UDC 628.972(045)
DOI:10.18372/1990-5548.67.15579

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

IMAGE DEPTH EVALUATION SYSTEM BY STREAM VIDEO

[1,2]Aviation Computer-Integrated Complexes Department, 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 paper considers the method of estimating the depth of streaming video. An algorithm for obtaining a depth map using the method of image separation is proposed, which can be used in various fields of technology and industry to determine the object and calculate the distance to it. The debugging algorithm and the process of its adaptation to specific used external devices and software have been developed. Two Urchin Tracking Module Webcams (SJ-922-1080) were used for the experimental setup with the following characteristics: video resolution – FullHD (1920x1080), sensor – complementary metal-oxide-semiconductor, field of view – 90°, autofocus, frame rate per second – 20. Developed program code for these cameras in the MatLab environment and its adaptation algorithm for any other cameras of similar resolution. An experimental study of the algorithm.*

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

I. INTRODUCTION

Videos and photos are closely intertwined with our lives. Almost every mobile phone is equipped with a camera. Almost every camera can record video. 3D graphics are ubiquitous. With the development of possibilities, the need for "cheap" construction of 3D scenes increases. The most obvious of these methods is stereo vision – obtaining a three-dimensional picture of the world from a video sequence or several images.

Currently, the consumer has access to stereoscopic 3D image display technologies. At the same time, the number of available devices that allow consumers to use 3D image content is still extremely limited due to the high cost of acquiring and processing stereoscopic content. Consumer electronics companies appear to have strong demand for technology that can automatically convert existing 2D images to stereoscopic 3D in real time or near real time via consumer display devices. The main problem facing 2D to 3D conversion methods is the misconception that given 2D image information, several different 3D configurations can be obtained. In particular, automated 2D to 3D conversion algorithms operate based on image characteristics such as color, position, shape, focus, tint, and motion. They do not perceive "objects" within the scene as the human eye does. Optical motion analysis provides the most promising video analysis techniques that have led to the development of structure-of-motion techniques.

Such methods and algorithms are known but their practical use requires some additional research directed to adapt them to the exact task.

II. REVIEW OF EXISTING METHODS

A depth map [1], [5] is an image where for each pixel, instead of a color, its distance to the camera is stored.

In computer 3D graphics and computer vision, a depth map is an image or image channel containing information about the distance of the surfaces of objects in a scene from a point of view.

An image depth map contains information about the distance between various objects or parts of objects represented in a given image. This information can be useful in many areas.

1) Creating 3D sensors. They are able to build a three-dimensional picture of their environment, are used to orient the autonomous robot in space.

2) For systems that use augmented and virtual reality technologies. For example, cameras that capture user actions in video games using virtual reality technology.

3) In unmanned vehicles, which also use depth maps for road orientation.

4) For photo processing. For example, depth maps are used to blur the background in a photo so that the person stands out more clearly.

There are several methods [3] how to achieve this goal, namely:

- building with special depth chambers (ToF chambers, Structured light chambers);
- depth map building on a stereopair;
- using neural networks.

At the moment, active and passive methods of recovering information about the depth of a real scene are known. Active methods use ultrasonic transducers or laser illumination of the workspace to

provide fast and accurate depth information. However, these methods have limitations with respect to the measurement range and the cost of the hardware components. Passive methods based on computer vision are usually implemented with simpler and more inexpensive distance sensors. Such methods are capable of generating depth information from the obtained pair of images and the parameters of the two cameras.

III. Problem Statement

It is necessary to develop system that will allow to obtain the information about the distance to surrounding objects without the limitations of active sensors.

To achive this we must solve the following problems:

- determine the type of used sensor and method that will be used;
- develop the structure of the system;
- choose the main components of the system;
- develop the software part of the system.

