Scientific Visualization, 2022, volume 14, number 2, pages 1 - 17, DOI: 10.26583/sv.14.2.01

Method for Generating Synthetic Images of Masked Human Faces

M.A. Letenkov[1,A,B], R.N. Iakovlev[2,A,B], M.V. Markitantov[3,A,B], D.A. Ryumin[4,A,B], A.I. Saveliev[5,A,B], A.A. Karpov[6,A,B]

[A] St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS), Russia
[B] St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Russia

[1] ORCID: 0000-0001-5745-5354, 01prime@yandex.ru
[2] ORCID: 0000-0002-6721-9707, iakovlev.r@mail.ru
[3] ORCID: 0000-0001-7987-1025, m.markitantov@yandex.ru
[4] ORCID: 0000-0002-7935-0569, ryumin.d@iias.spb.su
[5] ORCID: 0000-0003-1851-2699, saveliev.ais@yandex.ru
[6] ORCID: 0000-0003-3424-652X, karpov@iias.spb.su

Abstract

This study is devoted to the topical problem of generating synthetic images of human faces. The paper presents a new method for generating images of human faces in protective masks. The proposed method is based on the combined use of a neural network method for detecting three-dimensional facial landmarks (3D-FAN) and three-dimensional modeling tools. Approbation and quality assessment of the proposed method was conducted on a test dataset, which includes 3836 images. The dataset included human faces images of different gender and age, taken at different distances and at various angles relatively to the camera lens. To assess generation results, the method of multi-criteria assessment was used with the involvement of an expert group. For each generated image final scores were formed by averaging the obtained ratings, both by criteria and by experts. During the experiment, the developed method has demonstrated a high and stable quality for the following ranges of face orientations [-20; +55], [-60; +60] and [-70; +80] along the OX, OY and OZ axes, respectively. The final proportion of correctly generated images of masked human faces turned out to be 95.9%.

Keywords: synthetic face generation, synthetic visual data corpus, masked face generation, 3D modeling, 3D-FAN, Blender.

1. Introduction

Today, existing approaches aimed at solving the problems of face detection and recognition both on individual images and on video sequences [1-7] have a number of significant limitations in terms of abilities for their application. In particular, partial overlap or partial face occlusion is one of the most significant factors that negatively affect the quality of predictions for existing solutions.

Taking into account the global spread of the COVID-19 coronavirus pandemic [8], as well as the established measures to curb the growth of the infected citizens number, wearing personal protective equipment is urgent [9-11]. In this connection, there is an increasing demand on specialized biometric identification systems capable of recognition in conditions of partial overlapping of a person's face. The development of specialized datasets or corpuses plays a critical role both in the development of such systems and at the stage of validation of recogni-

tion results. Thus, this study is devoted to urgent problem: the development of an effective method for generating synthetic images of masked human faces.

2. Related works

Considering the issue of generating synthetic images of human faces a large variety of methods have been proposed. The earliest approaches in this field were based on the idea of combining various adjacent areas of human faces from existing sets of images [12]. Later studies considered the possibility of generating new images by sequentially superimposing images of faces in similar orientations [13]. Also, there are several well-known methods where the generation of synthetic images is based on changes in emotions expressed by informants [14] [15].

For most modern methods of generating synthetic images of human faces, the following classification can be proposed:

1. Methods based on three-dimensional (3D) models of faces [16-18];
2. Methods based on the application of generative deep learning models [19].

In general, methods included in the first group are focused on improving the quality of predictions of machine vision models trained on the generated data. In these methods sets of images usually are generated from 3D models of human faces or heads, built in various modeling environments. In particular, in [17, 20], 3D face models are used to generate synthetic high-resolution image sets containing human faces of a small scale. Such datasets can be extremely useful for training neural network face detectors. In another group of solutions 3D models are used to generate synthetic images, where faces are presented in non-standard conditions. Datasets obtained in this way can be used for training high-quality models of face and emotion recognition [21] [22] [23]. The most significant disadvantages of generation methods based on the use of 3D models are highlighted below:

- The integrity of texture and spatial structure of an object in the generated image directly depends on the quality of the 3D model. When 3D models are obtained using photogrammetry methods, texture and polygonal artifacts may appear, which can lead to inability of the subsequent use of the generated images, since the presence of even minor artifacts directly affects the anthropometric characteristics of the modeled object.
- To obtain high-quality generation results, you need to use high-polygonal 3D models, as well as textures with a sufficiently high resolution, which is not always possible.
- Using high-quality models and textures, as well as rendering images in high resolution, the generation time non-linearly increases, since the rendering process is quite resource-intensive.

Despite the above disadvantages, methods for generating synthetic images of human faces based on the 3D models have a number of undeniable advantages. Due to the wide functionality of modern software, it becomes possible to simulate absolutely any environmental conditions. Particularly, in addition to the ability of controlling the spatial arrangement of light sources relative to a 3D face model, managing the textures of the 3D models and the surrounding scene is also available. One of the most valuable features of modern software is the ability to define and set up low-level materials types of the 3D models being used. This functionality can be especially useful when generated datasets are used in the processes of development and testing of anti-spoofing software [24-26]. However, the generated dataset representativeness highly depends on the number of 3D models involved, in this connection the full-fledged use of the described group of methods in current study is not possible.

Methods included in the second group predominantly involve the use of generative adversatial neural network models (Generative Adversatial Networks - GAN) [27]. One of the first successful applications of GAN in the problem of generating images of human faces is described in [28]. However, the quality of the generated samples, demonstrated in the given solution, remained at a rather low level, moreover synthetic images of faces themselves had a low resolution and in most cases did not preserve the integrity of graphs, built on the basis of anthropometric points.

In recent years, with the development of GAN, solutions have emerged that shows better results. Currently existing models from only one frontal face image allow generating photore-

alistic images at different viewing angles [29, 30, 31]. The use of GAN makes it possible to influence not only the spatial orientation of faces, but also to change such characteristics as gender, skin color, and even age [32, 33]. The StyleGAN network presented by a group of researchers from NVIDIA [34] can generate high-resolution photorealistic images of faces, providing the ability to influence the values of such high-level attributes as skin color, hair style, angle, and the presence of glasses.

