

Video Based Face Recognition Method

Anna V. Pyataeva^{1,2}, Maria V. Verkhoturova¹

¹ Siberian Federal University, Krasnoyarsk, Russia, anna4u@list.ru

² Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russia

Abstract. In current report, method for face detection and recognition based on visual data is proposed. For face detection Viola-Jones algorithm with Haar like features estimation based on cascade architecture were studied. The local binary patterns descriptor used for face recognition stage.

Keywords: face recognition; face detection; local binary pattern; Viola-Jones algorithm; Haar-like features.

1 Introduction

Face recognition technique based on visual processing is significant for many applications. For example in the personal information protection, human-machine interaction, proctoring providing in distance learning platforms, access on the territory of objects with high security level e.t.c. Facial recognition approaches also vary considerably. At the initial stage of development of approaches to face recognition, geometric features were used to highlight the characteristic facial features [1, 2]. Nowadays, to solve this problem, deep learning technologies [3, 4], evolution algorithms [5], particle swarm method [6] and other approaches. On face recognition efficiency may influence various factors like varying expressions and illumination poor [7, 8], subject's pose variations [9], the own-age, -gender, and -ethnicity [10-12] etc.

2 Face identification method

The first step of the proposed user identification algorithm is face detection based on Viola-Jones algorithm. The next algorithm step is face recognition using local binary pattern features estimation.

2.1 Face detection

In first face identification algorithm step one-against-all classification was used. This classification divides image objects into two classes "face" and "no-face objects". The Viola-Jones algorithm is identified as one of the classic approaches to solving the problem of face recognition [13]. The main area of application of the Viola-Jones method is the problem of face detection [14, 15]. The basis of the work of the Viola-Jones method is the identification of the Haar like features and the use of a cascade classification model. A feature of the Viola-Jones method is to work with an integral way of representing the image. The integral image representation is a matrix with the same size as the original image. Each of its elements contains the sum of the pixel intensities locating to the left and above the current element. integral image representation elements for each original image pixel are calculated by the Eq. 1.

$$L(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j), \quad (1)$$

where $I(i, j)$ is the original image pixel intensity, (i, j) its coordinates. Thus, each element of the matrix L is the sum of the intensity values of pixels in a rectangle from a pixel $(0, 0)$ to a pixel with coordinates (x, y) .

The scanning window consisting of adjacent rectangles - the Haar primitives - is moved across the video image being examined to calculate the Haar-like features. It becomes possible to separate face from other image objects by selection such characteristic of the pixels intensity change features. By moving the scanning window over the entire image, the Haar-like features are calculated. These features show the intensity difference value on the region of interest. In the present work basic and additional Haar masks were used (Fig. 1). Use of additional Haar masks allows detecting faces with different angles of rotation to the camera, even faces facing the camera with a rotation angle of

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

more than 30 degrees [16]. At the next stage the Haar-like features in the Viola-Jones algorithm are organized into a cascade classifier. The result of the Viola-Jones classification algorithm is a set of attributes for each area, consisting of 200 values of differences in intensity, allowing separating images containing a face from images without it.

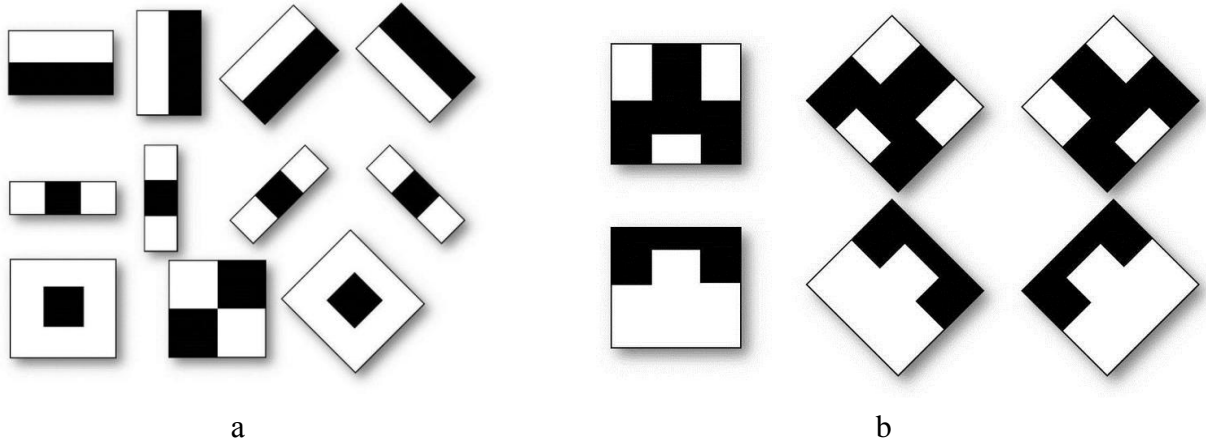


Figure 1. (a) basic Haar-like features; (b) additional Haar-like features.

