DOI 10.1007/s10559-024-00655-w
Cybernetics and Systems Analysis, Vol. 60, No. 1, January, 2024

NEW MEANS OF CYBERNETICS, INFORMATICS, COMPUTER ENGINEERING, AND SYSTEMS ANALYSIS

MULTILEVEL FACE RECOGNITION SYSTEM*

V. M. Opanasenko,[1] Sh. Kh. Fazilov,[2] S. S. Radjabov,[3] and Sh. S. Kakharov[4] UDC 004.93:004.89

Abstract. *The problem of biometric person identification on the basis of component-based face recognition is considered. It is shown that the face recognition system can be represented as a hierarchically organized multilevel system, in which an ensemble of local classifiers forms "soft" decisions about the images of individual components of a face belonging to given classes. Then, based on the integration of these decisions, the final decision on whether the recognized face belongs to one of the given classes is formed. The problems of constructing a local classifier model, as well as choosing an integrator of intermediate solutions of local classifiers, are formulated and solved.*

Keywords: *pattern recognition, multilevel recognition system, classifier ensemble, classifier combination rule, decision making.*

INTRODUCTION

In recent years, one of the promising computer vision directions, namely, face recognition, has been intensively developed. Here, we have to distinguish [1–6] among scientific research and development performed withing this direction.

In the case of practically significant face recognition programs, such as security applications, incorrect classification can cause negative consequences. One of the approaches to increasing reliability of such systems is the use of an ensemble of face recognition algorithms with the subsequent combination of their results on the basis of any integration rule.

The concept of multilevel recognition of objects that have a complex hierarchical structure is used as the basis for implementing the above approach. Such objects are a set of elements of different hierarchy levels. A combination of different elements consists of a set of elements from the previous level, which, in turn, is an integral part of the combination of elements of the higher level.

Taking into account that a human face can be characterized as a composite object represented by such main components as eyebrows, eyes, nose, mouth, the right eye, etc., face recognition system can be represented as a multilayer system.

PROBLEM STATEMENT

Figure 1 presents a simplified scheme of a two-level component-based facial recognition system consisting of the following elements: B_j is a block distinguishing the jth facial component and forming the vector of features

* Schientific work was performed within the framework of the project No. ФЗ-20200929308 of the Ministry of Innovative Development of the Republic of Uzbekistan.

[1]V. M. Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine, Kyiv, Ukraine, *vlopanas@ukr.net.* [2]Digital Technologies and Artificial Intelligence Research Institute, Tashkent, Uzbekistan, *sh.fazilov@gmail.com*. [3]Institute of Basic and Applied Research at the TIIAME National Research University, Tashkent, Uzbekistan, *s_radjabov@yahoo.com*. [4]Kokand University, Kokand, Uzbekistan, *sh.kaxarov93@gmail.com*. Translated from Kibernetyka ta Systemnyi Analiz, No. 1, January–February, 2024, pp. 175–181; DOI 10.34229/KCA2522-9664.24.1.15. Original article submitted February 13, 2023.

 1060-0396/24/6001-0146 ©2024 Springer Science+Business Media, LLC

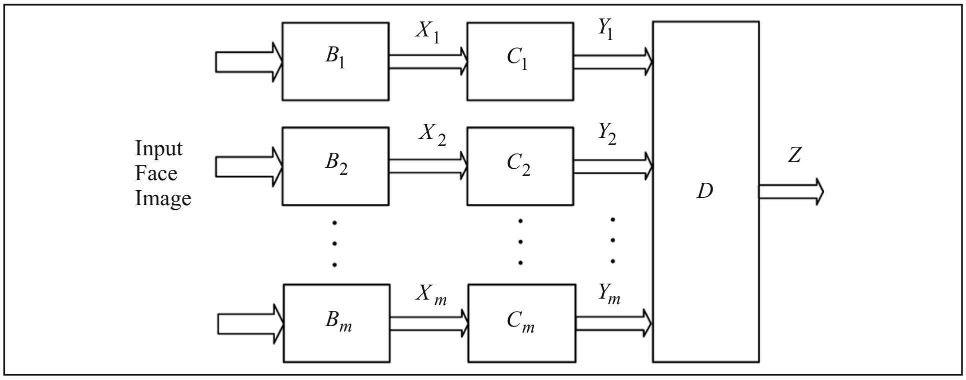

Fig. 1. A scheme of the two-level system of component-based face recognition.

