[1]**M. P. Vasylenko,**
[2]**O. S. Sych**

# IMAGE DEPTH DETECTION SYSTEM

Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and
Telecommunications, National Aviation University, Kyiv, Ukraine
E-mails: [1]m.p.vasylenko@nau.edu.ua     ORCID 0000-0003-4937-8082, [2]sychanna33@gmail.com

*Abstract*—*The article will be intended to explain the Image Depth Detection System, which in turn should perform all the necessary functions for detecting objects on a 3D scene, obtaining information about the sizes of bodies, and determining distances to certain bodies. Since such systems already exist in the modern world, the main goal was to create a model that meets the requirements regarding price-quality, since systems of this type currently on the market have a high cost price. Also, the improvement of the method of determining the depth of the image in this system was created thanks to the analysis of filtering methods available in the Matlab environment. An analysis of operation in different conditions of use, i.e. in the presence of external disturbances, such as daylight on the street, has been created.*

**Index Terms**—Stereo vision; disparity map; depth map; calibration; rectification; image filtering.

## I. INTRODUCTION

This article investigates the impact and quality of 3D scene construction in imaged rather than streamed video, under different weather conditions and at different observation points, which makes it possible to feel in more detail the effect of various phenomena on the model itself during operation and to improve accuracy by considering a single pair of images, and not a large amount of flow at a certain frequency.

The design consists of two cameras, which were selected from the principle of price-quality, and a box to fix and protect the model itself, thus forming protection from the environment in various conditions of use. The design is connected to a computer that performs the software part, which consists in creating a stereo pair - artificial adjustment of cameras, image analysis at the initial stage and after filtering, which as a result gives an opportunity to see the difference in the accuracy of constructing a 3D image, which can be used for various goals, for example to find out the size or distance to the target object.

## II. PROBLEM STATEMENT

Computer vision [1] is an area of artificial intelligence (AI) that enables computers and systems to extract meaningful information from digital images, videos, and other visual data, and to take actions or make recommendations based on that information. If artificial intelligence allows computers to think, then computer vision allows them to see, observe and understand.

Computer vision works just like human vision, except humans are one step ahead. Human vision has the advantage of being able to learn to distinguish objects, determine their distance, know if they are moving and if there is something wrong with the image.

Computer vision teaches machines to perform these functions, but it must do so in much less time with cameras, data, and algorithms than with retinas, optic nerves, and visual cortex. Because a system trained to inspect products or monitor a manufacturing asset can analyze thousands of products or processes per minute, spotting subtle defects or problems, it can quickly overwhelm human capabilities.

Computer vision is used in a variety of industries, from energy and utilities to manufacturing and automotive, and the market continues to grow.

In other words, computer vision, which is also a field of machine learning, studies the technologies and tools that allow computers to perceive and interpret information from the real world. AI gives machines cognitive ability, and cameras allow them to see. Together, they form computer vision that allows you to observe and understand.

Taking into account all of the above, a system of image depth research [2] was developed, with the help of which it is possible to find out the distances and sizes of this or that object. Currently, many analogs of this technology are used, but usually the price of such a system exceeds the expected result. The developed installation primarily performs the function of recognizing the position of the object in space. Let's highlight standard examples of recognition tasks.

- *Content image search:* a specific unit with a defined content is selected from the array based on certain characteristics. The nature of the content can

have different interpretations and selection parameters: similarity parameters, multi-criteria approach in text format, etc.

- *Position estimation:* determining the orientation of the object according to the position of the sensor, computer vision information reading camera. For example, determining the orientation of a product in a warehouse or on a conveyor line by a robot.

The problem of motion characterization is also popular. The speed of movement of certain points and their trajectory is determined on the sequence of images. The following tasks are solved:

- monitoring of camera movement;
- tracking the movement of the object.

In addition to object recognition and detection, computer vision can reconstruct images.

