

# Wavelet transform based optimization method for Three-Dimensional computer vision

Svetlana Antoshchuk<sup>1</sup>, Galina Shcherbakova<sup>1</sup>, Sergey Kondratyev<sup>1</sup>, Daria Koshutina<sup>1</sup> and Oleksandr Usov<sup>1</sup>

<sup>1</sup> Odessa Polytechnic National University, Shevchenko Ave., 1, Odessa, 65044, Ukraine

## Abstract

This work aims to develop approximate depth estimation methods for three-dimensional computer vision in unmanned vehicles (UVs) based on a developed wavelet transformation-based optimization method. This method can assess informative features for matching and/or optimizing costs in image analysis, refining discrepancies, and more. Possible solutions are demonstrated for obtaining an approximate depth map by simplifying the calculation of disparity values, traditionally used for forming a depth map in intensity space and using edge description with adjustable detail based on wavelet transformation. The advantage of the developed optimization method over existing algorithms in the wavelet space is the increased speed due to the rational selection of the Haar wavelet support length in the extremum search area. Modeling confirmed the effectiveness of the proposed approach for constructing depth maps and allowed for recommending the proposed method for unmanned vehicles operating under limited computational and energy resources.

## Keywords

Three-dimensional computer vision, contour detection, wavelet transform, optimization, image analysis

## 1. Introduction

The primary source of information about the surrounding environment for solving a multitude of tasks in various fields such as unmanned vehicles (quadcopters, unmanned cars), robotics, and vision for the visually impaired (UVs) is video cameras [1]. Recent advancements in video sensor technologies, as well as image and video analysis and processing methods, have enabled machine vision to confidently advance into the realms of automation and security systems, including household, industrial, and military applications.

Modern machine vision systems not only extract video information but also present it in a form and quantity that allows for the identification of significant, informative features of objects and processes. These systems recognize and assess their position and condition, select rational trajectories for UV movement, and, if necessary, generate corresponding control actions.

A distinctive feature of machine vision systems is not only the identification and recognition of an object and its position but also the consideration of the scene and object depth, i.e., the transformation of a two-dimensional image into a three-dimensional one, where information about the object is represented not just in units of brightness but with pixel/distance parameters. The analysis of the evolution of three-dimensional computer vision and the corresponding hardware and software designed for obtaining images and forming depth parameters has shown that the main problem is the high cost of the sensors and image processors used, making them largely inaccessible to the broader consumer market. Therefore, finding solutions for obtaining 3D depth information about objects in an image characterized by low cost, low power consumption, and acceptable

---

ICST-2024: Information Control Systems & Technologies, September, 23 – 25, 2024, Odesa, Ukraine

✉ asg@op.edu.ua (S. Antoshchuk); galina.sherbakova@op.edu.ua (G. Shcherbakova); kondratiev@op.edu.ua (S. Kondratyev); d.v.koshutina@op.edu.ua (D. Koshutina); a.usoff@stud.op.edu.ua (O. Usov)

🆔 0000-0002-9346-145X (S. Antoshchuk); 0000-0003-0475-3854 (G. Shcherbakova); 0000-0003-4975-5757 (S. Kondratyev); 0009-0004-1326-8775 (D. Koshutina)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

performance is an important and relevant scientific and practical task for a wide range of the aforementioned UVs.

## 2. Leveraging LiDAR and Stereo Cameras for Efficient Depth Information Retrieval in UVs

When solving navigation tasks for UVs or collecting depth information about objects for them, the acquisition of this information can be redistributed and/or transferred to backup subsystems and/or systems operating on other physical principles [2]. For instance, one such approach could be the combination of data from LiDAR (laser sensors) and video sensors. Currently, the LiDAR-SLAM technology is actively being researched by Suzuki [3], Lee, Savkin, and Vuletic [4], Tripicchio et al. [5], Bi et al. [6], Mansouri et al. [7], Vong et al. [8], and others.

These researchers have demonstrated the feasibility of using LiDAR for path planning, collision avoidance, orientation, and navigation.

However, image sensors (video cameras) are characterized by their lightweight, potential decrease in accuracy under low lighting conditions, and loss (absence) of depth information, which increases computational load during data processing. LiDAR allows for direct distance measurement, but the result depends on the reflective properties of the surface and is energy-intensive [9]. Therefore, a less energy-intensive, interesting, and promising solution for obtaining depth information about objects and/or scenes in an image is the stereoscopic vision approach. This approach involves the use of two video cameras with known optical characteristics and parameters of their relative positioning [4, 7, 8]. Typically, video cameras with similar optical characteristics are used. These cameras are directed in the same direction, with the distance between the optical centers of the cameras being much smaller than the distance to the observed objects. Depth information is extracted by comparing and analyzing this pair of images. This approach allows for modeling image processing similarly to how it occurs in the vision of many living beings. Just as a given control point in space is located in different positions in each human eye, this allows the system to calculate the position of this point in space. As a result, a height map is obtained where objects closer to the observer are displayed in lighter shades, and those further away are displayed in darker shades. Several solutions enable the creation of depth maps using local block matching (StereoBM function) or semi-global block matching (StereoSGBM function) [10-11].

