International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11077-11081

Grain Wagon Fill Detection using Camera and Deep Convolution Network

AleksandrYurievich Shkanaev[1], Aleksei Vladimirovich Panchenko[2] and Nikita Igorevich Kodosov[3]

[1,2,3]Cognitive Technologies LTD, 60-letya Oktyabrya Avenue, 9, Moscow, 117312, Russia.

Orcid: 0000-0002-5148-5795

Abstract

The article provides a description of the system of grain stream automatic control when grain is loaded into the grain box. The system uses the data from only one camera and the deep convolutional network for the box elements and grain stream recognition. A rigid fixation of the camera above the grain box allows to use the bird's-eye viewmodel of transformation for generating signals about the location of the box relative to the auger pipe.

Keywords: grain harvester, semantic segmentation, convolutional neural networks, conventional loading process.

INTRODUCTION

Grain harvesting is carried out with grain harvesters. Loading the grain harvested with the combine harvester is usually carried out into the tractor box or vehicle box. The process is very particular about operating the combine harvester with the tractor close to it because the grain flow must fill the box evenly and accurately, figure 1.

Figure 1: The task of filling the tractor box

The result of the process of filling the box with grain depends on two operators: a combine harvester operator and an agricultural machine operator and his ability to keep the attitude of the machine. Good alignment leads to reducing loss level and allows to use the grain box space efficiently. That guarantees an ideally level profile in the box. To achieve this result the combine harvester operator should concentrate on two things at once: - to align the machine with the tractor; - to control the grain stream in order to fill the tractor box efficiently.

The automation of the process allows the operator to steer the machine correctly while the system is automatically filling the tractor box. That assures more safety and comfort. To automate the process, the system needs to understand the auger pipe position in relation to the tractor box and issue warning signals. In case the auger pipe is deflected towards the grain box edges or even beyond them the harvester operator receives a warning signal. This information can be useful for the operator in order to provide the control of the auger pipe, the grain flow and the proportional filling of the tractor box. To solve the problem in this application domain, there are several alternatives, which are briefly described in the following section.

RELATED WORKS

To solve the problem of loading the grain the following business and academic solutions are used. One of the solutions is using additional GPS equipment and high-accuracy sensors to control the auger pipe position [1]. Constructing the mathematical model allows to compute the quantity of grain in the vehicle box. The computation of the grain stream position is performed on the basis of the information received from different sensors and GPS data. However, the accuracy of the grain profile construction is very low due to the low accuracy of GPS and the auger pipe sensors.

The other solution is using the system of grain stream automatic control by Class company. The system uses the stereocamera data [2]. The stereocamera allows us to receive 3D image of the grain volume in the vehicle box. When the vehicle approaches to the harvester, the system senses its attitude and detects the trailer, as shown in figure 2. In figure 2, the green lines indicate the box boundaries and the direction for the auger pipe movement. The system predicts the auger position from measurements from the camera and rotation sensors. The accuracy of positioning is greatly reduced in conditions of vibration and dust [3].

Figure 2: The work of the system Autofill by Class company

Case New Holland recently has proposed the IntelliFillsystem, that is based on the Time-Of-Flight (TOF) camera with 3D images. This camera allows the system to work both in the dark and in the daylight [4]. The camera is fixed on the auger pipe of the harvester, and measures the distance to all the grain box corners and the grain quantity in the container. This data allows to control the process of filling the container automatically.

The other solution is the approach [5], based on using air drones and quadcopters equipped with a camera. The advantage of the approach is the absence of noise and dust, as well as a good overview of the tractor box and the possibility of obtaining measurements from any side. Disadvantages are connected with high requirements to the accuracy of the air vehicle positioning, its additional economic expenditures and wind influence.

The analysis of the works shows that in most of the existing solutions the cameras data is used. However, the systems under consideration consist of two cameras or they are equipped with special volume sensors [6]. The innovation of the proposed work is the use of only one camera to determine the auger pipe position in relation to the container edges. That reduces the price of the system and simplifies its work. Moreover, in the given approach the deep convolutional networks image segmentation is used for controlling the auger pipe. This provides high reliability of the solution of the problem. Determination of the container edges allows to predict the point of grain falling in the box and, as a result, to generate the data signals for the harvester operator.

Formulation of the task

There given a coordinate system associated with the earth's surface, which we denote as OXYZ. We denote the camera coordinate system by *oxyz*. The camera is rigidly fixed to the auger pipe, figure 3. The position of the auger pipe is time-invariant. The coordinates of the point C - the extreme point of the position of the auger pipe in the OXYZ coordinate system are equal (x_c,y_c,h_c). The internal and external parameters of the camera are known, given by the matrices ***K***and ***R***,***T***. The external parameters of the camera, ***R*** and ***T***, are determined with respect to the ground plane OXYZ. The image formed by the camera is denoted by I_t.

