

UDC: 656.073: 004.93

## The development of an intelligent system for recognizing the volume and weight characteristics of cargo

V. D. Shepelev<sup>1,a</sup>, N. V. Kostyuchenkov<sup>2,b</sup>, S. D. Shepelev<sup>3,c</sup>, A. A. Alieva<sup>1,d</sup>,  
I. V. Makarova<sup>4,e</sup>, P. A. Buyvol<sup>4,f</sup>, G. A. Parsin<sup>4,g</sup>

<sup>1</sup> South Ural State University, 76 Lenin Avenue, Chelyabinsk, 454080, Russia

<sup>2</sup> S. Seifullin Kazakh AgroTechnical university, 62 Zhenis Ave, Nur-Sultan

<sup>3</sup> South Ural State Agrarian University, 13 Gagarin st., Troitsk, 457100, Russia

<sup>4</sup> Kazan Federal University, 10a Syuyumbike prosp., Naberezhnye Chelny, 423812, Russia

E-mail: <sup>a</sup> shepelevvd@susu.ru, <sup>b</sup> kostyuchenkov\_NV@mail.ru, <sup>c</sup> shepelev2@yandex.ru,

<sup>d</sup> albishka150123@gmail.com, <sup>e</sup> kamivm@mail.ru, <sup>f</sup> skyeyes@mail.ru, <sup>g</sup> Inf801.parsin@gmail.com

*Received 14.09.2020, after completion — 26.01.2021.*

*Accepted for publication 28.01.2021.*

Industrial imaging or “machine vision” is currently a key technology in many industries as it can be used to optimize various processes. The purpose of this work is to create a software and hardware complex for measuring the overall and weight characteristics of cargo based on an intelligent system using neural network identification methods that allow one to overcome the technological limitations of similar complexes implemented on ultrasonic and infrared measuring sensors. The complex to be developed will measure cargo without restrictions on the volume and weight characteristics of cargo to be tariffed and sorted within the framework of the warehouse complexes. The system will include an intelligent computer program that determines the volume and weight characteristics of cargo using the machine vision technology and an experimental sample of the stand for measuring the volume and weight of cargo.

We analyzed the solutions to similar problems. We noted that the disadvantages of the studied methods are very high requirements for the location of the camera, as well as the need for manual operations when calculating the dimensions, which cannot be automated without significant modifications. In the course of the work, we investigated various methods of object recognition in images to carry out subject filtering by the presence of cargo and measure its overall dimensions.

We obtained satisfactory results when using cameras that combine both an optical method of image capture and infrared sensors. As a result of the work, we developed a computer program allowing one to capture a continuous stream from Intel RealSense video cameras with subsequent extraction of a three-dimensional object from the designated area and to calculate the overall dimensions of the object. At this stage, we analyzed computer vision techniques; developed an algorithm to implement the task of automatic measurement of goods using special cameras and the software allowing one to obtain the overall dimensions of objects in automatic mode.

Upon completion of the work, this development can be used as a ready-made solution for transport companies, logistics centers, warehouses of large industrial and commercial enterprises.

Keywords: cargo measurement, computer vision, machine learning

Citation: *Computer Research and Modeling*, 2021, vol. 13, no. 2, pp. 437–450 (Russian).

The reported study was funded by RFBR, project number 19-29-06008 \ 20.

© 2021 Vladimir D. Shepelev, Nikolai V. Kostyuchenkov, Sergey D. Shepelev,  
Albina A. Alieva, Irina V. Makarova, Polina A. Buyvol, Gleb A. Parsin

This work is licensed under the Creative Commons Attribution-NoDerivs 3.0 Unported License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by-nd/3.0/>  
or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

УДК: 656.073: 004.93

## Разработка интеллектуальной системы определения объемно-весовых характеристик груза

В. Д. Шепелев<sup>1,a</sup>, Н. В. Костюченков<sup>2,b</sup>, С. Д. Шепелев<sup>3,c</sup>, А. А. Алиева<sup>1,d</sup>,  
И. В. Макарова<sup>4,e</sup>, П. А. Буйвол<sup>4,f</sup>, Г. А. Парсин<sup>4,g</sup>

<sup>1</sup> Южно-Уральский Федеральный университет, Россия, 454080, г. Челябинск, пр. Ленина, д. 76

<sup>2</sup> Казахский агротехнический университет им. С. Сейфуллина,  
Казахстан, 010011, Нур-Султан, пр. Женис, д. 62

<sup>3</sup> Южно-Уральский Аграрный университет, Россия, 457100, г. Троицк, ул. Гагарина, д. 13

<sup>4</sup> Казанский Федеральный Университет, Россия, 423812, г. Набережные Челны, пр. Сююмбике, д. 10а

E-mail: <sup>a</sup> shepelevvd@susu.ru, <sup>b</sup> kostyuchenkov\_NV@mail.ru, <sup>c</sup> shepelev2@yandex.ru,  
<sup>d</sup> albishka150123@gmail.com, <sup>e</sup> kamivm@mail.ru, <sup>f</sup> skyeyes@mail.ru, <sup>g</sup> Inf801.parsin@gmail.com

*Получено 14.09.2020, после доработки — 26.01.2021.*

*Принято к публикации 28.01.2021.*

Промышленная обработка изображений или «машинное зрение» в настоящее время является ключевой технологией во многих отраслях, поскольку эта технология может использоваться для оптимизации различных процессов. Целью настоящей работы является создание программно-аппаратного комплекса измерения габаритно-весовых характеристик груза на базе интеллектуальной системы, основанной на нейросетевых способах идентификации, позволяющих преодолеть технологические ограничения аналогичных комплексов, реализованных на ультразвуковых и инфракрасных измерительных датчиках. Разрабатываемый комплекс будет производить измерения грузов без ограничения на объемные и весовые характеристики груза, который необходимо тарифицировать и сортировать в рамках работы складских комплексов. В состав системы будет входить интеллектуальная компьютерная программа, определяющая объемно-весовые характеристики груза с использованием технологии машинного зрения и экспериментальный образец стенда измерения объема и веса груза.

