Identification of persons using stereo image pairs and perturbation functions

Sergey Vyatkin, Alexander Romanyuk, Sergey Pavlov and Oksana Romanyuk

October 16, 2019

Identification of persons using stereo image pairs and perturbation functions

Sergey Vyatkin[1], Alexander Romanyuk[2], Sergey Pavlov[2], Oksana Romanyuk[2]

[1] Institute of Automation and Electrometric of the Siberian Branch of the RAS, Russian Federation

[2] Vinnytsia National Technical University, Ukraine

Abstract. 3D recognition is one of the most progressive methods. The problem of face recognition is considered. In this model, special points, which form a feature vector, are identified. The method offers the following advantages: continuous and secret identification of the object; it is impossible to use a fake object; twins can be distinguished; weak dependence on head turning (the range of head deflection is substantially increased); weak dependence on external illumination, hair, and face turgidity in the case of a correct choice of the light range. Three-dimensional identification can be used in darkness, and it remains effective even in the case with head turning up to 90◦. For this purpose, a method based on scalar perturbation functions and set-theoretic operation of subtraction is proposed. It is shown that all surface points and the mask volume are used in the process of sample testing for more accurate identification. Using the calibrated stereo pair for the face, the depth map is calculated by the correlation algorithm. As a result, a 3D mask of the face is obtained. Using three antropomorphic masks, a coordinate system that ensures a possibility of superposition of the tested masks is constructed; finally, certain parts are cut off by a clipping plane for equalization of the volumes. Applying the set-theoretic operation of subtraction the set of 3D points (voxels) belonging to the object is determined. 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.

Keywords: reconstruction, stereo pair, depth map, height map, scalar perturbation functions, correlation algorithm, operation of subtraction, voxelization.

1. Introduction

For a considerable period, image analysis and computer graphics have evolved independently without finding common ground. This is not surprising: in fact, computer graphics and image analysis are two symmetrical methods of processing visual information. Computer graphics operates with formal descriptions of objects to create their visual images. On the other hand, image analysis systems work with images to produce formalized object models in one form or another. 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, such as computer games, various training systems, simulators for pilots, drivers and astronauts.

In theory, virtual reality is an absolute interface between a person and a computer, which uses all or almost all systems of human interaction with the outside world. In this sense, the user of the virtual reality system experiences the sensation of interaction with some model environment that has the features of the real world, without contact with it. The model environment and its perception are usually fully modeled on a computer. One of the most important problems in the development of a virtual reality system is the construction of a model environment. The world in which the interaction between the system and the user takes place. The modeling environment may consist of objects created entirely in the computer with the hands of the designer, and of the objects "digitized" with real. In other words, descriptions of objects in the form of their geometric models can be obtained by various methods of analysis of real objects. One of the most common methods of obtaining such a description is based on the analysis of image pairs.

The field of study of this work is a set of solutions to the problem of determining the spatial parameters of the object by the stereo pair of its images. Obtaining three-dimensional models of real objects by analyzing pairs of their images is necessary in order to add virtual objects to them for subsequent use in virtual reality systems.

The aim of the work is to develop a method and implement a program based on it to obtain a real-time three-dimensional model of the visible part of the surface of the scene observed by two video cameras.

2. Restoration of three-dimensional coordinates of a scene on stereo pairs of images

Under the reconstruction of the geometry is understood as the restoration of the height map of the entire scene. Height map - a two-dimensional array (x, y), which in each cell is the value of the coordinate Z. so, our goal - to get a map of the heights of the scene by binocular vision.

Consider a situation where two cameras at different points register the same scene. A pair of images obtained in this case is called a stereo pair. Let us first turn to the simplest case. Let the same cameras are arranged so that their optical axes are parallel, and the line passing through the optical centers is perpendicular to the optical axes (this line is called the baseline, and its segment enclosed between the optical centers - the base). Put the length of the base equal to B. Choose a global coordinate system whose origin O is located on the baseline in the middle between the optical