2.1 Face recognition

The second step is face recognition using local binary pattern (LBP) texture features estimation. The LBP descriptors are computing for person identification. Classification based on local binary pattern operator is widely uses in many applications [17-19]. To personal identification a face image is broken into non-intersecting blocks. The LBP was introduced by Ojala et al. [20] as a binary operator robust to lighting variations with low computational cost and ability of simple coding of neighboring pixels around the central pixel as a binary string or decimal value. The use of local binary patterns for solving the face recognition task is shown in Fig. 2.

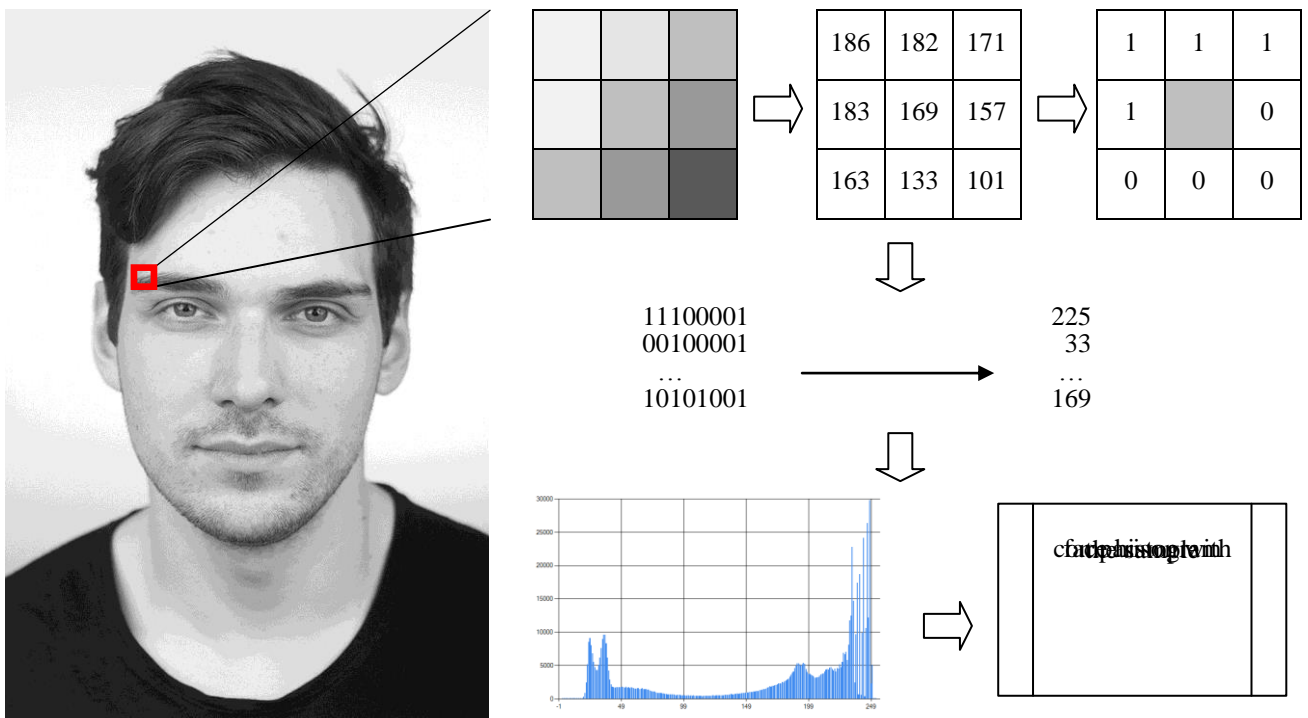


Figure 1. Local binary pattern for face recognition.