Let's consider the method for constructing a depth map, implemented by the StereoBM function. The algorithm involves the following stages of data processing:

1. Loading images from both video cameras (often using the OpenCV library).
2. Converting images from color to grayscale.
3. Creating a StereoBM object with specified parameters.
4. Calculating the disparity map (the magnitude of the difference in the localization of corresponding image elements obtained from the right and left cameras). Traditionally, this is done by matching pixels in the left and right images to calculate a disparity map.
5. Outputting and saving the disparity map.

In the StereoBM algorithm, stages 1, 2, 3, and 5 are standard and do not provide opportunities to increase execution speed and thereby reduce energy consumption. Therefore, stage 4, which is aimed at calculating the disparity map, is more promising. This stage involves the following steps:

1. Selecting blocks. Blocks of a fixed size are highlighted in the image.
2. Comparing blocks. For each block found in the left image, the algorithm tries to find a matching block in the right image.
3. Calculating disparity. For each block in the left image, a displacement (disparity) is calculated, indicating how many pixels the corresponding right image is shifted. This disparity is stored in a disparity map.

4. Normalizing the disparity map. The resulting disparity map may contain values other than depth. Therefore, it is normalized so that the values are within [0, 255] for convenient display.

In this part of the algorithm, the most complex and crucial for the quality of the depth map is the block comparison stage. The block comparison algorithm searches for a pattern on the left frame corresponding to a block on the right frame. This is done by algorithmically moving along the horizontal axis of the right image and comparing it with the corresponding block of the left image. Various approaches can be used to assess the degree of block matching [10-14]:

1. Based on assessing the proximity of pixel intensity values in blocks.
2. Based on assessing the minimum sums of squared differences of intensities (Sum of Squared Differences, SSD).
3. Based on determining the minimum normalized cross-correlation coefficient (Normalized Cross-Correlation).
4. Based on the Semi-Global Matching (SGM) method, which takes into account global and local features of low-quality images with different intensities or difficult or complex image acquisition conditions.
5. Based on adaptive window matching. Instead of a fixed block size, algorithms can use adaptive window mapping to account for changes in texture and intensities in images.
6. Based on the Graph Cuts method, which uses graph structures to calculate the disparity map. These algorithms can take global and local image properties into account and improve accuracy in challenging imaging conditions (e.g., low-light conditions).
7. Based on machine learning methods. These methods rely on machine learning models trained to predict disparity based on feature detection in images. They can handle images with uneven texture and variations in illumination.
8. Based on deep learning methods such as Convolutional Neural Networks (CNN). These neural networks can be trained on large datasets and perform well in complex lighting conditions.
9. Joint use of sensors implemented on various physical principles, such as LiDAR or infrared cameras, to improve the accuracy of the disparity map.

Note that the approaches are numbered and sorted by the degree of increase in computational costs. Approaches N2 to N6 require high energy consumption, which may be critically unacceptable for BTS. Approaches N7 and N8 rely on neural network calculations, requiring computational resources and a training "database," which can be challenging to provide for some BTS applications. Approaches involving the joint use of sensors implemented on different physical principles (e.g., N9) lead to increased energy consumption. Therefore, for further research, the method of comparing image blocks in the right and left images (N1) was chosen as the base method, which can ensure the speed of calculations. However, this method is characterized by low noise immunity, determining its low quality. This property is especially noticeable in changing light conditions and/or low light conditions, often encountered in applied problems solved by BTS. Additionally, the analysis allowed us to establish the following. It should be noted that when developing algorithms for obtaining depth information about objects and/or scenes, two assumptions about the structure of the observed scene are used [9]. The first assumption considers that neighboring pixels have similar disparity values (the difference between the positions of an object in the left and right images) because the scene objects are piecewise smooth. The second assumption is based on the idea that homogeneous areas of the scene correspond to flat surfaces in 3D. The algorithm for constructing a depth map involves measuring the displacement along the x-axis of each point in the right frame and correlating it with the corresponding point in the left frame [9, 15]. The search for the corresponding point is strictly along the horizontal line of each frame.

Therefore, to correctly determine the distance to objects during vertical calibration (along the y-axis in the image), the position of the cameras is set such that the horizontal lines of both cameras

coincide. Horizontal calibration (along the x-axis in the image) is done by rotating the cameras relative to each other until the x-coordinates of points at a distance of more than 10 meters coincide. Consequently, this method requires pixel-precise positioning of the cameras both horizontally and vertically. This complicates alignment and reduces the quality of object positioning [9, 12]. A significant argument in favor of using stereoscopic systems in UVs is the availability of numerous open-source algorithms in the OpenCV computer vision library (the StereoBM class exists) for implementing stereo vision in languages such as C/C++, Python, Java, Ruby, Matlab, and others [12]. The main disadvantages of stereoscopic systems for constructing depth maps for mobile navigation systems and UVs include the following [9]:

- The need for camera calibration. It should be noted that even with perfectly accurate camera positioning, obtaining a depth map in these systems (e.g., when avoiding obstacles) poses difficulties due to the necessity of pixel-by-pixel image matching.
- Dependence on the quality of the initial images and/or incorrectly set camera parameters, lighting, and illumination for each camera.
- Dependence of the number of computational operations and, consequently, performance on the size and quality of the image.