Grain at high speed flies out of the auger pipe along a parabolic trajectory and falls into the box. The box dimensions ***a***,***b***and its height ***h*** are known in advance. It is required to determine the position of the point of incidence of grain relative to the box boundaries $\Delta \boldsymbol{x}$ and$\Delta \boldsymbol{y}$. Based on the parameters $\Delta \boldsymbol{x}$ and$\Delta \boldsymbol{y}$, it is necessary to work out a solution for a) normal, b) permissible and c) an impermissible deviation of the drop point relative to the box.

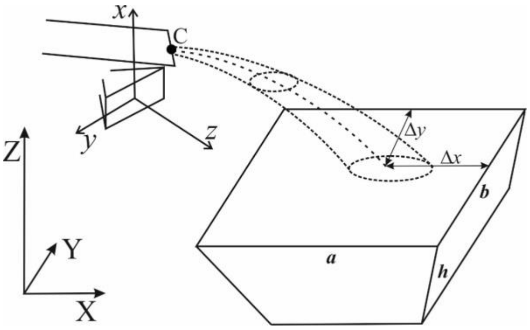


Figure3. Scheme for the placement of auger pipe, camera and box of harvesting machine

Solution

This task is proposed to be solved in several stages: -preprocessing; -segmentation; -projective transformation (transformation to top view); - calculation of the point of grain fall; - issuing messages to the operator. Let's consider these stages in more detail.

At the first stage, the image is normalized and preprocessed. At this stage, no light is eliminated, color normalization is performed, distortion is eliminated and the image is scaled. An example of the original image I_t is shown in the figure. As a rule, all images include a box fragment (its outer and inner part), a stream of falling and poured grains.

Figure 4: Image$I_t$formed by the camera

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 21 (2017) pp. 11077-11081

At the second stage, a convolutional neural network operates. It implements the segmentation of the image and finds the following objects of classification: falling grain, fallen grain, interior and exterior of the box. A more detailed description of this stage, the network architecture and the features used for learning the data will be given below. An example of the image of the convolutional network formed at the output is shown in the figure.

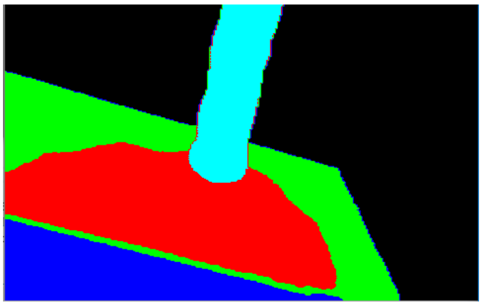


Figure 5: The image formed at the output of the convolutional network

In the third stage, it performs transformation of the image to the top view from the known internal $\boldsymbol{K}$ and external $\boldsymbol{R},\boldsymbol{T}$ parameters of the camera. That provides a "correct" geometric interpretation of the box. It is assumed in the work that the upper edge of the vehicle box for loading the grain is horizontal and does not have a significant inclination, which does not lead to large projective distortions. Bird's-eye view transformation is carried out proceeding from the parameters of the homography matrix $\boldsymbol{H}$ obtained in accordance with the formulas [8, 9] starting from the matrices $(K \cdot [R \quad T])^{-1}$.

Figure 6: Example of projective image conversion

In the next stage it is assumed that the fall of grain into the box can be described by a parabolic trajectory using a model of the form

$$z = -k_1 x^2 - k_2 x + h_c$$

Where k_1, k_2 –the coefficients determined experimentally based on the capacity of the auger, h_c –the height of the auger pipe.

For the upper edge of the body with a height h, the point of incidence of the grain can be found by substituting $z = h$. Thus, the point of intersection of the plane of the upper edge of the body and the grain flow found in the OXYZ coordinate system has the coordinates (x_{corn}, y, h). The definition of the unknown parameter y can be carried out on the basis of the following formula

$$y = y_c - T_y$$

where T_y – the second component of the vector $\boldsymbol{T}$.