The operator $LBP_R(P)$ is calculated in a surrounding relatively a central pixel with intensity I_c by Eq. 2, where P is a number of pixels in the neighborhood, R is a radius, I_c and I_n Y component values is from YUV color space. If $(I_n - I_c) \geq 0$, then $s(I_n - I_c) = 1$, otherwise $s(I_n - I_c) = 0$. Binary LBP code is computing as follows: the current LBP bit is assigned the value “1” if for the current pixel the Y intensity is less than the central pixel intensity, “0” otherwise. In

this manner is calculated P-bit binary LBP code that describes a pixel neighborhood. In this paper, we take into account 8 intensity values of neighboring pixels, that is, the radius R=1, to construct LBP binary code. The pixels are traversed clockwise, the bit width of the LBP binary code is 8.

$$LBP_R(P) = \sum_{n=0}^{P-1} s(I_n - I_c) \cdot 2^n \quad (2)$$

Then a binary code is transformed to decimal code. For histogram computing amount of equal numbers is calculated defining a position and a height of histogram columns. The constructed histograms for different parts of the face are concatenated into one histogram. Chi-square distance, histogram intersection distance, Kullback-Leibler divergence, and G-statistic are usually used during classification stage. In this research, the Euclidian Distance by Eq. 3 was chosen for histogram comparison as the most recommended metric.

$$D = \sqrt{\sum_{i=1}^n (hist1_i - hist2_i)^2}, \quad (3)$$

where $hist1_i$ – column with number i of studied face image histogram, $hist2_i$ – column with number i of the face image histogram from available facial dataset, n - the number of histogram columns. The block diagram of the algorithm for using the local binary pattern operator for the face recognition task is shown in Fig. 3.