Проведен анализ исследований, посвященных решению аналогичных задач. Отмечено, что недостатком изученных способов являются очень высокие требования к расположению камеры, а также необходимость ручной работы при вычислении размеров, автоматизировать которую не представляется возможным без существенных доработок. В процессе работы исследованы различные способы распознавания объектов на изображениях с целью проведения предметной фильтрации по наличию груза и измерения его габаритных размеров. Получены удовлетворительные результаты при применении камер, сочетающих в себе как оптический способ захвата изображений, так и инфракрасные датчики. В результате работы разработана компьютерная программа, позволяющая захватывать непрерывный поток с видеокamer Intel RealSense с последующим извлечением из обозначенной области трехмерный объект и вычислять габаритные размеры объекта. На данном этапе выполнено: проведен анализ методик компьютерного зрения; разработан алгоритм для реализации задачи автоматического измерения грузов с использованием специальных камер; разработано программное обеспечение, позволяющее получать габаритные размеры объектов в автоматическом режиме.

Данная разработка по завершении работы может применяться как готовое решение для транспортных компаний, логистических центров, складов крупных производственных и торговых предприятий.

Ключевые слова: измерение грузов, компьютерное зрение, машинное обучение

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-06008 \ 20.

## 1. Introduction

Industrial imaging or “machine vision” is currently a key technology in many industries as it can be used to optimize various processes. It is successfully used for determining the traffic flow parameters [Shepelev et al., 2020a; Shepelev et al., 2020b], in logistics for recognizing cargo types [He et al., 2020], automatic driver assistance systems [Orlovska et al., 2020], in the automotive industry for quality management and robotic assembly [Lückenhaus, 2020].

The development of this technology is justified by the transition of the world economy to Industry 4.0, which provides for the formation of a cyber-physical system, within which all elements are active objects that participate in the information exchange and make appropriate decisions without human participation [Schiele, Torn, 2020; Hozdić, Butala, 2020]. In “smart factories”, machines will understand their environment and will be able to communicate using a single network protocol with each other, as well as with the logistics and business systems of suppliers and consumers. Manufacturing equipment and products will become active system components that govern their production and logistics processes, so their quality and speed will depend on the efficiency of processing and control algorithms. Therefore, the search for and optimization of computer vision and machine learning techniques to accelerate and improve the reliability of object recognition is more urgent than ever.

## 2. Study of computer vision and machine learning methods, the choice of the optimal methods for objectives to be achieved in the work

Using the results of 2D image recognition has already become a standard task. 3D identification technology opens up a number of completely new opportunities, including the optimization of processes in the logistics system, as one of the most comprehensive one covering the entire product life cycle [Makarova et al., 2016; Makarova et al., 2020]. Thus, the use of 3D vision technologies helps to increase the degree of automation of production scenarios; however, several issues should be dealt with to solve this problem.

First, it is necessary to receive a video stream from a recording device (video camera), extract a frame from it and send it for processing. Further, it is necessary to reliably determine the presence of the load, as well as its position in space, on the received frame. The area of the space where the load is defined will be of interest for determining the coordinates of points on the surface of the load, by which it will be possible to determine its size.

There are several works of other authors dealing with similar problems.

For example, in [Salih, Malik, 2012], the authors used the concept of triangulation to determine the distances between two specified points in a flat photograph, and hence the linear dimensions. The disadvantage of this method is the very high requirements for the location of the camera, as well as the need for manual work when calculating the dimensions, which cannot be automated without significant modifications.

In [Kar et al., 2015; Gadelha et al., 2015; Girdhar et al., 2016], neural networks were used for the primary recognition of an object in an image, which can later be transformed into a three-dimensional body with known dimensions.

In [Brown, Lowe, 2005], a three-dimensional body is reconstructed from a series of photos obtained when shooting an object from different angles.

To solve the problem of identifying the presence and position of cargo in space, it is obviously not enough to use ordinary RGB cameras (i.e., cameras from which we receive only a two-dimensional image). It is also necessary to obtain the information on the depth (distance) of each point of the image from the camera lens. To obtain the information on the depth of objects, several types of solutions are currently known:

- using cameras with two lenses. Such cameras allow one to simulate human binocular vision and stereoscopic photography, giving three-dimensional photographs [Gui et al., 2020];

- using TOF cameras (Time of Flight). In such cameras, the calculation of depth and distance is provided using the “time of flight” measuring technology, which originates from algorithms used in radars. This creates a long-range image similar to radar portraits, except that it uses a light pulse instead of a radio frequency signal [Baek et al., 2020].
- LIDAR cameras (Light Detection and Ranging), based on the technology of obtaining and processing the information on objects based on active optical systems [Yamada et al., 2020].

### **2.1. Study of the possibility of measuring loads with a SICK Visionary-T industrial camera**

The Sick Visionary-T (AG V3S110-2x model) is a time-of-flight camera with a resolution of  $176 \times 144$  pixels typical of this type of cameras. The characteristics of this equipment are shown in Table 1.