Image restoration can be interpreted as cleaning unnecessary noises (results of sensor defects, blurring, etc.). The simplest and most common solution to this problem is the application of certain filters of different frequencies. There are also more complex methods of solving this problem, based on the approach of determining "how it should be" and comparing it with "how it is". A more complex process of noise removal [3] is based on the analysis of input data, the determination of certain structural elements, and then the use of more highly specialized filters based on the received information. Features of the practical implementation of computer vision systems strongly depend on the structure and perform an auxiliary or intermediate function. The use of these systems can also vary depending on whether the tasks of the system are predetermined or can be supplemented or changed during its operation. field of application. Systems can be separate functional solutions, or be part of a more complex.

## III.   CAMERA CALIBRATION

The cameras must be positioned parallel to the structure's *X* and *Y* axes in order to get an appropriate result. For this, much as the entire project, the code was created based on the MATLAB program. Algorithm for showing parallel axes from video cameras step by step:

1) Use the proper equipment to enter the program's required number of inputs.

2) Create an unending image capture.

3) Register streaming images for pre-showing.

4) Design a feature that allows for the on/off of image capturing.

5) The center axes for the left and right cameras are presented in a cycle for processing photos with a delay of half a second.

6) Manually align the central axes as parallel as it can be done.

7) Put the program to an end.

An illustration of this may be seen in Fig. 1, where the cameras must be manually adjusted and the axes are depicted with red lines.

After achieving the artificial centering of the cameras, software calibration of the cameras is carried out.

A template calibration approach [4] was used in this work. The list of actions below explains the calibrating procedure.

*1) Figuring out the correct world coordinates for 3D template points.*

The checkerboard cells' corners are 3D points. As the starting point for the global coordinate system, you can choose any corner of the board. The $Z_w$ axis is parallel to the board and perpendicular to the $X_w$ and $Y_w$ axes. As a result, $Z_w$ 0 and all chessboard points are on the *XY* plane.

A checkerboard with given dimensions is shot in various orientations to create 3D points (Fig. 2). The chessboard is connected to the global coordinate. All corner points are on a plane, therefore the observer can select $Z_w$ at random for any point that is equal to 0. The chessboard's points are evenly spaced apart, therefore it is simple to calculate the coordinates ($X_w$ and $Y_w$) of each 3D point by choosing one point as the origin (0, 0) and measuring the other points in relation to it.
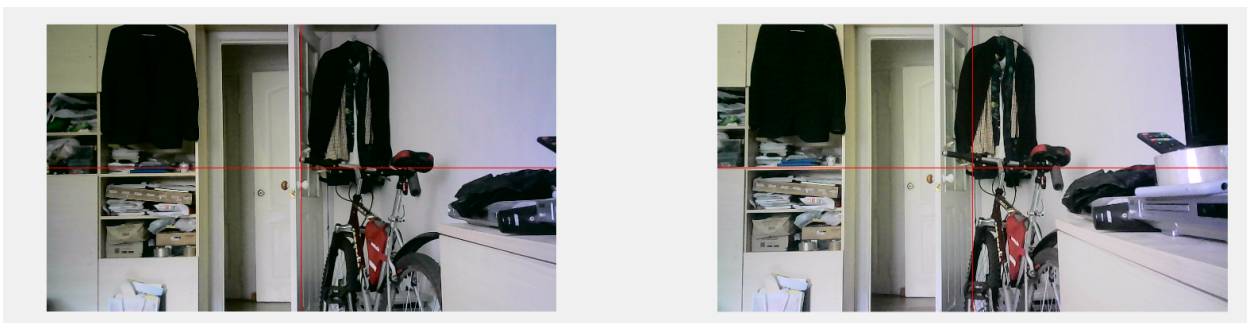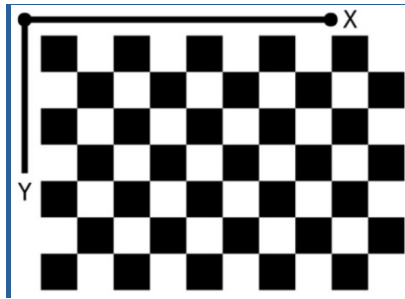


Fig. 1. Central axes on two cameras

Fig. 2. World coordinate system: A pin-attached checkerboard pattern represents world coordinates

The graphic makes it simple to identify and locate the checkerboard squares. The checkerboard square corners are also perfect for localizing them because they have steep slopes in both directions. The junction of the checkerboard lines is also where the corners are located. The accuracy of the search for the corners of staggered squares is determined by all of these variables.

