best arg.	Cor-resp. max. differ.
1	-0.1	-0.1	0.45	[-0.2,-0.1]	-0.1	0.28	[-0.2,-0.1]	0.1	0.9
3	[-1.3,-0.1]	-0.7	0.5	[-2.3,-0.1]	-1.3	0.85	[-3.6,-0.1]	-0.9	1.48
4	0.9	0.9	0.53	[0.8,0.9]	0.8	0.78	[0.7,0.9]	0.8	1.28
5	[-1.8,-0.1]	-1	0.48	[-2.6,-0.1]	-0.8	0.73	[-3.4,-0.1]	-1.4	1.53
6	[0.6,0.8]	0.8	0.43	[0.5,0.9]	0.7	1.03	[0.4,0.9]	0.7	1.55
9	[0.1,10.0]	7.9	0.53	[2.9,10.0]	5.4	0.6	[2.1,10.0]	4.5	1.38
10	[0.1,10.0]	0.2	0.55	[0.1,10.0]	2.2	0.93	[0.1,10.0]	0.5	1.55
11	0.1	0.1	0.05	0.1	0.1	0.43	0.1	0.1	1.05
12	-	-	-	-	-	-	0.3	0.3	0.05
14	0.1	0.1	0.33	0.1	0.1	0.28	0.1	0.1	1
15	[0.6,0.9]	0.7	0.5	[0.5,0.9]	0.6	0.55	[0.3,0.9]	0.6	1.43
20	[-3.8,-1.1]	-2.6	0.53	[-5.1,-1.1]	-3.5	0.75	[-1.9,-1.8]	-1.9	0.08
25	[0.1,0.2]	0.1	0.53	[0.1,0.2]	0.2	0.45	[0.1,0.3]	0.2	1.3
med. voting	-	-	2.43	-	-	4.15	-	-	3.83
	-	-	-	-	-	0.93	-	-	2.18

Table 6.8 The results obtained with the FERET dataset and the Fisherfaces method and three norms applied to six face segments.

T-norm fam. num.	Correlation measure			Cosine measure			Euclidean measure		
	Range of the param.	Value of the best arg.	Cor-resp. max. differ.	Range of the param.	Value of the best arg.	Cor-resp. max. differ.	Range of the param.	Value of the best arg.	Cor-resp. max. differ.
1	[0.1,0.4]	0.1	0.63	[0.1,0.5]	0.2	0.78	[0.1,0.4]	0.2	1.08
3	[0.1,0.7]	0.3	0.33	[0.1,0.6]	0.2	0.45	[0.1,0.6]	0.5	0.8
4	[1.1,1.2]	1.1	0.18	[1.1,1.2]	1.1	0.15	[1.1,1.2]	1.2	0.43
5	[0.1,1.7]	0.5	0.33	[0.1,1.6]	1.1	0.45	[0.1,1.9]	0.9	0.6
6	-	-	-	-	-	-	1.3	1.3	0.05
10	-	-	-	-	-	-	[5.4,10.0]	10	0.35
12	[0.4,0.6]	0.5	0.78	[0.4,0.7]	0.5	0.75	[0.4,0.6]	0.5	1.15
14	-0.1	-0.1	0.2	-0.1	-0.1	0.4	-0.1	-0.1	0.03
15	[1.1,1.9]	1.3	0.5	[1.1,2.0]	1.4	0.68	[1.1,1.9]	1.4	0.93
18	0.1	0.1	0.8	[0.1,0.3]	0.1	1.03	0.1	0.1	0.43
20	[-0.9,1.8]	-0.1	0.4	[-0.9,1.7]	0.2	0.6	[-0.9,2.0]	0.7	0.83
23	[0.1,0.2]	0.2	0.75	[0.1,0.2]	0.1	0.73	0.1	0.1	1.2

Table 6.9 Various methods applied to the whole images of face with respect to different datasets.