centers of the cameras, the OZ axis is parallel to the optical axes, and the OX axis is directed along the baseline.
Let the coordinates in the camera image planes coincide with the main points $U_0 = V_0 = 0$, and the coordinate units in the global system and in the camera image planes are the same ($w = h = 1$).
Select a point with global coordinates (X, Y, Z). The coordinates of its projection in the image plane of the first (left) camera are denoted by (x', y'), and in the image plane of the second (right) camera – by (x'', y'') . (Projections of the same point M in the planes of images of different cameras are called conjugate points.) It is not difficult to verify that

$$x' = f(X + b/2)/Z \,,\, x'' = f(X - b/2)/Z \,,\, y' = y'' = fY/Z$$

Note that in the direction perpendicular to the direction of the baseline, the coordinates of the conjugate points (y -coordinates) are the same. This fact is of great importance in the automated search for conjugate points on the stereo pair, allowing to significantly reducing the size of the search area. From the first two relations it follows that

$$Z = fb/(x' - x'')$$

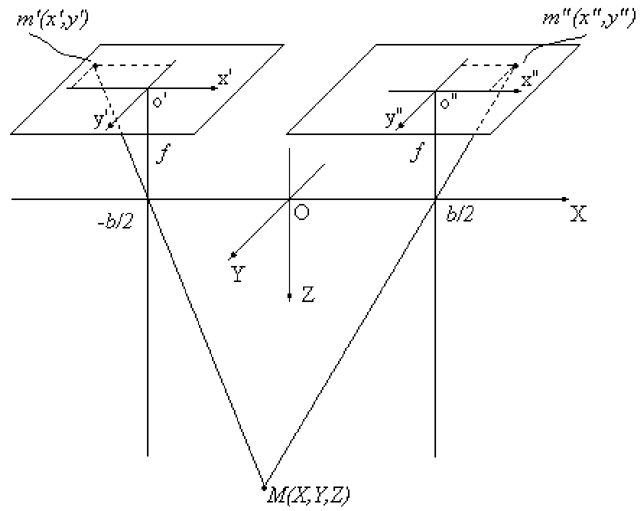

Fig. 1. Stereo-pair

This means that, knowing the geometry of the shooting and measuring the coordinates of the projections of the same point in the planes of the camera image, you can calculate the depth (coordinate Z) of this point. Moreover, thus generated correlations enable calculating all the three-dimensional coordinates of the point:

$$X = b\frac{(x' + x'')}{2(x' - x'')}, Y = b\frac{(y' + y'')}{2(x' - x'')}$$

The difference $d = x' - x''$ is called disparity. From the formulas, it becomes obvious that errors in the coordinates of the projections are more affected by low variance and, consequently, the distances to distant objects are measured less accurately than to close ones. On the other hand, at a fixed range, the variance is proportional to the size of the base; hence, the accuracy of measurements increases with increasing base. Further, we will see, however, that the increase in the base can lead to errors that are not compensated by the increase in the accuracy of measurements.

From all the above it follows that to estimate the three-dimensional coordinates of a point on the stereo pair:

1) it is necessary to know the internal parameters of the cameras (the task of their calibration),
2) it is necessary to know the parameters of the relative position of the cameras (the problem of mutual orientation),
3) it is necessary to find and determine the coordinates of the corresponding conjugate points on the images (the task of finding conjugate points).

3. Conditions and assumptions

From the theory described above it is clear that the main difficulty of the development described in this Chapter of the algorithm is the correct choice of the mechanism for finding the corresponding points. Therefore, additional conditions and assumptions are introduced that limit the search space and thus reduce the likelihood of false matches.

Epipolar restriction. As mentioned, the matching space is narrowed to one-dimensional (along the epipolar), which significantly increases the efficiency of matching and reduces the probability of error. In the case of a standard binocular configuration epipoles, correspond to the scan lines of the image, which further facilitates the work of search algorithms.

The smoothness condition States that scenes can only contain smooth and opaque objects. Thus, for the depth (and, accordingly, the disparity) along the surfaces of objects, the condition of continuity is fulfilled. Namely, if the features of p1 and ql on the target and offset image are corresponding. Then any feature of p2 from the neighborhood of p1 on the target image can correspond only to such a feature of q2 on the offset image that the difference between the norms of the disparity pairs (p1, ql) and (p2, q2) does not exceed the modulo in advance of the given small value. It is obvious that the smoothness condition is not satisfied near the object boundaries. Therefore, many algorithms determine the boundaries of objects (and, accordingly, the area of the barrier) precisely by a sharp change in the disparity.

