DOI: 10.7251/JTTTP2402069S
UDC: 628.971.03:004.421.6

ORIGINAL SCIENTIFIC PAPER

Component-Based Object Recognition Algorithm

Pavel Slivnitsin
Perm National Polytechnic University, Russia, slivnitsin.pavel@gmail.com

Leonid Mylnikov
HSE University, Russia, lamylnikov@hse.ru

Egor Efimov
HSE University, Russia, eaefimov@edu.hse.ru

Received: April 29, 2024
Accepted: November 25, 2024

Abstract: The paper presents an approach to object recognition based on the hypothesis of representing objects using a set of geometric primitives and relations between them. The goal of the paper is to develop a method for object recognition in the environment, which allows to recognize objects based on their description. For this purpose, the following tasks are solved: the recognition of a set of geometrical objects (primitives), the estimation of relations between primitives and the search of correspondences between the found primitives and relations and the defined templates (descriptions objects). The set of geometric primitives is selected taking into account the nature of the subject area of the objects to be recognized. The paper presents object recognition examples through the use of the method proposed. As a result, the operability of the proposed object recognition method is confirmed. An object description method has been developed. For experiments, the images of primitives were used generated in the Blender 3D, as well as photos of primitives from the kid's toy constructor. The primitive detection model was trained on a training sample consisting of 1000 artificial images and 50 real images. The research results can be applied in algorithms for recognizing traffic participants as well as traffic signaling objects

Key words: object recognition, object detection, recognition by components, computer vision, relation encoding, recognition algorithm.

INTRODUCTION

Recognition algorithms have found wide applications, including industrial applications such as equipment condition monitoring, product quality control [1], [2], text digitization [3], staff authentication [4], etc. Object recognition approaches are based on feature extraction of recognition objects, their search in new data and class assignment [5]. Based on the extracted features, an object prototype is built. A prototype provides an "average" representation of an object. In modern deep learning-driven algorithms object features are automatically extracted based on labeled examples from the training dataset.

I. Biederman proposed the theory that humans use a set of geometric features and the relations between them to recognize objects [6]. Biderman's theory is based on the assumption that every object can be represented by a set of geons (a set of geometric shapes). Each geon is described by a set of non-accidental properties that remain the same when the angle of view is changed [5], [6]. The evolution of such recognition approaches can be traced in [7], [8]. The recognition task can be defined as determining the necessary set of geometric primitives and their relations to recognize objects, selecting and training a model to recognize primitives and developing rules for relation estimation and recognition based on primitives and relations.

A set of geometric primitives can be defined taking into account the nature of the subject area and the task to be solved. A distinctive feature of the approach is that the number of geometric features and relations is finite. Three-dimensional shapes represent the examples of primitives, i.e. prism, sphere, torus, etc. (shape features). The examples of relations are spatial relations between primitives (above, below, farther, closer, etc.), contact, scale, distance, etc.

Described below object recognition approach is based on two assumptions: 1) any object can be represented as a finite set of geometric primitives connected to each other by a finite set of relations; 2) the recognition of a known (described) object can be performed as a search for a set of object primitives, connected with each other by a set of known relations.

Recognition performance depends on a number of parameters, such as metrics, the quality and size of the training dataset, and the nuances of the recognizable objects. In order for a sufficient dataset to be collected, sev-

eral iterations of data collection are typically taken, incl. dataset preparation, training and model comparison. [9]. When the number of recognizable classes increases, a decrease in recognition accuracy can be observed, which is discussed in [10] and confirmed by empirical observations. Based on that, we can make assumptions about the application restrictions of recognition algorithms and, hence, develop an algorithm that solves some recognition problem for a new application, which requires time and resources.

At the moment, the most common tool for object recognition is convolutional neural networks [11], which have replaced the non-neural-based recognition algorithms (such as Haar cascades [12], HOG [13], DPM [14]). In this case, it is not quite reasonable to talk about the advantages of some methods over the others, because the recognition algorithm (as well as other parameters such as loss function, optimizer, etc.) is selected for a particular problem [15] (same method can show good and poor performance depending on its application).

PRIMITIVE DETECTION

In order for the objects to be recognized in the image, sets of primitives and relations will be used. The set of primitives may depend on the subject area of objects and take into account the nature of this area, for example, outdoor lighting maintenance [16] (a variety of mounts, shapes, textures, etc.).

To investigate the possibility of recognition by components, it is assumed that objects can be constructed from a set of primitives that includes different prisms and cylinders.

For the experiment, we use the SSD300 VGG16 neural network model. The base network is VGG16. On top of the base network, convolutional layers are added to extract features at different scales. Anchors are placed on feature maps and allow the model to generate bounding box predictions for objects at different scales and aspect ratios. Anchor frames are predefined rectangles of different sizes and ratios placed at different scale levels of feature maps. Synthetic and real images are used to train the neural network.