However, despite all the capabilities of GANs, their application in the context of the described problem is complicated because of the following disadvantages: GANs training requires large annotated datasets (corpuses). Training visual corpuses should contain high-resolution images of human faces, both in conditions of partial overlapping with masks or other personal protective equipment (PPE), and without any overlapping. Additional requirements for representativeness and minimal volume of training dataset also are arise from the wide variety of existing PPE: various colors textures, and geometric characteristics. Another significant disadvantage of GANs is the high probability of the occurrence of graphic artifacts at the texture and structural levels of the face, which directly affects the anthropometric graph and respectively the ability of using the obtained datasets for training face recognition models. In addition, the change in the characteristics of generated faces in existing solutions is achieved by influencing the representations of faces in the latent space, which implies the absence of both the transparency of the generation process and full-fledged control over this process.

Thus, it can be concluded that it is inexpedient to use GAN based methods for generating synthetic human face images in the context of the current research.

Another method for generating synthetic facial images should also be considered. This stand-alone method called MaskTheFace [35, 36] does not fit the criteria of the proposed classification and directly allows synthetic images of PPE to be imposed over images of human faces.

MaskTheFace basely relies on the models presented in the computer vision library Dlib [2], namely the face and facial landmarks detectors. This solution also proposes a proprietary algorithm for impose PPE. In accordance with the algorithm, firstly, the position and orientation of the face on the examined image is estimated. Secondly, the PPE image is selected from the annotated set of PPE images. The PPE image selection algorithm ensures the selection of such a PPE image, where the orientation most closely matches the face orientation in the image being considered. At the final stage, the selected PPE image is scaled and stretched to the face area by applying affine and perspective transformations (face area is represented by a heuristically specified group of facial landmarks). The most significant advantages of this method for generating synthetic images include: high speed of the image generation process, relatively low resource intensity of the solution, deterministic nature of proposed algorithms, does not require training on large visual corpuses. Moreover, the ability of using this approach for augmentation of already existing datasets in some cases allows to improve the quality of predictions of neural network recognition models trained on corpuses, extended in this way. Examples of PPE impose using MaskTheFace are shown in Figure 1.

Fig. 1. Examples of generating synthetic face images with impose of PPE using MaskTheFace [36].

Despite all the advantages of MaskTheFace, this method of generating synthetic images has a number of significant disadvantages:

- For faces and facial landmarks detection Dlib models are used, which in some cases demonstrate an extremely low prediction quality level.
- Correct impose of PPE is possible only in cases where the facial orientation is close to a frontal. The quality of PPE overlay is rapidly deteriorating with face rotation angles increase. Method incorrectly operates in such cases because of the general logic's imperfection of the heuristic algorithm for determining facial orientations and the facial landmarks detector Dlib imperfection, which is unable to correctly calculate the facial landmarks coordinates at large face rotation angles (relatively to the frontal position).
- Inability to correctly impose masks on faces oriented at complex angles (sequential rotation of an object relative to two or more axes of a three-dimensional coordinate system basis).
- In the process of imposing PPE, affine and perspective transformations are used, which significantly affects both the texture quality of the applied PPE and its structural integrity.

Examples of incorrect PPE imposing using the MaskTheFace method are shown in Figure 2.

Fig. 2. Examples of incorrect generation of synthetic face images with overlaid PPE, obtained using MaskTheFace.

Based on the analysis results it can be concluded that at the moment there are no solutions that would allow implementation the synthetic face image generation with imposed PPE and at the same time would ensure high quality of PPE overlay in case of complex face orientations in the processed image.

Thus, within this study, the most promising solution to the problem of generating synthetic images of human faces imposed with PPE is the development of a combined solution, where 3D modeling techniques are used for image generation process and deterministic algorithmic basis – for PPE imposing process. Further, the proposed method for generating synthetic images of human faces with PPE imposed will be considered.

3. The proposed approach

The proposed method for generating synthetic images of human faces in PPE is based on a combination of a synthetic images' generation method using 3D modeling, and a proprietary algorithm for imposing PPE. In general, the proposed method can be represented by the following stages:

1. Finding the region of interest on an image containing a person's face.
2. Determination of 3D facial landmarks coordinates on the region of interest using a neural network detector FAN [37].
3. Determination of face spatial orientation and size based on a set of 3D facial landmarks coordinates.

4. Scaling and spatial orientation control of the target PPE model on a pre-prepared scene in a 3D modeling environment. Scaling and orientation control are carried out in accordance with face orientation and previously determined PPE scaling factor.

5. Rendering the oriented PPE model by means of a 3D modeling tool.

6. Imposing the PPE image obtained at the previous stage on the original face image according to the developed algorithm.

To implement the proposed generation method, FAN neural network detectors from the repository [38] were used. This repository includes both face detectors designed to determine region of interest coordinates, and detectors of two-dimensional (2D-FAN) and three-dimensional (3D-FAN) facial landmarks on images. Within the proposed method, a 3D facial landmarks detector was chosen. The rejection to use the 2D-FAN detector was due to the low accuracy of assessing the face orientation on the basis of 2D facial landmarks, especially for cases of complex face orientations in the studied images. The 3D-FAN 3D facial landmarks detector, taking as input an image bounded by the coordinates of the region of interest, returns a 68 × 3 matrix **D**, each row of which defines the coordinates (X, Y, Z) of some facial landmark in a 3D non-metric coordinate system with an orthonormal basis, where the abscissa and ordinate axes are parallel to the upper and left sides of the region of interest, respectively, and the applicate axis is directed from the image plane towards the observer. The origin of the described coordinate system coincides with the upper left corner of the region of interest. Examples of detecting 3D facial landmarks are shown in Figure 3. The 3D editor Blender was chosen as a modeling environment [39].

Fig. 3. Examples of detecting 3D facial landmarks using 3D-FAN model.

