FACE RECOGNITION TECHNIQUES

Olexandr N. Romanyuk[1], Sergey I. Vyatkin[2], Sergii V. Pavlov[1], Pavlo I. Mykhaylov[3], Roman Y. Chekhmestruk[4], Ivan V. Perun[4]

[1]Vinnytsia National Technical University, Vinnitsa, Ukraine, [2]Institute of Automation and Electrometry SB RAS, Novosibirsk, Russia, [3]3D GNERATION GmbH, Dortmund, Germany, [4]3D GENERATION UA, Vinnitsa, Ukraine

Abstract. The problem of face recognition is discussed. The main methods of recognition are considered. The calibrated stereo pair for the face and calculating the depth map by the correlation algorithm are used. As a result, a 3D mask of the face is obtained. Using three anthropomorphic points, then constructed a coordinate system that ensures a possibility of superposition of the tested mask.

Keywords: methods of recognition stereo pair, depth map, correlation algorithm, perturbation functions, operation of subtraction

TECHNIKI ROZPOZNAWANIA TWARZY

Streszczenie. Omawiany jest problem rozpoznawania twarzy. Rozważane są główne metody rozpoznawania. Użyta zostaje skalibrowana para stereo dla twarzy oraz obliczanie mapy głębokości poprzez algorytm korelacji. W wyniku takiego, uzyskiwana jest maska twarzy w wymiarze 3D. Użycie trzech antropomorficznych punktów, a następnie skonstruowany systemu współrzędnych zapewnia możliwość nakładania się przetestowanej maski.

Słowa kluczowe: metody rozpoznawania pary stereo, mapa głębi, algorytm korelacji, funkcje perturbacji, działanie odejmowania

Introduction

Work on the identification of persons has been going on for a long time. Previously, this problem was based on two-dimensional (2D) images: their acquisition and comparison with the existing set. However, in most cases, small changes in the position of the object of observation made the system ineffective. The test samples could not be matched to any of the databases and the error rate was too high.

Known methods of identification of individuals. In the video, the face of a man is captured by a video camera and a special filter processes the image. Next, the task of selecting special points (FFE – Face Feature Extraction) on the image is automatically solved, after which these points (and the distances between them) form a standard by which the comparison is made.

The advantages of the method include the ability to carry out continuous identification covertly. The disadvantages of the method – dependence on the rotation of the head, and external characteristics of the face.

An alternative to video is a thermogram – the heat emitted by the human body, which can be registered by an infrared camera and this image can be processed. The method is convenient because a person can be registered in complete darkness, which increases the secrecy of registration. However, at the same time there is a dependence on external sources of thermal noise. The disadvantage can also be attributed to the use of special equipment.

Three-dimensional recognition (3D) is one of the most advanced methods. The head is highlighted with a special slide consisting of lines. By distorting the lines on the surface of the head three-dimensional model of the head is restored. In this model, special points are allocated, which form a feature vector. Advantages of the method - continuous identification of the object and the possibility of hidden identification. The system can also operate in an invisible range. Three-dimensional recognition lacks a number of drawbacks. It is almost impossible to fake a fake. Twins are different. Small dependence on the rotation of the head (the deviation range is significantly increased). With the right choice of light range small dependence on ambient light, from the hair. Reduced dependence on facial tumors (because there are anthropomorphic points on the face, almost not susceptible to swelling). The quality of recognition competes with most of other methods. 3D identification can be used in a dark environment; it remains effective even when the face is rotated, up to 90 degrees (up to the "profile" position). Disadvantages of the method – special equipment, high computational power requirements (often requiring hardware implementation of algorithms, which, accordingly, increases the cost of the system). The recognition system performs the identification process by performing a number of actions.

1. General face image processing in recognition

It is possible to identify a common pattern of the face recognition process. The first step is the detection and localization of faces on the image. At the recognition stage, the alignment of the face image (geometric and brightness), the calculation of features and directly recognition is performed-comparison of the calculated features with the standards incorporated in the database. The main difference of all presented algorithms will be the calculation of features and comparison of their sets among themselves.

1.1. Elastic graph matching

The essence of the method is reduced to elastic comparison of graphs describing images of faces [24]. Faces are represented as graphs with weighted vertices and edges. At the recognition stage, one of the graphs – the reference graph – remains unchanged, while the other is deformed to best fit the first one. In such recognition systems, graphs can be either a rectangular lattice or a structure formed by characteristic (anthropometric) points of the face.

Feature values are calculated at the vertices of the graph, most often using complex values of the Gabor filters or their ordered sets-Gabor wavelets, which are calculated in some local area of the graph vertex locally by convolution of pixel brightness values with the Gabor filters.