*2) Take pictures of the template from various angles.*

Reading the pictures from the left and right cameras and then creating a cycle with the goal of obtaining 50 photos as the final condition. A time delay, which can be adjusted to meet your needs and requirements, was introduced to allow you to adjust the position when capturing a photograph. Consequently, two pictures were shot at once (Fig. 3).
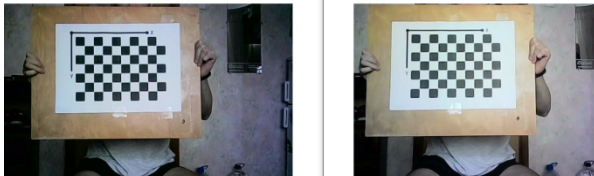


Fig. 3. The left and right photos being captured simultaneously

*3) Using the app for calibrating stereo cameras.*

MATLAB software component calibrates stereo cameras so it may be used to capture 3D sceneries, collect depth information from images, and other tasks.

The user must first input photos from the left and right cameras, as well as the precise dimensions of the checkerboard squares, to the interface before beginning work. The distance is then measured with millimeter-level accuracy. The examples that work may then be examined, and the radial distortion by 2 and 3 coefficients computation, as well as the skew and tangential distortion parameters, can be included. In the context of the work that was done, the skew and tangential distortion parameters were selected as the two coefficients. Figures 4 and 5

show the outcome; during processing, certain frames were eliminated because their effects had a detrimental impact on the average inaccuracy per pixel.
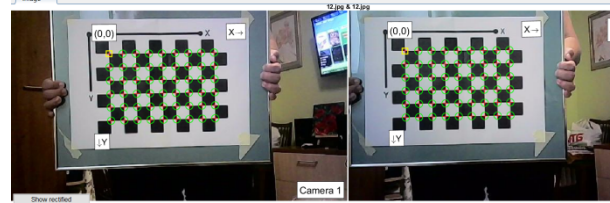


Fig. 4. The coincidence of the beginning points and axes of the template

*4) Outputting the data that was received to the script code.*

In order to avoid having to do it manually every time the "Load" command is written in the program code, which loads all the used data, all incoming data will be added to the script code, which must be stored beforehand and manually invoked.
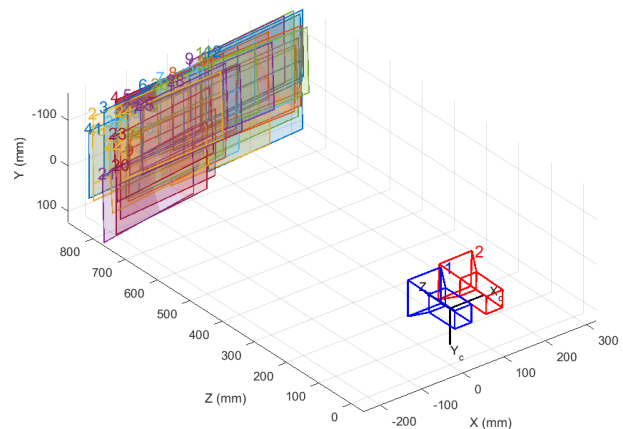


Fig. 5. Location of the cameras in space and according to each other

## IV. RECTIFICATION OF STEREO IMAGES

Image rectification (Fig. 6) is the process of combining two image planes into one plane so that all of the epipolar lines are perpendicular to the abscissa axis and have the same ordinates in both pictures [5]. Then, using the previously prepared stereo parameters, taking into account bevels, transformations, and the like, undistorted and corrected versions of the input images are returned. Typically, this sub-clause addresses the problem of camera placement errors by making the cameras' focal positions parallel.