The condition of limitation on the value of the disparity states that for any pair of corresponding features on the target and offset image, the disparity does not exceed the norm of the specified value. Thus, only objects that lie at a certain distance from the camera are considered.

Condition of orderliness states that the order of the corresponding features on epipolar does not change. Although this restriction does not apply to narrow objects that are close to the camera, its awareness has allowed creating a number of effective algorithms.

The uniqueness condition States that any feature in the key image can only match one feature in the offset image. Thus, transparent objects and areas lying along the direction of view of the key camera are not considered.

The photometric compatibility condition states that for any pair of corresponding singularities (pl,ql) and (p2,q2) on the target and offset image of the intensity difference, the absolute intensity difference does not exceed the modulo pre-set value. Thus, it is assumed that the corresponding points on the target and offset image have approximately the same intensity.

4. System calibration

The task of calibration of stereo cameras includes the calculation of parameters such as focal lengths, the relative position of the cameras in space, the parameters of optical curvature, which have all cameras, the parameters of brightness distortion, etc. Calibration methods are divided into photogrammetric [1, 3],[4-7] and self-calibration [2, 8, 9,10,12]:

Photogrammetric calibration. It is based on the observation of an object whose geometry is known with very good accuracy. A calibration object usually consists of two or three planes orthogonal to each other. Information about its geometry and the elevation maps taken by the cameras are used to determine the system parameters.

Self-calibration. This technology does not use any calibration objects. Camera movement is used in a rigid static scene, which makes it possible to determine the internal parameters of the cameras. If the internal parameters of the cameras are fixed, it is possible to determine all the parameters of the scene when moving. This approach is very promising,

but it does not always give the right results, because a very large set of parameters needs to be defined.
To solve the problem calibration can be carried out by any photogrammetric method. The next Chapter will offer one of the easiest to apply to the task.

5. Match search task

Studies of the mechanisms of stereopsis and binocular optical systems have been carried out for many years, and different authors have developed a significant number of algorithms that allow solving the problem of finding corresponding features with one or another efficiency. In general, according to the type of processed features, all algorithms for constructing depth maps can be divided into two large groups:
The algorithms for geometric features operate on the geometric features, such as angles, straight lines, etc., highlighted in the image with the help of filters. The geometrical features with the most similar parameters are declared offset. The advantages of the algorithms for geometric features are low probability of false matches and high accuracy of finding the disparity (modern filters easily find geometric features at the subpixel level). The main drawback of this kind of algorithms can be considered the relative sparsity of the resulting depth maps.
The algorithms for regions operate on pixels and their intensities. A feature in the General sense is considered a pixel, the parameter features-intensity. Such algorithms were first developed for processing aerial photography data of undeveloped areas. Such data are characterized primarily by the smoothness of the surface to be restored and the lack of a sufficient number of geometric features. In addition, the condition of the "plane of the world" is fulfilled - in local areas the target surface can be considered flat. The algorithms for the regions allow us to obtain sufficiently dense depth maps (ideally, the matching problem is solved for each visible pixel on the target and offset image). The disadvantages of such algorithms arise from the orientation to the "plane of the world" condition (as a consequence, unstable work in the areas of depth gap) and the use of only pixel intensity as parameters (as a consequence, unstable work with different camera parameters and in cases of lighting changes).
It is clear that to solve the problem of finding a match for areas, you need to set the criterion of "similarity" of pixels, which allows you to select the only one corresponding to the fixed pixel on the target image among the set of candidate pixels on the offset image. It is obvious that simple criteria like the intensity difference module are not suitable. Therefore, it is common to do the following: the pixels to be tested on the target and offset

image are surrounded by small Windows of a fixed size and determine the "similarity" of the Windows based on the calculation of the correlation coefficient. Check window instead of the immediate check of the intensities allows avoiding errors associated with noise and local changes of illumination. On the other hand, in the case of choosing a "bad" correlation criterion, the use of Windows leads to unstable operation of the algorithm in the areas of sharp changes in depth (they do not meet the condition of the "plane of the world").