Graph structure for face recognition: a) regular lattice b) graph based on anthropometric points of the face (Fig. 1).

The edges of the graph are weighted by the distances between adjacent vertices. The difference (distance, discrimination characteristic) between two graphs is calculated using some deformation price function, which takes into account both the difference between the feature values calculated at the vertices and the degree of deformation of the edges of the graph.
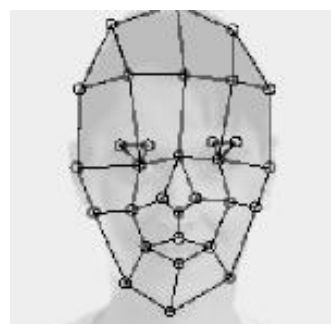

Fig. 1. The graph based on anthropometric points of the face

Deformation of a graph occurs by shifting each of its vertices by some distance in certain directions relative to its original location and choosing such a position that the difference between the values of the features (responses of the Gabor filters) at the vertex of the deformable graph and the corresponding vertex of the reference graph is minimal. This operation is performed alternately for all vertices of the graph until the smallest total difference between the features of the deformable and the reference graphs is reached. The value of the deformation price function at this position of the deformable graph will be a measure of the difference between the input face image and the reference graph. This "relaxation" deformation procedure must be performed for all reference persons included in the system database. The result of the system recognition is the standard with the best value of the deformation price function.

Some publications indicate 95–97% recognition efficiency even in the presence of different emotional expressions and changing the angle of the face to 15 degrees. However, the developers of elastic comparison systems on graphs refer to the high computational cost of this approach. For example, for comparison of the input image of the person with 87 reference approximately 25 seconds were spent at work on the parallel computer with 23 transputers [8] (note: the publication is dated 1993). Other publications on the subject either do not indicate the time or say that it is long.

Disadvantages: high computational complexity of the recognition procedure. Low-tech when memorizing the new standards. Linear time dependence on the size of the database.

1.2. Neural-networks

Currently, there are about a dozen varieties of neural networks (NN). One of the most widely used options is a network built on a multilayer perceptron, which allows you to classify the input image/signal in accordance with the pre-setup/training of the network.

NN are trained on a set of training examples. The essence of the training is to adjust the weights of interneuron connections in the process of solving the optimization problem by gradient descent. In the process of NN training, there is an automatic extraction of key features, determination of their importance and building relationships between them. It is assumed that the trained NN will be able to apply the experience gained in the learning process to unknown images due to generalizing abilities.

Convolutional Neural Network (hereinafter – CNN) showed the best results in the field of facial recognition (based on the results of the analysis of publications) [9], which is a logical development of the ideas of such NN architectures as cognitron and neocognitron. The success is due to the possibility of taking into account the two-dimensional topology of the image, in contrast to the multilayer perceptron.

Distinctive features of CNN are the local receptor fields (provide local two-dimensional connectivity of neurons), common weights (provide detection of some traits anywhere in the image) and hierarchical organization with spatial sampling (spatial subsampling). With these innovations, the CNN provides partial resistance to changes in the scale, offsets, rotations, changes in the angle and other distortions.

CNN testing on the ORL database, which contains images of faces with small changes in lighting, scale, spatial rotations, position and various emotions, showed 96% recognition accuracy.

CNN received its development in the development of Deep Face [18] (Fig. 2), which was acquired by Facebook for facial recognition of the users of its social network. All features of the architecture are closed.

Disadvantages ofNN: adding a new reference person to the database requires a complete retraining of the network on all available set (quite a long procedure, depending on the sample size from 1 hour to several days). Problems of mathematical nature related to training: getting into a local optimum, choosing the optimal optimization step, retraining, etc. Difficult to normalize the stage of choosing the network architecture (number of neurons, layers, nature of connections). Summarizing all the above, we can conclude that the NN is a "black box" with hard-to-interpret results.
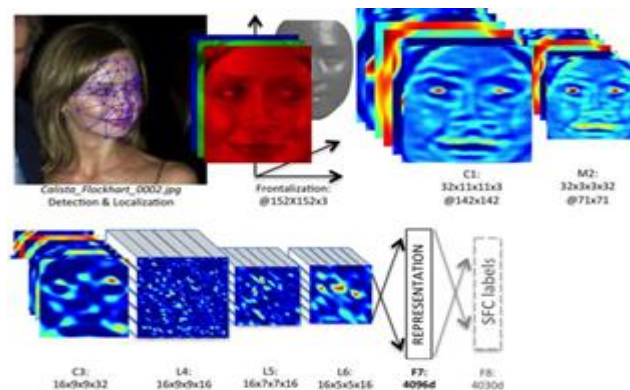

Fig. 2. Principle of operation of Deep Face

1.3. Hidden Markov's models

One of the statistical methods of face recognition is the hidden Markov models (HMM) with discrete time [12]. HMM uses statistical properties of signals and takes into account directly their spatial characteristics. The elements of the model are the set of hidden states, the set of observed states, the matrix of transition probabilities, the initial probability of states. Each has its own HMM. When recognizing an object, HMM generated for a given object database are checked and the maximum observed probability that the corresponding model generates the sequence of observations for a given object is searched.