As the above analysis showed, a number of methods based on the assessment of local and global image characteristics have been developed to evaluate disparity. Algorithms based on estimating local image characteristics are fast but low in accuracy, while global methods can improve accuracy and (often) noise immunity, but are characterized by low performance. The search for a compromise between accuracy and speed can be aimed at developing methods for assessing disparity based on optimization algorithms. When assessing disparity based on changes in the intensity of the pixels in an image line, the coordinates of the extremum with a signal-to-noise ratio below 6 are estimated. Known methods for searching for an extremum based on estimating the value of the first derivative, which are traditionally used at this stage, do not work when the signal-to-noise ratio is less than 15. To search for the minimum under such conditions, the authors developed and investigated an optimization method based on the wavelet transform (WT) [16]. However, due to the large number of stages using several types of wavelet functions, this method is characterized by low performance. Therefore, improving the method to increase its performance for the application task described above is relevant.

### 3. Positions Assessment of Depth Information for Object Positions

To reduce the impact of these disadvantages when constructing depth maps from stereo images, an approximate approach to depth information assessment is proposed. This approach considers the results of identifying areas of similar intensity or object contours in the images or points of maximum curvature in the contour description.

Figure 1 illustrates the main steps of in-depth assessment by identifying areas of similar intensity in a test image (see Figure 1, a). The proposed approach involves the following:

- For a segment of the intensity row, determine the region of maximum intensity on the image from one of the cameras, for example, the left one (1) (see Figure 1, b).
- Create a brightness envelope template: the intensity value corresponding to the extremum coordinate of the row and the intensity values of neighboring pixels (Figure 1, c), and locate a similar fragment in the image from the right camera (2).
- Determine the distance between similar elements and the corresponding disparity value.
- Processing is performed row-wise and column-wise. If necessary, the results are combined using a logical OR scheme.

To illustrate, brightness envelope templates representing a block of 3 pixels are provided for the left and right frames of the image (see Figure 1, c). The templates show the intensity distribution around the extremum in the given image row for the left and right cameras (see Figure 1, b).

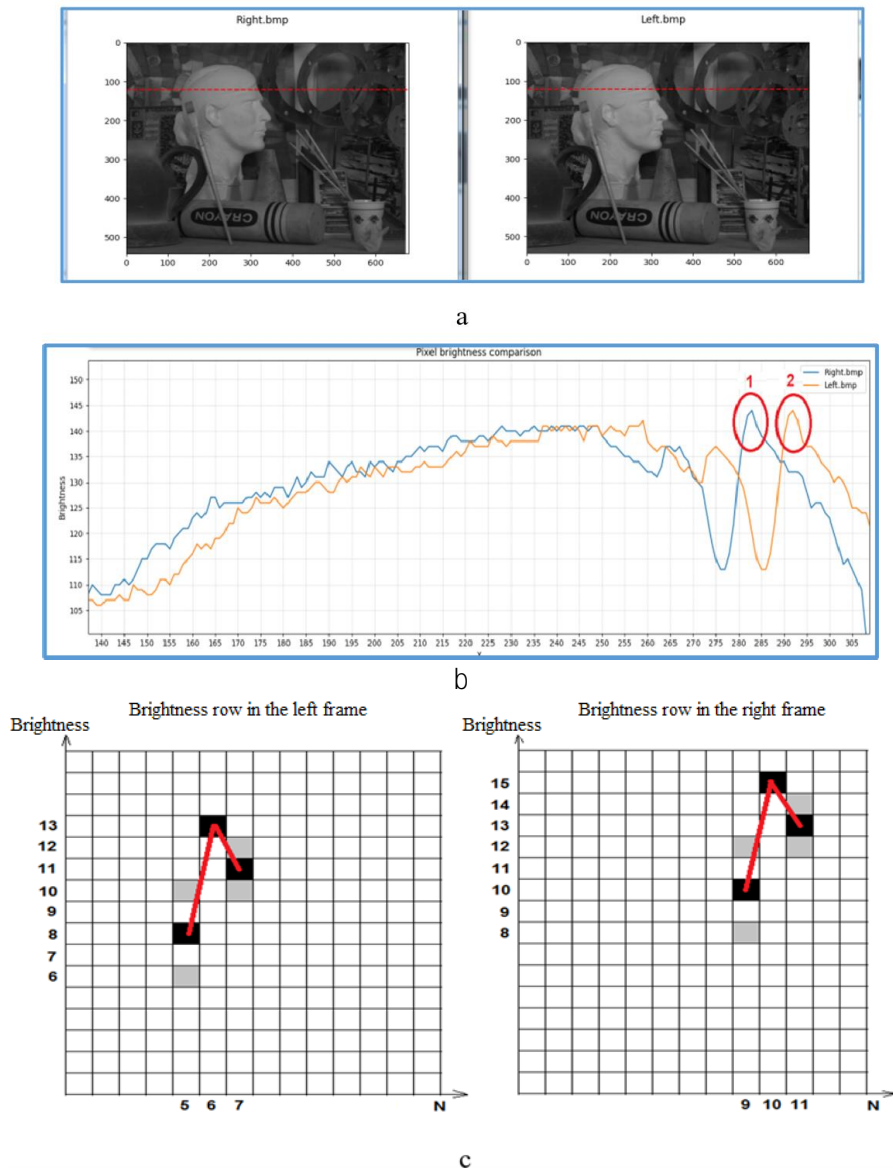


Figure 1: a - Test image: right and left frames; b - Fragments of intensity rows from the right and left frames; c - Template showing the distribution of intensities around the extremum in the given image row.