Before applying the developed method, it is necessary to carry out a series of preparatory actions and obtain a number of primary assessments. At the first step of the preparatory stage, a 3D scene is built inside the modeling environment, as well as several light sources are placed. The next step is to place a set of basic 3D models of human heads inside the 3D scene (Fig. 4).

Fig. 4. Example of a basic 3D model of a human head.

All models from the aforementioned set of basic head models are scaled down (normalized). For each basic head model, 3D-FAN detector determines a matrix of 3D facial landmark coordinates $\mathbf{D_{head}}$. Based on the resulting array of coordinates, the geometric center of facial landmarks $\mathbf{C_{head}}$ is calculated. $\mathbf{C_{head}}$ is defined as the unweighted average of coordinates of all facial landmarks along each axis (1). Each base head model from the aforementioned set is positioned in the 3D scene so that facial landmarks geometric center $\mathbf{C_{head}}$ coincides with the coordinates center of the scene. All basic head models are reduced to the same scale.

$$C(x, y, z) = \begin{cases} x = \frac{\sum_{i=1}^{N} y}{N} \\ y = \frac{\sum_{i=1}^{N} y}{N}, \\ z = \frac{\sum_{i=1}^{N} z}{N} \end{cases} \quad (1)$$

where N is the number of facial landmarks, in our case equal to 68, x, y, z are the coordinates of the $\mathbf{C_{head}}$ along the corresponding axes.

For each basic head model, the linear face size $\mathbf{L_{head}}$ (2) is calculated, which is equal to the length of the $\mathbf{S_{head}}$ vector between 1 and 17 3D facial landmarks of the basic head model $\mathbf{D_{head}}$ (Fig. 5).

$$L = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}, \quad (2)$$

where (x2, y2) - coordinates of facial landmark No. 17, (x1, y1) - coordinates of facial landmark No. 1.

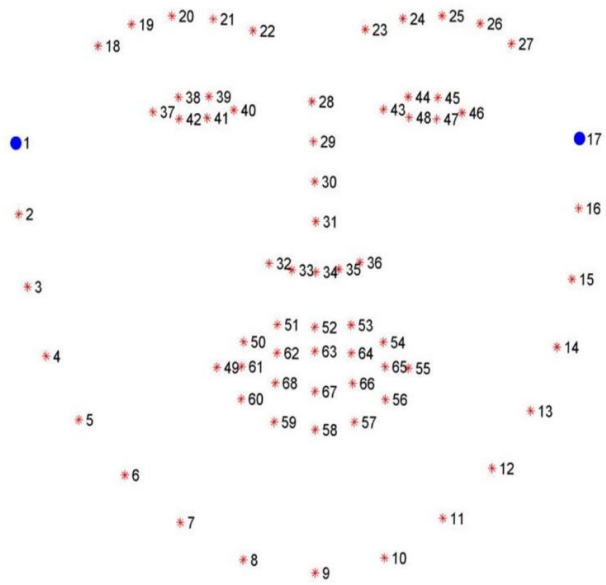


Fig. 5. Numeration of 3D facial landmarks retrieved using the 3D-FAN model.

At the next step of the preparatory stage, 3D models of PPE are placed inside the modeling scene. To form its own database of PPE models, publicly available 3D models of PPE of various types were used.

The process of adding each PPE model to a 3D scene at the preparatory stage is accompanied by the following operations. The PPE model is manually positioned and scaled relative to a set of basic 3D head models placed inside the simulation

scene. The PPE model is positioned as naturally as possible in each individual case. After positioning and primary scaling of the PPE model, a set $\mathbf{D_{mask}}$ of 3D landmarks of PPE is empirically selected (Fig. 6). Based on the array of coordinates of PPE landmarks, the geometric center of the entire PPE model $\mathbf{C_{mask}}$ is calculated. $\mathbf{C_{mask}}$ is calculated as an unweighted average of coordinates of all PPE landmarks along each axis (1).

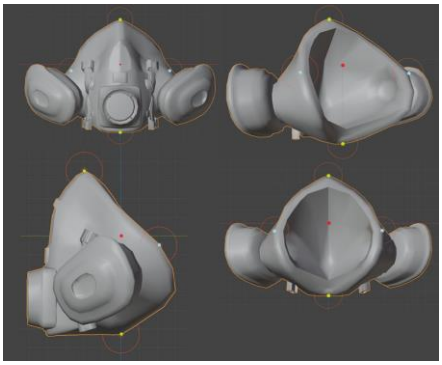


Fig. 6. Visualization of a PPE landmarks set D_{mask}. The geometric center C_{mask} is highlighted in red, the D_{mask} points are marked in light green. Points m1 and m2 are highlighted in blue.

From the $\mathbf{D_{mask}}$ PPE landmarks set, two equidistant key points are selected $\mathbf{m1(x, y, z) = D_{mask}[p]}$ and $\mathbf{m2(x, y, z) = D_{mask}[q]}$ (symmetric about the vertical axis). Based on the coordinates of the selected points **m1** and **m2**, the linear size $\mathbf{L_{mask}}$ of the PPE model is calculated. $\mathbf{L_{mask}}$ corresponds to the length of the $\mathbf{S_{mask}}$ vector between points **m1** and **m2** (2). Using the obtained value of the PPE model linear size, as well as previously calculated linear dimensions of basic head models $\mathbf{L_{head}}$, the average scaling factor $\mathbf{K_{mask}}$ for the PPE model is calculated (3).

$$K_{mask} = \frac{\sum_{i=1}^{n} \frac{L_{mask}}{L_{head(i)}}}{\mathrm{n}}, \tag{3}$$

where N is the total number of basic head models within the modeling scene, L_{mask} is the linear size of the PPE model, L_{head} (i) is the linear size of the i-th basic head model.

Upon completion of this step, the preparatory stage of the proposed method is considered completed.

The implementation of the first stage of the developed method is performed using the FAN face detector. Receiving an arbitrary resolution image containing N human faces, FAN detector returns an (N x 4) matrix, each row of which contains coordinate values (x1, y1) and (x2, y2) that define the region of interest for each face detected in the image.