To date, no example of commercial use of HMM for facial recognition has been found.

Disadvantages:
- It is necessary to select model parameters for each database.
- The HMM has no discriminating ability, i.e. the learning algorithm only maximizes the response of each image to its model, but does not minimize the response to other models.

1.4. Principal component analysis

One of the most well known and developed is the principal component analysis (PCA) method based on the transformation [7].

Initially, the principal component analysis was used in statistics to reduce feature space without significant loss of information. In the problem of face recognition it is used mainly to represent the image of the face by a vector of small dimension (principal components), which is then compared with the reference vectors embedded in the database.

Main purpose of the method of PCA is a significant reduction of dimensionality of the feature space so as better described the "typical" images, owned by many persons. Using this method it is possible to identify the different variability in the training set of facial images and describe this variation in a few orthogonal basis vectors, called own (eigenface).

The set of eigenvectors obtained once in the training sample of face images is used to encode all other face images, which are represented by a weighted combination of these eigenvectors. Using a limited number of eigenvectors, a compressed approximation of the input face image can be obtained, which then can be stored in the database as a coefficient vector that serves as a search key in the face database at the same time.

The essence of the PCA is as follows. Initially, the entire training set of faces is converted into one common data matrix, where each row is a single instance of the face image decomposed into a row. All faces of the training set should be reduced to the same size and with normalized histograms.

Then the data is normalized and the rows are reduced to the 0-th mean and 1-th variance, the covariance matrix is calculated. The problem of determining eigenvalues and corresponding eigenvectors (eigenvalues) is solved for the obtained covariance matrix. Then the eigenvectors are sorted in descending order of eigenvalues and only the first k vectors are left according to the rule:

$$\frac{\sum_{i=1}^{k} l_i}{\sum_{i=1}^{n} l_i} \rangle Threshold \ (0.9 \ or \ 0.95)$$

where l_i are ordered eigenvalues.

1) Zero mean:

$$x_{ij} = x_{ij} - \rho_i \quad \rho_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$$

2) Unit variance:

$$x_{ij} = \frac{x_{ij}}{\xi_j} \quad \xi_i = \frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \rho_i)^2$$

3) Covariance matrix: $\Sigma = XX^T$

4) Compute eigenvectors of $\Sigma : W^T$

5) Project X onto the k principal components (Fig. 3).

Fig. 3. An example of building (synthesis) of a human face using a combination of Eigen-faces and principal components

The PCA is well established in practical applications. However, in cases where the image of the face presents significant changes in light or facial expression, the effectiveness of the method decreases significantly. The point is that the PCA chooses a subspace with the goal of approximating the input dataset as closely as possible, rather than discriminating between classes of individuals. In [2] it was proposed to solve this problem with the use of linear discriminant Fisher (in the literature there is a name ″Eigen-Fisher″, ″Fisherface″, LDA). LDA looks for a data projection in which classes are as linearly separable as possible. For comparison, the PCA looks for a projection of the data that maximizes the spread across the person database (excluding classes). According to the results of experiments [2] in conditions of strong tank and lower shading of face images, Fisherface showed 95% efficiency compared to 53% of Eigenface.

1.5. Active appearance models

Active Appearance Models (AAM) are statistical image models that can be adjusted to the real image by means of various deformations. Tim Cootes and Chris Taylor proposed this type of two-dimensional model in 1998 [5]. Initially, AAM were used to evaluate face image parameters.

The AAM contains two types of parameters: shape-related parameters (shape parameters) and image pixel statistical model or texture-related parameters (appearance parameters). Before use, the model must be trained on a set of pre-labelled images. Marking of images is done manually. Each label has its own number and defines a characteristic point that the model will have to find during adaptation to the new image.

The layout of the face image of 68 points forms the shape of AAM. The AAM training procedure begins by normalizing the shapes on the labelled images to compensate for differences in scale, slope, and offset. The so-called generalized analysis is used for this purpose. From the entire set of normalized points, the principal components are then distinguished using the PCA method. The AAM shape model consists of a triangulation lattice s_0 and a linear combination of displacements s_i with respect to s_0 (Fig. 4).

