Engineering Journal issn: 0125-8281

Article

Developing an Algorithm for Real-Time 3D Identification of Images

Islam A. Alexandrov[a], Vladimir Zh. Kuklin[b], Alexander N. Muranov[c,*], and Dmitry V. Polezhaev[d]

Institute of Design and Technology Informatics, Russian Academy of Sciences, Vadkovsky Lane 18, building 1A, Moscow, 127055, Russian Federation
E-mail: [a]islam.alexandrov@rambler.ru, [b]kuklin_vladimir_ran@mail.ru, [c,*]alexander.n.muranov@yandex.ru (Corresponding author), [d,*]dm.polezhaev@bk.ru

Abstract. Particular methods exist to compare images based on comparing knowledge about images as a whole; however, for reliable operation of such algorithms, the image should have significant brightness jumps in most areas, heterogeneous scene details, and minimal distortions caused by affine transformations. This study aims to improve the efficiency of video information computing systems of machine vision in data processing by using methods and organization algorithms that allow real-time estimation of the moving object's location and 3D identification. For that, this paper solves a set of base theoretical problems, combining the choice of hardware for obtaining and processing information, determining the coordinate origin, defining the reference plane of the underlying surface, dividing images into levels to achieve higher processing speed, and determining the spatial coordinates of image points from the stereo system. This paper reviews existing image acquisition systems and considers correlation functions for the 3D identification problem. In developing an algorithm that performs real-time 3D identification, the problem at hand is formalized, and the number of levels of the image pyramid is selected, considering the problems arising in 3D identification.

Keywords: 3D identification, correlation functions, image pyramid, coordinate system, machine vision.

ENGINEERING JOURNAL Volume 28 Issue 5
Received 30 January 2024
Accepted 15 May 2024
Published 31 May 2024
Online at https://engj.org/
DOI:10.4186/ej.2024.28.5.83

DOI:10.4186/ej.2024.28.5.83

1. Introduction

In recent years, the number of tasks requiring automation of visual information processing performed by digital computing machines in real-time and related to various application areas has increased. Specialized computing systems of visual information processing are created to solve these tasks. Their design is currently a complex and urgent problem. One of the main directions of developing such systems is the construction of ground-based onboard video information computing complexes designed to estimate moving objects' location [1, 2].

One of the most urgent problems associated with developing onboard vision systems is detecting and selecting objects in sight of the image sensor. These are mobile equipment, airplanes, helicopters, motor vehicles, ships, and people. However, the information on the characteristics of objects to be selected usually includes only approximate dimensions of the object or area.

Some sources of variation in the conditions of image observation exist. First, it is the refraction of light rays in the atmosphere; second, motion and changes in the spatial orientation of the image sensor because of the sensor's placement on a moving object such as mobile equipment, an all-terrain vehicle, or a positioning device. In such a case, transformation parameter estimation methods based on video sequence image analysis are used [3-5]. By calculating the estimates of bias and distortion parameters, it is possible to compensate for their influence, but only partially because the estimates of these parameters will always have some error.

The problem of video information processing is known as image understanding. Image understanding means the transfer from a low-level, brightness-based description to a high-level, meaning-based description. The system that realizes this transition is called an image understanding system [6]. The evolution of brightness and geometric structure detection methods has gone on for about twenty years, from simple to complex [7]. The most developed area is the detection of simple structures like "spots, points, edges, corners, lines." Here, one of the central problems that distinguish the theory of image processing, particularly the theory of signal processing, are methods for detecting objects that are weakly sensitive to various types of variability [8, 9]. Such methods are characteristic only for images with distortions, such as optical sensor distortions, glares, shadows, occlusions, shape distortions, angular distortions, and noise components.

Video information processing should obtain data on the object geometry (coordinates of vertices and normals), further linking to the texture coordinates of the image. This study aims to improve the efficiency of video information computing systems of machine vision using methods and algorithms for organizing special data processing that makes it possible to determine the real-time location of a moving object [10].

2. Literature Review and Analysis

The problem of detection and selection of objects and pattern recognition has not yet been fully solved. However, within significant constraints, some methods make it possible to approach its solution. Particular methods for image comparison use the comparison of knowledge about images in general. Generally, it looks as follows: after calculating the value of a specific function for each image point, it is possible to assign a particular characteristic to the image. After that, the image comparison problem is reduced to comparing such characteristics [11].

Measurement of three-dimensional coordinates of points implies the calculation of spatial characteristics of an object using information of another kind. Photogrammetric methods have achieved great success in indirect measurements [12, 13], which make spatial measurements from the object images obtained in a specific way. Here, the triangulation principle is used to obtain numerical values of coordinates. To apply the triangulation principle, it is necessary to have at least two multiview (taken from different points) images of the object. As soon as the positions of points corresponding to one 3D point on the object are indicated in two images, it is possible to calculate the coordinates of this 3D point. However, to solve the problem, it is necessary to specify a common 3D coordinate system and know all the system parameters, such as camera positions and their internal parameters (Fig. 1).

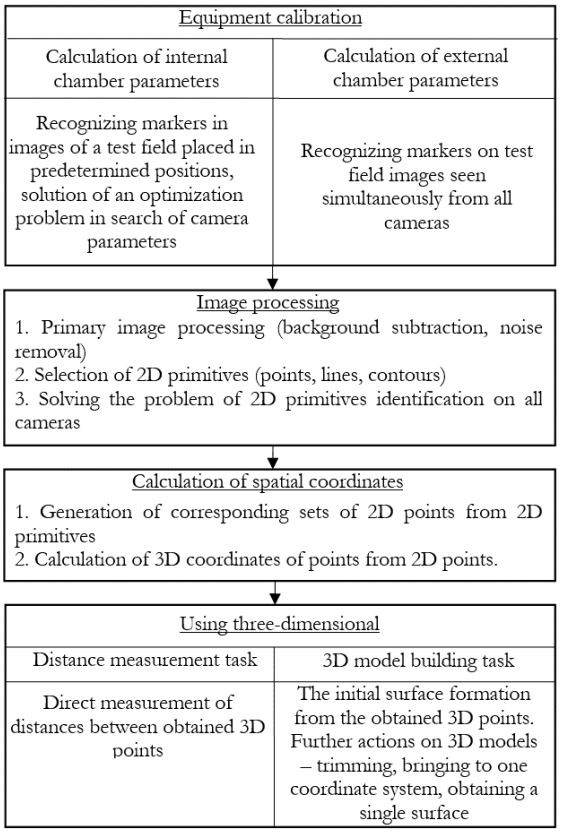

Fig. 1. Scheme of point coordinates determination in 3D space by photogrammetric system.