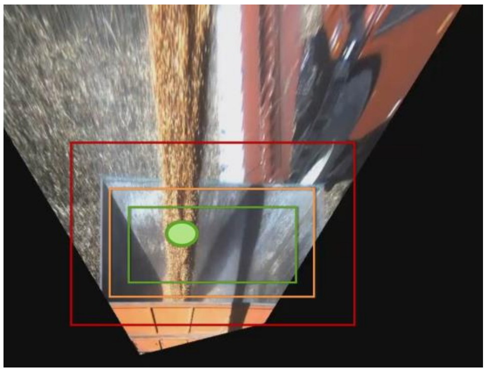


Figure 7: Example of an image of the point of incidence of a grain stream

Delivering messages to the operator is based on the following zones: normal operation zone (ok), attention zone (warning), forbidden zone (fail).

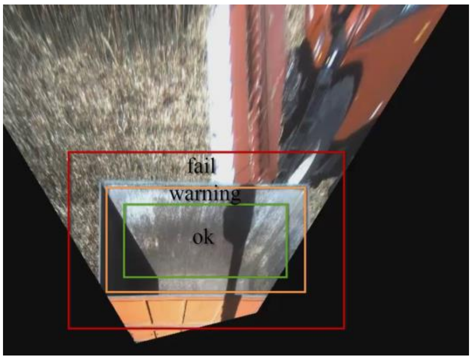
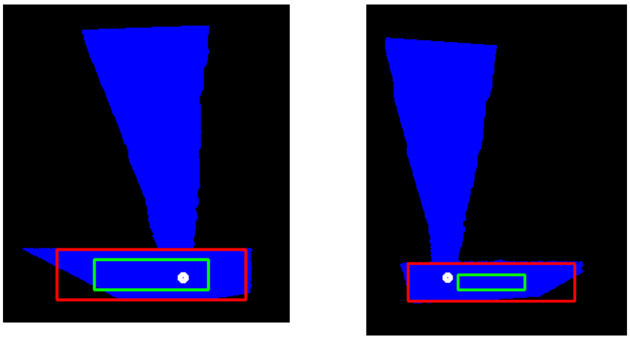


Figure 8: Zones for issuing warning messages to the operator

The algorithm described above is implemented in the form of special software and is currently running in test mode on one of the combines. Fragments of the windows of the software are shown in the figure 9. In the figure, one color shows the grain flow and the interior of the box. Areas for tracking are represented by two rectangles. The point of grain fall is represented by white color.



A) Grain entering the first area. Okmessage.

B) Grain entering the second area.

Warningmessage.

Figure 9: Fragment of special software windows

Segmentation of images using the U-net network

In relation to work [10], a reduced image with a resolution of 224x224 is used to build the network, with a reduced by half number of filters on each convolution layer, the figure is below. The structure of the network and its description is given below.

It consists of a contracting path (left side) and an expansive path (right side). The contracting path follows the typical architecture of a convolutional network. It consists of the repeated application of two 3x3 convolutions, each followed by a rectied linear unit (ReLU). Also ending each convolution blocks has 2x2 max pooling operation with stride two for downsampling. At each down sampling step we double the number of feature channels. Every step in the expansive path consists of an up sampling of the feature map followed by a 2x2 convolution (up-convolution) that halves the number of feature channels, a concatenation with the correspondingly feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. At the final layer a 1x5 convolution is used to map each 32-component feature vector to the desired number of classes. In total the network has 23 convolutional layers.

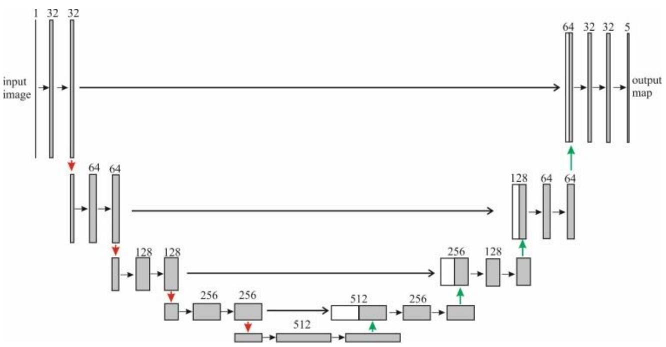


Figure 10: The architecture of the modified

U-netnetwork

In the figure 10, the gray blocks show a feature map. The number of filters used is shown in the figures at the top. Long thin arrows denote the copy operation received at the initial stages of the feature map and transferring them to later stages, the joining operation is shown with white rectangles. Red vertical arrows operation max pool 2x2, short arrows operation conv 3x3, green vertical arrows deconvolution 2x2.