It is important to note that because of this feature, the 3D impression is already produced. If the spectator takes special 3D glasses, will see the image in three dimensions, just like being watching a movie in a theater.
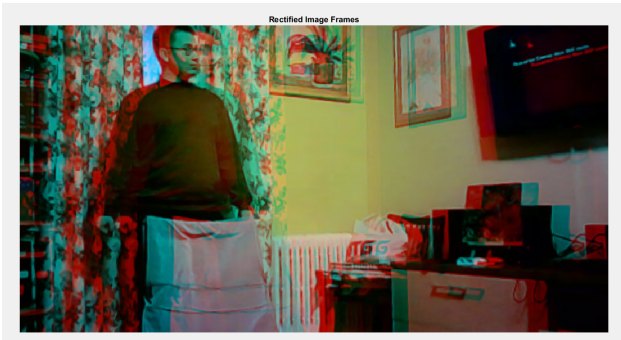
Fig. 6. Rectified Stereo Image

The final rectified image is reached with some image cropping. This is because, when trying to correct images, an appropriate position and angles of rotation of the camera are pursued, but black zones or holes in the plane appear and have a negative impact on subsequent work; as a result, the image is cropped to the size that is accurate (valid parts).

## V.    DISPARITY MAP

The disparity map [6] cannot be used with a color picture; instead, it operates on grayscale, which is the color mode of images that are shown in a table as standards of white brightness. This necessitates the conversion of RGB to Gray (grayscale). Most frequently, a stepped picture of a consistent succession of neutral-gray field optical densities is employed.

Disparity Map only displays a basic colored analog of an object's position [7] at a distance at this point. In other words, regardless of the position, the item is highlighted in a single hue according to the distance, i.e., the closer, the lighter its shade will be (Fig. 7).
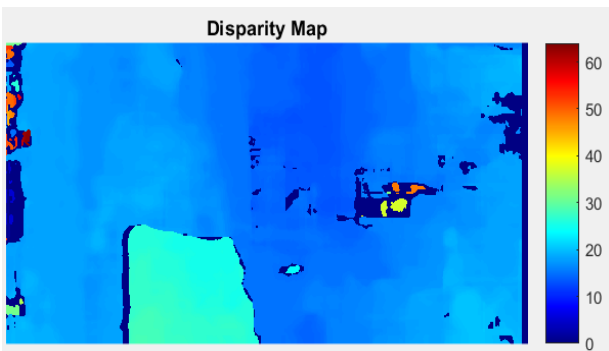


Fig. 7. Colored disparity map

## VI.    DEPTH MAP

A depth map is the end product. In reality, the relationship between the disparity values and pixel offset is inverse. The link between disparity and depth [8] may be stated as follows using the notation from the left side of Fig. 8.

$$\frac{T - d}{Z - f} = \frac{T}{Z} \rightarrow Z = \frac{f \cdot T}{d}. \qquad (1)$$
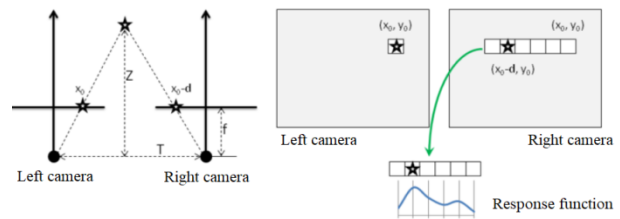


Fig. 8. Basic principle of depth map construction

The resolution of stereo vision systems that operate on this principle is better at close distances and worse at vast distances because of the inverse connection between depth and displacement.

Scene reconstruction is done by attaching stereo parameters to the disparity map; in other words, if a typical disparity map model is received, it is used to produce a scene since it already has all of the world coordinates for the points that are saved in the picture. The parameter input in this situation must be the same as that used for correction, accounting for any stereo image adjustments that were made prior to the matching disparity maps. The data must then be recast in meters because the initial data was stored in millimeters after receiving the data with each point's depth.