The most well-known algorithms for regions can be divided into three classes:

Local algorithms. In this case, the matching problem is solved independently for individual pairs of points (pixels) on the target and offset image. When solving the matching problem for each fixed pair of points, the results of the solutions for the other pairs are not taken into account. Each pair of corresponding points is searched independently and in parallel with other pairs.

Search algorithms in the epipolar space. In this case, the matching problem is solved for epipolar pairs on the key and offset image. Each pair of epipolar forms an epipolar space, and although the points within the space are compared independently based on the correlation criterion, the result of solving the matching problem for a given pair of points will be used to solve the problem for the next pairs of points.

The global algorithms compare the images. In this case, the target image and the offset image are compared as a whole, and the result of solving the matching problem is determined simultaneously for all points of the target image. It should be noted that although binocular visual systems predominate in nature, developers of algorithms for calculating depth maps usually use a larger number of cameras. Redundancy of information (three images instead of two), appearing in this case, allows to cope more effectively with the ambiguities characteristic of the stereopsis algorithms.

Criteria for the pixels:

Among the criteria for comparing pixels, two large groups can be distinguished. These are parametric and nonparametric criteria. Parametric explicit use of the values of the intensities of the pixels in the window. Use non-parametric topology data pixels.

6. Parametric tests

Among the parametric criteria, the most widely used is the sum of squares of SSD intensity differences. For the window of radius R, the sum of the squares of the

$$SSD(x_0, y_0, d) = \sum_{x=-R}^{R} \sum_{y=-R}^{R} (I_l(x + x_0, y + y_0) - I_r(x + x_0, y + y_0))^2$$

intensity differences is defined as

The SSD criterion uses the order of pixels in the window (which allows you to successfully discard symmetric areas) and integer calculations. It can be successfully implemented in linear form, so its complexity will be O(width*height*(disparity_map-disparity_map)). You can also use its modification, which is insured against brightness errors of video cameras, when the average brightness may differ from the cameras on constant or constant times:

$$SSD(x_0, y_0, d) = \frac{\sum_{x=-R}^{R}\sum_{y=-R}^{R}((I_l(x+x_0, y+y_0) - \bar{I}_l) - (I_r(x+x_0, y+y_0) - \bar{I}_r))^2}{\sqrt{\sum_{x=-R}^{R}\sum_{y=-R}^{R}((I_l(x+x_0, y+y_0) - \bar{I}_l) * \sum_{x=-R}^{R}\sum_{y=-R}^{R}(I_r(x+x_0, y+y_0) - \bar{I}_r))}}$$

7. Non-parametric criteria

Nonparametric criteria do not directly use intensity values, but try to somehow track the topology of the space that falls into the window. Among nonparametric criteria, the census criterion is the most common [13][14]:

$$SSD(x_0, y_0, d) = \sum_{x=-R}^{R}\sum_{y=-R}^{R}(I_l(x+x_0, y+y_0) > I_l(x_0, y_0)) \wedge (I_r(x+x_0, y+y_0) > I_r(x_0, y_0))$$

Here operation " > " takes the values 1 if the left operand is greater than the right one and 0 otherwise. The "^" operation takes the values 1 if the left operand is not right and 0 otherwise.

This criterion carries protection against glare and brightness errors of video cameras, when the average brightness may differ from the cameras on constant or constant times. However, it has big disadvantage – complexity O(R*R*width*height*(disparity_map-disparity_map)).

In addition, the authors [14] proposed an interesting criterion that can be used when the preliminary segmentation of images on the area is carried out. Each area has its own set of parameters that define its distortion. These parameters are found using the correlation coefficient. Two methods of cross-correlation and morphological correlation are considered. In both cases, the problem of finding parameter corrections is reduced to the problem of generalized eigenvalues. Unfortunately, this criterion is too time-consuming to use in solving the problem.

For calibration, it was decided to use the photogrammetric method, since the parameters of the cameras are assumed unchanged and it is enough to determine them once. To find matches, it was decided to use the search algorithm in the epistolary space as fast and efficient enough. During the comparative test of SSD and census criteria, SSD was selected.