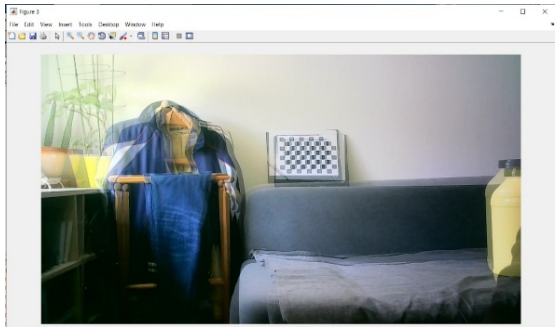
IV. Problem Solution

Since the main task was to develop a more acceptable result in terms of quality and price, the second method (passive) was chosen. This required two Urchin Tracking Module (UTM) Webcam models (SJ-922-1080) with FullHD video resolution (1920x1080) which were used as sensors. They were selected with the requirement for specific tasks due to their characteristics and acceptably low price. Since the task at hand requires stable settings for these cameras, a fixed link was made, which will be a very important moment for calibration, the rest of the work was done using a (personal computer) PC and the MatLab application package installed on it for solving technical computing problems, in which it was written code for calibrating cameras, calculating rectification and stereoanaglyph of images, filtering streaming video, calculating disparity and final moment, output image of a depth map. Further, it is shown step by step how it was designed and what was obtained during development.

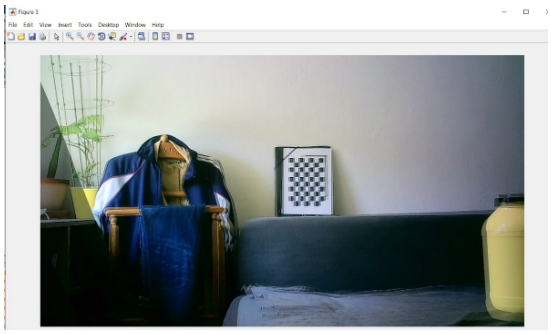
V. Camera Calibration

One of the most important points in the creation of this algorithm is the setting of two stereo cameras [2], since if the setting is not correct, the results will not give a meaningful answer.

Difficulties in using this method are in correct installing the two cameras: the axes of the cameras must be parallel to each other, as well as perpendicular to the line connecting the centers of the cameras. Due to improper installation of cameras, a very significant measurement inaccuracy can arise (a difference of one degree can lead to an error of more than two times). To reduce the error, it is proposed to increase the base to a distance of the same order of magnitude as that of the measured plane. This is solved by superimposing images from two streaming cameras, as a result of which the object is exposed at a certain distance and brought to the maximum convergence along the *X*, *Y* axes [8] (Fig. 1).



a)



b)

Fig. 1. Image overlay: (a) *X*-axis; (b) *Y*-axis

Calibration of cameras is usually performed by multiple photographs of a certain calibration template [6], it is easy to select key points on the image, for which their relative positions in space are known. Further, systems of equations are compiled and solved (approximately) that connect the coordinates of projections, matrices of cameras and the position of the template points in space. Thus, a checkerboard-like pattern was chosen, which should not be square. One side must contain an even number of squares and the other side must contain an odd number of squares. Therefore, the template contains two black corners along one side and two white corners on the opposite side. This criteria allows the application to determine the orientation of the template. The calibrator assigns the longer side to be the *x*-direction (Fig. 2).

It is necessary to measure one side of the checkerboard square, in current test it was 20 mm. The size of the squares can vary depending on the printer parameters, in theory, as the size of the

template is increased, the quality of the calibration will be improved, since the points of convergence of the outline of the squares will be at a longer distance.

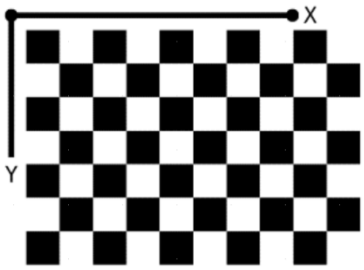


Fig. 2. Checkboard template

The main point of the algorithm is to find common points of the template, namely the extreme points of the square (Fig. 3), then the images are sorted independently by the found points for camera 1 and camera 2, the average overlay error is calculated and a visual representation of finding the study plane and video recorders. Below is shown Camera parameters configuration process through the Matlab environment is shown in (Fig. 4).