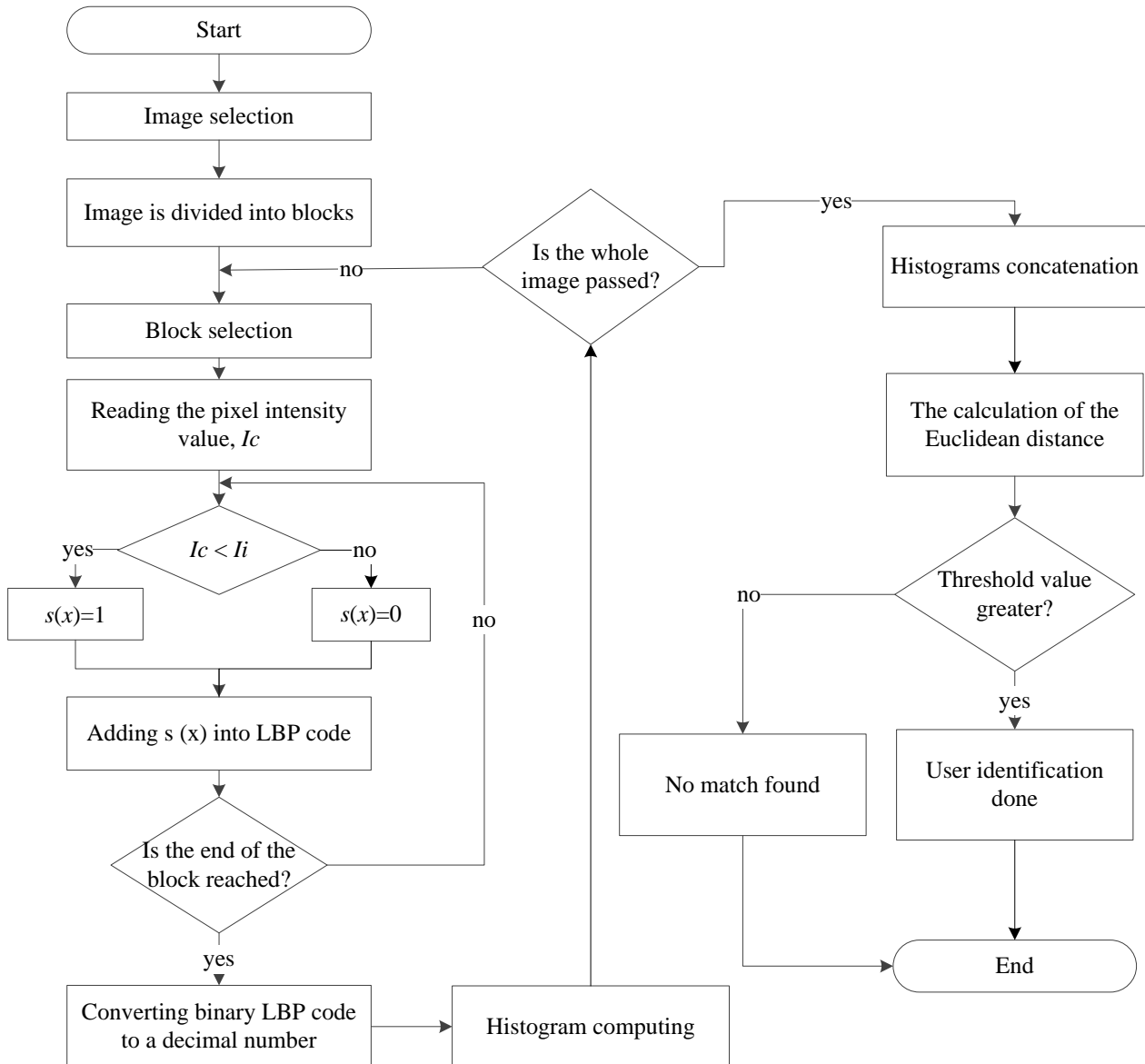


Figure 3. Face recognition algorithm.




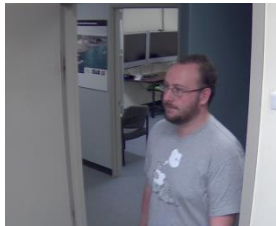




Thus, the combined histogram of facial fragments is compared on a threshold with each of the reference histograms, based on this comparison, user identification is performed.

2 Experimental and results

Experimental studies for the stages of face detection and recognition by video data were carried out separately. For face detection stage sample videos, including 4916 examples of individuals and 8,500 examples with no faces taken from the dataset Labeled Faces in the Wild Home [21] and the dataset Aberdeen [22] were applied. To verify the quality of the face recognition algorithm, the YouTubeFaces (YTF) [23], McGillFaces Database [24] and Db Fases Dataset [25] datasets were used. The dataset contains YouTubeFaces 3425 of video images 1595 different people, recruitment McGillFaces Database 60 videos with images of 40 different people set Db Fases Dataset 22 of the video with images of 38 different people. Video images have different levels of illumination, contain a different number of people of both sexes. At the same time, the number of people and the angle of rotation of the head of person to the camera differ in the videos. Images have different sizes from 160×120 pixels to 1280×720 pixels. The images contain both natural and human-made objects, people. In this case, video shooting was performed both indoors and outdoors. In addition, people were completely free in their movements, which led to arbitrary scale and facial expressions, the position of the head. The videos were in mp4 or avi format. Examples of frames of used videos are shown in Table 1.

Table 1. Description of some used videos.

Description of test video	Sample frame	Description of test video	Sample frame
YouTubeFaces\P1E_S1_C1. mp4. Number of frames: 1125. Resolution: 1280×720 . Number of faces: 2 female faces.		YouTubeFaces\P1E_S2_M6.mp4. Number of frames: 1025 Resolution: 640×480 . Number of faces: 3 female faces.	
YouTubeFaces\P1E_S1_C4.mp4. Number of frames: 1500. Resolution: 1280×720 . Number of faces: 1 female face, 1 male face.		YouTubeFaces\P1E_S2_M1.mp4. Number of frames: 875. Resolution: 640×480 . Number of faces: 1 female face, 1 male face.	
db\faces\crglad\00002.avi Number of frames: 300 Resolution: 160×120 , Number of faces: 1 female face.		YouTubeFaces\P1E_S2_M3.mp4. Number of frames: 750. Resolution: 1280×720 . Number of faces: 2 male faces.	
YouTubeFaces\P1E_S2_M13.mp 4. Number of frames: 850. Resolution: 640×480 . Number of faces: 5 female faces, 1 male face.		McGillFaces Database\mmdm2\ video\sx372.avi. Number of frames: 400 Resolution: 640×480 . Number of faces: 1 female face and 1 male face.	
YouTubeFaces\P1E_S2_D3.mp4. Number of frames: 575. Resolution: 480×320 . Number of faces: 1 female face, 2 male faces.		YouTubeFaces\P1E_S2_D5.mp4. Number of frames: 900. Resolution: 1280×720 . Number of faces: 2 male faces.	
McGillFaces Database\mmdm2\ video\si2028.avi. Number of frames: 575. Resolution: 640×480 .		YouTubeFaces\P1E_S1_C3.mp4. Number of frames: 1250. Resolution: 1280×720 . Number of faces: 1 male face.	