Next, a matrix is formed from the pixels inside the triangles formed by the points of the shape, in such a way that each column contains the pixel values of the corresponding texture. It is worth noting that the textures used for training can be both single-channel (grayscale) and multi-channel (for example, RGB color space or other). In the case of multichannel textures, pixel vectors are formed separately for each of the bands, and then their concatenation is performed. After finding the principal components of the texture matrix, the AAM model is considered trained. The AAM consists of a base view A0 defined by the pixels inside the base lattice s0 and a linear combination of offsets Ai relative to A0.
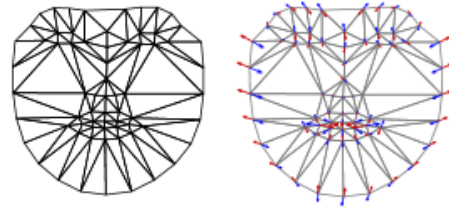

Fig. 4. AAM shape and displacement model

Fitting the model to a specific image of the face is performed in the process of solving the optimization problem, the essence of which is to minimize the functionality gradient descent method. The parameters of the model found in this case will reflect the position of the model in a particular image.

With AAM, you can model images of objects that are subject to both rigid and non-rigid deformation. AAM consists of a set of parameters, some of which represent the shape of the face; the rest set its texture. Deformations are generally understood as geometric transformations in the form of transport, rotation, and scaling compositions. When solving the problem of face localization in the image, the search for parameters (location, shape, texture) of AAM, which represent the synthesized image closest to the observed one, is performed. According to the degree of proximity AAM customized image is decided-there is a person or not.

2. Active Shape Models (ASM)

The essence of the ASM method [14] is to take into account the statistical relationships between the locations of anthropometric points on the available sample of full-face images. In the image, the expert marks the location of anthropometric points. In each image, the dots are numbered in the same order.

In order to bring the coordinates on all images to a single system, the so-called generalized Procrustean analysis is usually performed, because of which all points are brought to the same scale and centered. Then the average shape and covariance matrix are calculated for the whole set of images. Based on the covariance matrix, eigenvectors are computed and then sorted in descending order of their corresponding eigenvalues. The ASM model is defined by the matrix F and the mean vector s.

Localization of the ASM model on a new image that is not included in the training sample is carried out in the process of solving the optimization problem.

However, the main purpose of AAM and ASM is not a facial recognition, but the precise localization of the face and anthropometric points in the image for further processing.

In almost all algorithms, the mandatory step that precedes classification is alignment, which means alignment of the face image to the frontal position relative to the camera or bringing a set of faces (for example, in the training sample for training the classifier) to a single coordinate system. To implement this stage, it is necessary to localize anthropometric points typical for all faces on the image – most often; these are the centers of the pupils or the corners of the eyes. Different researchers distinguish different groups of such points. In order to reduce computational costs for real-time systems, developers allocate no more than 10 such points.

The AAM and ASM models are just intended to accurately localize these anthropometric points on the face image.

The main problems associated with the development of facial recognition systems

1) Illumination Problem,
2) Head Position Problem (face is a 3D object).

In order to evaluate the effectiveness of the proposed facial recognition algorithms, DARPA and the U.S. army research laboratory developed the FERET (face recognition technology) program.

The large-scale tests of the FERET program involved algorithms based on flexible comparison on graphs and various modifications of the PCA. The efficiency of all algorithms was approximately the same. In this regard, it is difficult or even impossible to make clear distinctions between them (especially if the test data are agreed). For frontal images taken on the same day, the acceptable recognition accuracy is typically 95%. For images taken by different devices and in different lighting conditions, the accuracy usually drops to 80%. For images taken with a difference of a year, the recognition accuracy was approximately 50%. It is worth noting that even 50 percent-is more than acceptable accuracy of the system of this kind.

Every year FERET publishes a report on a comparative test of modern facial recognition systems based on more than one million faces. Unfortunately, the latest reports do not disclose the principles of construction of recognition systems, and only the results of commercial systems are published. Today the leading system is NeoFace developed by NEC.

2.1. Method recognition based on scalar perturbation functions

The method of face recognition with the use of perturbation functions and the set-theoretic operation of subtraction was presented in [21, 23].

A calibrated stereo pair is used for calculating 3D points on the face. Let us assume that we have two projective matrices M_i

$$\begin{pmatrix} u_i s_i \\ v_i s_i \\ s_i \end{pmatrix} = M_i \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

where x, y, z are three-dimensional coordinates of the point, u_i and v_i are their projections in the image i, and s_i is the scale factor. The stereo pair is characterized by the following parameters: the points of the image planes $E_1 = (u_1, v_1)$ and $E_2 = (u_2, v_2)$, and the point of the world coordinate system $P = (x, y, z)$.