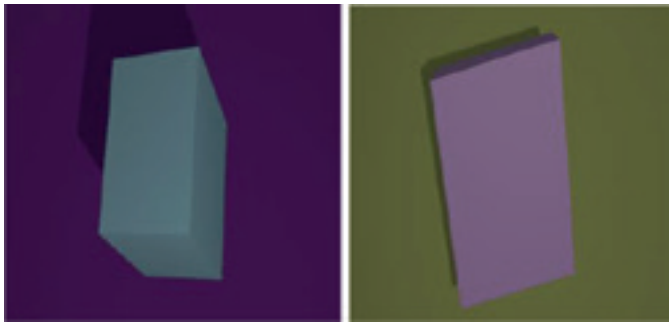


Fig. 1 – Example of synthetic images from the training dataset

The training was carried out for 18 epochs using 1000 synthetic images (800 - for training, 200 - for validation), after which an early stopping function was triggered.

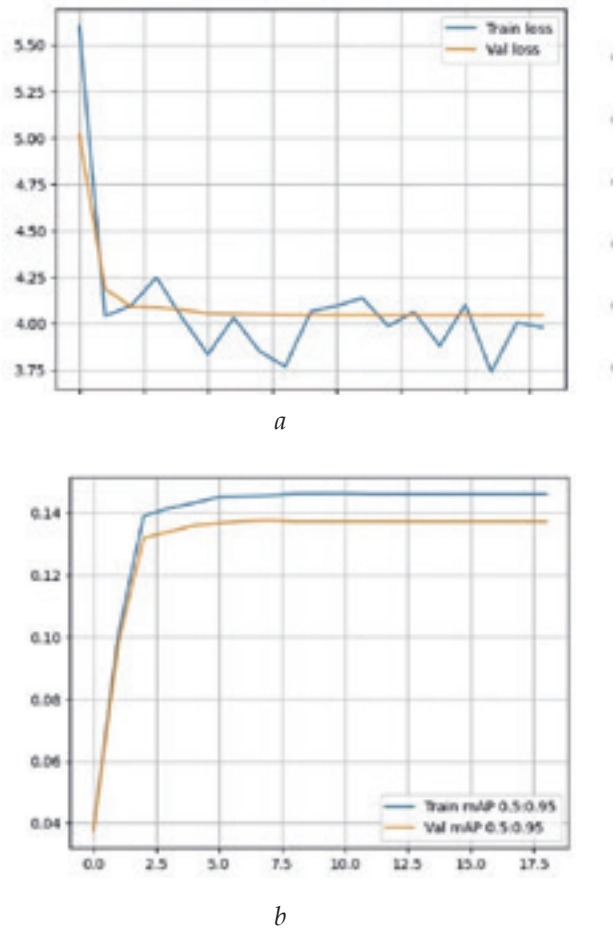
An example of synthetic images is shown in Fig. 1.

Then, the model was trained on mixed data for 14 epochs (50 images of real objects + 1000 synthetic images).

SGD optimizer was used for training. The learning rate was reduced to 0.0001 to get better convergence and avoid overtraining. Momentum is 0.9. Weight decay is 0.0005.

The loss function consists of two components: Bounding box regression loss and Classification loss.

The training plots are shown in Fig. 2 and 3.



a

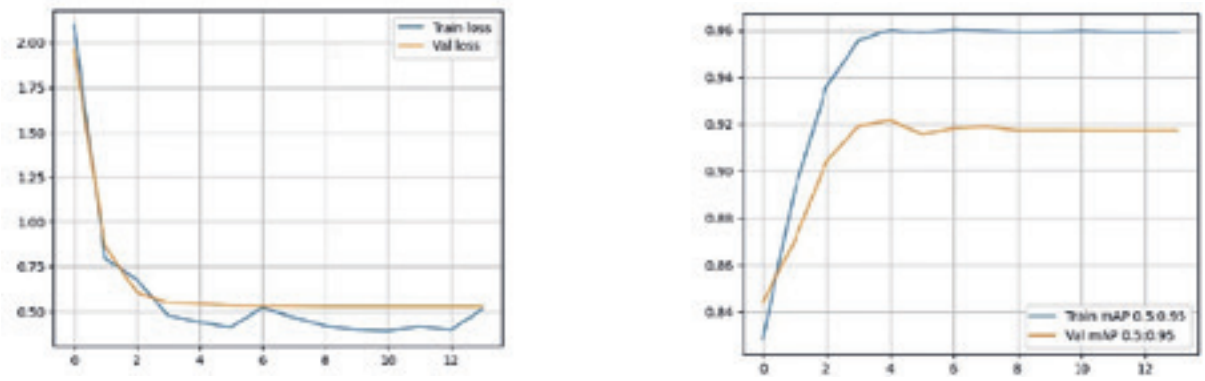
b

Fig. 2 – Training plots for 1000 synthetic images (***a*** – training loss, ***b*** – mAP)

Adding real data to synthetic data leads to an increase in accuracy.