T-norm fam. num.	FERET			LFW			Yale		
	Range of the param.	Value of the best arg.	Cor-resp. max. differ.	Range of the param.	Value of the best arg.	Cor-resp. max. differ.	Range of the param.	Value of the best arg.	Cor-resp. max. differ.
1	[-0.6,-0.1]	-0.3	0.11	-	-	-	-0.1	-0.1	0.15
2	-	-	-	[2.9,5.6]	3.9	0.03	[3.3,10.0]	5.8	2.13
3	[-10,-0.1]	-9.4	0.09	-	-	-	[-0.3,0.3]	-0.2	0.02
5	[-10,-0.1]	-5.8	0.1	-	-	-	[-10,-0.1]	0.5	0.03
6	-	-	-	-	-	-	2.3	2.3	0.02
7	0.1	0.1	0.05	-	-	-	-	-	-
8	[2.7,6.2]	4.8	0.08	[8.2,10]	10	0.02	[7.7,10.0]	10	0.95
9	[0.3,10]	1.3	0.11	[5.5,5.9]	5.5	0.01	[4.6,10.0]	7.9	0.09
10	[0.1,10]	0.5	0.06	-	-	-	[0.1,0.2]	0.2	0.03
11	[0.1,0.4]	0.2	0.11	-	-	-	-	-	-
13	0.1	0.1	0.01	-	-	-	-	-	-
14	[-0.1,0.5]	0.2	0.11	-	-	-	[-10,-6.5]	-7.4	1.39
15	[0.1,0.9]	0.1	0.11	-	-	-	[0.7,0.9]	0.9	0.04
17	2	2	0.01	[1.7,2.2]	1.8	0.01	[1.8,2.7]	2.1	1.36
20	[-10,-1.1]	-10	0.11	-	-	-	[-1.8,-1.1]	-1.8	0.04
21	-	-	-	[1.4,4.1]	2.2	0.02	-	-	-
24	-	-	-	[1.9,2.8]	2.2	0.02	[2.1,4.6]	2.9	1.37
25	[0.1,1.1]	0.5	0.11	-	-	-	[0.1,0.2]	0.1	0.09
med.	-	-	0.05	-	-	0.01	-	-	2.04
vot.	-	-	0.08	-	-	0.03	-	-	2.51

6.8. Further Generalizations of Choquet Integral

In this part of the study, we analyze the further generalizations and modifications of the Choquet integral. The motivation for this study is the series of works with propositions of modifications of the Choquet integral in relation to the classification problems. Hereafter, we assume that the general assumptions on the Choquet integral parameters and its generalization given by the formula (6.18) are still valid. The following Choquet-like functions were introduced:

$$C_M(x) = \sum_{i=1}^n M(h(x_i) - h(x_{i+1}), g(A_i)) \quad (6.21)$$

(Lucca et al. 2014, Lucca et al. 2015)

$$C_F(x) = \min(\sum_{i=1}^n F(h(x_i) - h(x_{i+1}), g(A_i)), 1) \quad (6.22)$$

(Lucca et al. 2018)

$$C_C(x) = \sum_{i=1}^n (C(h(x_i), g(A_i)) - C(h(x_{i+1}), g(A_i))) \quad (6.23)$$

(Lucca et al. 2017)

$$C_{O_b}(x) = \sum_{i=1}^n O_b(h(x_i) - h(x_{i+1}), g(A_i)) \quad (6.24)$$

(Lucca et al. 2016a) where $O_b(\cdot)$ is a so-called overlap function (i.e., commutative, increasing, continuous, $O_b(p, q) = 0$ for $pq = 0$ and $O_b(p, q) = 1$ for $pq = 1$), and

$$C_{Min}(x) = \sum_{i=1}^n (\min(h(x_i), g(A_i)) - \min(h(x_{i+1}), g(A_i))) \quad (6.25)$$