Using the calibrated stereo pair for the face, we calculate the depth map by the correlation algorithm. In this work, we use an area-based algorithm with correlation of image intensity levels.

$$s = \frac{\sum_{i,j} \left((I1(x+i, y+j) - \bar{I}1) - (I2(x+dx+i, y+dy+j) - \bar{I}2) \right)^2}{\sqrt{\left(\sum_{i,j} (I1(x+i, y+j) - \bar{I}1)^2 \right)\left(\sum_{i,j} (I2(x+dx+i, y+dy+j) - \bar{I}2)^2 \right)}}$$

Here $I1$ and $I2$ are the intensities of the left and right images, $\vec{I}1$ and $\vec{I}2$ are their mean values, dx and dy are the displacements along the epipolar line, and $s = \max(0.1 - c)$ is the correlation estimate.

There are two images of the stereo pair; scanning of these images provides information about the depth buffer (depth map).

In finding the perturbation peak, we calculate the characteristic size of the projection of the current interval, which is used as a basis for determining the detail level. For a larger interval, a rough approximation of the original function is taken. If a more detailed presentation is needed, then bilinear or bicubic interpolation of heights at the last detail level is performed. As a result, we obtain a 3D mask of the face (Fig. 5). Using three anthropomorphic masks, we construct a coordinate system that ensures a possibility of superposition of the tested masks; finally,

a clipping plane for equalization of the volumes cuts off certain parts. Applying the set-theoretic operation of subtraction

$$F3 = F1 \backslash F2$$

We determine the set of 3D points (voxels) belonging to the object f3 = **F**i (f1(x, y, z), f2(x, y, z)), **F**1: f1(x, y, z) ≥ 0, **F**2: f2(x, y, z) ≥ 0. To find 3D points, voxelization of the remaining part of the volume after the subtraction is needed.

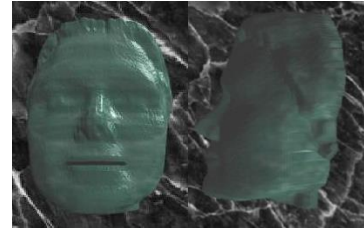The smaller the number of voxels left, the greater the similarity of the tested objects.

Fig. 5. 3D masks

2.2. Scanning technology

Computing three-dimensional (3D) data and information range is an important task in a variety of applications including computer graphics, medicine, multimedia, machine vision, navigation, automotive safety, computer interaction, tracking, image recognition, and more. Obtaining three-dimensional geometric data from a real and complex environment has been the subject of research for many years. Computer graphics and image analysis are the two methods of processing visual information. Computer graphics operate with formal descriptions of objects to create their visual images. Image analysis systems work with images to produce formalized models of objects. Recently, there has been a trend towards convergence and mutual integration of computer graphics and image analysis. This is primarily due to the development of virtual reality systems.

Scanning technologies are divided into contact and contactless. The first implies the presence of a mechanical device by means of which the coordinates of the selected points are transmitted to the computer.

Contactless three-dimensional scanners are much more complex devices, which have complex algorithms for creating objects. Some devices combine laser sensors and a digital camera, which is used for greater scanning accuracy.

New technologies are researched and developed with improved accuracy and reduced cost. In addition, systems, as well as methods suit for object perception, data acquisition and depth in three-dimensional space. These systems and technologies use various resources, including electromagnetic waves, sound waves, ultrasound, laser, light rays, etc. They have been used in various systems, including lasers, radars, sonars, ultrasonoscopes, imaging systems, etc. Among these systems the laser detection and rangefinder (LADAR) can be distinguished – a system of transmission, detection, processing and reception of electromagnetic waves reflected from targets. The laser sensors of these systems provide range information constructed from a set of point measurements from a single point of view. The laser rangefinder (LRF), working on the principle of transmitting a laser pulse to the surface of an object and receiving a reflected pulse, measures the time between sending and receiving a pulse [1]. On a 3D laser scanner, the concept is similar to LADAR. However, the system was created for the purpose of 3D digital measurement, visualization and documentation, it is able to capture 3D digital information and obtain images with high accuracy and speed [6]. Another typical system is the light detection and ranging system (LIDAR), a system for mapping terrain, measuring forests and vegetation. Lidar has the ability to detect different reflections from only one laser pulse [15]. Radio frequency identification (RFID) is an automatic tracking and identification technology that uses small tags (transponders) that are attached to a physical object where the

tags contain store information. Radar detection and range (Radar) is a detection technology that uses radio waves transmitted and reflected back by objects to obtain the angle, speed, range, and properties of targets [26]. Thermal imagers are another technology for sensing objects based on the difference in temperature scales, these technologies are able to detect the heat given off by the desired targets. Global positioning systems (GPS and GLONASS), space systems that determine the exact location and provide timely information about the desired location anywhere on Earth and in all weather conditions.