Both templates have a "peak" shape (other possible shapes include "trough," "plateau," "rise," "fall," etc.). Black squares highlight the brightness levels of each pixel in the blocks, while gray areas indicate the brightness dispersion boundaries. The disparity between blocks is 4 pixels along the X-axis, and the absolute difference in brightness is 2. In real stereo images, the brightness difference can reach up to 30. This approach is simpler than the well-known StereoBM method [10, 12 - 14], which uses two computationally intensive procedures for similar searches: Sum of Squared Differences (SSD) and Normalized Cross-Correlation (NCC), both of which also exhibit lower performance. The computational complexity of finding templates can be reduced if object contours are used in the calculation of disparity. It should be noted that contours are the most informative part of object images, and their analysis can significantly reduce the number of computational operations [15].

However, it should be noted that in this case, the quality of the depth map depends on the effectiveness of the contour detection methods used, among which the most noise-resistant methods are those using the wavelet transform (WT) [15]. To extract the contours of objects using the wavelet transform, the summing component of the approximation (vertical) of the smallest scale ( $a_1$ ) of the discrete wavelet transform is removed by setting its values to zero. Calculation of the inverse transformation using the detailing component ( $d_1$ ) leads to the selection of contours (horizontally) in the image thus reconstructed. In a similar way, the vertical boundary of an object can be selected. The procedure for identifying contours using the wavelet transform can be presented as follows:

$$\begin{array}{l} f_0 \rightarrow a_1 \rightarrow 0 \\ \downarrow \\ d_1 \rightarrow f' \rightarrow K(f'), \end{array} \quad (1)$$

where  $f$  – is the row (column) of the original image;  $a_1$  – summing component of the wavelet transform;  $d_1$  – detailing component of the wavelet transform;  $f'$  – row (column) of the image after the inverse transformation;  $K$  – operator for obtaining a contour preparation.

Due to the frequency-selective properties of the WT, the ratio of the intensity of the contour of an object having a smaller size to the intensity of the contour having a larger size decrease as the scale of the WT increases [15]. This pattern makes this technique insensitive to changes in the intensity of the object and allows you to take its size into account (adjust the detail). This scheme is proposed to be used to estimate depth by identifying the contours of objects with adjustable detail in the wavelet transform space (see Figure 2):

- The rows of the left frame image (see Figure 2, b) and the right frame image are convolved with wavelet functions of a given scale and contour areas are detected (see Figure 2, c);
- The distance between object contours and the corresponding disparity value are determined;
- Processing is performed along rows and columns as needed. Results are combined using a logical OR scheme;
- If necessary, the process is repeated for different scales.

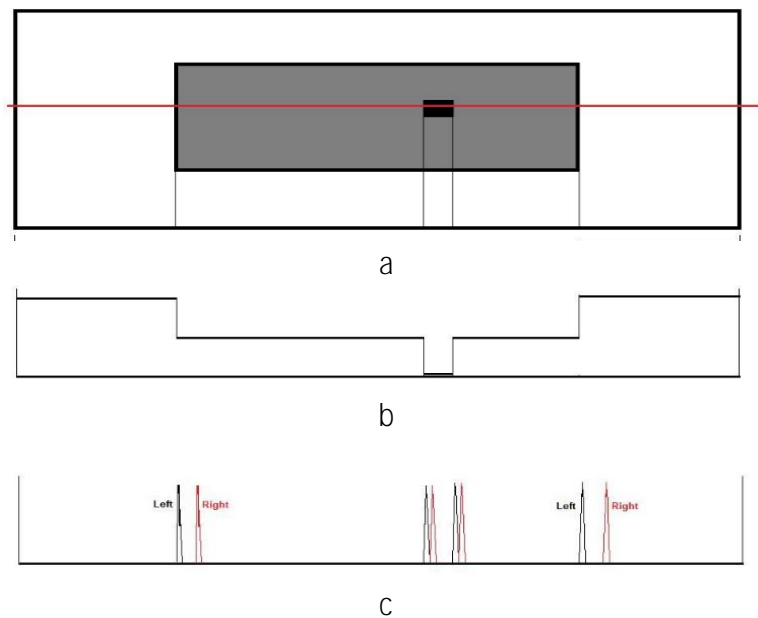


Figure 2: a - Test image, left camera; b - Fragment of intensity row, left camera; c - Wavelet transform result for the images from the left and right cameras with a wavelet function support length of  $s=1$

Research has also shown that with a wavelet function support length  $s = 1$ , both fine and coarse image details result in peaks of similar amplitude in the intensity drop area after processing.

As the scale increases, the relative size of the peaks for fine details decreases, while the amplitude of the peak at the boundaries of large-scale objects increases [15].