$X_j = (x_1^{(j)}, \ldots, x_{n_j}^{(j)})$, C_j is the local classifier of the "soft" recognition of the jth component forming the proximity estimates $Y_j = (y_1^{(j)}, \ldots, y_{n_j}^{(j)})$ of this component for each of the given classes, and D is the local solution integration block forming solution Z based on the face under recognition.

From the above scheme (see Fig. 1), it follows that the development of a multilevel face recognition system is based on constructing local classifiers, as well as an integrator based on the chosen rule of combining these classifiers.

Constructing a local teacher-trainable classifier is a classical recognition problem that can be stated as follows [7].

Let a feasible set Υ of objects be determined. Each object $S_i \in \Upsilon$ is described in a high-dimentiona feature space as follows: $S_i = a_{i1}, a_{i2}, \ldots, a_{in}$.

The set Υ is split into l classes that do not overlap, $K_1, K_2, \ldots, K_l$. Here, the partition of Υ is not determined fully, but is only a certain beginning information I_0 on the classes $K_1, K_2, \ldots, K_l$ (training sample)

$$I_0 = \{S_1, \ldots, S_i, \ldots, S_m;\ \tilde{\alpha}(S_1), \ldots, \tilde{\alpha}(S_i), \ldots, \tilde{\alpha}(S_m)\},$$

$$\tilde{\alpha}(S_i) = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in}),$$

where S_i is the training sample object ($i = \overline{1, m}$), $\tilde{\alpha}(S_i)$ is the informational vector of an object S_i, and α_{ij} is a marker that characterizes the fact that the object S_i belongs to the class K_j ($j = \overline{1, l}$). Here, $\alpha_{ij} \in \{0, 1, \Delta\}$ where 1 means that the object S_i belongs to the object class K_j, while 0 means the opposite, and Δ marks the situation when an algorithm A cannot determine the membership of the object S_i in any of the above classes.

Here, the main task is to calculate the values of the elementary predicate $P_j(S_i)$, $S_i \in K_j$, based on the information I_0 and the description $I(S)$ of the feasible object S. In other words, such an algorithm A has to be constructed that

$$A(I_0, I(S)) = (\alpha_1^A(S),\ \alpha_2^A(S), \ldots, \alpha_1^A(S)), \tag{1}$$

where $\alpha_j^{A(S)} \in \{0, 1, \Delta\}$, $j = 1, \ldots, l$.

The quality function $\varphi(A)$ of the algorithm A is given at the set $\{A\}$ of such algorithms. The specified problem is stated as follows: to find such an algorithm A^* among the algorithms determined in (1) that

$$\varphi(A^*) = \sup_{A \in \{A\}} \varphi(A). \tag{2}$$

In other words, we have to select an algorithm among recognition algorithms that can recognize objects from the control sample the most accurately.

The problem of constructing an integrator by taking integration rules into account is stated as follows.

Let us consider a certain integration rule E that depends on the vector of parameters Θ and that allows us to transform the outputs m of the local classifiers (candidate lists $Y_1, \ldots, Y_m$) into the solution Z. Here, $Y_j = (y_1^{(j)}, \ldots, y_l^{(j)})$ where $y_i^{(j)}$ is the estimation of proximity of the object S_j to the class K_i that is obtained by the jth classifier ($j = \overline{1, m}$; $i = \overline{1, l}$): $Z = E(Y_1, \ldots, Y_n, \Theta)$.

The integrator construction problem is to choose the adopted model E and to optimize its parameters Θ that ensure the highest recognition quality

$$\Theta^* = \arg\max \{\varphi(\Theta^*)\}. \tag{3}$$

SOLVING THE PROBLEM

To solve (1), a local classifier model is proposed to recognize face component images, and it is based on estimating the interconnection between the component features of the face being recognized. The stages of problem solution by the proposed local classifier model are presented in what follows.

Distinguishing highly cohesive feature sets of face component images. At this stage, n^* "independent" subsets of face component features are determined.