(Dimuro et al. 2018). Moreover, we propose to examine a set of several functions being modifications of the above formulas:

$$C_{MC}(x) = \sum_{i=1}^n (C(h(x_i), g(A_i)) - C(h(x_{i+1}), g(A_i)) + C(h(x_i) - h(x_{i+1}), g(A_i))) \quad (6.26)$$

$$C_{MMin}(x) = \sum_{i=1}^n M(\min(h(x_i), g(A_i)) - \min(h(x_{i+1}), g(A_i))) \quad (6.27)$$

$$C_{MMin2}(x) = \sum_{i=1}^n M(\min(h(x_i), g(A_i)), \min(h(x_{i+1}), g(A_i))) \quad (6.28)$$

$$C_{MinM}(x) = \sum_{i=1}^n (\min(M(h(x_i), g(A_i)), g(A_i)) - \min(M(h(x_{i+1}), g(A_i)), g(A_i))) \quad (6.29)$$

$$C_{Diff1}(x) = \sum_{i=1}^n M(h(x_{i-1}) - h(x_{i+1}), g(A_i)) \quad (6.30)$$

$$C_{Diff2}(x) = \sum_{i=1}^n M(h(x_{i-1}) + h(x_{i+1}) - h(x_i), g(A_i)) \quad (6.31)$$

and

$$C_{Diff3}(x) = \sum_{i=1}^n M((h(x_{i-1}) - h(x_{i+1}))/h(x_{i-1}), g(A_i)) \quad (6.32)$$

6.9. Experimental Results

Here, we discuss 27 experiments with settings similar to the described in the previous cases. AT&T, FERET, Yale, LFW, PUT, and MUCT datasets were in use. The details of each experiment are presented in Table 6.10. Our proposition, inspired by numerical methods formulas used in an approximation theory, namely, the $C_{MC}(x)$ function produces the best results (it has hit Choquet integral in 22 of 27 competitions). However, the functions $C_M(x)$, $C_F(x)$, and $C_C(x)$ produce also strong results (21 hits). Note that for a few series of experiments no winning function has been found.

Table 6.10 Summary of experiments with various generalizations and modification of Choquet integral serving as aggregation function. The symbol X is put in for the aggregation operators giving the accuracy higher than Choquet integral with product under the integral sign. The last two rows are the sum of Xs and the sum of global best results, i.e., the number of results giving the best result in case of specific classifiers.

Dataset, method of classification, partition of a face or specific methods	C_M	C_F	C_C	C_{Ob}	C_{Min}	C_{MC}	C_{MMin}	C_{MinM}	C_{Diff1}	C_{Diff2}	C_{Diff3}
AT&T, Eigenfaces (Euclidean distance) for eb, eo, n, m, lch, rch	X	X	X			X	X	X			
AT&T, Eigenfaces (cosine distance) for eb, eo, n, m, lch, rch											
AT&T, Eigenfaces (Canberra distance) for eb, eo, n, m, lch, rch	X	X	X			X					
AT&T, Fisherfaces (Euclidean distance) for eb, eo, n, m, lch, rch	X	X	X			X					
AT&T, Fisherfaces (cosine distance) for eb, eo, n, m, lch, rch	X	X	X			X					
AT&T, Fisherfaces (Canberra distance) for eb, eo, n, m, lch, rch	X	X	X			X					

distance) for eb, eo, n, m, lch, rch AT&T, LBP for eb, eo, n, m, lch, rch	X	X	X		X		
AT&T, MBLBP (3 px) for eb, eo, n, m, lch, rch	X	X	X		X		
AT&T, MBLBP (5 px) for eb, eo, n, m, lch, rch	X	X	X		X		
AT&T, MBLBP (7 px) for eb, eo, n, m, lch, rch	X	X	X		X		
AT&T, Full Ranking (3 px) for eb, eo, n, m, lch, rch	X	X	X		X		
AT&T, CCBLD (3 px) for eb, eo, n, m, lch, rch	X	X	X		X		
AT&T, Eigenfaces (Euclidean distanace) for ext. e, n, m, face							
AT&T, Eigenfaces (cosine distance) for ext. e, n, m, face							
AT&T, Eigenfaces (Canberra distance) for ext. e, n, m, face	X	X	X		X		
AT&T, Fisherfaces (Euclidean distanace) for ext. e, n, m, face	X	X	X		X		X
AT&T, Fisherfaces (cosine distance) for ext. e, n, m, face	X	X	X		X		
AT&T, Fisherfaces (Canberra distance) for ext. e, n, m, face	X	X	X		X		
AT&T, pca	X	X	X		X	X	X