The 3D-FAN facial landmarks detector used at the second stage of the developed method receives a face image limited by the region of interest (obtained at the first stage) and returns the $\mathbf{D_{real}}$ matrix.

The determination of the face spatial orientation, performed at the third stage, implies calculations of the rotation matrix $\mathbf{M_{face}(\alpha)}$ (4) and the subsequent determination of Euler angles X ', Y', Z'. Face spatial orientation determination results are shown in Figure 7. In addition to the face spatial orientation, this stage also includes linear face size $\mathbf{L_{real}}$ calculation (formula 2).

$$M(\alpha) = \begin{pmatrix} U(t) \\ V(t) \\ W(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha t & -\sin \alpha t \\ 0 & \sin \alpha t & \cos \alpha t \end{pmatrix} \begin{pmatrix} U(0) \\ V(0) \\ W(0) \end{pmatrix} \tag{4}$$

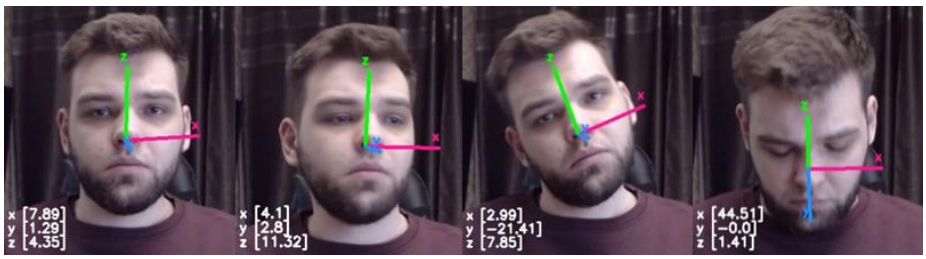


Fig. 7. Visualization of face spatial orientation determination results. In the lower left corner of each image, face rotation angles are presented relative to each axis. A C_{head} points were selected to position the orthogonal basis on each of images.

The fourth stage of the developed method includes following operations: scaling PPE model; control of PPE model orientation in simulated 3D scene. Since for each PPE model, linear dimensions and averaged scaling factors were calculated at the preparatory stage, then

for the final scaling of the PPE model it is necessary to stretch it along all axes by a value, calculated as the product of the target face's linear size $\mathbf{L_{real}}$ and the average scaling factor of the PPE model $\mathbf{K_{mask}}$ (formula 5).

$$\mu = L_{real} \cdot K_{mask}, \quad (5)$$

where L_{real} is the linear size of the target face, K_{mask} is the average scale factor of the PPE model.

The orientation control of the PPE model inside the modeling scene is performed by rotating PPE model relative to each of the scene axes by the angles X', Y', Z' respectively (rotation angles' equations have been presented above).

At the fifth stage necessary light sources are activated in simulated scene. At this step, it is possible to simulate absolutely any lighting options affecting the PPE model. Controlled parameters include the number and positions of involved light sources, as well as their types. One of options for the light sources location in a scene is illustrated in Figure 8. After successful activation and configuration of all light sources, the RGBA image of the selected PPE model is rendered.

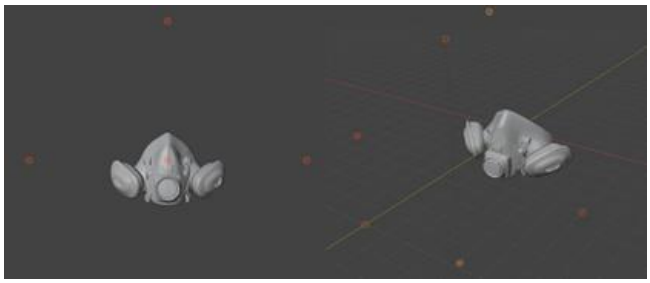


Fig. 8. An example of light sources location in a Blender 3D scene. Light sources are highlighted in orange.

PPE scaling and orientation control processes occur at the level of a 3D modeling environment so after rendering the PPE image it is only necessary to impose PPE image on the target face image. The imposing operation is the process of replacement target image pixels with pixels of the PPE image which was previously subjected to a parallel transfer operation (6).

$$(x,\ y) \rightarrow (x + \dot{x},\ y + \dot{y}), \quad (6)$$

where $\dot{x}$ is the x-axis displacement, $\dot{y}$ is the y-axis displacement.

To determine the values of $\dot{x}$ and $\dot{y}$, it is necessary to project onto the plane of the PPE image the 3D vector $\mathbf{H}$, the beginning of which coincides with the $\mathbf{C_{head}}$ point, and the end is at the $\mathbf{C_{mask}}$ point. Some examples of the proposed method operation results are shown in Figure 9.

Fig. 9. Examples of synthetic images of masked human faces generated using the developed method. The top and bottom rows of images represent generation results with different lighting configurations.

Further, let us assess the quality of the proposed method for generating synthetic images of human faces wearing the PPE on a test dataset.

4. Experiment results

At the moment, there are no algorithms or approaches that allow obtaining a qualitative assessment of the methods for generating synthetic images of human faces presented in conditions of partial impose of PPE. Thus, it was decided to assess the quality of image generation by means of an expert assessment of the generation results. For the experiments, a subset of the corpus of audiovisual Russian-language data of persons in protective masks (**BRAVE-MASKS** - Biometric Russian Audio-Visual Extended MASKS corpus), consisting of 7 informants, was used.

Each of the images used to form the original **BRAVE-MASKS-CL** dataset contains one face oriented at arbitrary angles relative to the OX, OY, and OZ coordinate axes, respectively. Images from the original dataset include non-PPE faces captured at two different distances from the fixing camera. General characteristics of the original **BRAVE-MASKS-CL** dataset are presented in Table 1.

Table 1. General characteristics of the original **BRAVE-MASKS-CL** dataset, represented by a subset of the **BRAVE-MASKS** visual corpus.