Table 1. The main characteristics of the Sick Visionary-T camera

Characteristic	Value
Measurement distance, m	0.5 ... 60
Measurement angles, degrees	$69 \times 56$
Resolution, pixels	$176 \times 144$
Light sensitivity, klx	< 50
Connection format	M12 17-pin (power supply), system plug digital IOs (24 V) M12 8-pin Gigabit Ethernet, X-coded
Power supply	24 V (+/-20%)
Power consumption, W	$\leq 22$
Weight, kg	$\sim 1.9$
Dimensions (Length $\times$ Width $\times$ Height), mm	$162 \times 116 \times 104$
Permissible ambient temperature, °C	0 ... +50
Impact protection	As per EN 60068-2-27:2009
Vibration protection	As per EN 60068-2-6 and 60068-2-64
Protection classes	III, IP67

As can be seen from this table, the camera can be used in a wide temperature range, in premises with dust and other harmful factors.

We carried out experimental measurements to determine the correspondence of the camera to the set task in terms of ensuring the required accuracy.

A white matte box with a picture was used as the object of measurements. The dimensions of the box were (W  $\times$  H  $\times$  D):  $640 \times 540 \times 220$  mm.

During the test, the object to be measured was placed opposite the camera at different distances. For each of the selected distances, the width and height of the object were manually measured, based on the corresponding points of the point cloud.

During the test, we noted that some of the points falling on the boundaries of the measured object were shifted along Z — the depth of the image. Such points were not taken into account in the measurements. For the convenience of measurements, the SOPAS software provides several settings, in particular, limiting the output of the point cloud by the image depth, filtering “floating” points, etc. In the course of the work, we used the parameters presented in Fig. 2.

During the tests, we found that the error in measuring the test object, in absolute value, was 10–36 mm for height and 9–24 mm for width (Table 2), which is not acceptable for the task at hand. The acceptable accuracy is 0.5 cm.

In our further work, we will use Intel RealSense d435 cameras.

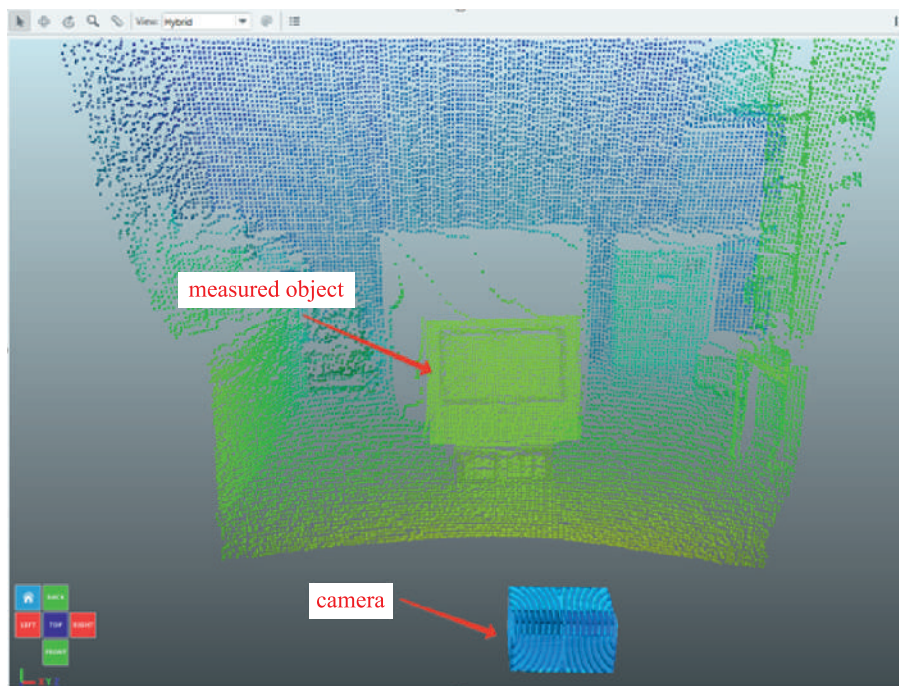


Fig. 1. The position of the camera relative to the measured object during the experiment



Fig. 2. Required parameters for adjusting the camera during the measurements: enable flying pixel filter — filtering of “floating” pixels — medium-weak; enable distance based filter — data filtering depending on the distance — no closer than 630 mm, no more than 3.5 m. We should also include other parameters, such as the dynamic mode, the confidence-based filter, the intensity-based filter, the smoothing filter

Table 2. Test results of the Sick Visionary-T camera

Distance to the object, mm	Received width, mm	Received height, mm	Absolute width deviation, mm	Absolute height deviation, mm
1,707	622	520	18	20
1,900	612	516	28	24
2,470	630	519	10	21
3,830	604	531	36	9

## 2.2. Study of the possibility of measuring loads using the Intel Realsense d435 camera

This camera has the following features:

- A built-in video processor, which supports the installation of up to 5 channels of the MIPI 2 serial interface in cameras for real-time processing of 3D images and accelerating data output;
- An advanced 3D stereo algorithm for a more accurate 3D perception and extended range.
- A set of video sensors allowing one to identify differences in images with resolutions up to  $1,280 \times 720$  pixels.
- Support for the open source Intel RealSense SDK 2.0 cross-platform;
- A special processor for color image signals, image adjustment and color data scaling;
- An active infrared projector to illuminate objects and enhance 3D data;
- A common shutter and wide angle of view ( $91.2^\circ \times 65.5^\circ \times 100.6^\circ$ ), which provides high-quality 3D perception of data while in motion or for moving objects, creating a wide view and eliminating the appearance of “blind” zones.
- The main technical characteristics of the camera are presented below [Intel RealSense, 2020].
- Working environment: indoor and outdoor;
- 3D technology: Active infrared stereo projector (IR);
- Image sensor technology: Common shutter: Pixel size —  $3 \mu\text{m} \times 3 \mu\text{m}$ ;
- Intel® RealSense™ Video Processor D4, Intel® RealSense™ Module D430;
- Depth of field — (H × V) for HD 16: 9:  $85.2^\circ \times 58^\circ$  (+/-  $3^\circ$ );
- 3D data output resolution: up to  $1,280 \times 720$ ;
- Frame rate of the output 3D-stream: up to 90 frames/sec;
- Min. focal length: 0.11 m;
- Max. range: about 10 meters. The accuracy depends on calibration, objects in the frame and lighting;
- RGB sensor resolution and frame rate:  $1,920 \times 1,080$ , 30 fps;
- RGB depth-of-field sensor (horizontal × vertical):  $69.4^\circ \times 42.5^\circ$  (+/-  $3^\circ$ );
- Camera dimensions (length × depth × height):  $99 \text{ mm} \times 25 \text{ mm} \times 25 \text{ mm}$ ;
- Connector: USB Type C Port;
- Fastening mechanism;
- One mount: 1/4-20 UNC. Two M3 mounts.

