

# Surface recognition of machine parts based on the results of optical scanning

M A Bolotov<sup>1</sup>, V A Pechenin<sup>1</sup>, N V Ruzanov<sup>1</sup> and E J Kolchina<sup>2</sup>

<sup>1</sup>Samara National Research University, Institute of Engines and Power Plants, Moskovskoe Shosse, 34A, Samara, Russia, 443086

<sup>2</sup>Stock company «RKC «Progress», Zemeca street, 18, Samara, Russia, 443009

e-mail: vadim.pechenin2011@yandex.ru

**Abstract.** To predict the quality parameters of products (in particular, the assembly parameters) mathematical models were implemented in the form of computer models. To ensure the adequacy of calculations, it is necessary to have information about the actual geometry of the parts, which can be obtained using noncontact measurements of parts of the assembly. As a result of measuring parts and components using optical or laser scanner, a large dimension array of measured points is formed. After standard processing (e.g. noise removal, combining the scans, smoothing, creating triangulation mesh), the recognition of individual surfaces of parts becomes necessary. This paper presents a neural network model that allows the recognition of elements based on an array of measured points obtained by scanning.

## 1. Introduction

The least automated step in the industry is the assembly of single and serial products characterized by medium and high complexity. These products include aircraft engines. The considered products are not made in large quantities as cars; they are characterized by a high degree of optionality and increased requirements for complexity and accuracy. The share of labour-consuming assembly in the total labour-intensiveness of products is up to 25% and largely determines their quality. There are several reasons that make it difficult to fully automate the assembly of these products. One of the significant reasons is the difficulty of determining the parameters of the operations performed by robots, which are guaranteed to ensure the specified accuracy and quality of products. The assembly of medium and high complexity products is a unique operation, during which the course of operations is changed according to the results of measurements and the geometric analysis of the assembled parts. Measurement of geometry is made by both contactless and contact methods.

To partially automate engine assembly processes, it is necessary to recognize both the individual parts and the surfaces of the parts along which the assembly will take place. Face recognition is possible using computer vision approaches [1,2,3]. The aim of this work is to create a model based on the use of neural networks, designed to recognize the surfaces of engineering parts after their measurement using an optical or laser scanner.

## 2. Object of research

To test the model, a real engine parts simulator was designed and manufactured: spacers in the turbine of an aircraft engine. A detail drawing is shown in Figure 1. The part contains cylindrical and flat edges, as well as threaded holes.