Characteristic	**Value**
Number of images, pcs.	4228
Image resolution, pixels	1920 x 1080
Number of informants, pcs.	7
Average number of images per informant, pcs.	604
Maximum angle of face rotation along the X axis, degrees	-80; +80
Maximum angle of face rotation along the Y axis, degrees	-90; +90
Maximum angle of face rotation along the Z axis, degrees	-80; +80

The original dataset includes images of persons of different sex and age to provide the necessary variability of data.

Based on the results of applying the developed method for generating synthetic images to the original dataset, the **FACE-3D-GEN** dataset was obtained, characteristics of **FACE-3D-GEN** are presented in Table 2.

Table 2. General characteristics of the **FACE-3D-GEN** dataset.

Characteristic	**Value**
Number of images, pcs.	3836
Image resolution, pixels	1920 x 1080
Number of informants, pcs.	7
Average number of images per informant, pcs.	548

The **FACE-3D-GEN** dataset contains synthetic images of faces rendered with overlapping PPE. The received dataset is slightly smaller than the original **BRAVE-MASKS-CL** dataset as during images processing, in a number of cases it was impossible to carry out face detection by means of the neural network detector FAN. The corresponding images were excluded from the final **FACE-3D-GEN** dataset. Fig. 11 shows example images from the **FACE-3D-GEN** synthetic dataset.

Fig. 10. Sample images from the **BRAVE-MASKS-CL** dataset.

Fig. 11. Sample images from the FACE-3D-GEN synthetic dataset.

In order to conduct an expert assessment of the synthetic images generation results from the FACE-3D-GEN dataset, a number of ordinal criteria were determined for the implementation of the procedure of evaluating the obtained images:

1. Natural orientation of PPE: PPE is oriented in accordance with the orientation of the face in the image;

2. Natural positioning of PPE: PPE on the image is located correctly relative to the position of the person's face;

3. Naturalness of the scale of PPE: the size of PPE in the image corresponds to the size of the person's face.

For each of the above criteria, an ordinal scale of the following type was set: 1 - does not correspond completely; 2 - mostly inconsistent; 3 - partially corresponds; 4 - mostly consistent; 5 - fully consistent. Together, these criteria facilitate making an unambiguous conclusion about the quality of synthetic images generation using the developed method.

In the generation results assessment process participated 5 independent industry experts with applied experience in the field of computer vision and image analysis. Each of the experts independently assessed the entire **FACE-3D-GEN** dataset against the criteria above. Fig. 12a, 12b, 12c show expert assessments distribution histograms according to criteria (1), (2) and (3), respectively.

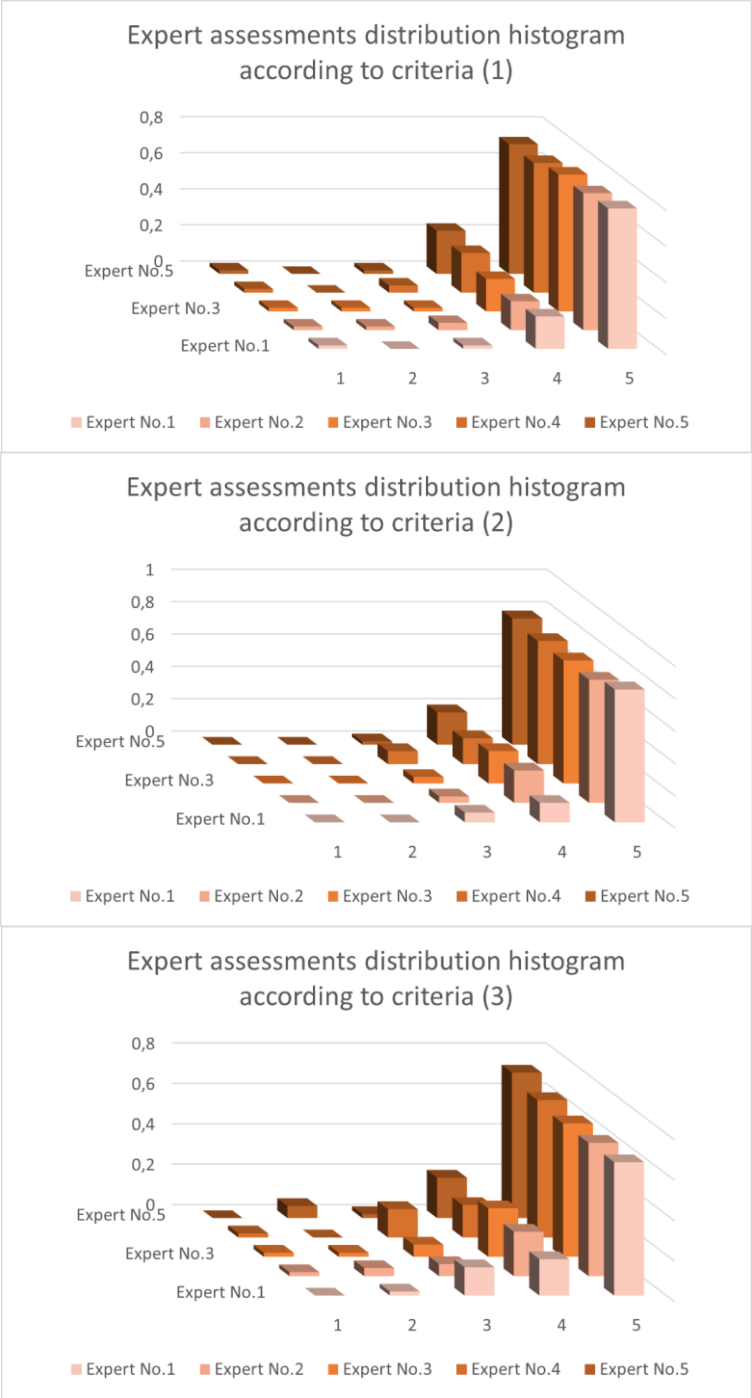


Fig. 12. Distribution histograms of expert assessments according to criteria (1), (2) and (3), respectively, based on the results of evaluating images from the **FACE-3D-GEN** dataset.