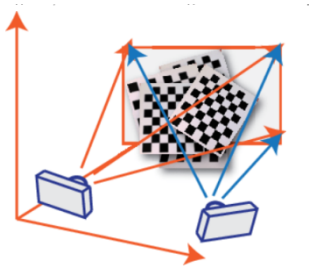


Fig. 3. General view of the system

This algorithm is useful because using a single template is necessary to investigate all the image by moving the pattern around the size of the window and preferably at different angles, but not more than 45 degrees as there will be distortion of the template object.

After the calibration processes, the data is imported for further processing, the data after this stage is very important because from them it is learned the focal length, the parameters of both cameras, movement, rotation along the axes, the number of convergence points, skew and different mismatches in both images.

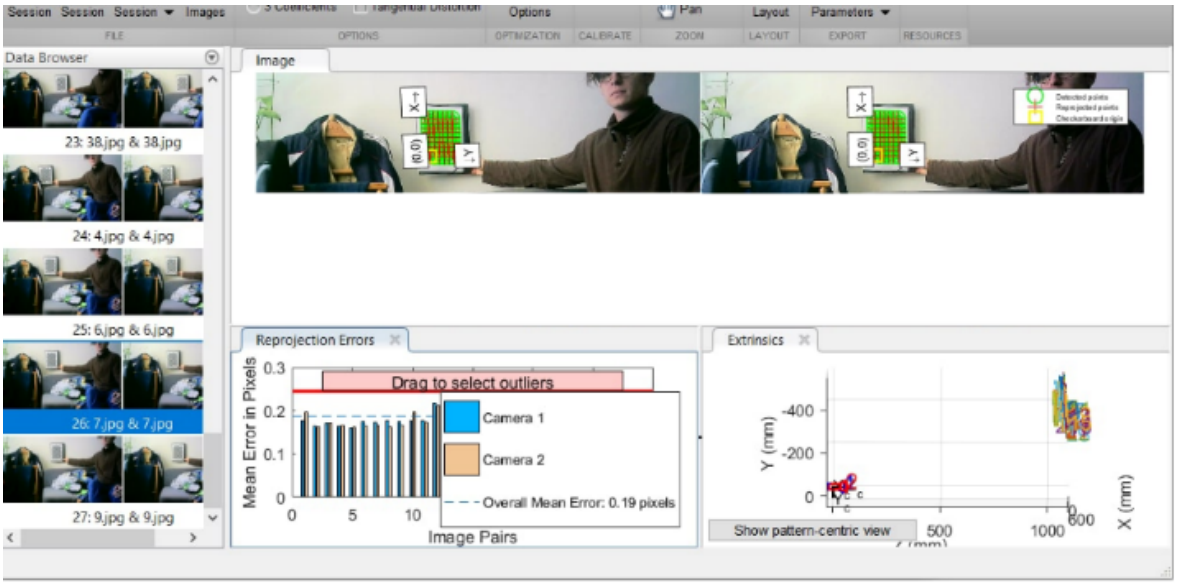


Fig. 4. Example for setting parameters

VI. Rectification of Stereo Images

Process of aligning images is called rectification. It is usually performed by remapping the image and is combined with getting rid of distortions. Since the input, although a calibrated image comes in, but this does not mean that it is aligned along the *Y*-axis absolutely accurately, because of this, a consistent improvement of the parameters under study is applied, if you pay attention, just to the basic formula for calculating the distance to the object [4], you can be sure of this.