(Canb.), lda (cos.), lbp, mblbp (5 px), full rank., ccbld FERET, Fisherfaces (Euclidean distasnce) for eb, eo, n, m, lch, rch	X	X	X			X					
FERET, Fisherfaces (cosine distasnce) for eb, eo, n, m, lch, rch	X	X	X			X					
FERET, Fisherfaces (correlation distasnce) for eb, eo, n, m, lch, rch	X	X	X			X					
FERET, pca (Canb.), lda (cos.), lbp, ccbld, mblbp (3 px) Yale, CCBLD, MBLBP (5 px), MBLBP (7px) LFW, CCBLD, LBP, MBLBP (5px)						X			X		
MUCT, pca (Canb.), lda (cos.), ccbld, lbp (7x7), mblbp (3px, 7x7) PUT, pca (Canb.), lda (cos.), ccbld, lbp, mblbp (3px), distances	X	X	X	X		X	X	X	X	X	
Sum of the results better than	21	21	21	1	0	22	3	3	4	1	0
Choquet integral Sum of global best results	2	2	11	1	0	9	1	1	3	0	0

Moreover, the best average calculated over all the experiments was obtained by the same integrand function as previously, see formula (6.20), with the parameter $\alpha=7.4$. However, among all the described tests, the following five families of t-norms hit the individual classifiers' results over 68 times (families no. 1, 3, 5, 15, and 20, respectively):

$$T_{\alpha}(x, y) = \left(\max(x^{-\alpha} + y^{-\alpha} - 1, 0) \right)^{-\frac{1}{\alpha}}, \alpha \in (-\infty, 0) \cup (0, \infty) \quad (6.33)$$

$$T_{\alpha}(x, y) = \frac{xy}{1 - \alpha(1-x)(1-y)}, \alpha \in (-\infty, -1] \quad (6.34)$$

$$T_{\alpha}(x, y) = -\frac{1}{\alpha} \log \left(1 + \frac{(e^{-\alpha x} - 1)(e^{-\alpha y} - 1)}{e^{-\alpha} - 1} \right), \alpha \in (-\infty, 0) \cup (0, \infty) \quad (6.35)$$

$$T_{\alpha}(x, y) = \exp \left(1 - ((1 - \log x)^{\alpha} + (1 - \log y)^{\alpha} - 1)^{\frac{1}{\alpha}} \right), \alpha \in (0, \infty) \quad (6.36)$$

$$T_{\alpha}(x, y) = \left(1 + \frac{((1+x)^{-\alpha} - 1)((1+y)^{-\alpha} - 1)}{2^{-\alpha} - 1} \right)^{-\frac{1}{\alpha}}, \alpha \in (0, \infty) \cup (0, \infty) \quad (6.37)$$

6.10. Conclusions

In this chapter, we have determined the most efficient aggregation operators in an application to the results of classifiers based on particular facial parts when the method of classification was Fisherfaces with 16 various similarity/dissimilarity measures utilized to compare the vectors representing the features. Choquet integral, median, voting, Hamacher function modifications and a few other operators appear as sound aggregation alternatives which may be considered in the systems based on several nearest neighbor classifiers.

Next, we have examined 25 classes of t-norms by replacing the product t-norm under the formula of the Choquet integral in the context of classifiers aggregation. Moreover, we have evaluated recently published new modifications of Choquet integral as well as with our own proposals. We based our tests on the facial parts and various methods utilizing the general nearest neighbor classification approach. In such an approach one of the families of t-norms exposed its well performance as the vehicle to build an effective classifier.