There are also three-dimensional scanning systems based on ultrasonic installations [10].

Ultrasonic technology based on sound wave is an approach to measuring distances and detecting objects. They transmit high-frequency pulses by using a sensor probe. Then the reflected waves are received back by the same sensor probe [16]. Another type of systems is the use of magnetic scanners, which use to determine the spatial coordinates of the object change its spatial magnetic field. However, ultrasonic and magnetic scanners are very sensitive to various kinds of noise.

Optical scanners are divided into active and passive. Passive devices are devices based on two cameras, devices used for reconstruction of silhouettes of objects, etc. In [3] an overview of passive scanning methods is given. Active systems have a matched light source and image receiver [11]. There are a number of methods of scanning the object: the method of structured light, time-of-flight cameras and methods of accumulation of information about the object (the method of assessing the shape of the movement). The essence of the methods of scanning the object is reduced to its illumination, for example, a template in the form of a grid, a projector can be used, projecting a regular grid of light lines on the surface of objects. The camera, when forming images, transmits the result of distortion of the grid due to the shape and orientation of the surface. As follows from the above, the method of object scanning requires specialized equipment (projectors, 3D scanners, radars, time-of-flight cameras, sonars or lidars [15]), whereas the proposed method of binocular stereo vision works with conventional web cameras. Methods of accumulation of information, in turn, analyze the local movement of parts of the scene over time. When the camera or object is moved, or both, the system receives a sequence of changing images. Surfaces and angles can be reconstructed from optical flow vectors or corresponding points in three-dimensional scenes. Identifying objects from motion presents a task similar to that of binocular stereo vision; only the images to be processed will be obtained at different times. This leads to the complexity and greater resource intensity of the problem of finding the corresponding points, which makes the method of binocular stereo vision more advantageous.

Modeling the shape of real-world objects from a series of images has been investigated in the recent years. One well-known approach to three-dimensional modeling is to create a shape from a silhouette that restores the shape of objects along their contours. This approach is popular due to its fast calculations and reliability. The first work on the construction of three-dimensional models from several points of view was described in [4]. The method [13] uses orthogonal projection to construct three-dimensional models. Numerous studies have been devoted to the creation of shapes from silhouettes to transform visible contours into visual form [17, 19].

3. Conclusions

During the experiments it was found, that the effective size of the comparison window in the construction of the search space is determined by the resolution and size of the objects depicted on it.

Quality testing of several criteria was carried out as the sum of modules of differences, sum of squared differences, and the criterion of the census. The first was discarded immediately, because its quality was almost the same as the second, but it is impossible to calculate a linear algorithm.

It was tested with a different set of permissions, using the SSD criterion. It is established that: on average, it takes approximately equal time to fill the search space with its values and search for the optimal path, the operating time depends linearly on the difference between the maximum and minimum disparities, the time depends linearly on the image area. While maintaining the width-to-height ratio.

The results of testing the method are encouraging. Both virtual objects from available databases and real persons were used. The 3D technology of face recognition provides effective operation; more than 98% of test objects were successfully recognized by using this method. Nevertheless, some factors result in failure of verification. These factors can be classified into two groups: incorrect position ahead of the camera and interferences in data readout. The first class includes situations where only some part of the face is visible for the camera. The face is not directed toward the camera, the head is turned downward or to the left or right from the camera, the person is located too close to the camera, or the person goes away from the camera too fast after the beginning of verification (less than one second). The method operates successfully if the recognized object moves uniformly, but the camera fails to capture the observed object exactly in the case of its sudden acceleration.

It should be noted that observation of only some part of the face in the camera is not completely unacceptable because fragments can be successfully verified by using the geometric operation of intersection. The proposed method allows for selective testing with the use of the geometric operation of intersection of a transparent cylinder or any other geometric shape with the surface.

Interferences of data readout occur if the facial expression is not neutral as required or if the headwear, mirror shades, or hair covers a major part of the face [22, 25].

Advanced methods are capable of recognition based on different facial expressions. Three-dimensional morphing is used for recognition in the proposed method.