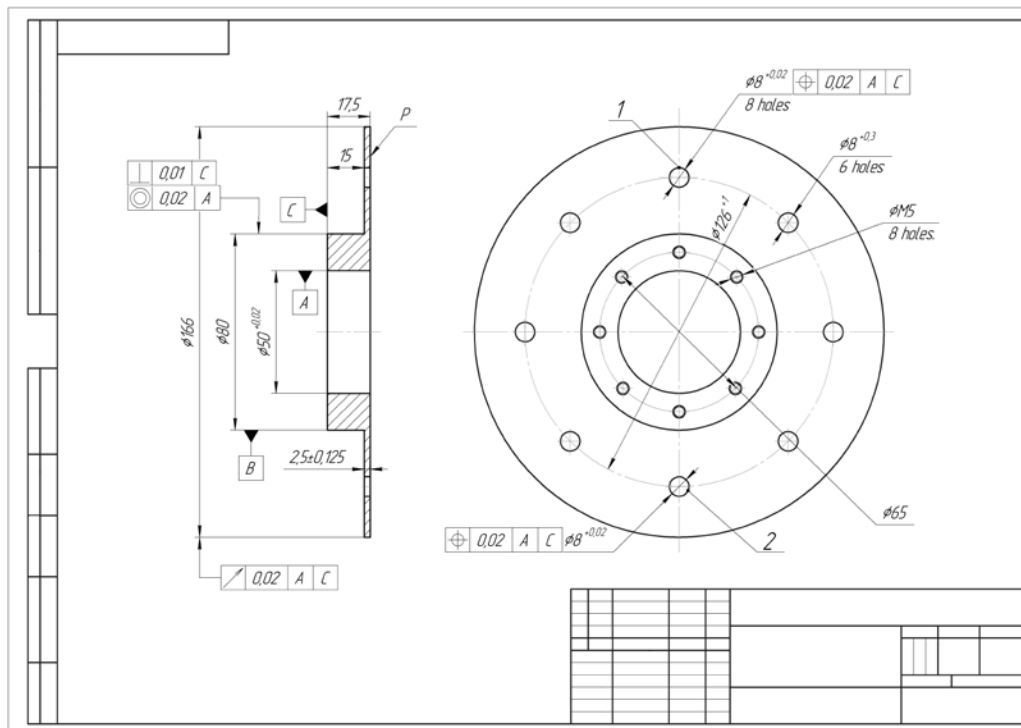


Figure 1. Drawing part "spacer simulator".

Automated element (surface) recognition of measured parts using neural networks solves two tasks: 1) segmentation of a part's components into types of surfaces (plane, cylinder, cone, etc.) 2) additional refinement of the boundaries of triangulation surfaces based on deviations of the facet normal vectors. To solve the first problem, a convolutional neural network was used.

## 3. Neural network model of surface recognition

Convolutional neural networks (CNN) are a very wide class of architectures which main idea is to reuse parts of the neural network to work with different small, local input areas [4]. The main area of application of convolutional architectures is image processing [5, 6].

At present, many approaches have been developed for recognizing three-dimensional objects in works devoted to computer vision. These approaches can be divided into two groups: recognition of elements, directly working on their own three-dimensional representations of objects, such as polygonal grids, voxel representations and arrays of points, and approaches based on signs and metrics that describe the shape of a three-dimensional object, "what it looks like" in the collection 2D projections [7].

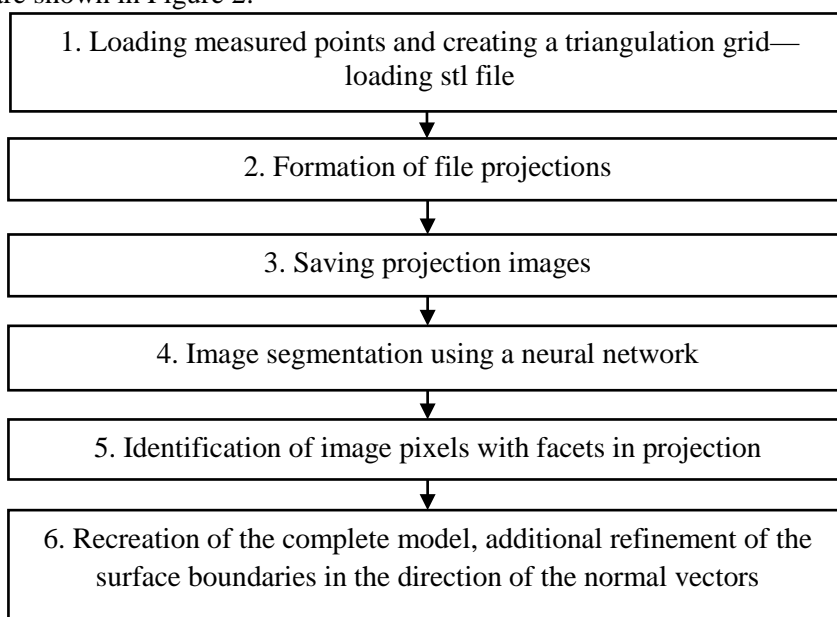
Except for the recent work by Wu et al. [8], who studied form descriptions from a voxel-based object representation through three-dimensional convolutional networks, the previous three-dimensional shape descriptions were mostly "manually developed" according to a specific geometric property of the shape surface or volume. For example, shapes can be represented by histograms or models with total signs of surfaces consisting of normals and curvatures [9], distances, angles, areas of triangles or volumes of tetrahedra calculated for sample points of surfaces [10], properties of spherical functions defined in volumetric grids [11], local shape diameters measured on tightly selected surface points [12], and thermal core signatures on polygonal grids [13, 14]. The development of controlled machine learning algorithms on top of such descriptions of three-dimensional figures creates several problems. First, the size of organized databases with annotated 3D models is rather limited compared

to image data sets. For example, ModelNet contains about 150 thousand objects. In contrast, the ImageNet database [15] already includes tens of millions of annotated images. Secondly, the additional features and metrics of three-dimensional figures tend to be very multidimensional, which makes the algorithms prone to retraining.