References

- Agrawal, U., Pinar, A. J., Wagner, C., Havens, T. C., Soria, D., Garibaldi, J. M.: Comparison of fuzzy integral-fuzzy measure based ensemble algorithms with the state-of-the-art ensemble algorithms. In: Medina, J. et al. (Eds.): Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations. IPMU 2018. Communications in Computer and Information Science, vol. **853**, pp. 329–341 (2018)
- Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with Local Binary Patterns. In: Proceedings of the 8th European Conference on Computer Vision, Lecture Notes in Computer Science **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.* **20**, 128–140 (2018)
- Alsina, C., Frank, M. J., Schweizer, B.: Associative functions. Triangular norms and copulas. World Scientific, New Jersey (2006)
- Baczyński, M., Bustince, H., Mesiar, R.: Aggregation functions: Theory and applications, part I. *Fuzzy Set. Syst.* **324** (2017a)
- Baczyński, M., Bustince, H., Mesiar, R.: Aggregation Functions: Theory and Applications, Part II. *Fuzzy Set. Syst.* **325** (2017b)
- Beliakov, G., Bustince Sola, H., A., Calvo Sánchez, T.: A practical guide to averaging functions. Springer, Cham (2016)
- Beliakov, G., Pradera, A., Calvo, T.: Aggregation functions: A guide for practitioners. Springer–Verlag, Berlin Heidelberg (2007)
- Bereta, M., Karczmarek, P., Pedrycz, W., Reformat, M.: Local descriptors in application to the aging problem in face recognition. *Pattern Recognit.* **46**, 2634–2646 (2013)
- Bharkad, S. D., Kokare, M.: Performance evaluation of distance metrics: Application to fingerprint recognition. *Int. J. Pattern Recognit. Artif. Intell.* **25**, 777–806 (2011)
- Brunelli, R., Poggio, T.: Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1042–1052 (1993)
- Bustince, H., Fernandez, J., Mesiar, R., Calvo T.: Aggregation functions in theory and in practice. In: Proceedings of the 7th International Summer School on Aggregation Operators at the Public University of Navarra, Pamplona, Spain, July 16–20, Advances in Intelligent Systems and Computing, **228**. Springer (2018)
- 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)
- Calvo, T., Mayor, G., Mesiar R.: Aggregation operators. New trends and applications. Physica-Verlag, Heidelberg (2014)
- Campomanes-Alvarez, C., Ibáñez, O., Cerdón, O.: Experimental study of different aggregation functions for modeling craniofacial correspondence in craniofacial superimposition. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 437–444 (2016)
- Chan, C.-H., Kittler, J., Messer, K.: Multi-scale Local Binary Pattern histograms for face recognition. In: Lee, S.-W., Li, S. Z. (Eds.): ICB 2007, Lecture Notes in Computer Science **4642**, pp. 809–818 (2007)

- Chan, C. H., Yan, F., Kittler, J., Mikolajczyk, K.: Full ranking as local descriptor for visual recognition: A comparison of distance metrics on s_n . *Pattern Recognit.* **48**, 1328–1336 (2015)
- Das, S., Guha, D.: Power harmonic aggregation operator with trapezoidal intuitionistic fuzzy numbers for solving MAGDM problems. *Irani. J. Fuzzy Syst.* **12**, 41–74 (2015)
- Dimuro, G. P., Lucca, G., Sanz, J. A., Bustince, H., Bedregal, B.: CMin-integral: A Choquet-like aggregation function based on the minimum t-norm for applications to fuzzy rule-based classification systems. In: Torra, V. et al. (Eds.): *Aggregation Functions in Theory and in Practice. Advances in Intelligent Systems and Computing* **581**, pp. 83–95 (2018)
- Dolecki, M., Karczmarek, P., Kiersztyn, A., Pedrycz, W.: Utility functions as aggregation functions in face recognition. In: *Proc. 2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–6 (2016)
- Ekenel, H. K., Stiefelhagen, R.: Generic versus salient region-based partitioning for local appearance face recognition. In: Tistarelli, M., Nixon M. S. (Eds.): *Advances in biometrics. LNCS* **5558**, pp. 367–375 (2009)
- Gottumukkal, R., Asari, V. K.: An improved face recognition technique based on modular PCA approach. *Pattern Recognit. Lett.* **25**, 429–436 (2004)
- Grabisch, M.: Fuzzy integral in multicriteria decision-making. *Fuzzy Set. Syst.* **69**, 279–298 (1995)
- Grabisch, M., Marichal, J.-L., Mesiar, R., Pap, E.: *Aggregation functions*. Cambridge University Press, Cambridge (2009)
- Haddadnia, J., Ahmadi, M.: N-feature neural network human face recognition. *Image Vis. Comput.* **22**, 1071–1082 (2004)
- Jarillo, G., Pedrycz, W., Reformat, M.: Aggregation of classifiers based on image transformations in biometric face recognition. *Mach. Vis. Appl.* **19**, 125–140 (2008)
- Karczmarek, P., Kiersztyn, A., Pedrycz, W.: On some aspects of an aggregation mechanism in face recognition problems. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada, J. (Eds.): *Artificial Intelligence and Soft Computing. ICAISC 2018. Lecture Notes in Computer Science* **10842**, pp. 148–156 (2018)
- Kiersztyn, A., Karczmarek, P., Pedrycz, W.: Multi-level aggregation in face recognition. In: Rutkowski, L., Scherer, R., Korytkowski, M., Pedrycz, W., Tadeusiewicz, R., Zurada J. (Eds.): *Artificial Intelligence and Soft Computing. ICAISC 2018. Lecture Notes in Computer Science* **10841**, pp. 645–656 (2018)
- Kim, T.-K., Kim, H., Hwang, W., Kittler, J.: Component-based LDA face description for image retrieval and MPEG-7 standardisation. *Image Vis. Comput.* **23**, 631–642 (2005)