Furthermore, the proposed methodology allows for adjusting the level of image detail, ensuring high noise resilience, resolution of object contour extraction, and providing an approximate depth information assessment that meets the requirements for a range of tasks.

Thus, depth map construction methods can be categorized into local and global groups. Local method-based algorithms are characterized by their speed and lower accuracy, while global methods enhance accuracy but are slower.

To find a compromise between accuracy and performance, a group of methods based on optimization algorithms is actively being developed.

These methods can be used for evaluating informative features, matching fragments, and/or optimizing costs in image analysis, as well as refining discrepancies and other related tasks [11].

#### 4. Development of the Optimization Method

Analysis has shown that the use of global optimization methods based on exhaustive search leads to high computational costs, while local methods based on first and second derivative estimates exhibit low noise resilience [16, 17].

Therefore, for solving a range of optimization problems in image analysis for UVs, the wavelet transformation (WT) - based approach has been developed, which enhances noise resilience while providing sufficient accuracy for these tasks. In known optimization methods using WT, the property of gradient estimation and processing with WT changes sign as it approaches an extremum [15].

Continuous wavelet transformation is defined by convolution:

$$W(s, x_0) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(x) \Psi^* \left( \frac{x - x_0}{s} \right) dx, \quad (2)$$

where  $f$  is the function being transformed (analyzed);  $\Psi_{s, x_0}(x)$  is a two-parameter basis function derived from the mother wavelet  $\Psi_0(x)$  through scaling with a scaling factor  $s \in \mathbf{R}^+$  and translation with a parameter  $x_0 \in \mathbf{R}$ ; and  $\Psi^*$  is the complex conjugate function with respect to  $\Psi$  (where  $s$  corresponds to the width of the wavelet and  $x$  defines the position of the wavelet on the x-axis). The factor  $\frac{1}{\sqrt{s}}$  introduced for normalization.

Iterative Extremum Search Method in Wavelet Transform Space. Let's define the one-dimensional partial wavelet transform:

$$h_n = \frac{1}{\pi} \int \frac{J(a)}{c_n - a_n} da_n \quad (3)$$

where  $J(a)$  is the objective function;  $a$  is the parameter vector;  $c_n - a_n$  represents the parameters.

In known optimization methods using wavelet transform (WT), the property of gradient estimation and wavelet processing is employed, where the sign of the gradient changes as it approaches an extremum. In other words, the condition for an optimum is considered to be the equality of all partial wavelet transforms:  $\mathbf{WT}(c) = \mathbf{0}$ .

Operator

$$\mathbf{WT}(c) = (h_1, \dots, h_N) \quad (4)$$

was used for synthesizing regular iterative algorithms. Here, the one-dimensional partial wavelet transform is: Based on this approach, several authors have developed and investigated algorithms for adaptation, optimization, segmentation, classification, and clustering [16 - 18]. For instance, a regular iterative optimization algorithm in the wavelet transform space [19] is as follows:

$$c[n] = c[n - 1] - \gamma[n] \cdot \mathbf{WT}c[n - 1] \quad (5)$$

where  $\gamma[n]$  – step size;  $n$  – iteration number;  $c$  – extremum coordinate.

It should be noted that due to the direction of the search being assessed using wavelet processing, the optimization requires significant computational resources.

This limits the applicability of the method for practical UV tasks, as it is necessary to ensure optimization efficiency, especially for multimodal objective functions with low signal-to-noise ratios [16, 17].

In the known method, the assessment of the search direction using Haar wavelet functions (WF) provides robustness against noise. The choice of the Haar wavelet among a significant number of existing wavelets is due to the fact that it is characterized by low computational complexity.

When implementing computational procedures using it, multiplication operations by +1 and -1 are performed, which can simplify the hardware implementation of this procedure and increase the speed of operation [17].

However, with a large Haar wavelet function length and a high signal-to-noise ratio of the objective function, the search may deviate in the wrong direction, moving away from the global minimum towards a local minimum. This results in reduced search efficiency.

The proposed work aims to improve search efficiency by selecting the Haar wavelet function length in the search area.

To achieve this, it is proposed that after localizing the minimum area of the functional during the Haar wavelet function search phase  $\Psi_1(i)$  and narrowing down the search area by defining constraints  $g_c \leq 0$ , the Haar wavelet function length should be selected judiciously, taking into account the characteristics of the extrema in disparity evaluation.

The method with iterative constraint evaluation using Haar wavelet functions is implemented in the following sequence:

Step 1. Execute steps 1-5 of the basic wavelet optimization method, taking into account the iterative scheme [17]. The coordinates of the extremum are determined as follows:

$$c[n] = c[n - 1] - \gamma[n] \cdot WT_k(Q(x[n], c[n - 1])) \quad (6)$$

where  $\gamma[n]$  – step size;  $n$  – iteration number;  $k$  – starting index;  $Q(x, c)$  – functional vector dependent on vector  $c = (c_1, \dots, c_N)$ ;  $x = (x_1, \dots, x_M)$  – minimum coordinate at iteration  $c[n - 1]$ ;  $WT_k$  – direction of movement towards the extremum, calculated as follows:

$$WT_k(Q(x[n], c[n - 1])) = \{G_{1k}, G_{2k}, \dots, G_{Nk}\} \quad (7)$$

where  $G_{jk}$  – the result of processing by the  $j$ -th variable:

$$G_{jk} = \frac{1}{s_k} \sum_{\substack{i=-\frac{s_k}{2} \\ i \neq 0}}^{\frac{s_k}{2}} Q(x[n], c_j + ia) \cdot \Psi_k(i) \quad (8)$$

where  $s_k$  is the length of the Haar wavelet support;  $\Psi_k(i)$  is the step size of the Haar wavelet discretization;  $\Psi_{1j}$  is the Haar wavelet at the  $k$ -th start;  $j = 1, \dots, N$  is the dimensionality of the parameter vector. This approach allows for determining the range of variation in the extremum [18] coordinates with  $\delta_1$  being the search error for the optimal start (determined during the preliminary functional quality studies);  $\delta_2$  being the search error for the practical task's optimum. The search is conducted considering the following parameters:  $c[0]$  as the initial approximation to the optimum coordinate;  $\gamma[1]$  as the step size;  $a$  as the Haar wavelet discretization step size;  $s_1$  as the length of the Haar wavelet support for the first start  $\Psi_k(i)$  (determined during the preliminary functional quality studies);  $\Delta_s$  as the step size for changing the length of the Haar wavelet support when determining the range of extremum coordinates;  $k = 1$  as the start number;  $n = 1$  as the iteration number;  $A_1$  as the value of the minimal height of the registered extremum. The search continues until the sign of processing  $WT_k(Q(x[n], c[n - 1]))$  changes when estimating the direction of movement towards the extremum.

Step 2. The constraints of the search zone for the next algorithm start  $g_1(c[n])$  and  $g_2(c[n])$  are determined:

$$g_1(c[n]) \geq c^*[n - 1]; g_2(c[n]) \geq c^*[n] \quad (9)$$

where  $c^*[n - 1]$  is the extremum coordinate at step  $[n - 1]$  with the previous sign, and  $x_n$  is the extremum coordinate at step  $n$ , where the sign changed (see Figure 1).



Step 3. The length of the Haar wavelet support  $s_k$  for the subsequent start is determined by selecting from the range  $s_k = \{2,4,6\}$  according to the condition  $s_k \leq \frac{|g_2(c[n]) - g_1(c[n])|}{a}$ .

Trajectory of Minimum Search Using the Proposed Optimization Method (Step  $\gamma$ , Haar Wavelet Support Length  $\Psi_k(i)s_1 = 10$ , Haar Wavelet Discretization Step  $\alpha = 10$ ) and Search Area Boundaries Defined Using Haar Wavelet  $\Psi_k(i)$  (see Figure 3). When the SNR (signal-to-noise ratio) is reduced to 2, the proposed method's performance (measured by timer) is on average 1.1 times higher (see Figure 4a), and the relative error in finding the minimum is 1.2 times greater due to the introduced stage of selecting the length of the Haar WF carrier (Fig. 2, b). The sensitivity of the developed optimization method to local extrema and the starting point of the search using the test Schwefel function

$$f_s(x) = 418,9829 + (-x \cdot \sin \sqrt{|x|}) \quad (10)$$

The sensitivity of the developed optimization method to local extrema and the starting point of the search using the test Schwefel function has been studied. This function has a false global minimum. The function at  $x \in (-500; 500)$  has a global minimum  $f_s(x) = 0$  at  $(x = 420.9829)$ . During the research, the starting point was chosen randomly. The gradient descent method made it possible to find the minimum closest to the start, while the optimization method with the Haar wavelet function achieved the global minimum with an error  $\delta \leq 10^{-2}$  in 128 out of 150 cases.

This results in a probability of finding the extremum coordinate of 0.85. The global minimum was not found when the starting point values were chosen outside the interval  $x \in (-420; 470)$ .

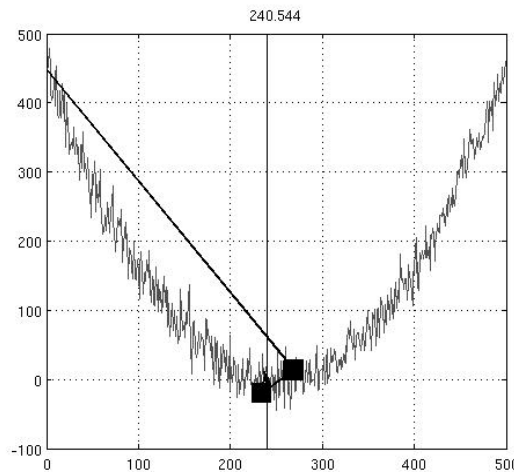
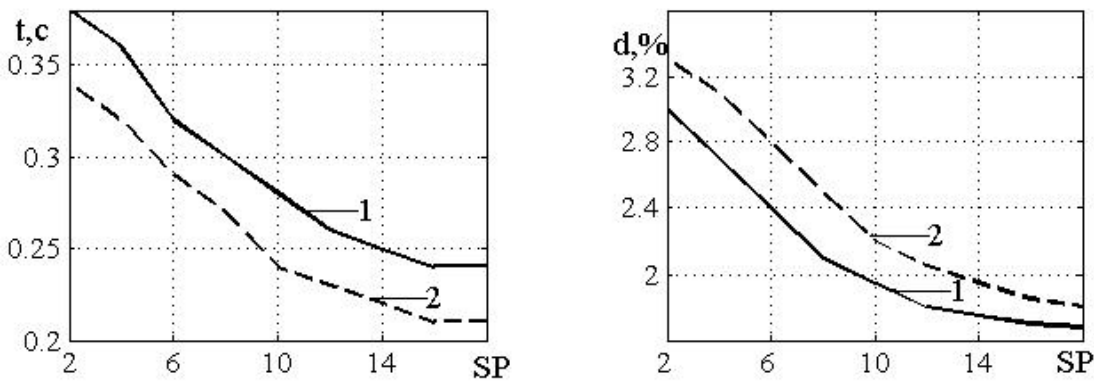