For all histograms shown in Fig. 12, the results of evaluating the dataset are similar regardless of the expert, no critical differences between distributions of ratings are observed for any of the criteria. It is important to note that for all criteria high assessments of the quality of the generated images prevail: the share of maximum assessments averaged over experts is 74.8%, 77.6% and 67.6% according to criteria (1), (2) and (3), respectively. Nevertheless, it is important to note that the **FACE-3D-GEN** dataset was assessed slightly lower according to criterion (3) by all experts: the averaged proportion of low assessments (less than or equal to 3) in this case, was 12.4% versus 5.6% and 4.8% for the criteria (1) and (2) respectively.

At the next step of this experiment, the resulting quality estimates of the generated synthetic images were determined and the analysis of the obtained results was carried out. The final estimates for each image from the **FACE-3D-GEN** dataset were formed by averaging the ratings given by experts for this image, both by experts and by criteria. As a measure of the central tendency within the experiment, it was decided to use: the truncated mean at the level of each individual criteria (the proportion of cutoff estimates was chosen equal to 40%), and the arithmetic mean when combining image ratings by various criteria.

Based on the obtained results, a distribution histogram of final estimates was formed (Fig. 13) and also the final proportion of incorrectly generated images was determined.

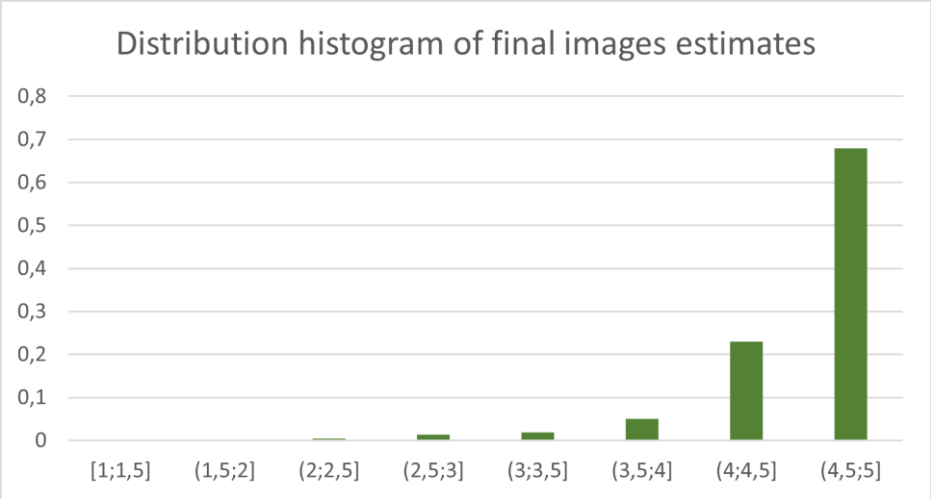


Fig. 13. Distribution histogram of final estimates for images from the **FACE-3D-GEN** dataset.

According to the histogram presented above, the proportion of images for which the value of the final estimates was greater than or equal to 4 was more than 90% of the total size of the **FACE-3D-GEN** dataset, which confirms high quality of the developed method for generating synthetic images of masked human faces. However, it should be noted that the information presented in Fig. 13 is not enough to make a final conclusion regarding the proportion of correctly generated images, since a high value of the image assessment according to two of the three specified criteria, with a low assessment according to the third criterion, can result in a high value of the final average assessment of this image and lead to false acceptance of such an image as correct.

As mentioned earlier, the specified criteria only in aggregate facilitate making an unambiguous conclusion about the quality of the generated image. Thus, during the evaluation process the image was considered as correctly generated if the following conditions were met simultaneously: in the context of all criteria, the resulting assessment of this image turned out to be greater than or equal to 3; the final score of this image, averaged over the criteria, turned out to be greater than or equal to 3.5. In accordance with these conditions, the proportion of correctly generated masked human face images was equal to 95.9%.

Based on the assessment results, samples with correct and incorrect generation results were identified. Out of 3836 images, 3679 were marked as correct, the remaining 157 were found to be incorrect (Fig. 14). Thus, the proportion of incorrect generation results is 4.11%.

Figure 14a shows generation examples, marked by experts as not meeting the criterion of natural orientation of PPE relative to face orientation. Figure 14b illustrates examples of non-compliance with the PPE positioning criterion. The generation examples shown in Figure 14c do not meet the PPE scale criterion. Most of the generated images recognized as incorrect are characterized by low evaluation results according to several criteria at once. The general characteristics of final sets with correct and incorrect generation results are presented in Table 4.

Fig. 14. Examples of incorrect PPE imposing according to criteria (1), (2) and (3), respectively.

Table 4. General statistical characteristics of image sets with correct and incorrect generation results.

Characteristics	**Axis**					
	OX		OY		OZ	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Average value, deg.	9,36	-32,60	-1,81	-3,99	0,86	1,07
Median	13,79	-35,29	-2,40	-4,74	2,37	1,41
Dispersion	199,09	393,67	305,35	935,76	492,65	3997,91

Table 4 presents statistical characteristics of sets with correct and incorrect generation results in the context of face orientation angles in the images processed. According to the above results, the sample with incorrect generation results has a higher level of dispersion over face orientation angles, which indicates that it contains a larger number of images in which human faces are oriented at large angles along one or another coordinate axis relative to the image plane.

Fig. 15 presents a scatter distribution diagram of correct and incorrect generation results depending on the orientation of the face in the image.