- Klement, E. P., Mesiar, R.: Logical, algebraic, analytic, and probabilistic aspects of triangular norms. Elsevier, Amsterdam (2005)
- Klement, E. P., Mesiar, R., Pap, E.: Triangular norms. Kluwer Academic Publishers, Dordrecht (2000)
- Kurach, D., Rutkowska, D., Rakus-Andersson, E.: Face classification based on linguistic description of facial features. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., Zurada, J. M. (Eds.): Artificial Intelligence and Soft Computing 2014, Part II. LNAI **8468**, pp. 155–166 (2014)
- Kwak, K.-C., Pedrycz, W.: Face recognition using fuzzy integral and wavelet decomposition method. IEEE Trans. Syst. Man Cybern. B Cybern. **34**, 1666–1675 (2004)
- Kwak, K.-C., Pedrycz, W.: Face recognition: A study in information fusion using fuzzy integral. Pattern Recognit. Lett. **26**, 719–733 (2005)
- Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S. Z.: Learning multi-scale block Local Binary Patterns for face recognition. In: Advances in Biometrics, International Conference, ICB 2007, Lecture Notes in Computer Science **4642**, pp. 828–837 (2007)
- Liu, W. L., Song, X. Q., Zhang, Q. Z., Zhang, S. B.: (T) fuzzy integral of multi-dimensional function with respect to multi-valued measure. Irani. J. Fuzzy Syst. **9**, 111–126 (2012)
- Lucca, G., de Vargas, R. R., Dimuro, G. P., Sanz, J. A., Bustince, H., Bedregal, B. R. C.: Analysing some t-norm-based generalizations of the Choquet Integral for different fuzzy measures with an application to fuzzy rule-based classification systems. In: ENIAC 2014, Encontro Nac. Intelig. Artificial e Computacional. SBC, São Carlos, pp. 508–513 (2014)
- Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., Asiain, M. J., Elkano, M., Bustince, H.: CC-integrals: Choquet-like Copula-based aggregation functions and its application in fuzzy rule-based classification systems. Knowl.-Based Syst. **119**, 32–43 (2017)
- Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., Bustince, H.: Pre-aggregation functions constructed by CO-integrals applied in classification problems. In: Proceedings of IV CBSF, pp. 1–11 (2016a)
- Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., Bustince, H., Mesiar, R.: C_F -integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems. Inf. Sci. **435**, 94–110 (2018)
- Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., Mesiar, R., Kolesárová, A., Bustince, H.: The notion of pre-aggregation function. In: Torra, V., Narakawa Y. (Eds.): MDAI 2015, LNAI **9321**, pp. 33–41 (2015)