Object detection tasks rely on image processing algorithms, which, in turn, must receive information from video sensors, one, two (stereo pair), or a whole group. To obtain such information, machine vision systems are developed. These systems use technology to create devices similar in structure to the human eye, endowing them with functions necessary for their work. Note here that a single video sensor in a machine vision system provides only monocular vision, which makes it possible to analyze changes in a flat image. Binocular vision provides stereoscopic vision, the ability to see the surrounding world in three dimensions, determine the distance between objects, and perceive the depth and physicality of the surrounding world.

The image processing process depends on the desired result and what we want. It can be a point cloud, a surface, a set of sections, a plan, a complex ST model, and a set of measurements (lengths, perimeters, diameters, areas, and volumes).

One of the examples of the algorithms of the systems performing 3D identification with the subsequent calculation of the 3D object point coordinates for two cameras is the searching algorithm for a point along the epipolar line. For each point m_1 from the left image, its paired point in the right image can lie only on some line l_2 determined by the shooting geometry (epipolar), which intersects the image of the illuminated line in point m_2. Thus, points m_1 and m_2 are a pair of corresponding points. 3D coordinates of the point M on the object surface are calculated from m_1 and m_2.

When the epipolar passes through the desired point on the right and left camera images, it is necessary to compute the coefficient of concordance. Correlation systems can be used in image processing and correspondence search tasks [14].

Correlation-extreme systems (CES) combine and realize random functions to determine motion coordinates. In particular, correlation-extreme navigation systems (CENS) use working information about one or other fields with spatially random structures. It is convenient to classify correlation-extreme systems by the type and volume of operational a priori (initial) information. According to the operational information type perceived by sensors, CESs are subdivided into two classes: CES I and CES II.

Class I of CEC includes systems where operational information about the navigation field is taken at a single current point. Such CECs use both surface and spatial fields. Class of CES II includes systems where the operational information is taken from some area (frame), i.e., the information sensor gives some image at any time. CES II class includes, in particular, navigation systems operating on the principle of terrain image superposition [15].

According to the volume of initial information used in the system, there are systems "without memory" and systems "with memory". The former, unlike the latter, do not have a "reference" image of the field in the navigation area and can only determine the rate of change of the aircraft coordinates relative to the landmarks, using any surface fields, including those unstable in time (e.g., cloud cover). According to the method of information storage and processing there are classes of analogue and digital systems.

Systems with minimal a priori information store information about the coordinates of discrete process points as "landmarks" in a memory block. Such systems require less memory space than systems with complete a priori information (with equal observation area) but have more strict accuracy requirements and more complex calculation algorithms. All CES subclasses differ in storing and processing a priori and operational information. From this point of view, CESs belong to analog (continuous), digital, and analog-to-digital (mixed). Analog systems store a priori information about the process and perform calculations in analog form. Digital CESs perform these operations by a digital computer, using, in many cases, a specific block to increase the amount of external memory.

Image processing techniques to select characteristic details such as lines, dashes, corners, and objects defined by reference images are well developed [16, 17]. Various image filtering algorithms, Hough transforms [18, 19], morphological operations, and others are used to detect them. However, most of these techniques have a significant disadvantage – either the accuracy of object localization does not exceed one pixel, or these use a priori information about the geometric properties of the object (search for straight lines and circles using the Hough method). This study uses an approach based on the search for correspondence of points of the left image to the right one using correlation algorithms using epipolar geometry. This approach is based on the fact that when searching for a point along an epipolar line using a correlation algorithm, the template moves pixel by pixel given by the search window, and the correlation coefficient is calculated at each position. When the correlation coefficient reaches its highest value, it is the position of the best correspondence.

Correlation algorithms are often used in the construction of machine vision systems based on the photogrammetric principle, so most of the published works consider the problem of selecting the characteristic features of the left image on the right one [20].

The correlation coefficient is a mathematical measure of the correlation between two random variables. The correlation coefficient or pairwise correlation coefficient in probability theory and statistics is an indicator of character. When two random variables change, correlation can be positive and negative (the absence of a statistical relationship is also possible, such as for independent random variables).

To perform the 3D identification procedure and to find the 3D coordinates of a point on scale images comparable to the one used in this study (CCD cameras have a resolution of 640×480 pixels), correlation algorithms are applied, which often take from several seconds to several minutes. Thus, as an example of using correlation functions in image processing, this study demonstrates the action of the algorithm of statistical

correlation of the sum of absolute differences $SAD_{ij}(u,v)$ of two parameters u and v (the size of the computation area is in pixels), where the most appropriate value of the right image is chosen for the value of the left image. The correlation function takes the following form:

$$SAD_{i,j}(u,v) = \sum_{(x,y)\in B_{i,j}} |g^t(x,y) - g^{t+1}(x+u, y+v)| \quad (1)$$

where $g^t(x,y)$ is the brightness of pixel t at the point (x,y),

$g^{t+1}(x,y)$ is the brightness of pixel t+1 at the point (x,y).

The summation in this approach is performed over all points of an object (e.g., a rectangular block). Using this technique to construct a 3D surface point cloud takes from 2 to 4 minutes of personal computer processor time, depending on the size of the (x, y) point search window. This example demonstrates the capabilities of 3D machine vision for a stationary object when processing time is not crucial. However, the use of this technique with the limited processing time does not make it possible to use it when solving the problem of real-time image processing set in this study.

3. Research Methods and Results

Video surveillance uses digital CCD cameras with a resolution of 640×480 pixels because this resolution should make it possible to find a 3D model of the underlying surface relief at a distance of 5 to 150 meters, after which the array of 3D coordinates of the obtained points will be checked for the margin with the given reference plane.

The image of one digital camera 640×480 pixels carries about 307,000 points in one frame, so the number of points obtained from one stereo pair will be approximately 614,000. After the image reduction, the number of points involved in processing will become about 370,000. At a processing rate of 15 frames/sec, the number of points will be around 5 million. To achieve the minimum processing time, it is necessary to develop such a 3D identification algorithm, which will determine the 3D coordinates of the points of the observed scene for the time T=1/15, i.e. for 65 ms (at 30 frames/s).

As a proposed solution to reduce the time of finding 3D coordinates of the underlying surface points, the method of finding points by image pyramid is used with the original image represented as N-images of different accuracies.

However, this type of task should presumably utilize a gradient map at each level [21], with the following conditions determining the algorithm of the corresponding point search module to speed up the search procedure:

- The size of the point search window,
- The frame size in pixels at the lower level of the pyramid,
- The search time of a point at the lower levels should be equal in total to the search time of a point at the upper (initial) level,
- The map of gradients should consider their directions (this procedure will reduce the algorithm's running time and increase the finding reliability, which will be described below).