Let us consider all feasible non-intersecting subsets from the symbol subset $\{x_1, \dots, x_p, \dots, x_n\}$. Let us denote the total set of these subsets by W. From the set W, let us distinguish n^* of "independent" subsets consisting of highly cohesive features. These subsets are a total set $W_{\Re}$ ($W_{\Re} \in W$) that is determined as $W_{\Re} = \{\Omega_1, \dots, \Omega_q, \dots, \Omega_{n^*}\}$. Moreover, the elements Ω_q ($q = \overline{1, n^*}$) depend both on n^* and the initial information. That is why they are different for each class, K_j ($j = \overline{1, l}$).

Creating a set of identifiable features of face component images. At this stage, a set of identifiable features of face component images n^* is formed.

To form an identifiable feature set of an image of each set component, let us distinguish from Ω_q a single typical member ψ_{i_q} ($\psi_{i_q} \in \Omega_q$). A set of such members determines a set of identifiable features that form a space with the dimension n^*. Here, n^* is much smaller than the input space dimension n ($n^* << n$). Next, let us denote the identifiable feature space as Ψ^*: $\Psi^* = (\psi_{i_1}, \dots, \psi_{i_q}, \dots, \psi_{i_{n^*}})$.

Determining the function of differences $d(F_u, F)$ between the face component images F_u and F. At this stage, the function of difference between face component images F_u and F that describes the difference degree between images and is given in the form

$$d(F_u, F) = \sum_{q=1}^{n^*} \lambda_q (a_{ui_q} - a_{i_q})^2, \tag{4}$$

where a_{ui_q} is the value of the i_q-identifiable feature of the image F_u and λ_q is an unknown parameter that characterizes the importance of the corresponding feature of the face component image is determined.

Determining the generalized function of differences between the image of the component of the face F and the class K_j. At this stage, the generalized difference function is determined that describes the differences between the image of the component of the face F and the class K_j. The generalized difference function has the form

$$D(K_j, F) = \sum_{F_u \in K_j^*} \gamma_u d(F_u, F), \tag{5}$$

where γ_u is the parameter that characterizes the importance of the image F_u in the training sample.

Determining the function proximity $R(K_j, F)$ between the image of the component of the face F and the class K_j. Here, the function of proximity between the image of the component of the face F and the class K_j is determined using radial distribution functions

$$R(K_j, F) = \frac{1}{1 + \tau D(K_j, F)}, \tag{6}$$

where τ is the recognition algorithm parameter. Using radial functions based on (6), in the case of the image of the component of the face F, we can calculate the membership estimations $F: R(K_1, F), \dots, R(K_j, F), \dots, R(K_l, F)$. The membership estimate sequence determines "soft" solutions to the local classifier.

Thus, the above stages form the local classifier model that is based on evaluating connections between the features of the above face component under recognition.

Solving optimization problem (2) in relation to the parameters n^*, Ω_q, γ_u, τ allows us to determine the specific local classifier that corresponds to the component of the face under recognition. Let us consider the solution to the problem of local solution integrator construction that is stated in form (3). There is a number of classifier combination technologies (rules) at the decision-making level that are based on three approaches according to [3], namely, the abstract, the ranking, and the scoring ones.

In the case of the first approach, a single class label is delivered to the classifier input from each classifier. In the case of the second approach, each classifier has a couple of labels ranked from the most to the less likely ones. In the case of the third approach, each classifier derives n^* best labels with accuracy estimations. The above classifier combination approaches are connected by the same idea, namely, the integration of information after each classifier presents its local solutions. Such an integration is justified by the fact that it allows us to combine a set of classifiers with some of them being weak. Classifier integration is aimed at ensuring the necessary level of productivity and recognition quality.

The analysis of scientific works from this field, for example, [5–9], stipulates the insufficient development of the theory of constructing methods of classifier integration and of their efficiency analysis. The following main problems (whose solution necessitates purposeful research) are presented in [5]:

— problems with two classes and low-dimensional feature spaces (many classifier integration methods take effect under certain constraints that are placed upon the number of classes and features in particular);

— theoretical proof that integrating weak classifiers or weak and strong classifiers can lead to higher integrator accuracy and efficiency;

— large-number classifier integration (one of the feasible approaches is the organization of hierarchical classifier structure where the solution is developed gradually, beginning with lower lever classifiers and ending with higher lever classifiers):

— classifier combination if a large number of classes (more than 500) is available.