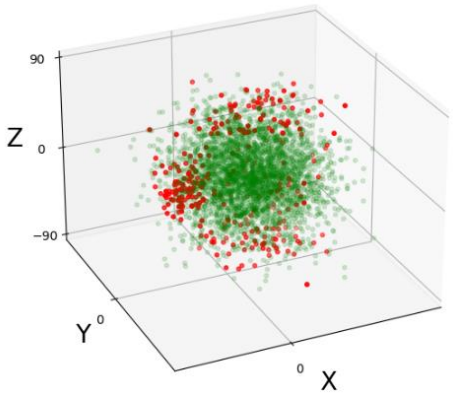
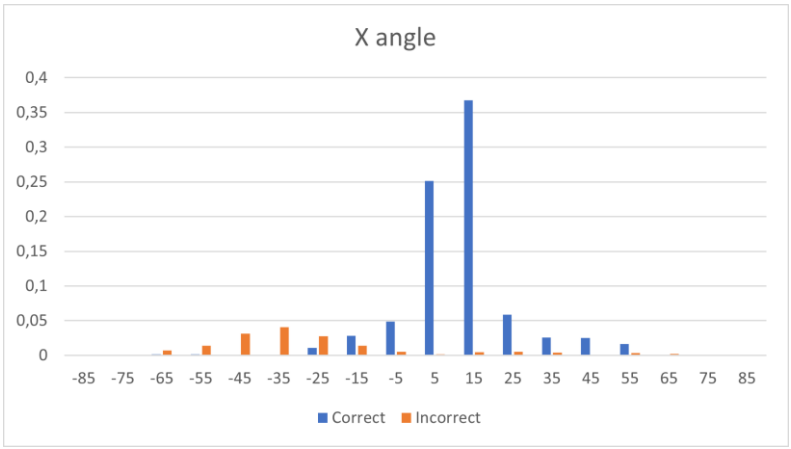
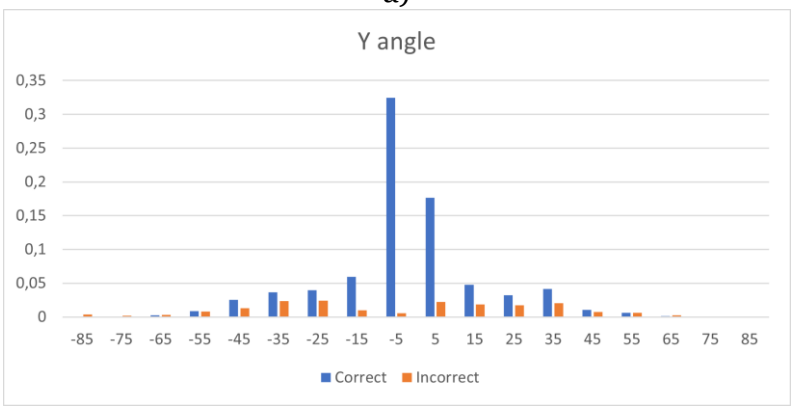


Fig. 15. Scatter distribution diagram of correct and incorrect generation results.

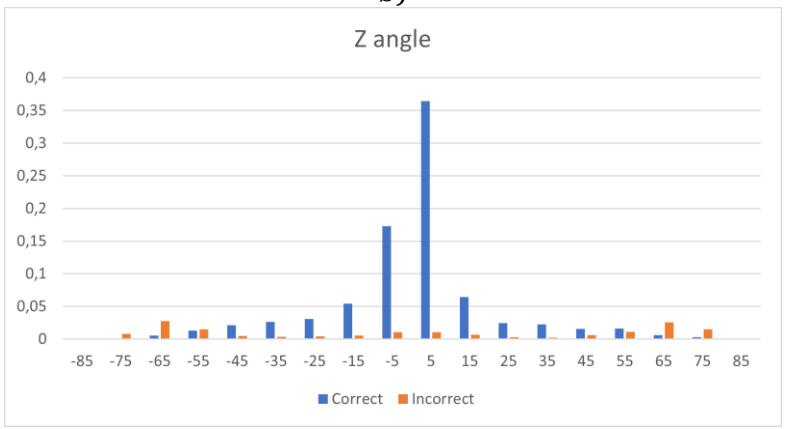
Each point of the diagram above is associated with a single image, and its position is determined by the face rotation angles along the OX, OY and OZ axes, respectively. On the diagram, three groups of points can be distinguished, where the main number of incorrect generation results is concentrated. For further analysis, we will construct distribution histograms of the face rotation angles along the OX, OY and OZ axes within the sets, containing correct and incorrect generation results (Fig. 16).



a)



b)



c)

Fig. 16. Distribution histograms of target face rotation angles along the OX (a), OY (b) and OZ (c) axes within the sets, containing correct and incorrect generation results.

According to the results presented in Figures 15 and 16, it can be concluded that incorrect generation results are mainly observed in the case of: extremely large and small values of face rotation angle around the OZ axis; significant negative values of face rotation angle around the OX axis. The obtained results can be explained by the fact that at large face rotation angles along the OZ axis, a significant area of the investigated face is not observed in the image, which distorts facial landmarks detector results and introduces an error in the PPE imposition process.

Similarly, in the case of significant negative face rotation angles along the OX axis, the upper third of the face, where the concentration of facial landmarks is the highest, ceases to be fully observed in the image, which also distorts facial landmarks detector results and, accordingly, leads to errors when imposing PPE. Thus, the developed method has demonstrated a stable quality of work when target face orientations on the original images were in the intervals [-20; +55], [-60; +60], [-70; +80] for the OX, OY and OZ axes, respectively.

5. Conclusion

In this study, a method for generating synthetic images of masked human faces was developed and its practical approbation was carried out. With this method can be obtained synthetic images of faces in PPE of various shapes, colors and textures, which can be presented in different lighting conditions. During the experiment devoted to the synthetic images generation of human faces in PPE with a wide range of target faces orientations the developed method demonstrated high and stable quality of work for the following ranges of face orientations [-20; +55], [-60; +60] and [-70; +80] along the OX, OY and OZ axes, respectively. The proportion of correctly generated face images turned out to be 95.9% according to the results of applying the developed method to a dataset of 3836 unique images.

Thus, the proposed method for generation synthetic masked human faces images is able to successfully generate synthetic images in a wide range of target face orientations. The developed solution can be used to form datasets for training face recognition systems, specialized on recognition faces in protective masks.

This research was supported by the RFBR project № 20-04-60529 "Analysis of voice and facial features of a human in a mask", as well as partially by the Grant of President of Russia № NSH-17.2022.1.6.