The simplest binocular system is when the optical axes of the cameras are assumed parallel, the cameras are fixed, the observation planes coincide, and the internal parameters of the cameras coincide. Due to the simplicity of the calibration implementation, an approach similar to that described in [1] was chosen. Several shots of a flat object with points on the surface are taken. Points are plotted in a certain order and knowledge of this order is used to find matches on stereo images. After finding the matches, the knowledge of the distance between the points is used, a system of equations for the internal and external parameters of the scene is compiled and the least squares method is solved, minimizing errors in the measurement of these parameters.

It is also possible to simplify the method-the definition of not all parameters, but only the focal length and the size of the base. Alternatively, more precisely, you need to find the product of these two quantities to use them to obtain Z:

$$Z = fb/(x' - x'')$$

The authors [11] conducted extensive testing of matching algorithms on pairs of images and found that the fastest is the sum-of-squared-differences algorithm - the difference of the sum of squares. In addition, indeed he the most simple. This algorithm can be briefly described as:

1. Calculate the price of matching a pair of pixels as the square of the pixel intensity difference with the given disparity.
2. The summation of the prices on the window with a specified constant size.
3. The disparity is selected from at least all values for each pixel.

Unfortunately, this algorithm shows too little stability because many conditions are not checked in it. Therefore, the following algorithm with the use of dynamic programming was chosen.

Consider a standard binocular system and an epipolar plane uniquely defined by the optical centers of the cameras and an arbitrary point in space. A couple of epipolar formed by the intersection of the epipolar plane and the picture planes of the target and offset of the image determines a discrete space, called an epipolar search space. Consider a couple of these epipolar (el, er) as the coordinate axes. The elements of the discrete epipolar search space B(i,j) are the values of the correspondence criterion for the I-th point el(i) epipolar on the target image and the j-th point er (j) epipolar on the offset image. The smaller the value of the criterion, the more likely it is that the points el(i) and er(j) form a corresponding pair. To solve the matching problem, let's move to a more convenient form of representation of the search space. For a standard binocular system, the pairs of epipolar will correspond to the pairs of scan

lines of the target and offset images. Consider a couple of these epipolar (el(y) er(y)). We introduce a discrete search space as follows: on one axis, we lay the values of the coordinate x along the epipolar el(y) on the target image, and on the other - the values of the disparity. The elements of the space ESSP (x,d,y) will be the match criterion values for the point(x) of the El (y) epipolar on the target image and the point (x-d) of the epipolar on the offset image. The smaller the value of the criterion, the more likely it is that the points el(x, y) and er(x-d,y) form a corresponding pair.
Consider all possible paths from the lower left corner to the right edge of the search space ESSP(x,d). Obviously, each of these routes is a solution to the matching problem for all epipolar points on the target image. If there are restrictions on the "form" of the path, using the method of dynamic programming, we can quickly choose from a variety of routes optimal in terms of cost. Thus, by correctly setting the cost and constraint function; we can effectively solve the matching problem for all points of the epipolar pair at once. Obviously, the desired optimal route in the search space cannot be of arbitrary shape. Make important assumptions:

a) A scene consists of objects that may exist in physical reality. Therefore, we consider only one-connected routes.
b) All objects present in the scene are opaque. Thus, the visible point can correspond to only one point. If it corresponds to more than one point, it is considered to be blocked.
c) The condition of order is fulfilled. From these assumptions, it immediately follows that the desired path cannot change direction, i.e. for the route point (i,j) only the point (i+l, j), (i, j+1) or (i+1 ,j+l) can be next. This fact allows you to successfully apply the method of dynamic programming.

We introduce the cost function as the arithmetic mean of the cost of the pixels for which we have already built a route. Moreover, we will look for a route that minimizes this function. The transformation is specified by the rule:
m=i - j, n=j
In this space, the rule (C) be reworded: route point (m,n) the following may be only the point (m+1, n), (m, n+1) or (m-1, n+1). To build the optimal route we will use a modified Dijkstra algorithm. Below is its description:
Task. It is known that all prices are non-negative. Find the minimum cost function path from the lower boundary of the space to the top.
We put the entire lower bound of the space in the list of points sorted by the cost function, for which the optimal path has already been found, and start the cycle.