The stage of system development should start with the 3D identification of points of the left image on the right one according to the generated algorithm, then calculate the spatial coordinates of the found points, and end with the information file creation with the data on the height of the obstacle points relative to the given reference underlying surface and the distance to the moving platform (Fig. 2).

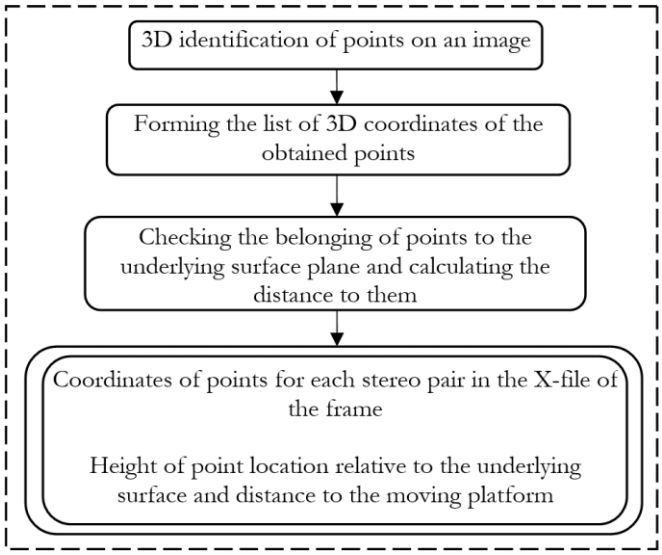

Fig. 2. Sequence of operations to obtain the final result.

The need to use correlation methods is because their main advantage is the high accuracy and reliability of the selection of objects in a complex background with a low signal-to-noise ratio. The increased computational complexity of correlation methods is surmountable using modern FPGA and microprocessors. However, another disadvantage, consisting of increased requirements for the availability of a priori information, can be overcome only partially using the image pyramid and gradient representation of the underlying surface features. Using the correlation method can be justified only when initial object detection is at a low-resolution level.

Internal orientation parameters define the position (x_p, y_p, f) of the projection center in the image coordinate system. Parameters of external orientation define the position of the projection center in the scene coordinate system (X_0, Y_0, Z_0) and the rotation of the image coordinate system relative to the scene coordinate system. The labels' coordinates serve as observations to solve the problem of calculating the vector of estimated parameters. Solving the system of normal equations uses the Gauss method [22].

Real-time estimation of the angular position parameters of the left (α_L,ω_L,κ_L) and the right (α_R,ω_R,κ_R) cameras in the pixel coordinate system includes the procedure of their mutual orientation necessary for

reliable distance estimation to an obstacle. Estimating the orientation parameters is performed under specified distances between some reference points of the test scene measured with the accuracy of 5 mm.

The applied calibration method makes it possible to obtain the resulting accuracy at the level of approximately 5 mm at a 15 m distance:

- The estimation accuracy of spatial coordinates of reference points of the test scene is 5.14 mm at a distance of 15 m;
- The estimation accuracy of angular parameters of the external orientation of cameras is 0.35.

After capturing a sequence of stereo pairs, reference points are automatically recognized on images with further calculation of their coordinates with subpixel accuracy and estimating mutual orientation parameters using the obtained sequence of stereo images.

Data flow in the camera calibration module is:

$$P_{\text{int}}^{T} = (f, x_0, y_0, m_x, m_y, k_{0x}, k_{1x}, k_{2x}, k_{0y}, k_{1y}, k)$$
$$P_{\text{int}}^{T'} = (f, x_0, y_0, m_x, m_y, k_{0x}, k_{1x}, k_{2x}, k_{0y}, k_{1y}, k)' \quad (2)$$

where P_{int}, P'_{int} – vectors of internal orientation parameters for the left and the right cameras, f – the focal length, (x_0, y_0) – the selected point, m_x, m_y – scale parameters, and k – distortion parameters.

$$P_{\text{rel}} = (\alpha, \kappa, \alpha', \omega', \kappa')_{\text{rel}} \quad (3)$$

where P_{rel} – the vector of mutual orientation parameters, α, ω, κ – left camera rotation angles, α', ω', κ' – right camera rotation angles.

At 3D identification of one stereo pair (left and right images), it is necessary to determine the 3D correspondence of image points for a time interval of 1/15 second. Methods of image reduction, division into levels of different scale and informativeness, and correlation by epipolar line with gradient refinement are developed as tools for searching the left image point on the right one. These operations and techniques allow 3D identification and search for image point coordinates from a real-time video sequence [23].

The 3D identification task includes the following subtasks:

1. Selection of a reference in one image;
2. Detection of an image corresponding to the reference in another image;
3. Subpixel refinement of the position of the image corresponding to the reference;
4. Estimation of 3D identification quality.

3D pair images differ from each other due to a set of factors of four categories [24]:

1) Global factors that uniformly distort the intensity level of characteristics of all scene (field) elements and cause geometric distortions of different natures;

2) Regional factors that uniformly distort the intensity level of characteristics only within homogeneous areas of the scene, for example, changes in contrast or brightness;

3) Local factors that independently affect each elementary component of the scene or their grouping, e.g., additive or multiplicative noise;

4) Non-structural factors that change the characteristic features of the scene, e.g., partial covering of the scene by the cloud, distortion of the scene by shadows and fades, and others.

The main problems in the 3D identification of two images are the difference in foreshortening when comparing the left image with the right one and the low detail of the observed area. There are no universal ways to overcome these problems, so the quality of the final result directly depends on how successfully they will be overcome [25].

The following main methods restrict the search area for 3D correspondence of a left image point to a right image one:

1) Application of epipolar geometry. If the orientation parameters of a stereo pair are known, then corresponding points should lie on epipolar lines of left and right images. The epipolar line for some point P of the object space is the line of intersection of the image plane and the plane passing through the camera projection centers and the point P. The use of epipolar geometry makes it possible to significantly reduce the search area since the search area here is a straight line rather than the whole overlapping area of images [26].

2) A priori relief height estimation, which imposes limitations. When there is a priori knowledge about the elevation of relief in each pixel of the main image, and their matching can be done by converting the planar coordinates of each pixel in the main image into geodetic coordinates and then converting them into planar coordinates in the supplementary image. Here, the more accurate a priori data on the range of relief heights are, the smaller the range of acceptable parallaxes for a given stereo pair is [27].