The data set used for training and testing consists of images obtained from a camera mounted on an agricultural combine . The images were manually marked, which allowed to formground truth images, in the number of 907 pieces. The marked images contained the following object classes:

- Flow – a stream of falling grain from an auger into a box;
- Seeds – grain, which is already in the body;
- Interior – the internal observable surface of the box;
- Exterior – the external observable surface of the box;
- Background – all other objects: the sky, the ground, trees, etc.

The areas of hand-marked classes in the image are presented in the form of closed figures 11. Image analysis showed that approximately 51% of the image area was occupied by the background, 5% by the exterior, 17% by the interior, 17% by the stream and 8% by the seeds. In addition, approximately 2% of the image area is occupied by the «Do not care» layer which is not used in training.

Figure 11: Example of ground truth images

Due to the small amount of images in the training set, the method of increasing (augmenting) the data was used [11]. The increase in the volume of data was carried out on the basis of four methods:

- horizontal image displacements (flips);
- rotation of images clockwise and counterclockwise at small angles (4-9 degrees);
- random image cropping with the preservation of 80% of the original area;
- minor changes in each color channel of the image.

After applying the above techniques, the number of images was 24 times increased. For training and network testing, the initial sample was divided into two groups: test data 868 and validation stage 39. The accuracy of the network was assessed using several characteristics of Precision, Recall, intersection over Union (IoU) and weighted IoU. Table 1 shows the network accuracy data for each of the characteristics separately.

Table 1: Networkaccuracy

	Background	Flow	Seeds
Precision	0.936	0.959	0.903
Recall	0.968	0.912	0.893
IoU	0.909	0.879	0.816
Weighed IoU	0.907	0.880	0.81

	Interior	Exterior	Average
Precision	0.921	0.962	0.937
Recall	0.93	0.930	0.928
IoU	0.868	0.898	0.874
Weighed IoU	0.867	0.899	0.874

CONCLUSION

In this work it is shown that it is sufficient to use one camera to solve the problem of autofilling the box with the grain . The solution of the problem of grain detection and interior of the box can be solved using the convolution neural network. The network U-net used in the work showed high recognition accuracy. The network is trained on its own image database. Using NVIDIA Jetson TX2 and high-performance deep learning inference optimizer and runtime NVIDIA TensorRT allows to achieve a network speed of 60 fps. To determine the position of the box, the information about the calibration parameters of the camera and the conversion of bird's-eye viewwas used.

ACKNOWLEDGEMENTS

This work was financially supported by the Ministry of Education and Science of Russian Federation (Agreement № 14.579.21.0122, Applied Research and Experimental Development individual identifier RFMEFI57915X0122).

REFERENCES

[1] Happich, G.; Harms, H.H.; Lang, T. Loading of Agricultural Trailers Using a Model-Based Method/ Agric. Eng. Int., CIGR J. 2009, XI, p.1–13.

[2] Moller, J. Computer Vision—A Versatile Technology in Automation of AgriculturalMachinery. J. Agric. Eng. 2010, 47, 28–36.

[3] A. V. Panchenko, N. V. Reshetnyak, A. M. Samohin, V. V. Postnikov, "The automated testing facility based on machine vision for optimizing grain quality control technology", Proc. SPIE 10341, Ninth International Conference on Machine Vision (ICMV 2016), 103411Q (17 March 2017).

[4] Posselius, J.; Foster, C. Autonomous self-propelled units: What is ready today and to come in the near future. Proceedings of the 23rd Annual Meeting Club of Bologna, Bologna, Italy, 9–10 November 2012.

[5] A.Hernandez ,H.Murcia, C. Copot and R. De Keyser. Towards the Development of a smart flying sensor: Illustration in the field of precision agriculture/ Sensors, 2015, 15(7), p.1-20.

[6] J.D.Gaard Grain wagon fill detection using ultrasonic sensors, 2012, 15(7), p.120.

[7] Shkanaev A.V., Polevoy D.V., Panchenko A.V., Borisov Y.Y. and Kolomeychenko M.I. Straw row position and orientation reconstruction through image segmentation. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 24 (2016) pp. 11803-11810.

[8] Berkovich, S.B., Kotov, N.I., Lychagov, A.S. et al. Vision system as an aid to car navigation // Gyroscopy Navigation, 2017, Volume 8, Issue 3, pp 200–208.

[9] Berkovich S.B., Sholokhov A.V., Sadekov R.N., Kotov N.I. Using vision systems to determine the vehicle position on the road /24th Saint Petersburg International Conference on Integrated Navigation Systems, ICINS 2017 - Proceedings 2017. p. 32-35.

[10] U-Net: Convolutional networks for biomedical image segmentation. arXiv preprint arXiv:1505.04597v1,p.1-8, 2015.