- 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 (2016b)
- 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. Springer, pp. 437–449 (2014)
- 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. Springer, pp. 59-70 (2015)
- 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)
- Mirhosseini, A. R., Yan, H., Lam, K.-M., Pham, T.: Human face image recognition: An evidence aggregation approach. *Comput. Vis. Image Unders.* **71**, 213–230 (1998)
- Naveena, C., Manjunath Aradhya, V. N., Niranjan, S. K.: 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, pp. 781–787 (2012)
- Oh, S.-K., Yoo, S.-H., Pedrycz, W.: 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 Syst. Appl.* **40**, 1451–1466 (2013)
- Pedrycz, W., Ekel, P., Parreiras, R.: Fuzzy multicriteria decision-making. Models, Methods and Applications. Wiley, Chichester (2011)
- Pedrycz, W., Gomide, F.: An introduction to fuzzy sets: Analysis and design. The MIT Press, Cambridge (1998)
- Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: Computer Vision and Pattern Recognition. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on, pp. 84–91 (1994)
- Perlibakas, V. Distance measures for PCA-based face recognition. *Pattern Recognit. Lett.* **25**, 711–724 (2004)
- Radtke, P. V. W., Granger, E., Sabourin, R., Gorodnichy, D. O.: Skew-sensitive boolean combination for adaptive ensembles – An application to face recognition in video surveillance. *Inf. Fusion* **20**, 31–48 (2014)
- Smiatacz, M.: Similarity measures for face images: An Experimental Study. In: Chmielewski, L., Datta, A., Kozera, R., Wojciechowski, K. (Eds.): Computer

Vision and Graphics. ICCVG 2016. Lecture Notes in Computer Science **9972**, pp. 341–352 (2016)

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

Torra, V., Narukawa, Y.: Modeling decisions. Information fusion and aggregation operators. Springer-Verlag, Berlin Heidelberg (2007)

Xue, Y., Tong, C. S., Zhang, W.: Survey of distance measures for NMF-based face recognition. In: Wang, Y., Cheung, Y., Liu H. (Eds.): CIS 2006, LNAI **4456**, pp. 1039–1049 (2007)

Yager, R. R., Kacprzyk, J.: The ordered weighted averaging operators: Theory and Applications. Springer Science+Business Media, New York (2012)

Yale Face Database. [online] <http://vision.ucsd.edu/content/yale-face-database> (Accessed 6 April 2017)

Conclusions and Future Work

In this book, solutions of selected important face recognition and decision-making theory problems on a basis of chosen Computational Intelligence techniques have been proposed. In the first chapter we have recalled the most important decision-making theory and Computational Intelligence techniques. In the next chapter we have analyzed the efficiency of local descriptors to cover the problems of aging and age differences in a face recognition aspect. The descriptors were also considered in a combination with Gabor wavelet images. The accuracies obtained during the experimental experiments show their potential efficiency. In particular, the Multi-Scale Block Local Binary Pattern descriptor has demonstrated its stability and invariance to the various kinds of problems (age gaps, age groups, similarity/dissimilarity measure applied to the nearest neighbor classifier, etc.). Moreover, a novel and efficient Chain Code-Based Local Descriptor was introduced as a valuable vehicle to facial recognition. Future work can focus on an enhancement of local descriptors theory or analyzing them in an application to significant face areas such as periocular region. An interesting may be an in-depth examination of combinations various LBP-based transforms and deep learning architectures.

In the same chapter, we have thoroughly analyzed the performance of the Chain Code-Based Local Descriptor and its various extensions based on an application of pixel blocks instead of single pixels. CCBLD has demonstrated its robustness to various problems appearing in face recognition tasks such as aging, noise, occlusion, pose, illumination, or preprocessing level. In a series of tests CCBLD has shown its potential and outperformed other important local descriptors such as Local Binary Pattern or Full Ranking. Future studies may focus on the consideration of other than the Levenshtein method of word comparison, an examination of the rotation invariance problem, and an application of the descriptor to the fields of image analysis such as texture analysis or noise detection.