The multilevel face recognition system considered in this article has a number of possibilities allowing us to reduce the impact of local classifiers when choosing combination rules. First of all, it is feature space dimensionality reduction by using local classifiers where each of them recognizes a certain face component. This allows for omitting the operation of concatenating the face component features under consideration, the use of which leads to the formation of a single feature vector for the whole facial area of high dimensionality. Second of all, it is the use of a limited classifier number that is determined by a small number of components used in facial recognition.

Note that the chosen classifier integration rule has to ensure the solution to the face recognition problem if multiple classes (for example, over 500) are present, which happens when searching for a large face database. As noted in [5], neural networks can successfully adapt to such a situation, justifying the expediency of their use in the role of an integrator in the above two-level face recognition system. When using a simple neural network [12] in this system, the local classifier output values m are transferred to the $m \times l$ input layer neurons of the neural network where l is the number of given classes. At the output integrator level, each of the l neurons generate an estimation of membership of the image of the face being recognized in one of the given classes. Then, the final decision regarding the face being recognized is made based on the estimations using Softmax function. A backpropagation algorithm is selected for neural network learning. The categorical cross-entropy loss on the Loss training samples [9] is used as an error function in the process of neural network learning as follows:

$$\text{Loss} = -\sum_{j=1}^{l} z_j \times \log(z_j^*), \tag{7}$$

where z_j^* is the output value that corresponds to the j-class at the model output (forecasting value), z_j is the previously known value that corresponds to the j-class, and l is the number of classes, $j = -\overline{1, l}$.

The final decision is made based on the probability estimates obtained at the output of the neural network integrator by using Softmax function that is an output function recommended to be used with the categorical cross-entropy error function.

TABLE 1. Quality of Different Integration Variants of Local Classifiers

Integrator	Recognition Accuracy		
	First Face Area (Eyes + Nose)	Second Face Area (Nose + Mouth)	Third Face Area (Eyes + Nose + Mouth)
Neural Network	0.951	0.895	0.973
Bayesian	0.923	0.886	0.945
Log-Linear	0.938	0.891	0.954

FACE RECOGNITION EXPERIMENT

When conducting experiments, the FERET facial imagery database [10] was used, which contains 1187 images of 69 human faces. The size of each image is 256 pixels in widths and 384 pixels in lengths.

To study the efficiency of different classifier models, a face recognition experiment taking into account a component-based approach was performed. Therefore, to conduct experiments in each of the 690 face images chosen from the FERET database (10 images for each of 69 persons), local areas, each of them representing a face component under study, namely, the right eye, the left eye, the nose, and the mouth, were distinguished using the algorithm that implements the modified Viola–Jones method [15]. The images of each component were included into the corresponding input sample. Thus, four input sample types were formed, each of them including 690 images of the corresponding face component. The descriptors determined by the method of local binary templates, which are described in [15] in detail, were used as features of the recognized face components.

The experiments were performed by using the considered neural network integrator, Bayesian integrator [6], and log-linear integrator [3]. The integrator implementation results are presented in Table 1.

As we can see in Table 1, the neural network integrator has a better quality compared to the other two. Here, the log-linear integrator quality is higher than that of the Bayesian one. Thus, the latter is oriented towards independent classifier integration. Taking into account that the local classifiers under integration are dependent as they are developed within a single model, this fact impacted the quality of the Bayesian integrator. Compared to the second face area including the nose and the mouth, the higher face recognition quality of the first area including the eyes and the mouth testifies to the fact that the upper part of the face is more informative than the lower one.

CONCLUSIONS

The existing problems of creating reliable recognition systems incentivized the researchers to propose new approaches to multiclassifier combination. To implement the approaches, combination adaptability has to be researched and combination levels and threshold value types have to be determined when making intermediary and final decisions. The article shows that classifier combination methods can be successfully implemented for the face recognition problems.

The article is mainly focused on stating and solving the problem of constructing local classifier models and choosing the correct rules for their combination. The proposed local classifier model is based on transitioning from the high-dimensional output system of dependent features to an independent identification feature system of a significantly lower dimension. A sufficiently high quality of recognizing main face component combinations (eyes, nose, and mouth) testifies to the efficiency preservation of the multilevel face recognition system considered by the article in the presence of partial occlusion. This feature of the above system broadens its practical application range, especially in the on-board systems responsible for active video surveillance [13].