This camera was tested according to the method described for the Sick camera; therefore, below we present only the test results that showed good applicability of this equipment for solving the problem.

Table 3. Test results of the Intel Realsense d435 camera

Distance to the object, mm	Received width, mm	Received height, mm	Absolute width deviation, mm	Absolute height deviation, mm
1,900	639	538	1	2
2,100	638	538	2	2
2,490	637	535	3	5
4,000	632	533	8	7

### 2.3. Application of machine learning to solve the problem of measuring loads

Currently, the most common computer vision methods for classifying objects in images are neural networks, which can, for convenience, be divided into “one-stage” and “two-stage”.

The first ones include networks with the following architecture:

- Yolo (You look only once);
- SSD;
- RetinaNet;

The second type includes networks with CNN architecture and its derivatives: Fast CNN, Faster CNN, R-CNN, etc.

Taking into account the high requirements for the measurement speed (no more than 3 s per one load), we used the Yolo v3 architecture [Darknet, 2020], which is currently the fastest way to recognize objects, while not inferior in recognition quality [The YOLOv3, 2020]. Figure 3 shows a comparative analysis of the quality and speed of common neural networks [Lin et al., 2017; Redmon and Farhadi, 2018]. As one can see from the figure, YOLOv3-608 is only two percent inferior to the leading quality FPN FRCN (57.9 versus 59.1), but at the same time, it processes images more than three times faster.

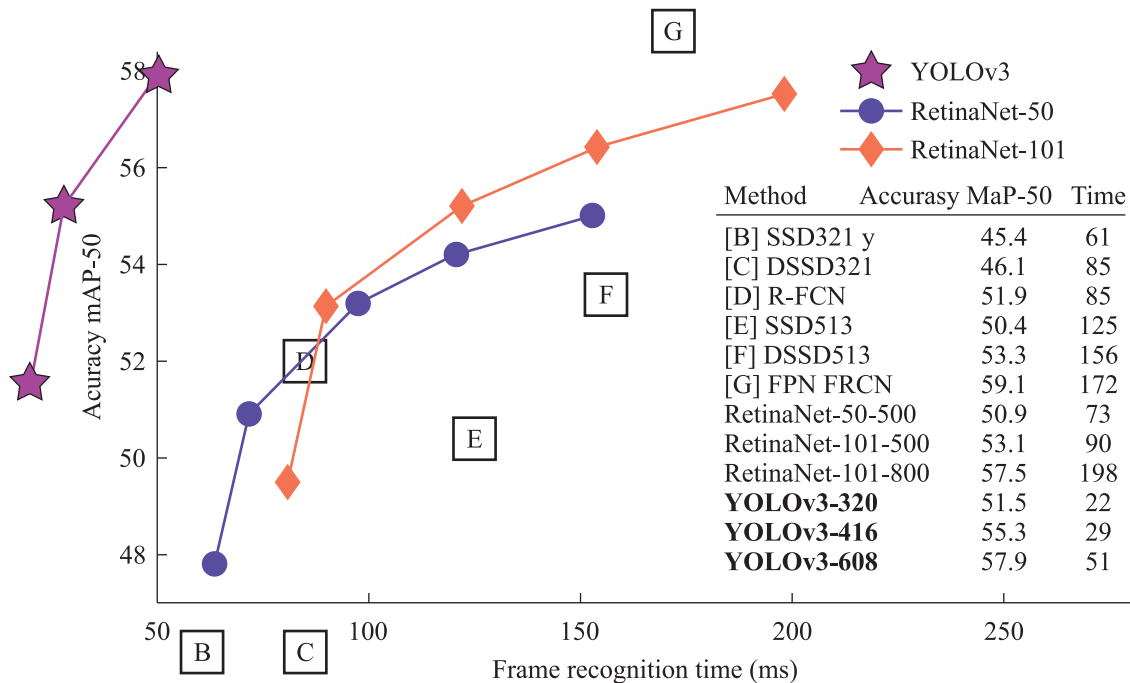


Fig. 3. A comparison of the accuracy of the object recognition time on the image of neural networks of various topologies

## 3. Selecting objects' images for machine learning

### 3.1. Collecting data to train the model

At the initial stage, we collected data for training (formed a dataset). The target data set consisted of images obtained from warehouses of freight companies containing goods of various sizes. About 200 images were taken. To achieve the best result in training the model, we included negative examples in the dataset, i.e., images that do not contain loads. Examples of images are shown in Fig. 4.





Fig. 4. Examples of images for training a neural network. The center of each image contains a possible object for measurement

### 3.2. The measurement algorithm

After all the devices (cameras) are connected, we can start our measurements. The first step is camera calibration. For correct calibration, we need to set the parameters of the checkerboard, such as the width of the cells and their number. Next comes the assessment of the checkerboard position in world coordinates and the calculation of the area of interest, in which the load will be measured. After the cameras are successfully calibrated, the calibration results are saved.