If we compare 2D systems and the proposed 3D method of recognition, we can see that the false response probability in the first case is 0.12% and the false rejection probability is 9.8% for the recognition threshold being set at 70%. In the second case, the recognition threshold was set at 90%, and the method provided the false response probability of 0.004% and the false rejection probability of 0.1%.

In all tests performed simultaneously for both technologies with the use of the same images, the 3D technology of face recognition turned out to be more efficient than the 2D technology.

An example of 3D recognition methods is the well-known method of fitting for reconstructing the shape and parameters of the texture. This method is based on a system of linear equations. Recognition is performed based on comparisons of the reconstructed shapes and texture of the image.

However, manual initialization is needed in the Face Identification by fitting a 3D Morphable Models method. The recognition time (approximately 1 minute on the Pentium III processor with a frequency of 800 MHz) does not satisfy the requirements of most real systems.

As compared to previously available methods, the proposed method offers the following advantages. 3D morphing allows recognition of faces with different facial expressions. Face identification based on some part of the image is possible. Texturing of the face surface is not needed; the method is completely automatic and fast (about 200 ms for one face image with a resolution of 640×480 pixels with the use of the Intel Core i7-2700K processor (8 MB cache memory, 3.90 GHz)), which is faster than the fitting method approximately by two orders of magnitude. The measurement error is no more than 0.8 mm (for each point of the 3D surface).

For real-time visualization, a binary method of searching for image elements with the use of graphics processing units adapted for calculating perturbation functions can be used. Therefore, a method of face recognition based on perturbation functions and

the set-theoretic operation of subtraction is proposed. Three-dimensional masks were used for face recognition. This method differs from available 3D methods by the fact that it involves not only all points of the surface in the recognition procedure, but also the volume of the tested mask. The method offers the following advantages: manual initialization of the process is not needed; three-dimensional morphing solves the problem of face recognition based on different facial expressions; face recognition based on only some part of the image is possible; face reconstruction is completely automated. The computation time is approximately 200 ms with a resolution of 640×480 pixels.

The method can be used in various situations where intellectual video monitoring of specially protected objects is needed: defense complex enterprises, heavily crowded areas, etc.

References

[1] Amann M. C., Bosch T. M., Lescure M., Myllylae R. A., Rioux M.:Laser ranging: a critical review of unusual techniques for distance measurement. Optical Engineering 40(1)/2001, 10–19, [http://doi.org/10.1117/1.1330700].

[2] Belhumeur P. N., Hespanha J. P., Kriegman D. J.: Eigen faces vs. Fisher faces: Recognition using Class Specific Linear Projection. IEEE Transactions on pattern analysis and machine intelligence 19(7)/1997, 711–720, [http://doi.org/10.1109/34.598228].

[3] Butime J., Gutierrez I., GaloCorzo L., Flores C.: Espronceda. 3D reconstruction methods, a survey. Proceedings of the First International Conference on Computer Vision Theory and Applications, 2006, 457–463, [http://doi.org/0.5220/0001369704570463].

[4] Chien C. H., Aggarwal J. K.: Identification of 3D Objects from Multiple Silhouettes Using Quadtrees / Octrees. Computer Vision Graphics And Image Processing 36(2–3)/1986, 256–273.

[5] Edwards G. J., Cootes T. F., Taylor C. J.: Face recognition using active appearance models. European Conference on Computer Vision, 1998, 581-595. [http://doi.org/10.1007/BFb0054766].

[6] Jecić S., Drvar N.: 3D Shape Measurement Influencing Factors. NDT – Competence & Safety, Zagreb 2004, 109–116.

[7] Jolliffe I. T.: Principal component analysis, second edition. Springer, New York 2002.

[8] Lades M., Vorbruggen J.C., Buhmann J., et al.: Distortion Invariant Object Recognition in the Dynamic Link Architecture. IEEE Transactions on Computers 42(3)/1993 300–311, [http://doi.org/10.1109/12.210173].

[9] Lawrence S., Giles C.L., et al.: Back. Face Recognition: A Convolutional Neural-Network Approach. IEEE Transactions on Neural Networks 8(1)/1997, 98–113, [http://doi.org/10.1109/72.554195].

[10] Lipton L.: Foundations of the Stereoscopic Cinema – A Study in Depth. Van Nostrand Reinhold, New York 1982.

[11] Martin W. N., Aggarwal J. K.: Volumetric Descriptions of Objects from Multiple Views. IEEE Transactions on Pattern Analysis and Machine Intelligence 5(2)/1983, 150–158.

[12] Nefian A. V.: A hidden Markov model-based approach for face detection and recognition. A Proposal for a Doctoral Dissertation. Georgia Institute of Technology 1998.