One of the latest works on the problems of object classification and the segmentation of individual parts, in which an array of measured points is used directly at the entrance to the network, is [16]. Developed by authors from Stanford University, the network is named PointNet, the main idea of the approach is to memorize the spatial features of each point and then merge all the individual features into a marked general point cloud. The network is based on a convolutional architecture. The main disadvantage of the reduced network is that at the input it is necessary to always have the same number of points for all objects, which in practice is not possible and you will have to resort to an artificial “distortion” of the data.

Based on the literature review and the specifics of the problem being solved, the decision was made to use an approach based on the use of 2D projections of objects in solving the current problem. In this approach, a convolutional neural network, U-net [17], is used for segmentation.

We will reveal the main idea and stages of the developed model for the segmentation of individual faces. The main idea is to create projections (pictures) for the faceted model of measured data, segmentation of faces on projections and identification of facets by segmented images. The stages of the approach are shown in Figure 2.



**Figure 2.** A flowchart of the face recognition model using projections of measured data.

Consider the face recognition steps in more detail.

### *3.1. Loading measured points and creating a triangulation grid—loading stl file*

As noted in the introduction, the model is designed to recognize geometry after measurement using optical and laser scanners. After measurement using the scanner and preliminary data processing, a file is created with the coordinates of points united into a facet surface of the \*.stl format. The file contains the following data:  $\mathbf{V}_{g \times 3}$  (matrix of coordinates of the vertices of the grid stl-model),  $\mathbf{F}_{m \times 3}$  (matrix of combinations of three vertices forming the facets of the surfaces),  $\mathbf{N}_{m \times 3}$  (matrix of coordinates of facet normals).

### *3.2. Formation of file projections*

To enable semantic segmentation of the facets into separate surfaces using deep neural networks, it is necessary to create projections of 3D surfaces on coordinate planes. To prepare the projections, Roberts' algorithm was used [18].

Roberts' algorithm is the first known solution to the problem of removing invisible lines. This is a mathematically elegant method that works in object space. The algorithm primarily removes from each body the edge or edges that are screened by the body itself. Then each of the visible edges of each body is compared with each of the remaining bodies to determine which part or parts, if any, are shielded by these bodies. Therefore, the computational complexity of Roberts' algorithm grows, theoretically, with the square of the number of objects.

The operation of Roberts' algorithm takes place in two stages:

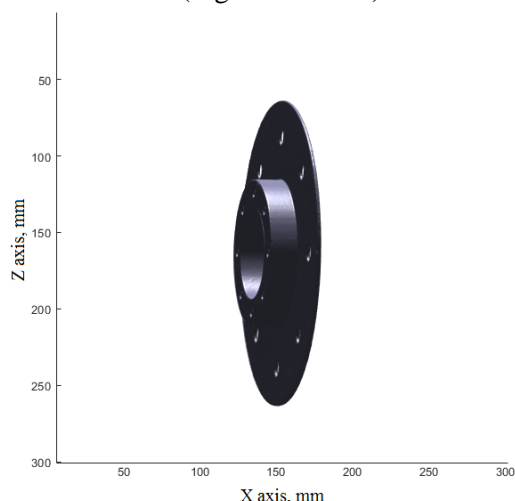
1. Definition of non-face faces for each body separately.
2. Identify and remove invisible edges.

To prepare the data, only the first stage of the algorithm was used. The second stage is not necessary for further decision; it is more complex and requires additional facets.