Figure 3: Constraints Identification Using Haar Wavelet Functions. The coordinates  $c^*[n - 1]$  and  $c[n]$  are marked with squares.



1 – basic method [16]; 2 – proposed method      1 – basic method [16]; 2 – proposed method  
 Figure 4: Experimental results of the basic [16] and proposed methods: execution time (in seconds) (a); relative error in determining the extremum coordinates (b).

Future research in this direction will focus on evaluating the impact of wavelet function support length on the performance of the maximum search procedure, as well as the effects of the wavelet discretization step and the step size in iterative searches with Haar wavelets ( $\gamma$ ).

## 5. Testing

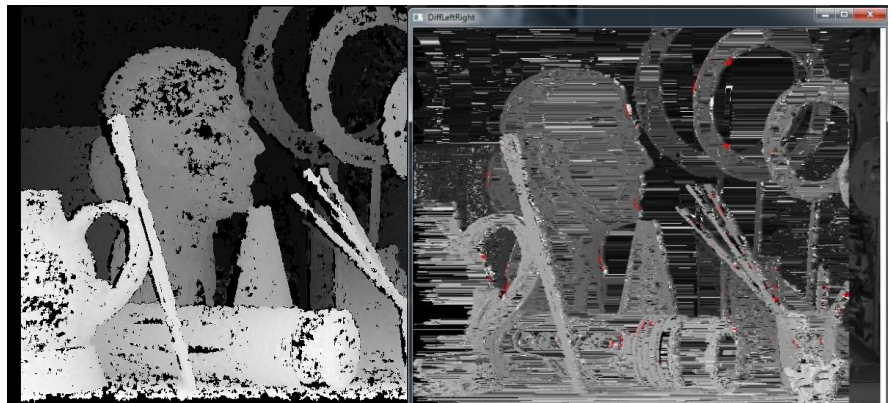
The developed optimization method on the wavelet transformation base was tested in the context of constructing a depth map from a test image in the image database [20, 21].

The length of the wavelet support used in the minimum search after investigation was chosen as 17, the wavelet discretization step was 1, and the step  $\gamma$  in the iterative Haar wavelet search was applied 0.9. The minimum image alignment error was found after 12 iterations (starting the search at [1; 1]).

Furthermore, simulation results indicated that modifying the known StereoBM method by incorporating wavelet transformation for disparity calculation enhances noise robustness, reduces error in locating the characteristic fragment on the intensity line, and also decreases energy consumption by nearly a factor of 2 (Table 1).

Table 1  
Results of Simulation

Algorithm		StereoBM	Developed
Power	Consumption,	950	620
mA			
Depth Map Obtained			



## 6. Conclusions

The paper addresses the development of approximate depth estimation methods for three-dimension computer vision in autonomous vehicles based on an improved optimization method utilizing wavelet transform. This method can be used for evaluating informative features for fragment matching and/or optimizing costs in image analysis, as well as for refining inconsistencies, among other applications. The study demonstrates possible solutions for obtaining an approximate depth map by simplifying the calculation of disparity values, traditionally used for depth map formation, within the intensity space and using image contour descriptions with adjustable detail based on wavelet transform. The advantage of the developed optimization method over existing algorithms on the wavelet transformation base is its increased efficiency due to the rational choice of Haar wavelet support length in the extremum search area.

Future research in this direction will focus on evaluating the impact of wavelet function support length on the performance of the maximum search procedure, as well as the effects of the wavelet discretization step and the step size in iterative searches with Haar wavelets.

This improvement makes the proposed method suitable for autonomous vehicles operating under constraints of computational and energy resources.