Distance to the object is calculated using folowing equation:

$$D = \frac{fd}{x_1 - x_2}, \quad (1)$$

where D is the distance to the object; f is the focal length of the cameras; x_1 and x_2 are the coordinates of the projections on the left and right images. This means that the convergence points converge as much as possible in height. Therefore, the corrected stereo image projects the images onto a common image

plane in such a way that the corresponding points have the same row coordinates (Y). This projection of the image makes the image appear as if the two cameras were parallel (Fig. 5). Using the disparity function for calculating the disparity map from the corrected images, a normalized reconstruction of a three-dimensional scene is obtained [7].

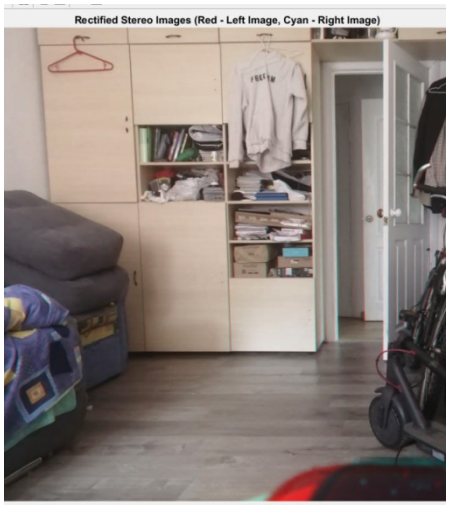


Fig. 5. Rectified Stereo Images

Obviously, the closer the object is located, the harder it is to align it, in the lower right corner (Fig. 5) such is the part of the table where the cameras are located, nevertheless this image is a stereo anaglyph, so 3-D glasses can be used to see the stereo effect.

VII. Disparity Map

Disparity Map is when two images (stereopair) are compared, in which it is known in advance that for any point of the first image it is necessary to find the corresponding point on the second image, but they must be searched for on a certain straight line (even a segment, i. e. an epipolar line), for this, the previous setting was made. As a result, we get for each point of the first image – the distance from the beginning of such a segment to the corresponding point on the second, after which the map compiled in this way is called the Disparity map.

More precisely, after the images are rectified, a search is performed for the corresponding pairs of points. The easiest way is as follows. For each pixel of the left picture with coordinates (x_0, y_0), a pixel is searched for in the right picture. It is assumed that the pixel on the right picture should have coordinates $(x_0 - d, y_0)$, where d is a value called disparity. The search for the corresponding pixel is performed by calculating the maximum of the response function, which can be, for example, the correlation of the neighborhoods of the pixels. The result is a disparity map.

It is worth noting that this is usually a black and white texture, and the values in it are used to determine the height of each point of the object's surface (values can be stored as 8-bit or 16-bit numbers), but for a more acceptable understanding of finding objects, a color filter was applied (Fig. 6).

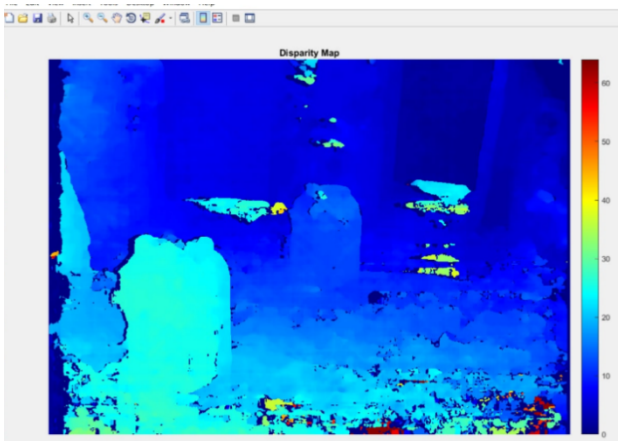


Fig. 6. Disparity map

It is worth noting that the closer the object is, the brighter the color becomes, some glitches are noticeable in the picture, but this can be solved when applying filters for the streaming image itself at the previous stages of processing, the image requires noise removal, light glare can negatively affect the resultant accuracy.