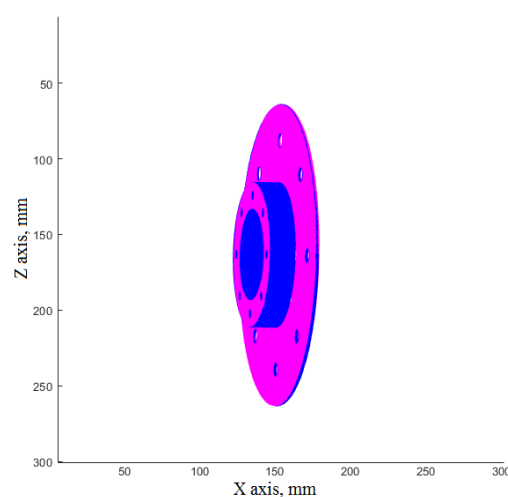
When creating projections on coordinate axes, orthogonal projections in the visual plane are obtained. For example, when projected on the  $XOY$  plane, the structure  $\mathbf{F}_{m \times 3}$  is preserved, and the matrix of the vertices  $\mathbf{V}_{g \times 3}$  is converted into  $\mathbf{V}_{g \times 2}$ , having only coordinates along the  $x$ - and  $y$ -axes. Bypassing the vertices of the obtained projections of the facets in the same sequence as in the original, you can divide them into two types: those that are oriented counter-clockwise, which means that we are looking at the facet from the outside of the body and those that are oriented clockwise, which means that we are looking at the facet from the inside. The order of the vertices determines the direction of the normal. Thus, if the component of the normal vector of the projection plane (in this example, the component along the  $z$ -axis) is negative, we look at the facet from the inside. Since the object is bounded by a closed surface, we cannot observe the faces from the inside—they are invisible. Thus, it is necessary to exclude the facets identified by the above method from the  $\mathbf{F}_{m \times 3}$  structure, obtaining the projection structure  $\mathbf{F}_{m_1 \times 3}$ , where  $m_1 < m$ .

### 3.3. Saving projection images

Image saving is performed using STL work libraries (functions of the *stltools* package by Pau Micó) and MATLAB graphics saving tools. The sizes of the pictures are saved, as are all projections. Using the same tools, training projection images are saved, where objects for recognition are highlighted in different colours (Figures 3 and 4).



**Figure 3.** Stl file details "spacer simulator".



**Figure 4.** Part projection used for segmentation.

Nevertheless, segmentation requires clear boundaries, so images stored in the \*.png format are imported into MATLAB (variable loading matrix RGB-colour) and converted to grayscale using the expression:

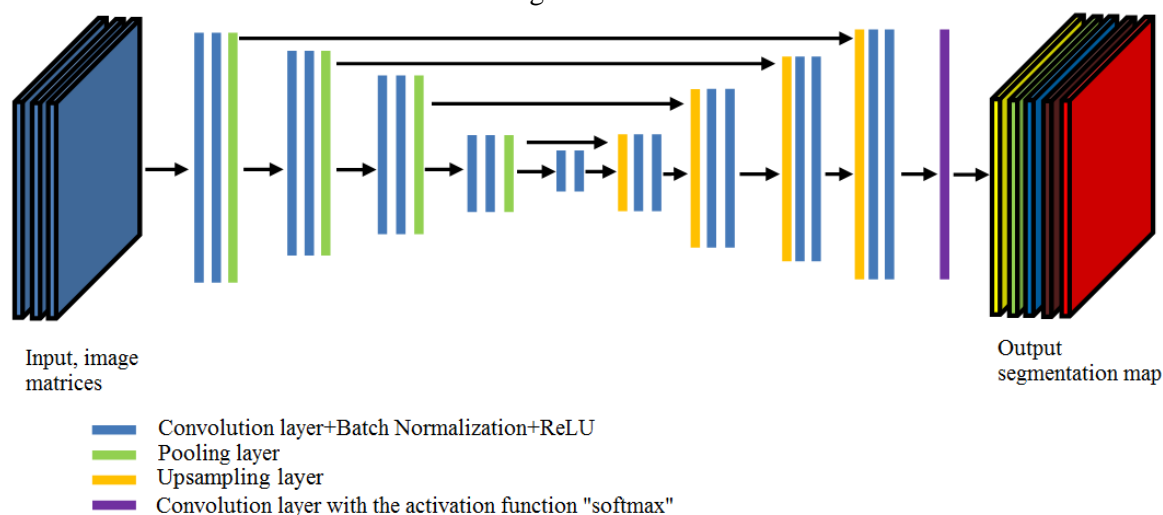
$$\mathbf{I}_{gray} = 0,299 \cdot \mathbf{R} + 0,587 \cdot \mathbf{G} + 0,114 \cdot \mathbf{B}, \quad (1)$$

where  $\mathbf{I}_{gray}$  is the grayscale image matrix and  $\mathbf{R}, \mathbf{G}, \mathbf{B}$  are the matrix components of the RGB system.