3) Application of image pyramid (hierarchical 3D identification). The result of the algorithm operation at a smaller scale level can be used to construct disparity constraints for image processing at the next scale level. This allows to reduce the algorithm running time (which is proportional to the value of the considered disparity range) and to reduce the number of errors due to cutting off a part of obviously false matches. This method uses possible image decimation by m times, where the 2D search area is reduced by m^2 times [28].

Figure 3 presents a set of tasks that form the problem of 3D identification of images.

DOI:10.4186/ej.2024.28.5.83

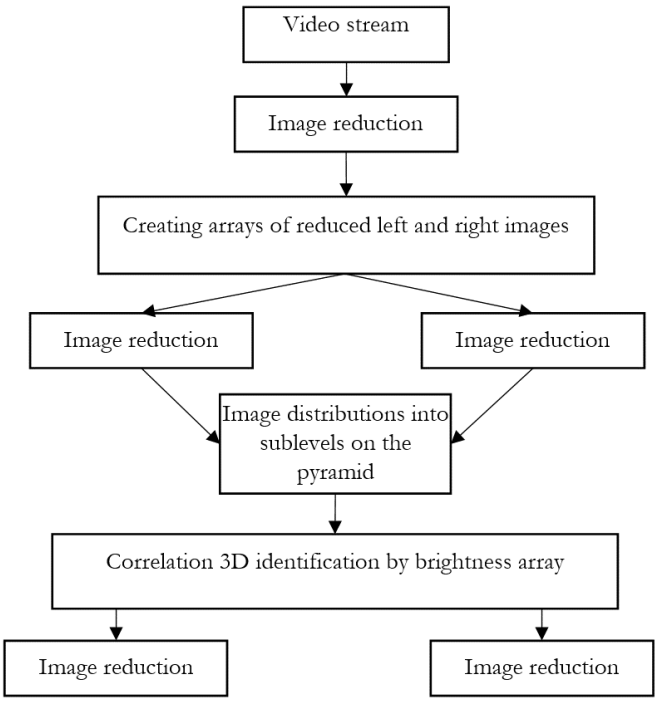

Fig. 3. Set of tasks that form the problem of 3D identification of images.

One of the procedures necessary to take advantage of the solved problem is image reduction. A video information system on a platform that moves along the surface observes the terrain. Obstacles to such a system are objects of the underlying surface. Since the video information system looks to the horizon, the upper part of the frame will not carry information about possible obstacles. Therefore, to save computational resources, it is necessary to exclude this area from the processing and point search.

Reduction means the process of reducing the size of the original image vertically and horizontally. This process is an element of the automated vision system operation, which makes it possible to select the part of the scene necessary for further recognition and analysis at the processing stage.

The complexity of applying the correlation method in the problem of searching for 3D correspondence of the left image point to the right image is in the choice of such sizes of the compared fragments, at which differences in the identical fragments are still small, and the estimation of the correlation coefficient remains reliable. The disadvantage of correlation methods is sensitivity to scale distortions in the identified fragments. The simplest way to reduce scale differences in the left and right fragments is to use preliminary affine image adjustment.

The recognition problem has an explicit complex hierarchical character and includes several main stages of visual field perception, reduction, normalization of selected objects, and recognition. To perform the 3D identification procedure on the whole observed image area, it is necessary to select the initial search area of the algorithm to start working.

Despite numerous attempts to create universal methods of searching for corresponding points in a stereo pair, this problem has not been completely solved due to its complexity corresponding to the difficulty of the general problem of image understanding. The first experiments in this field date back to the 50-60s. The basic idea of automatic 3D identification was that assuming sufficiently small, corresponding points, stereo pairs are similar, and it is possible to convert a photographic image into electrical signals by analyzing these signals for several points. Such an approach became possible by using digital images from CCD cameras of a video information system.

Many 3D identification methods fix one of the images and search (detect) the corresponding image on the other image using the selected 3D identification method. The fixed image will be called a reference. The area in the frame from the left camera, necessary for finding the corresponding area in the right frame, will be a reference.

The 3D identification strategy defines the general scheme of solving the problem of automatic 3D identification. The most common strategies are hierarchical identification and the neural network approach [29, 30]. Table 1 illustrates the differences between these methods.

Table 1. 3D identification methods.

Method of 3D identification	**A measure of proximity of images**	**Images**
Area identification	Correlation function, the sum of squares of brightness differences	Sections of the initial image
Feature identification	Target function	Edges and their attributes
Symbolic identification	Target function	Symbolic description

The best-known methods among the area methods are as follows.

1. Normalized correlation. This method is the simplest and was among the first developed. It describes one of the first automatic 3D identification systems based on the computation of the normalized correlation function of two images, which gives satisfactory results for images of uncomplicated scenes. Further improvements were introduced into the correlation scheme, such as an adaptive correlation window and correction of geometric distortions at foreshortening change.

2. Least squares identification [31]. It uses the sum of squares of brightness differences as a measure of image proximity. The features of this method are as follows:

Fulfilling the requirement of piecewise constant surface of objects in small neighborhoods;

– Using an iterative procedure;
– Adaptive removing geometric and brightness distortions of the images;

- Subpixel identification with accuracy estimation;
- The need for specifying an initial approximation.

3. Feature identification methods. The best-known are methods based on dynamic programming, relaxation, robust estimation, and graph identification [32].

To implement the algorithm of real-time module functioning, it is necessary to process several (15) frames per second. To reduce the time for processing one stereo pair, step-by-step 3D identification is performed. The developed multistage 3D identification algorithm at the initial stage reduces the image by reducing the size of the initial image vertically and horizontally.

Indeed, if the initial image size is $M \times S$ pixels, where M is the number of pixels in the image along the x-axis, and S – along the y-axis, and the area $M^* \times S^*$ does not contain obstacles due to low informativeness, then the image start for the algorithm will be a pixel with coordinates $\left(M_{x,y}, S_{x^i,y^i}\right)$, where I is the start of the search area along the y-axis. Thus, the region $M^* \times S^*$ will be excluded from further search. Therefore, the start of the area to be searched for 3D correspondence will be a point with coordinates $\left(M_1, S'_{N_{\text{нач}}}\right)$, where N_{start} is the image coordinate defined by the start of the search area for 3D correspondence of points, and $S'_{N_{\text{нач}}} = S - S^*$.

The practice of processing many images of the underlying surface obtained by CCD cameras mounted on a moving platform has shown that using only brightness features for identification can lead to incorrect determination of the point of the left image on the right one. In this regard, it is proposed to use images transformed to gradient form with the subsequent refinement of correlation dependence to determine the point position.