Volumetric representations [9] are filled simply to accomplish the display task, and this allows for numerous operations like increasing rotation and grabbing a point. The cameras ought to be perpendicular and not experience focus shifts during recording. Example of indoor use are depicted on Fig. 9.
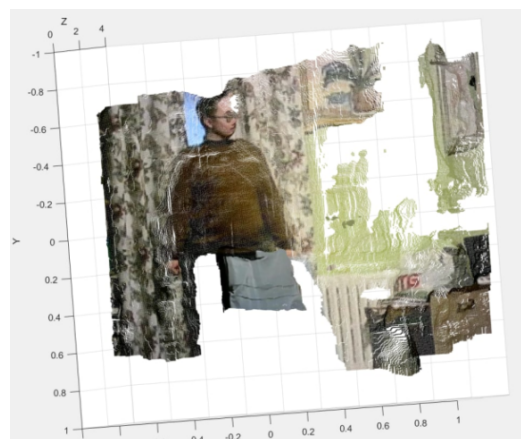


Fig. 9. Indoor use of the system

When the scene rotates, it is possible to see a three-dimensional image with cutouts in the areas where the bodies are obscured by one another or are not in the field of view, making it easy to determine the approximate distance to objects from these situations.

## VI. Image Filtering

If a number of shortcomings are observed during the shooting regarding lighting, light reflection from the surface, which directly reduces the image display result, there is a need to create internal filtering of the image before sending it for processing.

For this, the built-in image filtering functions in Matlab will be used and the best kind of filtering that can give the best result will be selected.

Since during experimental research it was found that the best accuracy results are characterized by the Gaussian and median filters, it was decided to combine them into one and find out whether this affects the final result.

The Gaussian blur filter is quite often used by itself or as part of other image processing algorithms. Next, a method will be described that allows to obtain blur with a speed that does not depend on the blur radius, using filters with an infinite impulse response.

So, let the original image be set to brightness $x(m, n)$. Gaussian blur with radius $r$ is calculated by the formula:

$$y(m,n) = \frac{1}{2\pi r^2} \cdot \sum_{u,v} e \frac{-(u^2 + v^2)}{2r^2} x(m+u, n+v). \quad (2)$$

The summation limits of $u$ and $v$ can be chosen plus or minus a few sigma, i.e. radius $r$, which gives the complexity of the algorithm of the order of $O(r^2)$ operations per pixel.

The median filter is the most significant interference suppression encountered in the development of this project. It is just right for eliminating all kinds of small scattered inclusions, but it is not an ideal tool because passes slightly larger areas of concentrated interference.

The median filter is a 3×3 pixel sliding window. At the input, it acquires 9 values (pixels), and at the output it gives one. The median filter works as follows: it sorts the input data (pixels) in ascending order and outputs the middle result (median).

The study confirmed that combining these methods with a sigma equal to one in the Gaussian filter gives the best result (Figs 10 and 11) and will be used in the future.
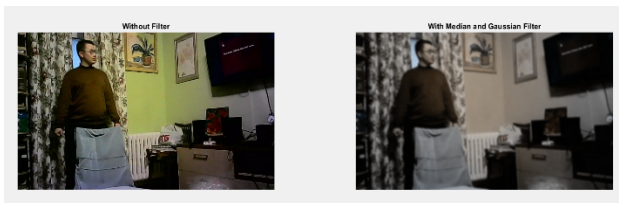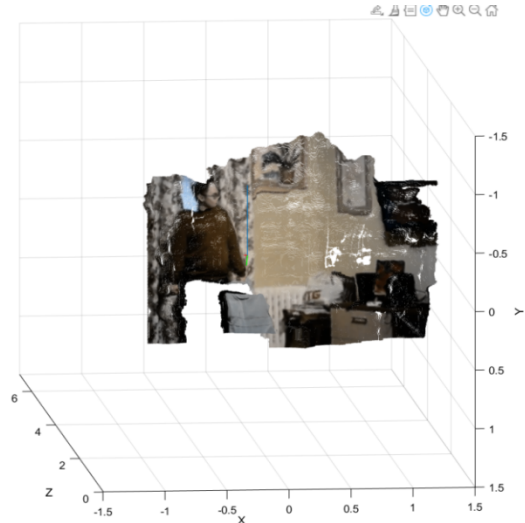