Multidimensional matrixes of images are saved to a \*.mat file. A total of six projections remain. The object is placed as if in a cube whose faces are parallel to the planes of the coordinates. The dimensions of the cube are such that it includes all the measured objects in the sample. Projections are accordingly made on the faces of the cube: two on the faces parallel to the *XOY* plane; two parallel to *XOZ* and two parallel to *YOZ*.

### 3.4. Image segmentation using a neural network

The U-Net network architecture is shown in Figure 5.



**Figure 5.** U-Net neural network architecture for image segmentation.

The network architecture is a sequence of layers of convolution and pooling, which first reduce the spatial resolution of the image, and then increase it by first merging it with the image data and passing it through other layers of the convolution. Thus, the network serves as a kind of filter.

The first half of the network contains layers of convolution with the activation function *ReLU*, normalization by mini-batch and layers of pooling (sub-sampling) and it is called a compressing path. The second part is an expanding path.

The upsampling layer is a reverse pooling layer that expands the feature map, followed by a convolution, which reduces the number of feature channels. Then comes the concatenation (“pasting” of linear objects) with an appropriately cut map of features from the compressive path and two convolutional layers.

On the last layer, convolution with a 1x1 core is used to bring each 64-component feature vector to the required number of classes. The activation function on the last layer is “softmax”.

The network was reproduced in the Python software environment.

### 3.5. Identification of image pixels with facets in projection

After the image is segmented, its pixels are compared to the coordinates of the corresponding projection. For matching, pixels are converted to points on the corresponding face of the cube described in Section 3.3. The values of the image pixels lie in the interval from 0 to 255. Accordingly, in order to identify the pixels responsible for a certain edge, values of a certain intensity are searched for in the image matrix. In this case, the search is made with a certain tolerance. The position of the found pixels in the matrix (row-column) is translated, taking into account the scale, into space coordinates (two coordinates from the image, the third—the coordinate of the corresponding cube face). In addition, in Section 3.2, the faces closed by others were not deleted. Therefore, when matching, a point of an object in an image can fall on two or more projection facets. In this case, the facet that is closest to the viewpoint is selected.

### 3.6. Recreation of the complete model, additional refinement of the surface boundaries in the direction of the normal vectors

The previous steps were necessary for preliminary automated recognition of only a fraction of the measured points. After identifying the vertices and facets to different surfaces, they are compared with the full set of vertices (search by equality of point coordinates). Points close to the geometric centre of the surface are selected. On the top in the geometric centre, a facet is selected that belongs to a specific face of the body. At the next stage, the algorithm for searching and refining facets of one face is used. We briefly describe the steps of the algorithm for searching for facets belonging to a specific face [19]:

1) input parameters are set:  $\mathbf{V}_{g \times 3}$ ,  $\mathbf{F}_{m \times 3}$ ,  $\mathbf{N}_{m \times 3}$ ,  $\vec{p}$  (coordinate vector of a point on the surface whose facets are to be selected),  $\vec{t}$  (normal at point  $p$ ), and  $\alpha$  (angle tolerance to find matching facets).

2) search for the facet that owns point  $p$  (intersection point of the beam and the facets).

3) search for all facets for which the normal vector is different from the normal vector  $\vec{t}$  by no more than a value  $\alpha$  (the scalar product of normal vectors is used to check the angles).

Facets are selected, from all the facets found that are suitable in the direction of the normal, that are associated with the first and among themselves common vertices.

## 4. Recognition results

To measure the details an optical 3D scanner RANGEVISION Pro2M was used. Figures 6 and 7 show photographs of the process of measuring the “spacer simulator” part (drawing in Figure 1).



Figure 6. One scan measurement of the part.