The analysis of many images of the underlying surface obtained by CCD cameras mounted on a moving object has shown that using brightness features for identification can lead to incorrect determination of the desired point of the left image on the right one. To clarify the correlation dependence and achieve better accuracy of point finding, the use of images transformed to gradient form is proposed. To realize the 3D identification algorithm, we create two arrays $\left\|\nabla_{x_L,y_L}\right\|$ and $\left\|\nabla_{x_R,y_R}\right\|$; one contains norms (lengths) of gradient vectors of points (x, y) in the left image, and another – in the right image.

The initial pixel value to start the search is selected experimentally. When the initial image size is 640×480 pixels, the search in cells from (0) to (199,640) is not performed. The pixel with coordinates (0, 200) is the initial pixel of the image.

The arrays Pix_{x_L,y_L} of the left image pixel and Pix_{x_R,y_R} of the right image pixel contain the pixel brightness values of the black and white camera image (Table 2).

Table 2. Arrays brightnesses of pixels of left and right images.

Pix_{x_L,y_L}		Pix_{x_R,y_R}	
0.2	Pix_{x_1,y_1}	0.2	Pix_{x_1,y_1}
…	…	…	…
640.48	$Pix_{xn,yn}$	640.48	$Pix_{xn,yn}$

The image processing algorithm can be represented by a general scheme where the input data matrix is $F(i,j)$ subjected to linear or nonlinear processing to enhance the brightness difference. The transformation result is an array of numbers $\nabla(i,j)$. Then, the comparison with the threshold is performed, determining the position of the image elements with pronounced gradients. If $\nabla(i,j) < TL(i,j)$, then there is a descending drop, and when $\nabla(i,j) \geq TU(i,j)$ – an ascending drop, where the values $\nabla(i,j) \geq TU(i,j)$ and $TU(i,j)$ are the lower and upper threshold values. The choice of threshold is one of the critical issues in differentiating between drops. If the threshold is too high, structural elements with low contrast will not be selected. Too low a threshold will cause noise to be mistaken for a drop.

To confidently classify a point as located on the brightness drop, the brightness change associated with that point must be substantially more significant than the brightness change at the background point. Since this is a local computation, the way to determine which value is significant and which is not is to set a threshold. The concepts of first and second derivatives are used to quantify the change in brightness. First-order derivatives in an image are computed using a gradient. To obtain the second-order derivatives, the Laplacian is applied.

The calculation of the first derivative of the digital image is based on various discrete approximations of the 2D gradient. The direction of the gradient vector coincides with the direction of the maximum change rate of the function fв at the point (x, y). One of the ways to find the first partial derivatives ∇_x and ∇_y in a particular point is to apply the following Sobel gradient operator [33]. The gradient values computing uses image convolutions with masks.

$$\nabla_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \quad (4)$$

$$\nabla_x = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad (5)$$

$$|\nabla| = \sqrt{\nabla_x^2 + \nabla_y^2} \quad (6)$$

The amplitude-frequency response of the operator is expressed by the relation:

$$|\nabla| = \sqrt{\nabla_x^2 + \nabla_y^2} \quad (7)$$

For the Sobel operator, which detects horizontal and vertical contours (brightness drops), we can present the corresponding masks for convolution of the initial image

and additional pairs of Sobel operator masks designed to detect gaps in diagonal directions. Each mask has a sum of coefficients equal to zero, i.e., these operators will give zero response in regions of constant brightness, as one would expect from a differential operator. The considered masks apply to obtain the gradient components ∇_x and ∇_y. To compute the gradient value, these components must be used together. The approach often applies when the gradient magnitude is calculated approximately through the absolute values of partial derivatives.

$$|\nabla(x,y)| = |\nabla_x(x,y)| + |\nabla_y(x,y)| \quad (8)$$

Since the brightness function is known only at discrete points, we cannot determine the derivatives until we assume that brightness is a continuous function to which these points belong. The derivatives at any single point are the brightness functions of all points in the image, but approximations of their derivatives can be determined with greater or lesser accuracy.

All points that are similar according to some predetermined similarity features described below connect and form a contour consisting of pixels that meet these criteria. Such an analysis uses the following two basic features to establish the similarity of the contour pixels:

- The value of the gradient operator response;
- The direction of the gradient vector.

The choice of the solution method for the 3D identification problem was based on analyzing real images considering the primary factors complicating 3D identification:

1) Significant brightness differences of stereo pair images arising when taking pictures of object surfaces at different angles;

2) Significant geometric distortions because of potentially complex relief;

3) Possibility of the areas with low brightness variation in images.

To obtain a detailed model of the underlying surface, it is necessary to identify many points, so an essential criterion in developing the algorithm of the module was its operation speed.

The scene image can be in different spatial scales. At that, the scene's large details are better visible on images with a fine (coarse) resolution. Fine scene details are visible only on images with high resolution. The informativeness of image sections also depends on their resolution. If you reduce the scale of the image on the *x*-axis which means using an image with a coarser resolution, then the dispersion grows higher at the same size of the area. This makes it possible to use images with a coarse resolution for selecting references with high informative value to improve the 3D identification quality. An image presented in several scales is a pyramid (Fig. 4).

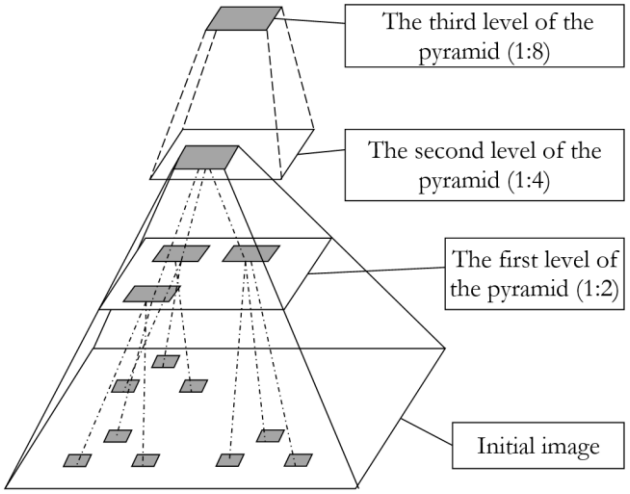

Fig. 4. Illustration of the image pyramid method.

To build a pyramid in the implementation of the algorithm of 3D identification module functioning, after receiving the first pair of images from the left and right cameras, their scaling is performed. The initial image from one of the cameras installed on the moving platform has an actual size of 640×480 pixels. In implementing the 3D identification method of the left image point on the right one, first, the correspondence is searched at the upper level of the pyramid, which is 1/20 of the original image (5%). Its size will be 40×30 pixels.