Fig. 10. Image with use of combined filter



Fig. 11. 3D scene using a combined filter

## VII. Model use on the External Environment

Now it will be considered what results the system gives when observing objects on the street. In the course of the research, a number of shortcomings were revealed, which are difficult to get rid of. Since most of the bodies located outside do not have a stable or reflective surface [10], the incorrectness of the initial data was found. For example, if it is necessary to display objects that are at a great distance, it is necessary to change the design model, since the distance between the camera lenses is directly proportional to the distance of the observed object, if it is necessary to find out the conditional distance to a body that is more than 20 meters away, it is necessary to anew recalibrate cameras with a pre-increased distance between optical lenses, the same algorithm occurs in the opposite case.

The object of observation was a car that was located at a distance of 3.20 meters.

For understanding, the initial image and the image after filtering are shown (Fig. 12).
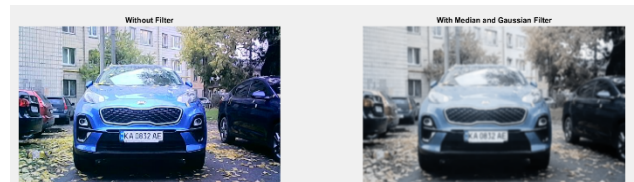


Fig. 12. Initial and filtered view of the image for further processing

Below is a 3D scene of the car (Fig. 13) and the surrounding objects, it should be noted that the covering of the car created significant obstacles for accurate reflection and as you can see in the image, some of the surfaces were completely cut off, because the mirror principle was formed, and the focal lens could not determine the point of the object.
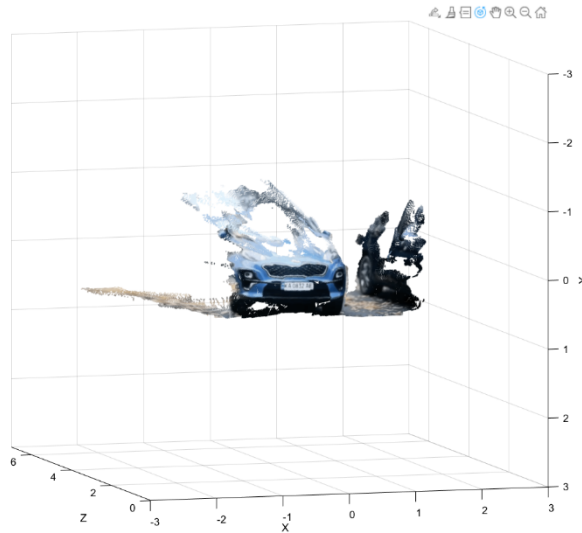
Fig. 13. 3D scene of the environment

It can also be seen that the shooting took place in a triangular shape, because of this, some of the objects could not be reflected, because the focal distance was not defined for them, for example, the rear part of the right neighboring car was well reflected, and the front part was not. From the received data, it can be noted that the cameras worked well at a distance of up to 5 meters, then the image began to cut off.

## VIII. CONCLUSIONS

During investigation, it was established that a useful stereo pair could be made for assessing the surroundings and creating 3D landscapes utilizing the MATLAB software environment and two web cameras.

The method's adjustment, namely calibration, getting a good 3D model, etc., are its key drawbacks. Another challenge was building a solid platform that would allow the cameras to be moved around and their location adjusted in relation to the axes.

Additionally, subtle differences were seen between various areas, projective planes, and the graph itself. Because the light is not steady, some glare, body transparency, or vice versa, specularity, are possible. Mirrors, which were excluded from consideration since the cameras couldn't comprehend and match positions with the actual world and things in the mirror didn't have a spatial location, revealed an important fact.

Given that the setup is made of inexpensive equipment and materials, the result can be viewed as being quite successful.

Under the condition of tracking outdoor objects, the system showed that with increasing distances the coefficient of relative error increases, and the system becomes unusable, but this problem can be solved with proper adjustment of the structure itself and calibration with adjustment for long distances.