The next step is load measurement. Upon receipt of the frames from all the devices, we calculate the point cloud for each camera. The parameters for calculating the point cloud are the frames from various cameras, the serial number of the device, calibration parameters of each camera, internal properties of the camera depth, the area of the interest, threshold for the depth value in world coordinates. The resulting point cloud is used as output parameters.

After that, we define a bounding box around the load and draw this box. The parameters for defining the bounding box are a point cloud array. The output parameters are the serial number of the device; the length, width, and height of the bounding box; a dictionary with corner points of a bounding box around the load. When calculating the dimensions, we receive a bounding box around the load in 2D using the  $X$ ,  $Y$  coordinates from the point cloud. We define a bounding box around the load in 3D. We receive the coordinates of the bounding box within the image. We transform the coordinates of the bounding box according to the coordinates received from the device. All the characteristics are calculated using the standard methods in the Python language.

The process of measuring the load volume is shown in Fig. 5.

### 3.3. Image tagging

At the next stage, we marked the training data. In our case, the marked objects are cargoes. All cargoes were marked on the image if their visibility reached approximately 80% or more.

The Yolo-Label software product was used for marking. As a result of marking for each image from the dataset, we obtained a text file with the identification of each cargo and its position in the frame. This will also allow us to further determine to which category the cargo belongs.

### 3.4. Model training

A specialized Darknet framework was used to train the model. At the same time, changes were made to the standard architecture of the Yolov3 network.

First, the maximum number of training epochs was reduced according to the recommendation “use at least 2,000 epochs for each type of object”. In our case, the maximum number of epochs for one class is set equal to 4,000.



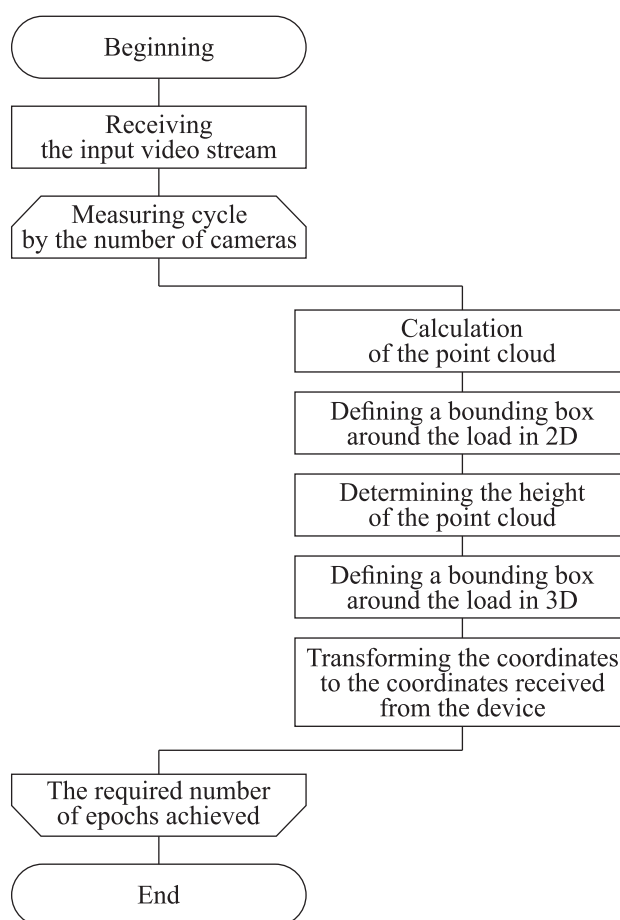


Fig. 5. An algorithm for measuring the volume of cargo demonstrating the processing of the input video stream for each used camera. The bounding box is determined in the cycle of each camera depending on the calculated point cloud

Further, the value classes = 1 is set for layers of the “Yolo” type neural network, which means that the network should recognize only one type of objects. The filters value for superfine mesh layers is set equal to 18.

We also modified the data file for the network, in which the actual values of the following data were set:

1. The number of classes equal to the number of object types;
2. The path to the file with the list of images, on which the neural network will be trained;
3. The path to a file with a list of images, on which the neural network will be tested;
4. The path to the folder where the network training results will be located.

At the end of the data pre-processing, we trained the neural network using the Nvidia RTX 2070 graphics accelerator. The training time was approximately 7 hours. The training accuracy for mAp-50 was about 92% (Fig. 6).

#### 4. The development of measurement algorithms

For convenience, the measurement process is divided into 5 stages (Fig. 7).

At the first stage (Realsense d435 camera), we identified the constants determining the maximum load size and camera parameters.

Depending on the size of the load, the camera should be calibrated using an appropriate calibration board representing alternating black and white squares (Fig. 8).