The choice of the number of levels depends on the size of the initial frame and features of the underlying surface image obtained from video sensors. Each level is reduced by two times compared to the previous one. Two conditions are essential when choosing the number of pyramid levels:

- Time required to process (perform 3D identification) one frame of the underlying surface obtained from the left and right cameras at all levels (starting from the highest and ending with the lowest, the initial one);
- Reliability of processing and amount of information contained in images of pyramid levels, depending on similarity coefficients between them.

Using the limitation of the search area by applying the initial approximation of the coordinates of the correlation function maximum, determining more accurate initial approximations for processing the lower levels based on the results of processing the upper levels is performed.

The number of pyramid levels is determined by the image detail, on which the algorithm of point 3D correspondence detection functions, and on each subsequent level, the image is reduced by two times. At the selected number of pyramid levels, a fast transition to another level and processing is performed only in the search area of this point.

Search for a characteristic point on the image from the left camera on the image from the right camera is performed by five levels of the pyramid:

1st level – 40×30 pixels;

2nd level – 80×60 pixels;

3rd level – 160×120 pixels;

4th level – 320×240 pixels;

The initial level is 640×480 pixels (original image).

Each cell (pixel) of an image at a higher level of the image pyramid is transformed into four, doubling in the *x*-axis and doubling in the *y*-axis.

To specify the similarity measure of the correspondence search area defined at any level of the pyramid and consisting of some number of cells, the K_i indicator is used. The data in the cells are the pixel brightness values from 0 to 255. The cells of the search area 3×3 pixels on the left image are compared with the given search area 5×5 pixels on the right image. The developed 3D identification algorithm compares the brightness values for each cell of the left image search area (a_{Li},b_{Lj}) from the right image search areas (a_{Ri},b_{Rj}) along the epipolar line.

$$\frac{a_{L_i}}{b_{L_j}} \leq \frac{a_{R_i}}{b_{R_j}} \quad (9)$$

where *a* – search area line number, *b* – search area column number.

In the right image, a refinement is made around the desired point in a 3×3 pixel area, shifting the area by one pixel relative to the central one (the desired point). To determine the area containing the desired point, it is necessary to compare nine areas of the right image with the initial area of the left image. The above compares the nine regions of the right image with the initial area of the left image. Since the area values are pixel brightnesses, we search for a 3×3 pixel area on the right image, whose brightness differs least from the brightness values of the same dimension area of the left image containing the initial point.

The analysis of the figure illustrates the comparison process, where A_{Ln} is the matrix containing brightness values of the initial area, and aB_{Rn} is the matrix containing brightness values of the desired area. The comparison procedure is performed according to the criterion.

$$K = \sum_{n=1}^{9} det\left(A_{L_{n_i}} - B_{L_{n_i}}\right) \quad (10)$$

where *K* – similarity coefficient of areas;

n – the number of areas 3×3;

A_L – matrix of brightness values of the left image search area element;

B_{Rn} – matrix of brightness values of the right image search area element;

i, j – row and column numbers.

The more the number of matched brightness values of cells (pixels) of the image for each area is, the smaller the value of the coefficient K will be, which will be the solution to the problem of refining the coordinate of the desired point.

Figure 5 shows the scheme of the action sequence when implementing the algorithm of the 3D identification module functioning further implemented programmatically.

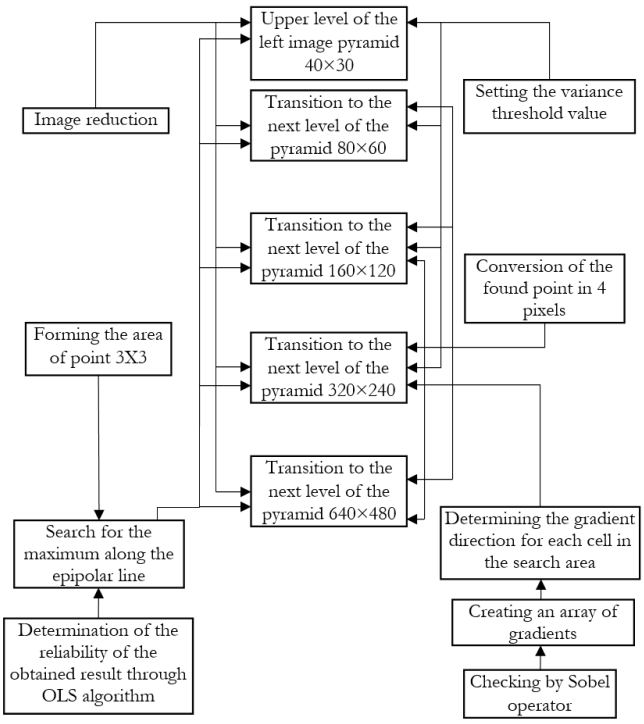

Fig. 5. Scheme of the search algorithm by brute force search.

The accuracy of the reference image is higher the more inhomogeneous the corresponding image fragments are. From the computational point of view, the simplest and most effective indicator of fragment heterogeneity is the dispersion of brightness within the fragment

$$\sigma^2(x_0, y_0, N) = \\ = \frac{1}{(2N+1)^2}\sum_{x=-N}^{N}\sum_{y=-N}^{N}\left(f(x+x_0, y+y_0)\right)^2 - \\ -\left(\frac{1}{(2N+1)^2}\sum_{x=-N}^{N}\sum_{y=-N}^{N}f(x+x_0, y+y_0)\right)^2 \quad (11)$$

The algorithm searches for 3D correspondence over a given area (1 to *N*). The problem of 3D identification is solved by a consecutive search of the values of the shift parameters from the permissible range. At more detailed levels, the initial approximation obtained from previous levels is used.

For the reference *f* of the left image, the image *f* must be searched for in the overlap region of the right image. This requires about MN_na operations, where *n* is the number of pixels in the reference. For real stereo pairs $a \approx 0.6$, $M > 1000$, so the total stereo identification time becomes unacceptably large.

To find stereo correspondence of points of the left image to points of the right image, the procedure of correlation stereo-identification of an image point using brightness features is carried out. To find some pixel *P* of the left image with coordinates (x_p, y_p) in the right image, the whole overlap region of the right image should be checked for the presence of pixel *P*.

4. Analysis and Discussion of the Results

The 3D identification module algorithm result is the determination of the point on the right image corresponding to the point on the left image. Figure 6 shows the dependence of time T (ms) to perform 3D identification of a point on the number of levels *N* in the image pyramid.