1. We get from the list the first point. Remove it from the list.

2. Check if it lies on the upper border of the exit cycle to point 4.
3. For each of its three neighbors, watching whether his sample from a list of previously. If so, go to start of loop point 2. Otherwise, we consider a new cost function for it, write the current point as a "parent" for this neighbor and add the neighbor to the list.
4. Restore the prescribed "parents" the desired path.

The hardest part of this algorithm is the constant sorting of the list. When you set it as a tree, it occurs in nLog (n) steps. The testing showed that of the point lying closer to the end of the list almost never contribute to the construction of the route. The introduction of an additional point to the algorithm is obvious: 2 (a) if the length of the list reaches the constant C2, and then delete all its elements following the number C1. (C1<C2).

Constant value C1 was chosen to be equal to the width of the search space, i.e. (disparity_max-deparity_min). This decision is based on the peculiarities of filling the search space – if a poorly textured surface occurs in the scene; the space is filled with small values close to each other. It is from such areas that the most points fall into the list. Obviously, such areas are bounded by the width D of the search space. The value of C2 was chosen as 3C1. Testing has shown that adding this small check reduces the length of the sorted list by several times without making noticeable distortions.

8. The method of face recognition

The method of face recognition with the use of perturbation functions and the set-theoretic operation of subtraction was presented in [16, 17].

A calibrated stereo pair (Fig. 1) is used for calculating 3D points on the face.

Let us assume that we have two projective matrices Mi

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

where x, y, z are three-dimensional coordinates of the point, ui and vi are their projections in the image i, and si is the scale factor. The stereo pair is characterized by the following parameters: the points of the image planes E1 = (u1, v1) and E2 = (u2, v2), 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 [18].

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

Here I1 and I2 are the intensities of the left and right images, ¯I1 and ¯I2 are their mean values, dx and dy are the displacements along the epipolal 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) (Fig. 2).

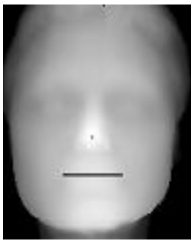

Fig. 2. Depth map

In finding the perturbation peak, we calculate the characteristic size of the projection of the currentinterval, 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. 3). Using three antropomorphic masks, we construct a coordinate system that ensures a possibility of superposition of the tested masks (Fig. 4); finally, certain parts are cut off by a clipping plane for equalization of the volumes.

Fig. 3. Reconstructed object

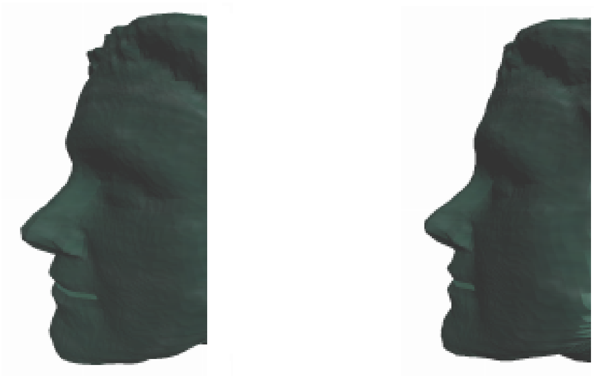

Fig. 4. The test sample and the object from the database

Applying the set-theoretic operation of subtraction

F3 = **F**1**F**2

We determine the set of 3D points (voxels) belonging to the object f3 = **Φ**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.

9. 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 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 data bases 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, there are some factors that 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 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 cover a major part of the face.

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 that 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 on the basis of 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 on the basis of 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 on the basis of different facial expressions; face recognition on the basis of 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: defence complex enterprises, heavily crowded areas, etc.

Further research areas:

1. Development of filters superimposed on the resulting elevation map.
2. Development of filters superimposed on the resulting three-dimensional search space (x, d, y) surface, after finding the surface heights.
3. Development of filters superimposed on the two-dimensional search space (x, d) before constructing the optimal route.
4. Development of criteria for comparison of pixels.
5. The development of methods of finding surface correspondences using a separate Y for each cost function, and the total for the entire scene.