Figure 7. Group of measurement scans of the part.

To assess the quality of segmentation, you can use the modified loss function given in [20]. It is proposed to calculate the error in determining the parameters of a rectangle when recognizing faces and images in a picture using the intersection over union (IoU) metric, which is equal to the ratio of the area of intersection of the rectangle obtained as a result of detection and the rectangle from the mark-up to the area of their union.

In our case, we work with body facets, so instead of squares, we can operate on the number of facets. Therefore, a coefficient  $\delta_{segm}$  is calculated that is equal to the ratio of the number of facets  $N_{p \cap \delta}$  that are the intersection of the set of facets of the surface, obtained as a result of recognition, and actually belong to the surface of the facet to the  $N_{p \cup \delta}$  set of facets, which is their union:

$$\delta_{segm} = N_{p \cap \delta} / N_{p \cup \delta}. \quad (2)$$

Thus, the value of the coefficient lies in the interval [0; 1]. If there are several recognizable surfaces, then a generalized coefficient can be calculated  $\delta_{segm}$  for all recognizable body faces.

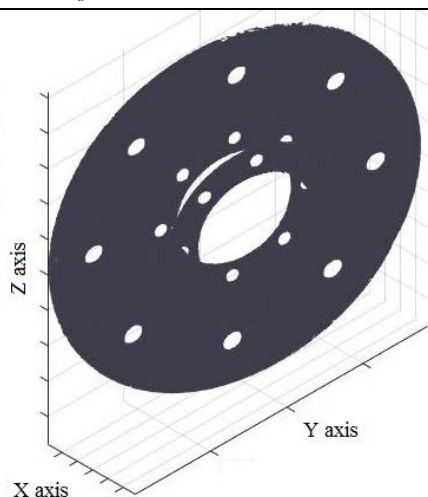
The number of measured facets of the part  $m$  was 622130. For training the neural network, a sample of 1000 cases of the stl-model of the considered part, aligned in different ways in space, corresponding to a total of 6000 projections, was formed. For formation of the training set, a nominal model of the part was used, which was saved in the \*.stl format.

Recognized flat and cylindrical faces of the “spacer simulator” detail measured with the scanner using all six projections are shown in Figures 8 and 9.

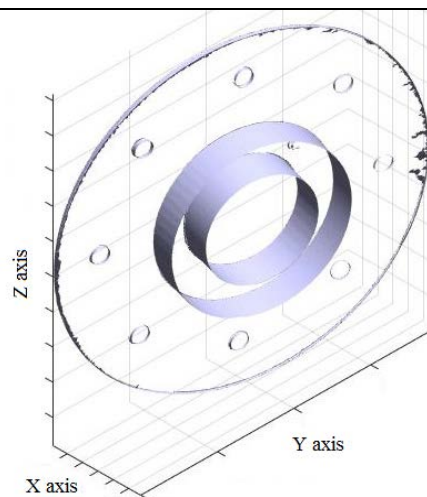
Table 1 shows the calculated coefficients  $\delta_{segm}$  of the measured cylindrical, flat and combined recognizable faces.

**Table 1.** Facet recognition errors.

Object of evaluation	Cylindrical faces	Flat edges	All recognizable faces
Value of $\delta_{segm}$	0.76	0.95	0.91



**Figure 8.** Recognized flat faces



**Figure 9.** Recognized cylindrical faces.

The recognition error of cylindrical faces is higher than flat ones. Although all cylindrical faces were identified, there were facets belonging to flat faces, as well as some facets from the threaded holes. The overall ratio exceeded 90%, which is associated with a much larger number of facets of flat faces compared to cylindrical.

## 5. Conclusion

The model presented in this paper allows you to quickly recognize cylindrical and flat surfaces of parts using a trained neural network and face facet search algorithm from previously recognized facets. The developed model is needed for further prediction of the assembly parameters of the product based on computer modelling [21]. Using digital prediction will allow robotic assemblers to function without human intervention.