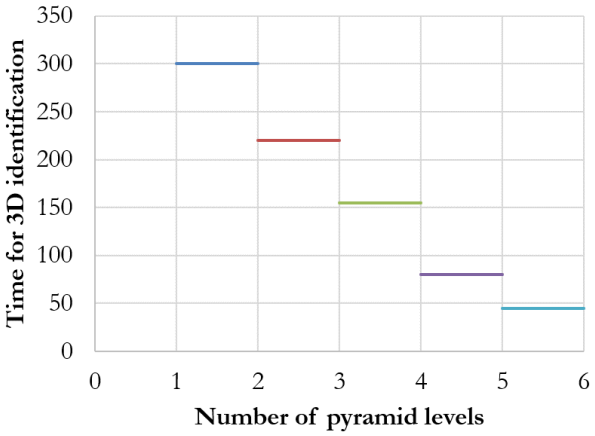

Fig. 6. Graph of dependence of time for 3D identification on the number of pyramid levels.

After finding the corresponding points of the right image for the left image, it is necessary to calculate their 3D coordinates for detecting obstacles on the underlying surface and determining the location of the moving platform. An important aspect is the time spent on the 3D identification, which leaves 8 ms for two subsequent algorithms of the functioning of modules.

5. Conclusion

This work is devoted to improving the efficiency of the video information computing system of machine vision using methods and algorithms for organizing data special processing, which makes it possible to estimate the location of a mobile object in real-time. Within its framework, the development of methods and algorithms of module functioning and software for processing the results of video measurements by the computing system was performed.

The most significant new scientific results consist in developing the structure of a multi-modular high-performance computing system for processing stereo image data in real time using information from two video sensors. Also, a new algorithm for image processing module functioning was developed, which makes it possible to perform their real-time sequential 3D identification (up to 15 times per second).

Acknowledgement

Some results of this research were obtained as part of the work under the Subsidy Agreement dated July 14, 2023 No. 075-15-2023-599 on the topic: "Development of intelligent, high-precision tools for analyzing terrain and managing transport systems intended for highly productive agriculture" with the Ministry of Science and Higher Education of Russia.

References

[1] J. D. B. Nelson, S. K. Pang, N. G. Kingsbury and S. J. Godsill, "Tracking ground based targets in aerial video with dual-tree wavelet polar matching and particle filtering," *11th Int. Conf. Inf. Fusion*, pp. 1-7, 2008.

[2] C. Fruh and A. Zakhor, "Constructing 3D city models by merging aerial and ground views," *IEEE Comp. Graph. Appl.,* vol. 23, no. 6, pp. 52–61, Nov. 2003, doi: 10.1109/mcg.2003.1242382.

[3] G. Liu, M. Z. Li, Z. Mao, and Q. S. Yang, "Structural motion estimation via Hilbert transform enhanced phase-based video processing," *Mech. Syst. Signal Proc.*, vol. 166, p. 108418, Mar. 2022, doi: 10.1016/j.ymssp.2021.108418.

[4] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," *Comp. Sci. Rev.*, vol. 28, pp. 157–177, May 2018, doi: 10.1016/j.cosrev.2018.03.001.

[5] A. P. Shukla and M. Saini, "'Moving object tracking of vehicle detection': A concise review," *Int. J. Signal Proc. Image Proc. Pattern Recognit.*, vol. 8, no. 3, pp. 169–176, Mar. 2015, doi: 10.14257/ijsip.2015.8.3.15.

[6] S. Seo et al., "Artificial optic-neural synapse for colored and color-mixed pattern recognition," *Nat. Commun.*, vol. 9, no. 1, Nov. 2018, doi: 10.1038/s41467-018-07572-5.

[7] C. N. Naga Priya, S. Denis Ashok, B. Maji, and K. Senthil Kumaran, "Deep Learning Based Thermal Image Processing Approach for Detection of Buried Objects and Mines," *Eng. J.*, vol. 25, no. 3, pp. 61–67 Mar 2021, doi:10.4186/ej.2021.25.3.61

[8] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Dig. Signal Proc.*, vol. 126, p. 103514, Jun. 2022, doi: 10.1016/j.dsp.2022.103514.

[9] T. Mahalingam and M. Subramoniam, "A robust single and multiple moving object detection, tracking and classification," *Appl. Comp. Inf.*, vol. 17, no. 1, pp. 2–18, Jul. 2020, doi: 10.1016/j.aci.2018.01.001.

[10] C. Dechsupa, P. Prasankok, W. Vattanawood, and A. Thongtak, "MorphoNet: A Novel Bivalve Images Classification Framework with Convolutional Neural Network," *Eng. J.*, vol. 27, no. 9, pp. 71–81, 2023, doi:10.4186/ej.2023.27.9.71

[11] V. Zh. Kuklin, A. A. Tatarkanov, and A. A. Umyskov, "Trainable regularization in dense image matching problems," *HighTech Innov. J.*, vol. 4, no. 3, pp. 617–629, Sep. 2023, doi: 10.28991/hij-2023-04-03-011.

[12] H. M. Yilmaz, M. Yakar, and F. Yildiz, "Digital photogrammetry in obtaining of 3D model data of

irregular small objects," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 37, pp. 125-130, 2008.

[13] R. Deli, E. D. Gioia, L. M. Galantucci, and G. Percoco, "Accurate facial morphologic measurements using a 3-camera photogrammetric method," *J. Craniofac. Surg.*, vol. 22, no. 1, pp. 54–59, Jan. 2011, doi: 10.1097/scs.0b013e3181f6c4a1.

[14] F. Walch, C. Hazirbas, L. Leal-Taixé, T. Sattler, S. Hilsenbeck and D. Cremers, "Image-based localization using lstms for structured feature correlation," *IEEE Int. Conf. on ICCV*, pp. 627-637, 2017, doi: 10.1109/ICCV.2017.75.

[15] A. Tantsiura, "Evaluation of the Potential Accuracy of Correlation Extreme Navigation Systems of Low-Altitude Mobile Robots," *Int. J. Adv. Trend. Comp. Sci. Eng.*, vol. 8, no. 5, pp. 2161–2166, Oct. 2019, doi: 10.30534/ijatcse/2019/47852019.

[16] J. Chaki and N. Dey, *Texture Feature Extraction Techniques for Image Recognition*. Singapore: Springer, 2020. doi: 10.1007/978-981-15-0853-0.

[17] I. Alexandrov, G. Malysheva, and T. Guzeva, "A Qualitative Visual Analysis of the Fractured Surfaces of Epoxy/Carbon Fibre Composite Prepared by the Melt and the Solution Technologies," A*dv. Comp. Mat. Technol. Aerosp. Appl.*, pp. 43-46, 2012.