[13] Niem W.: Robust and Fast Modeling of 3D Natural Objects from Multiple Views. Proceedings Image and Video Processing II 2182/1994, 388–397.

[14] Prabhu U., Seshadri K.: Facial Recognition Using Active Shape Models, Local Patches and Support Vector Machines, 2009.

[15] Reutebuch S. E., Andersen H., Mcgaughey R. J., Forest L.: Light Detection and Ranging (LIDAR): An Emerging Tool for Multiple Resource Inventory. J. For. 103(6)/2005, 286–292.

[16] Siudak M., Rokita P.: A survey of passive 3D reconstruction methods on the basis of more than one image. Machine Graphics & Vision 23(3/4)/2014, 57–11.

[17] Szeliski R.: Shape from rotation. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91), 1991, 625–630.

[18] Taigman Y., Yang M., et al.: Deep Face: Closing the gap to human level performance in face verification. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 1701–1708.

[19] Vedula S., Rander P., Saito H., Kanade T.: Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences. Proc. 4th Conference on Virtual Systems and Multimedia (VSMM98), 1998, 326–332.

[20] Vyatkin S. I., Romanyuk A. N., Gotra Z. Y., et al.: Offsetting, relations and blending with perturbation functions. Proc. of SPIE 10445/2017, 104452B.

[21] Vyatkin S. I., Romanyuk S. A., Pavlov S. V., Necheporyk M. L.: Face Identification Algorithms and its using. Modern Engineering and Innovative Technologies 5/2018, 111–115.

[22] Vyatkin S. I.: Complex surface modeling using perturbation functions. Optoelectronics, instrumentation and data processing 43/2007, 226–231.

[23] Vyatkin S. I.: Method of face recognition using of scalar perturbation functions and set-theoretic operation of subtraction. Optoelectronics, instrumentation and data processing 52(1)/2016, 1–7.

[24] Wiskott L., Fellous J. M., Kruger N., et al.: Face Recognition by Elastic Bunch Graph Matching. Proc. of International Conference on Image Processing 1/1997, 129–132, [http://doi.org/10.1109/ICIP.1997.647401].

[25] Wójcik W., Pavlov S., Kalimoldayev M.: Information Technology in Medical Diagnostics II. Taylor & Francis Group, London 2019.

[26] Rawal R., Yadav V., Sharma S.: Radar – a brief study. International Journal of Innovative Research and Technology 1(12)/2015, 1017–1020.

Prof. Olexandr N. Romanyuk
E-mail: rom8591@gmail.com

Doctor of Technical Sciences, Professor, Department of Software. Work of Vinnytsia National Technical University, head of Software Department
Research field: formation and processing of graphic images

http://orcid.org/0000-0002-2245-3364
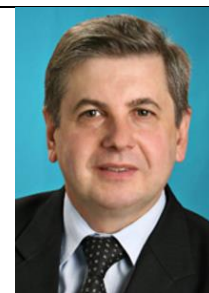
Ph.D. Sergey I. Vyatkin
E-mail: sivser@mail.ru

Candidate of Technical Sciences, senior scientific researcher. Research field: formation and processing of graphic images

http://orcid.org/0000-0002-1591-3588

Prof. Sergii Pavlov
e-mail: psv@vntu.edu.ua

Doctor of Technical Sciences, Professor, Academician of the International Academy of Applied Radioelectronic.Vice-rector of for Scientific Work of Vinnytsia National Technical University, professor of biomedical Engineering, Scientific direction – biomedical information optoelectronic and laser technologies for diagnostics and physiotherapy influence. Deals with issues of improving the distribution of optical radiation theory in biological objects, particularly through the use of electro-optical systems, and the development of intelligent biomedical optoelectronic diagnostic systems and standardized methods for reliably determining the main hemodynamic cardiovascular system of comprehensive into account scattering effects.

http://orcid.org/0000-0002-0051-5560

M.Sc. Pavlo I. Mykhaylov
E-mail:pm@3dgeneration.com

General director CEO 3D GNERATION GmbH (Germany). Research field: formation and processing of graphic images

http://orcid.org/0000-0001-5861-5970

Ph.D. Roman Y. Chekhmestruk
E-mail: Rc.ua@3dgeneration.com

Candidate of Technical Sciences. Technical Director 3D GENERATION UA, researchfield: formation and processing of graphic images.

http://orcid.org/0000-0002-5362-8796

M.Sc. Ivan V. Perun
E-mail: ip.ua@3dgeneration.com

Project manager 3D GENERATION UA, researchfield: formation and processing of graphic images.

http://orcid.org/0000-0001-7402-4417