## 6. References

- [1] Zakani F R, Bouksim M, Arhid K, Aboufatah M and Gadi T 2018 Segmentation of 3D meshes combining the artificial neural network classifier and the spectral clustering *Computer Optics* **42(2)** 312-319 DOI: 10.18287/2412-6179-2018-42-2-312-319
- [2] Blokhino Y B, Gorbachev V A, Rakutin Y O and Nikitin A D 2018 Development of an algorithm for semantic segmentation of real-time aerial photographs *Computer Optics* **42(1)** 141-148 DOI: 10.18287/2412-6179-2018-42-1-141-148
- [3] Nikitin M U, Konushin V S and Konushin A S 2017 Neural network model for video-based face recognition with frames quality assessment *Computer Optics* **41(5)** 732-742 DOI: 10.18287/2412-6179-2017-41-5-732-742
- [4] Nikolenko S I, Kadurin A and Arkhangelskaya E 2018 *Deep learning* (SPb: Peter)
- [5] Hubei D H 1988 *Eye, Brain, and Vision* (New York: Scientific American)
- [6] Webvision: The Organization of the Retina and Visual System URL: <http://webvision.med.utah.edu/book/>

- [7] Su H, Maji S, Kalogerakis E and Learned-Miller E 2015 Multi-view Convolutional Neural Networks for 3D Shape Recognition *ICCV* **7410471** 945-953
- [8] Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X and Xiao J 2015 3D ShapeNets: A deep representation for volumetric shape modeling *In Proc. CVPR* 1-9
- [9] Horn B K P 1984 Extended gaussian images *Proc. of the IEEE* **72(12)** 1671-1686
- [10] Osada R, Funkhouser T, Chazelle B and Dobkin D 2002 Shape distributions *ACM Transactions on Graphics* **21(4)** 807-832
- [11] Kazhdan M, Funkhouser T and Rusinkiewicz S 2003 Rotation invariant spherical harmonic representation of 3D shape descriptors *Proceedings of the Eurographics/ACM SIGGRAPH symposium on Geometry processing* 156-164
- [12] Chaudhuri S and Koltun V 2010 Data-driven suggestions for creativity support in 3D modeling *ACM Transactions on Graphics* **29(6)** 183
- [13] Bronstein A, Bronstein M, Ovsjanikov M and Guibas L 2011 Shape Google: Geometric words and expressions for invariant shape retrieval *ACM Transactions on Graphics* **30(1)** 1
- [14] Kokkinos I, Bronstein M, Litman R and Bronstein A 2012 Intrinsic shape context descriptors for deformable shapes *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- [15] Deng J, Dong W, Socher R, Li L-J, Li K and Fei-Fei L 2009 Imagenet: A large-scale hierarchical image database *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- [16] Qi C R, Su H, Mo K and Guibas L J 2016 PointNet: Deep learning on point sets for 3d classification and segmentation *ArXiv preprint arXiv:1612.00593* 652-660
- [17] Ronneberger O, Fischer P and Brox T 2015 U-net: Convolutional networks for biomedical image segmentation *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9351** 234-241
- [18] Roberts L G 1965 Machine perception of three-dimensional solids *Optical and Electro-Optical Information Processing* (Cambridge: MIT Press) 159-197
- [19] Stepanenko I S, Pechenin V A, Ruzanov N V and Khaimovich A I 2018 Technique of increasing the accuracy of GTE parts manufactured by selective laser melting *Journal of Physics: Conference Series* **1096(1)** 012143 DOI: 10.1088/1742-6596/1096/1/012143
- [20] Yu J, Jiang Y, Cao Z and Huang T 2016 UnitBox: An Advanced Object Detection Network *INNS*
- [21] Bolotov M A, Pechenin V A and Murzin S P 2016 Method for uncertainty evaluation of the spatial mating of high-precision optical and mechanical parts *Computer Optics* **40(3)** 360-369 DOI: 10.18287/2412-6179-2016-40-3-360-369

### Acknowledgements

The work was supported by the Ministry of Education and Science of the Russian Federation in the framework of the program to improve the competitiveness of the Samara University among the world's leading research and educational centers for 2013–2020 and partly by the Russian Federation President's grants (project code CII-262.2019.5). Experimental studies were carried out on the equipment of the center for the collective use of CAM technologies of the Samara University (RFMEFI59314X0003).