References

1. Zhang K. et al. Joint face detection and alignment using multitask cascaded convolutional networks //IEEE Signal Processing Letters. – 2016. – Т. 23. – №. 10. – С. 1499-1503.
2. Dlib C++ Library: [Электронный ресурс]. URL: http:// http://dlib.net. (Дата обращения: 12.12.2021).
3. Deng J. et al. Retinaface: Single-stage dense face localisation in the wild //arXiv preprint arXiv:1905.00641. – 2019
4. Zhang F. et al. Accurate face detection for high performance //arXiv preprint arXiv:1905.01585. – 2019.
5. Li J. et al. DSFD: dual shot face detector //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2019. – С. 5060-5069.
6. Schroff F., Kalenichenko D., Philbin J. Facenet: A unified embedding for face recognition and clustering //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2015. – С. 815-823
7. Parkhi O. M., Vedaldi A., Zisserman A. Deep face recognition. – 2015
8. Spinelli A., Pellino G. COVID-19 pandemic: perspectives on an unfolding crisis //The British journal of surgery. – 2020.
9. Feng S. et al. Rational use of face masks in the COVID-19 pandemic //The Lancet Respiratory Medicine. – 2020. – Т. 8. – №. 5. – С. 434-436.
10. Cheng K. K., Lam T. H., Leung C. C. Wearing face masks in the community during the COVID-19 pandemic: altruism and solidarity //The Lancet. – 2020.
11. Rab S. et al. Face masks are new normal after COVID-19 pandemic //Diabetes & Metabolic Syndrome: Clinical Research & Reviews. – 2020. – Т. 14. – №. 6. – С. 1617-1619.
12. W. Liu, D. Lin, and X. Tang, “Neighbor combination and transformation for hallucinating faces,” in Proc. IEEE ICME, 2005.
13. D. Bitouk, N. Kumar, S. Dhillon, S. Belhumeur, and S. K. Nayar, “Face swapping: Automatically replacing faces in photographs,” ACM Trans. on Graphics (TOG) - Proc. ACM SIGGRAPH, vol. 27, no. 3, 2005

14. F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas, “Expression flow for 3d-aware face component transfer,” ACM Trans. on Graphics (TOG) - Proc. ACM SIGGRAPH, vol. 30, no. 4, 2011

15. B. Samarzija and S. Ribaric, “An approach to the deidentification of faces in different poses,” in Proc. MIPRO, 2014

16. Malov D., Letenkov M. Synthetic Data Generation Approach for Face Recognition System //Proceedings of 14th International Conference on Electromechanics and Robotics “Zavalishin's Readings”. – Springer, Singapore, 2020. – С. 501-510.

17. Peng X. et al. Learning deep object detectors from 3d models //Proceedings of the IEEE International Conference on Computer Vision. – 2015. – С. 1278-1286.

18. Малов Д. А., Летенков М. А. Методика генерации искусственных наборов данных и архитектура системы распознавания лиц для взаимодействия с роботами внутри киберфизического пространства //Робототехника и техническая кибернетика. – 2019. – Т. 7. – №. 2. – С. 100-108.

19. Kim T. et al. Learning to discover cross-domain relations with generative adversarial networks //International Conference on Machine Learning. – PMLR, 2017. – С. 1857-1865.

20. WIDER FACE: A Face Detection Benchmark: [Электронный ресурс]. URL: http://shuoyang1213.me/WIDERFACE. (Дата обращения: 12.12.2021).

21. Yan W. J. et al. CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces //2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG). – IEEE, 2013. – С. 1-7.

22. Challenges in Representation Learning: Facial Expression Recognition Challenge: [Электронный ресурс]. URL: https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data. (Дата обращения: 12.12.2021).

23. Mollahosseini A., Hasani B., Mahoor M. H. Affectnet: A database for facial expression, valence, and arousal computing in the wild //IEEE Transactions on Affective Computing. – 2017. – Т. 10. – №. 1. – С. 18-31.

24. Boulkenafet Z., Komulainen J., Hadid A. Face antispoofing using speeded-up robust features and fisher vector encoding //IEEE Signal Processing Letters. – 2016. – Т. 24. – №. 2. – С. 141-145.

25. Liu Y., Jourabloo A., Liu X. Learning deep models for face anti-spoofing: Binary or auxiliary supervision //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 389-398.

26. Shao R. et al. Multi-adversarial discriminative deep domain generalization for face presentation attack detection //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2019. – С. 10023-10031.

27. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, “Generative adversatial nets,” in Proc. NIPS, 2014

28. Gauthier J. Conditional generative adversarial nets for convolutional face generation //Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester. – 2014. – Т. 2014. – №. 5. – С. 2.

29. Tran L., Yin X., Liu X. Disentangled representation learning gan for pose-invariant face recognition //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – С. 1415-1424.

30. Zou H., Ak K. E., Kassim A. A. Edge-gan: Edge conditioned multi-view face image generation //2020 IEEE International Conference on Image Processing (ICIP). – IEEE, 2020. – С. 2401-2405.

31. R. Huang, S. Zhang, T. Li, and R. He. Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. arXiv:1704.04086, 2017.

32. G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. arXiv:1702.01983, 2017.

33. Wan L. et al. Fine-grained multi-attribute adversarial learning for face generation of age, gender and ethnicity //2018 International Conference on Biometrics (ICB). – IEEE, 2018. – С. 98-103.

34. Karras T., Laine S., Aila T. A style-based generator architecture for generative adversarial networks //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2019. – С. 4401-4410.

35. Anwar A., Raychowdhury A. Masked Face Recognition for Secure Authentication //arXiv preprint arXiv:2008.11104. – 2020.

36. https://github.com/aqeelanwar/MaskTheFace

37. Bulat A., Tzimiropoulos G. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks) //Proceedings of the IEEE International Conference on Computer Vision. – 2017. – С. 1021-1030.

38. Detect facial landmarks from Python: [Электронный ресурс]. URL: https://github.com/1adrianb/face-alignment. (Дата обращения: 12.12.2021).

39. Blender: [Электронный ресурс]. URL: https://www.blender.org. (Дата обращения: 12.12.2021).