[18] A. S. Hassanein, S. Mohammad, M. Sameer, and M. E. Ragab, "A survey on Hough transform, theory, techniques and applications," *ArXiv preprint*, 2015. doi:10.48550/arXiv.1502.02160

[19] L. Chandrasekar and G. Durga, "Implementation of Hough Transform for image processing applications," *Int. Conf. Commun. Signal Proc.*, pp. 843-847, 2014. doi: 10.1109/ICCSP.2014.6949962.

[20] W. Förstner and B. P. Wrobel, *Photogrammetric Computer Vision*. Cham: Springer, 2016. doi:10.1007/978-3-319-11550-4

[21] J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht, "Gradient descent only converges to minimizers," *JMLR: Workshop Conf. Proc.,* pp. 1–12, 2016.

[22] P. Li, D. Wang, L. Wang, and H. Lu, "Deep visual tracking: Review and experimental comparison," *Pattern Recogn.*, vol. 76, pp. 323–338, Apr. 2018, doi: 10.1016/j.patcog.2017.11.007.

[23] M. Humenberger, C. Zinner, M. Weber, W. Kubinger, and M. Vincze, "A fast stereo matching algorithm suitable for embedded real-time systems," *Comp. Vis. Image Underst.*, vol. 114, no. 11, pp. 1180–1202, Nov. 2010, doi: 10.1016/j.cviu.2010.03.012.

[24] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Proc. Image Commun.*, vol. 28, no. 8, pp. 870–883, Sep. 2013, doi: 10.1016/j.image.2012.08.004.

[25] A. Tatarkanov, I. Alexandrov, A. Muranov, and A. Lampezhev, "Development of a technique for the spectral description of curves of complex shape for problems of object classification," *Emerg. Sci. J.*, vol. 6, no. 6, pp. 1455–1475, Dec. 2022, doi: 10.28991/esj-2022-06-06-015.

[26] A. Goldstein and R. Fattal, "Video stabilization using epipolar geometry," *ACM Trans. Graph.*, vol. 31, no. 5, pp. 1–10, Aug. 2012, doi: 10.1145/2231816.2231824.

[27] L. Polidori and M. El Hage, "Digital elevation model quality assessment methods: a critical review," *Remote Sens.*, vol. 12, no. 21, p. 3522, Oct. 2020, doi: 10.3390/rs12213522.

[28] J. Ji and J.-S. Zhao, "Increased plane identification precision with stereo identification," *Robot.*, vol. 41, no. 9, pp. 2789–2808, Jun. 2023, doi: 10.1017/s0263574723000681.

[29] I. A. Alexandrov, A. V. Kirichek, V. Z. Kuklin, and L. M. Chervyakov, "Development of an algorithm for multicriteria optimization of deep learning neural networks," *HighTech Innov. J.*, vol. 4, no. 1, pp. 157–173, Mar. 2023, doi: 10.28991/hij-2023-04-01-011.

[30] W. Nualtim, W. Suwansantisuk, and P. Kumhom, "Face Synthesis and Partial Face Recognition from Multiple Videos," *Eng. J.,* vol. 27, no. 4, pp. 29–44, 2023, doi:10.4186/ej.2023.27.4.29

[31] M. Verhaegen and V. Verdult, *Filtering and system identification*. UK: Cambridge University Press, 2007, doi: 10.1017/cbo9780511618888.

[32] H. Wang and X. Bi, "Person re-identification based on graph relation learning," *Neural Proc. Lett.*, vol. 53, no. 2, pp. 1401–1415, Mar. 2021, doi: 10.1007/s11063-021-10446-5.

[33] O. Vincent and O. Folorunso, "A descriptive algorithm for Sobel image edge detection," *InSITE Conf.,* pp. 97-107, 2009, doi: 10.28945/3351

DOI:10.4186/ej.2024.28.5.83

Islam A. Alexandrov was born in Russia, on 27.04.1991. In 2012 he completed his bachelor's degree and later - in 2014, he graduated from the master's program at Bauman Moscow State Technical University. Later, in 2020, he received the degree of candidate of technical sciences.

He currently works as a senior researcher at IDTI RAS. He has experience as an engineer and researcher, has experience in scientific leadership of teams of research projects. Area of scientific interests - automation of technological preparation of production, system analysis, decision-making methods, multi-parameter optimization and mathematical modeling. He has 31 Scopus-indexed documents over the last five years.

Mr. Alexandrov in 2020 was elected Chairman of the Council of Young Scientists of the IDTI RAS. For achievements in the field of science and technology, he was awarded the Moscow Government Prize for young scientists for 2023.

Alexander N. Muranov was born in Russia, on 28.03.1990. In 2012 he completed his bachelor's degree and later - in 2014, he graduated from the master's program at Bauman Moscow State Technical University. Later, in 2021, he received the degree of candidate of technical sciences.

He currently works as a senior researcher at IDTI RAS, Moscow. He has experience as an engineer and researcher. Area of scientific interests - automation of technological preparation of production, system analysis, decision-making methods, multi-parameter optimization and mathematical modeling. He has 23 scopus-indexed documents over the last five years.

Mr. Muranov has no memberships in professional societies. For achievements in the field of science and technology, he was awarded the 2023 Moscow Government Prize for young scientists.

Vladimir Zh. Kuklin was born in Russia, on 18.02.1952. In 1970 he graduated from the Kazan State University. Ulyanov-Lenin (Kazan, Russia) with a degree in mathematics, and in 1982 he received the degree of Candidate of Technical Sciences, defending his dissertation at the Leningrad Institute of Aviation Instrumentation (St. Petersburg, Russia). Later, in 2000, he received the degree of Doctor of Technical Sciences, having defended his dissertation at the St. Petersburg State University of Aerospace Instrumentation (St. Petersburg, Russia).

He works as a leading researcher, IDTI RAS, Moscow. He has 10 Scopus-indexed documents over the last five years.

Dr. Kuklin has no memberships in professional societies.

Dmitry V. Polezhaev was born in Russia, on 27.10.1986. In 2020 he graduated from Volgograd Technical University with a master degree in "Automated Design Systems". Currently, he is a postgraduate student at the Institute of Design and Technological Informatics of the Russian Academy of Sciences (Moscow, Russia). The field of scientific interests - automation, processing data, decision - making methods, optimization and modeling.

He is currently working as a junior researcher at IDTI RAS. He has 3 Scopus-indexed documents over the last five years.

Mr. Polezhaev has no memberships in professional societies.