Number of faces: 2 male faces.			
McGillFaces Database\mmdm2\ video\sx102.avi. Number of frames: 425. Resolution: 640×480. Number of faces: 1 male face.		db\faces\crglad\000046.avi. Number of frames: 300. Resolution: 160×120. Number of faces: 2 male faces.	
McGillFaces Database\mmdm2\ video\si1425.avi. Number of frames: 375. Resolution: 640×480. Number of faces: 1 male face.		db\faces\crglad\000050.avi. Number of frames: 300. Resolution: 160×120. Number of faces: 1 male face.	
McGillFaces Database\mmdm2\ video\s11.avi. Number of frames: 325. Resolution: 640×480. Number of faces: 3 male faces.		YouTubeFaces\P1E_S2_M7.mp4. Number of frames: 700 Resolution: 640×480. Number of faces: 2 female faces, 2 male faces.	
YouTubeFaces\P1E_S1_K1.mp4. Number of frames: 300. Resolution: 480×320. Number of faces: 1 child's face.		YouTubeFaces\P1E_S1_K2.mp4. Number of frames: 400. Resolution: 480×320. Number of faces: 1 child's face.	

The training sample was 80%, the test sample was 20% of the total sample. To evaluate the effectiveness of face detection and recognition algorithms, the indicators of detection accuracy (TR), false-positive (FAR) and false-negative (FRR) were used. Faces rotated relative to the camera by an angle of more than 55 degrees were not taken into account. The results of detection and recognition of faces are shown in Table 2.

Table 2. Experimental results.

Video	Face detection			Face recognition		
	TD, %	FRR, %	FAR, %	TR, %	FRR, %	FAR, %
YouTubeFaces\P1E_S1_C1	100	0,00	0,00	99,5	0,50	0,44
YouTubeFaces\P1E_S1_C3	100	0,00	0,00	100	0,00	0,00
YouTubeFaces\P1E_S1_C4	100	0,00	0,00	99,1	0,01	0,66
YouTubeFaces\P1E_S2_M1	100	0,00	0,00	97,5	2,50	1,10
YouTubeFaces\P1E_S2_M3	100	0,00	0,00	96,2	4,00	3,80
YouTubeFaces\P1E_S2_M6	100	0,00	0,00	100	0,00	0,00
YouTubeFaces\P1E_S2_M13	93,9	6,10	5,80	87,9	12,1	11,7
YouTubeFaces\P1E_S2_M7	100	0,00	0,00	100	0,00	0,00
YouTubeFaces\P1E_S2_D3	100	0,00	0,00	100	0,00	0,00
YouTubeFaces\P1E_S2_D5	95,3	6,66	4,70	88,9	11,1	11,1
YouTubeFaces\P1E_S2_D6	100	0,00	0,00	98,2	1,80	1,78

YouTubeFaces\P1E_S2_D8	100	0,00	0,00	100	0,00	0,00
------------------------	-----	------	------	-----	------	------

As shown by the results of experimental studies, gender and age of people do not affect the quality of the algorithm for detecting and recognizing faces. The quality of the algorithm is influenced by such factors as scene illumination level, video resolution, speed of people moving on the stage, face rotation angle and the face openness degree. Thus, an additional error in the algorithm of face detection and recognition is made by accessories worn on the face, such as glasses, scarves, hats. Negative impact is also closing part of the face with hair, beard or mustache. Emotional facial expression in most cases does not affect the results of the algorithm, but it can cause difficulties in recognition, for example, with a wide smile or closed eyes of a person. In addition, when shading a part of the face, the quality of the algorithm may decrease.

Thus, the solution of the problem of facial recognition today is relevant for the implementation of various practical tasks. In the present work, the Viola-Jones algorithm was used to face detection stage; local binary patterns were used for facial recognition. Experimental studies conducted on heterogeneous video data confirm the effectiveness of the proposed methods.