The third chapter has dealt with methods of evaluation of the saliency of facial features in the process of face recognition realized by humans, and after adjustment procedures, in the processes carried by computers. The presented approach is based on the expert knowledge. In particular, it has been shown that the Analytic Hierarchy Process can efficiently serve as a generic approach of obtaining the weights of the facial features and their groups to be used by people (experts or witnesses of a crime) who describe faces. The three-level AHP hierarchy has been developed along with the entropy-based method of finding the relevance and confidence of the assessment of subjects. It is novel and original approach since the fundamental from the point of human perception linguistic (not numerical) values have been used to build important and intuitively appealing results. Moreover, the AHP has been used to construct the

classifiers based on human (expert) knowledge. It demonstrated recognition rates varying between 94 and 100%. It leads to the conclusion that the AHP-based approaches can be important vehicles to improve the processes with important and dominating role of an expert or a group of experts. Future work may be focused on automatization of the process of the estimation of the both abstract and concrete facial features and applying the feature weights to the problems of aggregation of classifiers based on the facial regions, or an application of the entropy measure to estimate the crime witness.

Chapter 4 has been devoted to a novel and in-depth approach to the AHP and fuzzy AHP techniques based on the graphic interfaces. This way of acquisition of the experts' evaluations leads to very consistent results. A series of tests with real life problems (prediction of sports results, preferred movie genre, etc.) demonstrates the efficiency and potential applicability of the proposal. In addition, the examples of using show that the method can be easily applied to biometrics tasks, in particular, to the problem of description of biometric features. Finally, PSO-based optimization leads to marginalization of the inconsistency appearing in the experts' reciprocal matrices. Future studies may be focused on an application of uncertainty modeling, e.g., an application of another slider related to the level of uncertainty and Granular Computing methods related to an optimal allocation of the granularity of information. Finally, it is interesting to check another geometric graphical components (multi range slider, dial arc with a range $[-90^\circ, 90^\circ]$, etc.) or a time length (delay) of the decision-making by experts.

In the chapter no. 5 an application of Sugeno fuzzy measure as a vehicle to quantify a manner of aggregation of the important information contained in facial regions or parts has been discussed. Its property of monotonicity in the context of face classification with respect to the saliency of facial regions has been thoroughly analyzed. On the basis of the experiments one can conclude that the fuzzy measure appears as an efficient classification model. Next, a novel and original approach to obtain the optimal values of the fuzzy measure densities being the fundamental part of the important concept of classification on a basis of aggregation with the Sugeno λ -fuzzy measure and Choquet integral has been described. The densities relate to the saliency of facial regions or parts and the importance of particular classifiers, i.e., methods not related with particular facial features but with algorithms of classification. The proposed approach has demonstrated high efficiency in the series of experiments conducted for the classical face recognition methods such as Eigenfaces, Fisherfaces, and local descriptors. Future work may be focused on the analysis of other than Eigenfaces and Fisherfaces methods in relation to the application of fuzzy measure, analysis of the problem of increasing (high) number of facial features, or a comprehensive study of eye area along with its subareas and their influence on the recognition processes based on the fuzzy measures. Moreover, exploiting other than PSO optimization techniques or finding the optimal weights for other

than the Choquet integral-based aggregation methods such as OWA (Ordered Weighted Averaging) operators.

In the last chapter, the most efficient aggregation operators in an application to the results of classifiers based on particular facial parts when the method of classification was Fisherfaces or Eigenfaces with 16 various similarity/dissimilarity measures utilized to compare the vectors representing the features have been analyzed. A several operators such as Choquet integral, median, voting, or Hamacher function and others appear as sound aggregation alternatives which may be successfully considered in the biometric systems based on the nearest neighbor classifiers. In the second part of the chapter 25 classes of t-norms located in the place of the product t-norm under the formula of the Choquet integral in the context of aggregation of classifiers have been thoroughly checked. Moreover, various modifications of Choquet integral have been discussed in this aspect. We have based our tests on the facial parts and selected important methods utilizing the general nearest neighbor classification approach. One of the families of t-norms exposed its well performance as the vehicle to build an effective aggregation function for this group of classifiers. Future studies can be focused on the next generalizations of fuzzy integrals of Sugeno or Shilkret types. In addition, an interesting seems to be the question on the optimal classifiers number in relation to an aggregation operator.