## REFERENCES

[1] Where is computer vision used today. *World Trade Center.*URL:https://corp.wtcmoscow.ru/services/international-partnership/actual/gde-segodnya-primenyaetsya-kompyuternoe-zrenie/ [in Russian].

[2] L. A. Kotyuzhansky Calculation of the depth map of the stereo image on the GPU in real time. Basic research. no. 6–2, 2012, pp. 444–449. [in Russian].

[3] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. Journal of Computer Vision 47.* April-June 2002, pp. 7–42.

[4] A. T. Vakhitov, L. S. Gurevich, and D. V. Pavlenko, "Review of stereo vision algorithms," *Stochastic optimization in computer science*, vol. 4, no. 1-1, 2008, pp. 151–169 [in Russian].

[5] E. S. Ilyasov, "Calculation of the distance to the observed object from the images of the stereopair," *Young scientist. International scientific journal*, no. 14(118), 2016, pp. 146–151. [in Russian].

[6] I. Cabezas, M. Trujillo, "A Non-linear Quantitative Evaluation Approach for Disparity Estimation," *in: Proc. Intl. Joint Conf. on Computer Vision and Computer Graphics Theory and Applications*, 2011, pp. 704–709.

[7] H. Hirschmuller and D. Scharstein, "Evaluation of Stereo Matching Costs on Images with Radiometric Differences," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2009, 1582–1599. https://doi.org/10.1109/TPAMI.2008.221

[8] Richard Hartley and Andrew Zisserman, Multiple View Geometry in Computer Vision. Second Edition pp. 178–193, pp. 458–493.

[9] Maryna Mukhina, "Comparison of error metrics in matching algorithms of images by surf detector," *Proceedings of the National Aviation University*, 2014, no. 4, pp. 128–132. https://doi.org/10.18372/2306-1472.61.7603

[10] G. J. Iddan and G. Yahav, "3D imaging in the studio and elsewhere," *Proc. SPIE,* vol. 4298, 1994, pp. 48–55.

**Vasylenko Mykola.** ORCID 0000-0003-4937-8082.
Candidate of Science (Engineering). Senior lecturer.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: Kyiv National University of Technologies and Design, Kyiv, Ukraine, (2012).
Research interests: renewable energy sources, thermal noise based estimation of materials properties.
Publications: more than 20 papers.
E-mail: m.p.vasylenko@nau.edu.ua

**Sych Oleksii.** Bachelor.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2020).
Publications: 2.
E-mail: sychanna33@gmail.com

**М. П. Василенко. О. С. Сич. Система визначення глибини зображення**

У даній статті розглянуто систему визначення глибини зображення, яка в свою чергу повинна виконувати всі необхідні функції для виявлення об'єктів на 3D сцені, отримання інформації про розміри тіл і визначення відстаней до певних об'єктів. Оскільки подібні системи вже існують у сучасному світі, головною метою було створити модель, що відповідає вимогам ціна-якість, оскільки системи такого типу, які зараз представлені на ринку, мають високу собівартість. Також удосконалення методу визначення глибини зображення в цій системі було створено завдяки аналізу методів фільтрації, доступних у середовищі Matlab. Проаналізовано роботу в різних умовах експлуатації, тобто за наявності зовнішніх перешкод, наприклад денне світло на вулиці.
**Ключові слова:** відеоспостереження; автоматизація; ключові точки; виявлення характерних точок; контур; виявлення контуру.

**Василенко Микола Павлович.** ORCID 0000-0003-4937-8082.
Кандидат технічних наук. Старший викладач.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Факультет аеронавігації, електроніки і телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Київський національний університет технологій та дизайну, Київ, Україна, (2012).
Напрям наукової діяльності: відновлювальні джерела енергії, оцінка властивостей речовин та матеріалів за власними електромагнітними випромінюваннями.
Кількість публікацій: більше 20 наукових робіт.
E-mail: m.p.vasylenko@nau.edu.ua

**Сич Олексій Сергійович**. Бакалавр.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Факультет аеронавігації, електроніки і телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2020).
Кількість публікацій: 2.
E-mail: sychanna33@gmail.com