References

- [1] Turk M., Pentland A. Eigenfaces for recognition // J. Cognit. Neurosci. 1991. No. 3 (1). P. 71-86.
- [2] Belhumeur P.N., Hespanha J.P., Kriegman D.J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection // IEEE Trans. Pattern Anal. Mach. Intell. 1997. No. 19 (7). P. 711 - 720.
- [3] Taigman Y., Yang M., Ranzato M., Wolf L. Deepface: closing the gap to human-level performance in face verification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1701-1708.
- [4] Parkhi O.M., Vedaldi A., Zisserman A. Deep face recognition // Proceedings of the British Machine Vision Conference (BMVC). 2015. Vol. 1. P. 41.1-41.12.
- [5] Zhi H., Liu S. Face recognition based on genetic algorithm // Journal of Visual Communication and Image Representation. 2019. Vol. 58. P. 495-502.
- [6] Khan S.A., Ishtiaq M., Nazir M., Shaheen M. Face recognition under varying expressions and illumination using particle swarm optimization // Journal of Computational Science. 2018. Vol. 28. P. 94-100.
- [7] Nikan S., Ahmadi M. A modified technique for face recognition under degraded conditions // Journal of Visual Communication and Image Representation. 2018. Vol. 55. P. 742-755.
- [8] Ding C., Tao D. Pose-invariant face recognition with homography-based normalization // Pattern Recognition. 2017. Vol. 66. P. 144-152.
- [9] Liang Y., Zhang Y., Zeng X.X. Pose-invariant 3D face recognition using half face // Signal Processing: Image Communication. 2017. Vol. 57. P. 84-90.
- [10] Wu S., Wang D. Effect of subject's age and gender on face recognition results // Journal of Visual Communication and Image Representation. 2019. Vol. 60. P. 116-122.
- [11] Muikudi P.B.L., Hills P.J. The combined influence of the own-age, -gender, and -ethnicity bases on face recognition // Acta Psychologica. 2019. Vol. 194. P. 1-6.
- [12] Segal S.C., Reyes B.N., Gobin K.C., Moulson M.C. Children's recognition of emotion expressed by own-race versus other-race faces // Journal of Experimental Child Psychology. 2019. Vol. 182. P. 102-113.
- [13] Viola P., Jones M.J. Rapid Object Detection using a Boosted Cascade of Simple Features // Proceedings IEEE Conf. on Computer Vision and Pattern Recognition. 2001. Vol. 1. P. 511-518.
- [14] Irgens P., Bader C., Lé T., Saxena D., Ababei C. An efficient and cost effective FPGA based implementation of the Viola-Jones face detection algorithm // Hardware X. 2017. No.1. P. 68– 75.
- [15] Nguyen T., Hefenbrock D., Oberg J., Kastner R., Baden S. A software-based dynamic-warp scheduling approach for load-balancing the Viola-Jones face detection algorithm on gpus // J. Parallel Distrib. Comput. 2013. No. 73(5). P. 677–685.
- [16] Pyataeva A.V., Verkhoturova M.V. Face detection using the Viola – Jones algorithm. // Proceedings of the International Scientific Conference «Regional Problems of Earth Remote Sensing» RPEERS 2018, Krasnoyarsk, Russia, 2018, P. 188-191.
- [17] Yuan F., Shi J., Xia X., Zhang L., Li S. Encoding pairwise Hamming distances of Local Binary Patterns for visual smoke recognition // Computer Vision and Image Understanding. 2019. Vol. 178. P. 43-53.

- [18] Xu Z., Jiang Y., Wang Y., Zhou Y., Li W., Liao Q. Local polynomial contrast binary patterns for face recognition // *Neurocomputing*. 2019. Vol. 355. P. 1-12.
- [19] Hassaballah M., Alshazly H.A., Ali A.A. Ear recognition using local binary patterns: A comparative experimental study // *Expert Systems with Applications*. 2019. Vol. 118. P.182-200.
- [20] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition* 1996. No. 29. P. 51–59.
- [21] Labeled Faces in the Wild Home database. Available at: <http://vis-www.cs.umass.edu/lfw/>.
- [22] Aberdeen dataset. Available at: http://pics.psych.stir.ac.uk/2D_face_sets.htm.
- [23] YouTubeFaces dataset. Available at: <http://www.cs.tau.ac.il/~wolf/ytfaces/index.html#download>.
- [24] McGillFaces dataset. Available at: <https://sites.google.com/site/meltemdemirkus/mcgill-unconstrained-face-video-database>.
- [25] Db Fases dataset. Available at: <http://www.videorecognition.com/db/video/faces/cvglab/>.