REFERENCES

1. V. M. Kyyko, K. V. Kyyko, O. V. Nedashkyvskyi, V. V. Matsello, and M. I. Shlesinger, "Method for computer identification of a person by facial image," Patent of Ukraine No. 81514, Bul. No. 1, Appl. Jan 30, 2006, Publ. Jan 10 (2008),
2. Y. Anchishkin, V. Kyyko, Y. Musatenko, M. Schlesinger, I. Kovtun, and K. Milshteyn, "System and method for identifying object in video," United States Patent No. US 8,064,641 B2, Date of Patent: Nov. 22 (2011).
3. A. Tsariov, Y. Musatenko, M. Schlesinger, V. Kyyko, K. Kyiko, and O. Nedashkyvskyi, "Method and system of person identification by facial image," United States Patent No. US 8,184,914 B2, Date of Patent: May 22 (2012).
4. A. Gumede, S. Viriri, and M. Gwetu, "Hybrid component-based face recognition," in: 2017 Conf. on Information Communication Technology and Society (ICTAS), Durban, South Africa, 8–10 March (2017), pp. 1–6. https://doi.org/10.1109/ICTAS.2017.7920665.
5. H. Adam and V. Serestina, "Invariant feature extraction for component-based facial recognition," Int. J. Adv. Comput. Sci. Appl., Vol. 11, Iss. 3 (2020). https://doi.org/10.14569/IJACSA.2020.0110386.
6. S. K. Paul, S. Bouakaz, C. M. Rahman, and M. Shorif, "Component-based face recognition using statistical pattern matching analysis," Pattern Analysis Applic., Vol. 24, 299–319 (2021). https://doi.org/10.1007/s10044-020-00895-4.
7. M. M. Kamilov, Sh. Kh. Fazilov, N. M. Mirzaev, and S. S. Radjabov, Models of Recognition Algorithms Based on the Assessment of the Interconnectedness of Features [in Russian], Fan va Texnologiya, Tashkent (2020).
8. M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers combination techniques: A comprehensive review," IEEE Access, Vol. 6, 19626–19639 (2018). https://doi.org/10.1109/ACCESS.2018.2813079.
9. V. N. Opanasenko and S. L. Kryvyi, "Synthesis of neural-like networks on the basis of conversion of cyclic Hamming codes," Cybern. Syst. Analysis, Vol. 53, No. 4, 627–635 (2017). https://doi.org/10.1007/s10559-017-9965-z.
10. S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann, "Review of classifier combination methods," in: S. Marinai and H. Fujisawa (eds.), Machine Learning in Document Analysis and Recognition, Studies in Computational Intelligence, Vol. 90, Springer, Berlin–Heidelberg (2008), pp. 361–386. https://doi.org/10.1007/978-3-540-76280-5_14.
11. C. De Stefano, F. Fontanella, and A. Scotto di Freca, "A novel naive Bayes voting strategy for combining classifiers," in: Proc. Int. Conf. Front. Hand-Writing Recognit., Bari, Italy, (2012), pp. 467–472. https://doi.org/10.1109/ICFHR.2012.166.
12. Y.-D. Lan and L. Gao, "A new model of combining multiple classifiers based on neural network," in: Proc. 4th Int. Conf. Emerg. Intell. Data Web Technol., Xi'an, China (2013), pp. 154–159. https://doi.org/10.1109/EIDWT.2013.32.
13. S. Chitroub, "Classifier combination and score level fusion: Concepts and practical aspects," Int. J. Image Data Fusion., Vol. 1, No. 2, 113–135 (2010). https://doi.org/10.1080/19479830903561944.
14. Color FERET Database. NIST. URL: https://www.nist.gov/itl/products-and-services/color-feret-database.
15. P. Viola and M. J. Jones, "Robust real-time face detection," Int. J. Comput. Vision, Vol. 57, No. 2, 137–154 (2004). https://doi.org/10.1023/B:VISI.0000013087.49260.fb.
16. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 24, No. 7, 971–987 (2002). https://doi.org/10.1109/TPAMI.2002.1017623.
17. V. Opanasenko, A. Palahin, and S. Zavyalov, "The FPGA-based problem-oriented on-board processor," in: 2019 10th IEEE Intern. Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Vol. 1, Metz, France, 18–21 Sept (2019), pp. 152–157. https://doi.org/10.1109/IDAACS.2019.8924360.