## References

- [1] A. Karnati, D. Mehta, and Ks, Manu. "Artificial Intelligence in Self Driving Cars: Applications, Implications and Challenges," *Universal Journal of Business and Management* 21 (2022): 1-28. doi:10.12725/ujbm.61.1.
- [2] I. Konovalenko, E. Kuznetsova, A. Miller, B. Miller, A. Popov, D. Shepelev, and K. Stepanyan, "New Approaches to the Integration of Navigation Systems for Autonomous Unmanned Vehicles (UAV)," *Sensors* 18 (2018): 3010. doi:10.3390/s18093010.
- [3] S. Suzuki, "Integrated Navigation for Autonomous Drone in GPS and GPS-Denied Environments," *Journal of Robotics and Mechatronics* 30.3 (2018): 373–379. doi:10.20965/jrm.2018.p0373.
- [4] H. Li, A. V. Savkin, and B. Vucetic, "Collision Free Navigation of a Flying Robot for Underground Mine Search and Mapping," in *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1102–1106. IEEE, 2018. doi:10.1109/ROBIO.2018.8665108.
- [5] P. Tripicchio, M. Satler, M. Unetti, and C. A. Avizzano, "Confined Spaces Industrial Inspection with Micro Aerial Vehicles and Laser Range Finder Localization," *International Journal of Micro Air Vehicles* 10.2 (2018): 207–224. doi:10.1177/1756829318757471.
- [6] Y. Bi, M. Lan, J. Li, K. Zhang, H. Qin, S. Lai, and B. M. Chen, "Robust Autonomous Flight and Mission Management for MAVs in GPS-Denied Environments," in: *11th Asian Control Conference (ASCC)*, 2017. doi:10.1109/ascc.2017.8287144.
- [7] S. S. Mansouri, C. Kanellakis, D. Kominiak, and G. Nikolakopoulos, "Deploying MAVs for Autonomous Navigation in Dark Underground Mine Environments," *Robotics and Autonomous Systems* 126 (2020): 103472. doi:10.1016/j.robot.2020.103472.
- [8] C. H. Vong, R. Ravitharan, P. Reichl, J. Chevin, and H. Chung, "Small Scale Unmanned Aerial System (UAS) for Railway Culvert and Tunnel Inspection," in: *ICRT 2017*, 2018, pp. 1024–1032.. doi:10.1061/9780784481257.102.
- [9] H.-W. Choi, et al., "An Overview of Drone Applications in the Construction Industry," *Drones* 7.8 (2023): 515. doi:10.3390/drones7080515.
- [10] R. A. Hamzah and H. Ibrahim, "Literature Survey on Stereo Vision Disparity Map Algorithms," *J. Sensors* (2016): 8742920:1-8742920:23. doi:10.1155/2016/8742920.
- [11] J. Du and J. Okae, "Optimization of Stereo Vision Depth Estimation Using Edge-Based Disparity Map," in: *10th International Conference on Electrical and Electronics Engineering (ELECO)*, 2017, pp. 1171-1175.
- [12] M. Richter, M. Rosenberger, R. Illmann, D. Buchanan, and G. Notni, "Suitability Study for Real-Time Depth Map Generation Using Stereo Matchers in OpenCV and Python," in: *Engineering for a Changing World: Proceedings: 60th ISC, Ilmenau Scientific Colloquium, Technische Universität Ilmenau, September 04-08, 2023*. doi:10.22032/dbt.58859.
- [13] A. Aslam and M. S. Ansari, "Depth-Map Generation Using Pixel Matching in Stereoscopic Pair of Images," (2019) arXiv:1902.03471v3 [cs.CV]
- [14] A. A. Fahmy, O. Ismail, and A. K. Al-Janabi, "Stereo Vision Based Depth Estimation Algorithm in Uncalibrated Rectification," *International Journal of Video & Image Processing & Network Security* 13.2 (2013).
- [15] S. G. Antoshchuk, S. B. Kondratyev, G. Y. Shcherbakova, and M. A. Hodovychenko, "Depth Map Generation for Mobile Navigation Systems Based on Objects Localization in Images," *Herald of Advanced Information Technology* 5.1 (2022): 11-18. doi:10.15276/hait.05.2022.1.
- [16] G. Shcherbakova, V. Krylov, V. Abakumov, V. Brovko, and I. Kozina, "Sub Gradient Iterative Method for Neural Networks Training," in: *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications: 6th IEEE Int. Workshop IDAACS'2011, Prague, Czech Republic, 15–17 Sept. 2011: Proceedings*, pp. 361-364.
- [17] W. Huan, G. Shcherbakova, A. Sachenko, L. Yan, N. Volkova, B. Rusyn, and A. Molga, "Haar Wavelet-Based Classification Method for Visual Information Processing Systems," *Applied Sciences* 13.9 (2023): 5515. doi:10.3390/app13095515.

- [18] Y. Bodyanskiy, N. Lamonova, I. Pliss, and O. Vynokurova, "An Adaptive Learning Algorithm for a Wavelet Neural Network," *Expert Systems* 22.5 (2005): 235-240. doi:10.1111/j.1468-0394.2005.00314.x.
- [19] G. Shcherbakova, H.-S. Shi, V. Krylov, N. Bilous, and S. Antoshchuk, "Estimation of the Duration of RR-Intervals of Electrocardiograms by Means of Multi-Start Optimization Based on Wavelet Transformation," in: *IEEE 9th International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, 21-23 September 2017, Bucharest, Romania. doi:10.1109/IDAACS58523.2023.10348849.
- [20] Middlebury Stereo Vision Dataset, "Full-Size Stereo Data and Scene Information," Middlebury Stereo Vision Project. URL: <https://vision.middlebury.edu/stereo/data/scenes2005/FullSize/Art/Illum1/Exp1/>.
- [21] Middlebury Stereo Vision Dataset, "Stereo Data Archive," Middlebury Stereo Vision Project. URL: <https://vision.middlebury.edu/stereo/data/scenes2005/>.