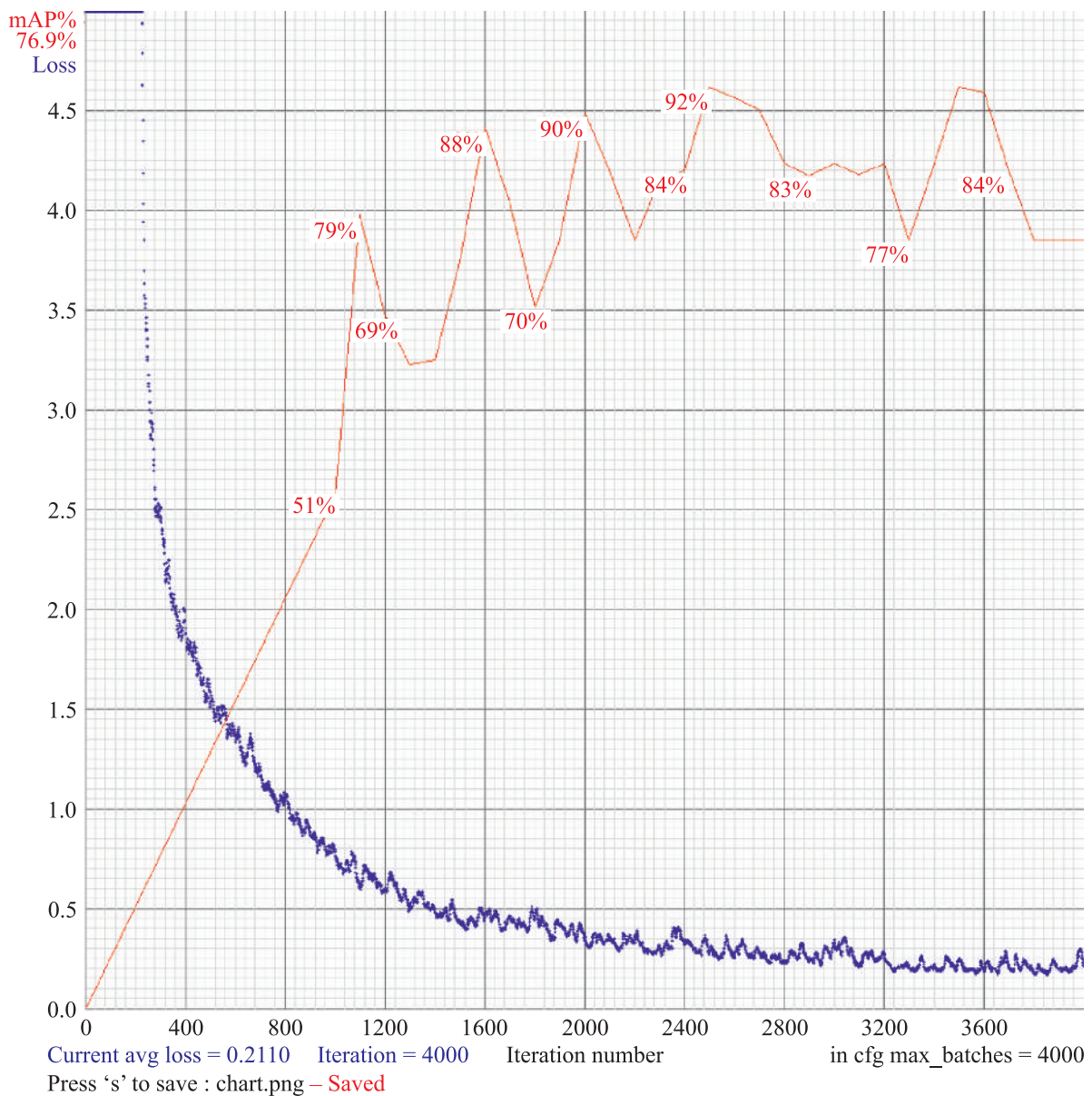


Fig. 6. The network training graph. The decreasing loss function is shown in blue. Its average value at the 4,000-th iteration is 0.2110. The average cargo recognition accuracy is shown in red - the maximum accuracy is 92%

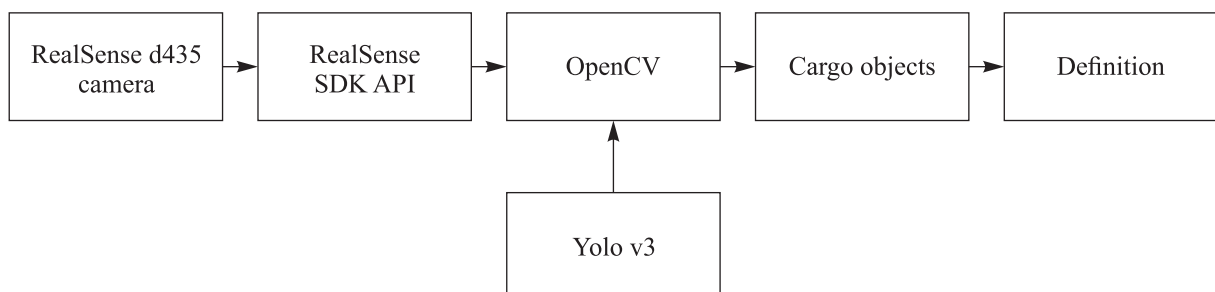


Fig. 7. Work stages to determine the size of goods. The process consists of the following items: determination of the camera parameters; saving the resulting image and the point cloud; determination of the cargo position using the OpenCV framework and the YOLOv3 neural network; filtering the point cloud; determination of the cargo size

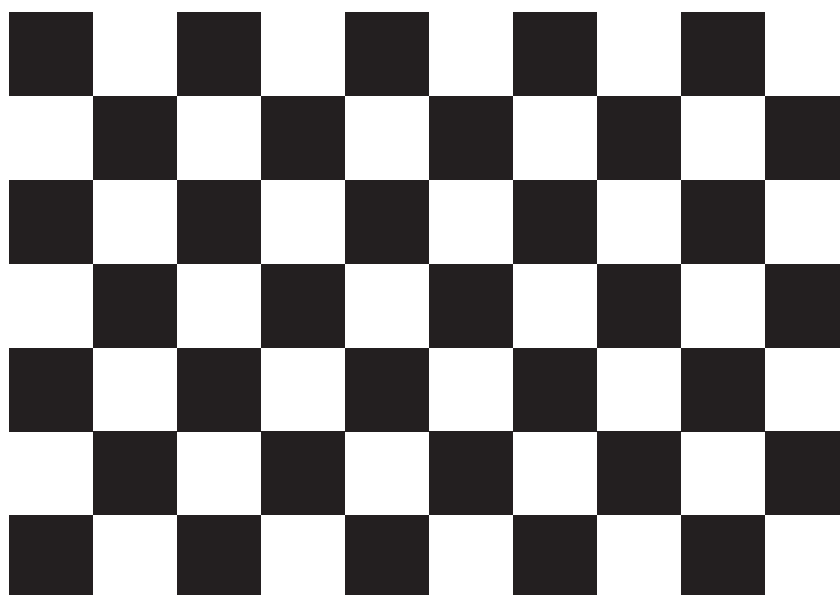


Fig. 8. The appearance of the calibration board

It has been experimentally determined that acceptable measurement results for small loads are given by calibration boards with a cell size of 0.026 m. In this case, the absolute measurement error is 1–2 mm.