VIII. Depth Map

Actually, the depth values are inversely proportional to the amount of pixel disparity. Calsulations for each point of the image are performed using equation (1). Visual representation of the result is shown in (Fig. 7).

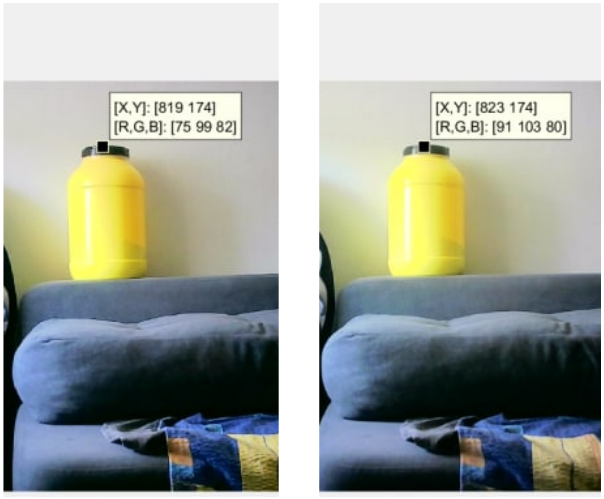


Fig. 7. Example

Having two images available, which are aligned along the axis and knowing the parameters of the cameras, a calculation is made for each of the points *X*_left_image *X*_right_image. Finally, get a point cloud [9], which returns an array of three-dimensional coordinates of world points (*X*, *Y*, *Z*), which reconstructs the scene from the disparity map. The StereoParams input must be the same input you use to correct stereo images that match the disparity map. In the example with a room (Fig. 8), it builds the internal structure of all found objects, there are some nuances about finding small bodies, but this already depends on how the cameras are configured for a specific task, or rather what type of dimensions should be looked for. It should also be remembered that when receiving a depth map, the so-called "refuse" appears when detailing bodies, while the layers of world points are demolished, this is corrected by truncating along the axes and removing these points from the main depth map. This requires downsampling the data using a box grid filter and setting the grid filter size to 10 cm. The grid filter divides the point cloud space into cubes. The points inside each cube are combined into one output point by averaging their *X*, *Y*, *Z* coordinates.

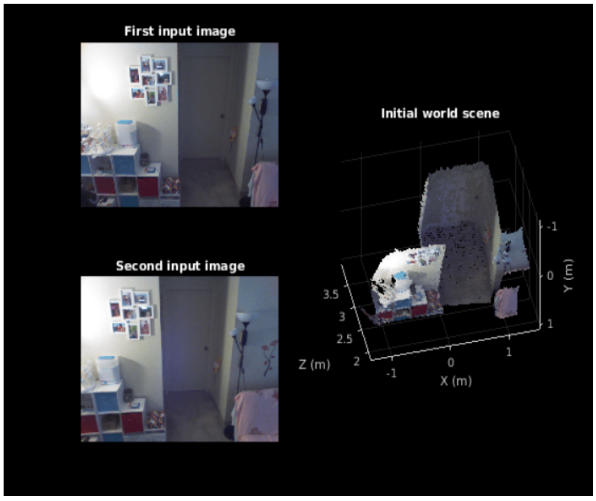


Fig. 8. Disparity map of the room

If there is a need to detect specified type of object a function to find the centroids must be written according to the specified parameters.

IX. CONCLUSION

Proposed system allows to measure the distance to certain objects by calculating the depth map of the scene by using two video cameras as sensors. It is suitable for navigation, collision avoidance and design tasks.

The main advantage of such system low cost when it's operation is approximately in the same quality range as LIDAR sensors for the same purpose.

The existing algorithm of depth map calculation has been improved by adding truncation filters and calibration processes for specific type of camera and will be suitable for any type of camera with the same resolution.