The most promising looks 2 and 5 points. Since after constructing the search surface, we also have a lot of similar in price function surfaces, we can choose another, based on additional filters – for example, the median. Moreover, the development of a global cost function would help to reduce the number of errors in the construction of the search surface.

References

[1] Zhengyou Zhang. A Flexible New Technique for Camera Calibration.// Published in: IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 22, Issue: 11, Nov 2000), Page(s): 1330 - 1334. DOI: 10.1109/34.888718

[2] Marina Kolesnik. Direct Recovery of the Camera Internal Parameters Using Known Angles. https://www.researchgate.net/publication/228895886_Using_Angles_for_Internal_Camera_Calibration_and_Calibration_Update

[3] O. Faugeras and G. Toscani. The calibration problem for stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 15–20, Miami Beach, FL, June 1986. IEEE.

[4] Peter F Sturm Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length. https://hal.inria.fr/inria-00525676/document

[5] Richard I. Hartley. Self-Calibration of Stationary Cameras.// International Journal of Computer Vision, February 1997, Volume 22, Issue 1, pp 5–23.

[6] Richard I. Hartley Estimation of Relative Camera Positions for Uncalibrated Cameras.// Published in: Proceeding ECCV '92 Proceedings of the Second European Conference on Computer Vision, Pages 579-587

May 19 - 22, 1992, Springer-Verlag Berlin, Heidelberg ©1992

[7] Paulo R. S. Mendonc.a and Roberto Cipolla. A Simple Technique for Self-Calibration.// Published in CVPR 1999 DOI:10.1109/CVPR.1999.786984

[8] Anton van den Hengel, Wojciech Chojnacki, Michael J. Brooks, Darren Gawley. A new constrained parameter estimator: experiments in fundamental matrix computation

https://www.researchgate.net/publication/221259813_A_New_Constrained_Parameter_Estimator_Experiments_in_Fundamental_Matrix_Computation

[9] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. The International Journal of Computer Vision, 8(2):123–152, Aug. 1992.

[10] Karsten Mühlmann, Dennis Maier, Jürgen Hesser, Reinhard Männer. Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation// Published in: Journal International Journal of Computer Vision archive Volume 47 Issue 1-3, April-June 2002, Pages 79-88. Kluwer Academic Publishers Norwell, MA, USA

[11] Daniel Scharstein1 and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. November 2001 Technical Report MSR-TR-2001-81.

[12] R. Zabih, J. Woodfill. Non-parametric Local Transforms for Computing Visual Correspondence. Proc. ECCV94. http://www.cs.cornell.edu/~rdz/Papers/ZW-ECCV94.pdf

[13] R. Zabih, J. Woodfill. A Non-parametric Approach to Visual Correspondence. Submitted to IEEE PAMI. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B9CCD0F771FC8E08BBADAF88ED8AD147?doi=10.1.1.17.143&rep=rep1&type=pdf

[14] Sergey Yu. Zheltov, Alexander V. Sibiryakov. Using an Image Preliminary Segmentation for Adaptive Subpixel Correlation.

https://www.academia.edu/2190839/Using_An_Image_Preliminary_Segmentation_For_Adaptive_Subpixel_Correlation

[15] Boykov, Y., Veksler, O., and Zabih, R. (1999). Fast approximate energy minimization via graph cuts. In Seventh International Conference on Computer Vision (ICCV'99), pages 377–384, Kerkyra, Greece.

[16] Vyatkin S.I. Method of Face Recognition with the Use of Scalar Perturbation Functions and Set-Theoretic Operation of Subtraction. Optoelectronics, Instrumentation and Data Processing, 2016, Vol. 52, No. 1, pp. 1–7. Allerton Press, Inc., 2016. ISSN 8756-6990

[17] Vyatkin S.I, Romanyuk S.A., Pavlov S.*V*, Necheporyk M.L Face identification algorithms and its using // Modern engineering and innovative technologies. Technical sciences. International periodic scientific journal. 2018. Issue 5. Part 1. (Germany). P. 111-115. DOI: 10.30890/2567-5273.2018-05-01-078

[18] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision Applications, 6(1)*, 1993. pp. 35–49.