The constants determining the parameters of the camera include:

- “own| camera parameters: focal length, resolution, and lens distortion;
- parameters determining the position of the camera in space.

In this work, the position of the camera was not constant, since it was not known at what angle and at what distance to place the camera from the measured object. Therefore, the algorithm for dynamic camera calibration was applied using a calibration board. If there is a calibration board in the frame, the board is no longer needed for further measurements and one can start measuring.

At the “Realsense SDK API” stage, the two-dimensional image received from the camera, as well as the three-dimensional representation of the image (“point cloud”) are saved in separate variables of the program code.

A two-dimensional image using the OpenCV framework is supplied to the input of the YOLOv3 neural network, which processes the frame and, if there is a cargo, returns its position in the frame.

Further, all points that do not belong to the cargo object, as well as points that lie below the position of the calibration board, are removed from the point cloud. With this filtering in mind, we get a point cloud that belongs only to the cargo.

For the point cloud obtained, the dimensions of the cargo are determined by calculating the distances between the most distant points along the  $x$ ,  $y$  and  $z$  axes.

## 5. The development of the program code

The program code was developed in Python.

During the development we used the following third-party open source libraries:

- OpenCV is a library for image manipulation and use of neural networks in computer vision;
- RealSense SDK API — methods for working with images for Intel Realsense cameras;
- Numpy is a library for mathematical calculations;
- Json is a library for working with data in JSON format;
- Pickle is a library for working with the pickle format, in which the camera calibration results are saved.

In addition to the existing libraries, we developed our own solutions:

- Methods for automatic determination of the calibration board in the frame and its position in space;
- Methods for converting the data from a camera to a point cloud;
- A method for combining the data received from several cameras into a single point cloud;
- A method for obtaining the overall dimensions of an object by searching for a parallelepiped of the minimum size that describes the cargo;
- A method of visualization of measurement results. Visualization is performed directly on the video stream received from the camera; at the same time, the dimensions of the load in mm are displayed in the upper corner, and the load itself is highlighted with a box in the form of a parallelepiped, which allows a visual comparison of the actual load with the calculated one (Fig. 9).

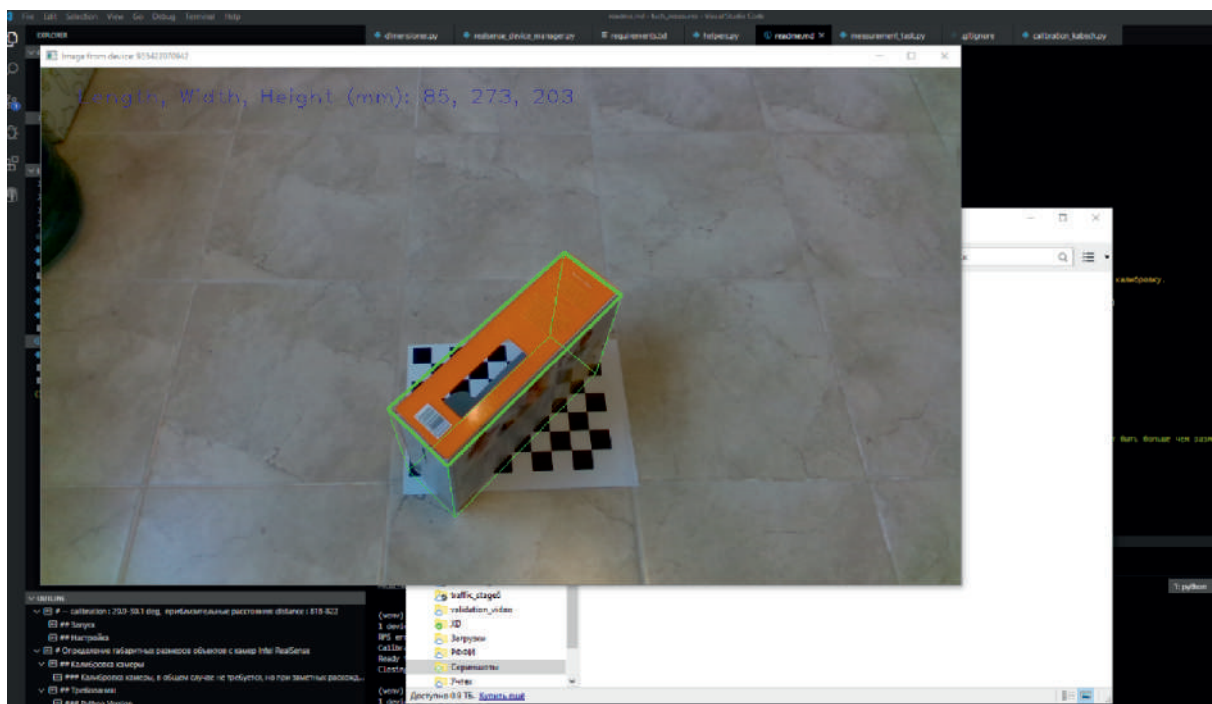


Fig. 9. Visualization of measurement results

## Conclusion

In the course of the work, we found that time-of-flight cameras do not provide the required measurement accuracy due to their low resolution. The Intel RealSense camera, using the stereoscopic imaging principle in conjunction with infrared sensors, showed good measurement results. The average absolute cargo measurement error was 4 mm.