REFERENCES

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

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

[3] 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, pp. 151–169, 2008. [in Russian].

[4] 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), pp. 146–151, 2016. [in Russian].

[5] I. Cabezas and 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.

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

[7] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*. Second Edition, issue 13, 2015, pp. 178–193, pp. 458–493. ISBN 978-0-521-54051-3

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

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

Received February 12, 2021

Vasylenko Mykola. orcid.org/0000-0003-4937-8082
Candidate of Science (Engineering). Senior lecturer.
Aviation Computer-Integrated Complexes Department, 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. Student.
Aviation Computer-Integrated Complexes Department, National Aviation University, Kyiv, Ukraine.
Publications: 1
E-mail: sychanna33@gmail.com

М. П. Василенко. О. С. Сич. Система оцінювання глибини зображення за потоковим відео

В роботі розглянуто метод оцінювання глибини за потоковим відео. Наводиться алгоритм отримання карти глибин за допомогою методу поділу зображень, який може бути використаний у різних сферах техніки та промисловості для визначення об'єкта і обчислення відстані до нього. Розроблено алгоритм налагодження та процес його адаптації під конкретні застосовувані зовнішні пристрої і програмне забезпечення. Для експериментальної установки були використані дві веб камери модуля відстеження Urchin Webcam (SJ-922-1080) з такими характеристиками: роздільна здатність відео – FullHD (1920x1080), сенсор – комплементарний метал-оксидний-напівпровідник, поле огляду – 90°, автофокус, частота кадрів в секунду – 20. Розроблено програмний код для даних камер у середовищеі Matlab та алгоритм його адаптації для будь-яких інших камер аналогічної роздільної здатності. Проведено експериментальне дослідження роботи алгоритму.
Ключові слова: стерео зір; карта несходження; карта глибини; калібрування; ректифікація.

Василенко Микола Павлович. orcid.org/0000-0003-4937-8082
Кандидат технічних наук. Старший викладач.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, 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. Студент.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Національний авіаційний університет, Київ, Україна.
Кількість публікацій: 1.
E-mail: sychanna33@gmail.com

М. П. Василенко. А. С. Сыч. Система оценки глубины изображения по потоковым видео

В работе рассмотрен метод оценки глубины по потоковому видео. Приводится алгоритм получения карты глубин с помощью метода разделения изображений, который может быть использован в различных сферах техники и промышленности для определения объекта и вычисления расстояния до него. Разработан алгоритм настройки и процесс его адаптации под конкретные применяемые внешние устройства и программное обеспечение. Для экспериментальной установки были использованы две веб-камеры модуля отслеживания Urchin Webcam (SJ-922-1080) со следующими характеристиками: разрешение видео – FullHD (1920x1080), сенсор – комплементарный металл-оксидный-полупроводник, поле обзора – 90°, автофокус, частота кадров в секунду – 20. Разработана программный код для данных камер в среде MatLab и алгоритм его адаптации для любых других камер аналогичного разрешения. Проведено экспериментальное исследование работы алгоритма.
Ключевые слова: стерео зрение; карта несхождения; карта глубины; калибровки; ректификация.

Василенко Николай Павлович. orcid.org/0000-0003-4937-8082
Кандидат технических наук. Старший преподаватель.
Кафедра авиационных компьютерно-интегрированных комплексов, Национальный авиационный университет, Киев, Украина.
Образование: Киевский национальный университет технологий и дизайна, Киев, Украина, (2012).
Направления научной деятельности: возобновляемые источники энергии, оценка свойств веществ и материалов по их собственным электромагнитным излучениям.
Количество публикаций: больше 20 научных робот.
E-mail: m.p.vasylenko@nau.edu.ua

Сыч Алексей Сергеевич. Студент.
Кафедра авиационных компьютерно-интегрированных комплексов, Национальный авиационный университет, Киев, Украина.
Количество публикаций: 1.
E-mail: sychanna33@gmail.com