We developed the software for automatic cargo measurement, which is part of the hardware and software complex. The hardware and software complex actually allows one to receive the overall dimensions of cargo within no more than 3 s. Measurements are carried out without restrictions on volume and weight characteristics of goods.

Upon completion of the work, this development can be used as a ready-made solution for transport companies, logistics centers, and warehouses of large industrial and commercial enterprises.

## References

- Baek E.-T., Yang H.-J., Kim S.-H., Lee G., Jeong H.* Distance Error Correction in Time-of-Flight Cameras Using Asynchronous Integration Time // *Sensors*. — 2020. — Vol. 20. — P. 1156.
- Brown M., Lowe D. G.* Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets // *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, Ottawa, Ontario, Canada. — 2005. — P. 56–63.
- Darknet [Electronic resource]. — URL: <https://github.com/AlexeyAB/darknet> (accessed: 22.01.2020).
- Gadelha M., Maji S., Wang R.* 3D Shape Induction from 2D Views of Multiple Objects // *2017 International Conference on 3D Vision (3DV)*. — 2015. — P. 402–411.
- Girdhar R., Fouhey D. F., Rodriguez M., Gupta A.* Learning a Predictable and Generative Vector Representation for Objects / B. Leibe, J. Matas, N. Sebe, M. Welling (eds.). *Computer Vision — ECCV 2016. Lecture Notes in Computer Science*, vol. 9910. Springer, Cham. — 2016. — [https://doi.org/10.1007/978-3-319-46466-4\\_29](https://doi.org/10.1007/978-3-319-46466-4_29)
- Gui Y., Gao H., Li Q., Xia T.* Method and apparatus for recognizing object, device, vehicle and medium // 2020. — Patent No. US20200074196A1. <https://www.freepatentsonline.com/20200074196.pdf>
- He P., Wu A., Huang X., Rangarajan A., Ranka S.* Video-based Machine Learning System for Commodity Classification // *Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems*. — 2020. — Vol. 1. — P. 229–236.
- Hozdić E., Butala P.* Concept of Socio-Cyber-Physical Work Systems for Industry 4.0 // *Tehnički vjesnik*. — 2020. — Vol. 27, Iss. 2. — P. 399–410.
- Intel RealSense Product Family D400 Series. Datasheet [Electronic resource]. — URL: <https://www.intelrealsense.com/wp-content/uploads/2020/06/Intel-RealSense-D400-Series-Datasheet-June-2020.pdf> (accessed: 14.06.2020).
- Kar A., Tulsiani S., Carreira J., Malik J.* Category-Specific Object Reconstruction from a Single Image // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. — 2015. — P. 1966–1974.
- Lin T.-Y., Goyal P., Girshick R., He K., Dollar P.* Focal loss for dense object detection // *arXiv:1708.02002*. — 2017.
- Lückenhaus M.* The Role of Machine Vision in the Automotive Industry. [Electronic resource]. — URL: [https://www.photonics.com/Articles/The\\_Role\\_of\\_Machine\\_Vision\\_in\\_the\\_Automotive/a58196](https://www.photonics.com/Articles/The_Role_of_Machine_Vision_in_the_Automotive/a58196) (accessed: 14.06.2020)
- Makarova I., Khabibullin R., Belyaev E., Gabsalikhova L., Mukhametdinov E.* Improving the logistical processes in corporate service system // *Transport Problems*. — 2016. — Vol. 11 (1). — P. 5–18.
- Makarova I., Shubenkova K., Buyvol P., Mukhametdinov E.* A Software module for multi-criteria suppliers' selection with respect to the spare parts logistic // *Transport Problems*. — 2020. — Vol. 15 (1). — P. 105–116.
- Orlovska J., Novakazi F., Lars-Ola B., Karlsson M. A., Wickman C., Söderberg R.* Effects of the driving context on the usage of Automated Driver Assistance Systems (ADAS) — Naturalistic Driving Study for ADAS evaluation // *Transportation Research Interdisciplinary Perspectives*. — 2020. — Vol. 4. — P. 100093.
- Redmon A., Farhadi A.* YOLOv3: An Incremental Improvement // *arXiv:1804.02767*. — 2018.
- Salih Y., Malik A. S.* Depth and geometry from a single 2d image using triangulation // *2012 IEEE International Conference on Multimedia and Expo Workshops*. — 2012. — P. 511–515.

- Schiele H., Torn R.-J.* Cyber-physical systems with autonomous machine-to-machine communication: Industry 4.0 and its particular potential for purchasing and supply management // *International Journal of Procurement Management*. — 2020. — Vol. 13, Iss. 4. — DOI: 10.1504/IJPM.2020.108617
- Shepelev V., Aliukov S., Nikolskaya K., Das A., Slobodin I.* The Use of Multi-Sensor Video Surveillance System to Assess the Capacity of the Road Network // *Transport and Telecommunication*. — 2020a. — Vol. 21 (1). — P. 15–31.
- Shepelev V., Glushkov A., Almetova Z., Mavrin V.* A Study of the Travel Time of Intersections by Vehicles using Computer Vision // *Proceedings of the 6th International Conference on Vehicle Technology and Intelligent Transport Systems*. — 2020b. — Vol. 1. — P. 653–658.
- The YOLOv3 Object Detection Network Is Fast! [Electronic resource]. — URL: <https://syncedreview.com/2018/03/27/the-yolov3-object-detection-network-is-fast/> (accessed: 25.01.2020).
- Yamada H., Ahn J., Mozos O. M., Iwashita Y., Kurazume R.* Gait-based person identification using 3D LiDAR and long short-term memory deep networks // *Advanced Robotics*. — 2020. — DOI: 10.1080/01691864.2020.1793812