Primitive recognition examples are shown in Fig. 4. As can be seen from the figure, false positives are observed. Therefore, the results require filtering by confidence.

As a result, we get primitive coordinates, which will be used to estimate relations.



a *b*

Fig. 3 – Training plots for 50 real images + 1000 synthetic images (***a*** – training loss, ***b*** – mAP)

Table 1 – Models mAP (mean average precision)

Data	mAP 0.50:0.95	mAP 0.50	mAP 0.75
Synthetic only	0.038	0.077	0.038
Synthetic and Real	0.157	0.336	0.126

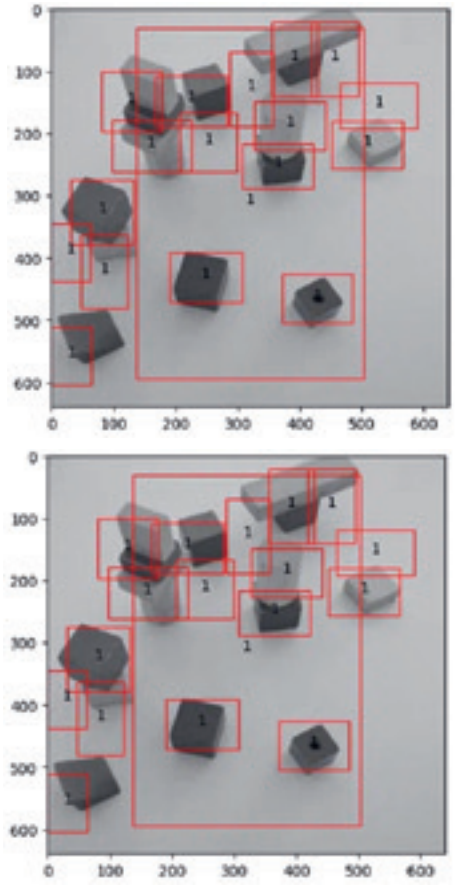
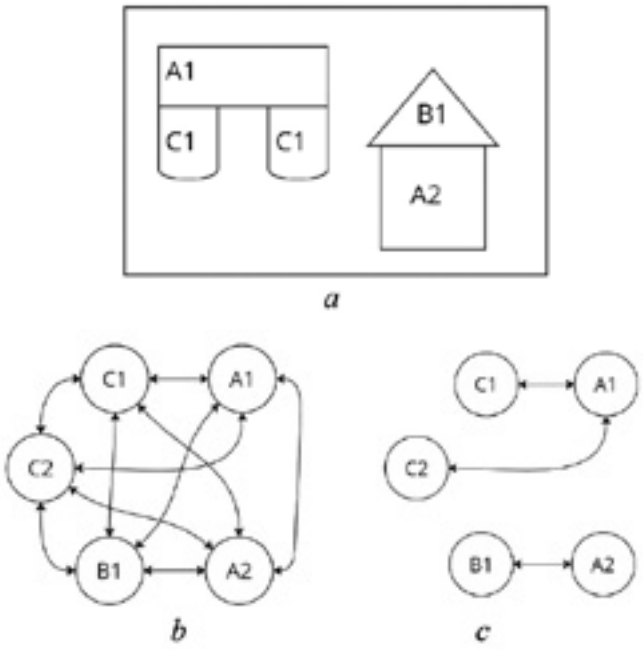


Fig. 4 – Example of recognized primitives

ESTIMATION OF RELATIONS BETWEEN PRIMITIVES

To specify relations between primitives, the representation of objects in the form of graphs is well applicable, where primitives are nodes of the graph, and relations are edges of this graph [5]. In practical implementation, it will allow to store information about recognized objects in explicit form in graph databases.

We use the following labels: A – rectangular prism, B – triangular prism, C – cylinder. In this case, the scene description can be represented as in Fig. 5a.



a *b* *c*

Fig. 5 – Schematic representation of the scene (***a*** – scene, ***b*** – hypothesis generation, ***c*** – relation estimation)

Fig. 5b shows that the list of possible relations between all primitives may be redundant for recognizing an object; therefore, we limit relation set to the relations between contacted primitives.

Using the example of Fig. 5a, we use the following set of relations in two-dimensional space: on the left, on the right, above, below, and intermediate states. As a result, we obtain the following possible variant of the recognized object shown in Figure 6a.

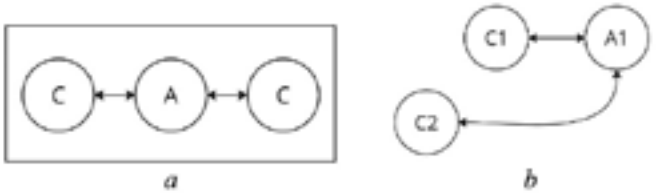


Fig. 6 – Known object and hypotheses of possible objects in the scene (a- known object, b- possible variant of the object in the scene)

Table 2. Relation encoding.

Relation	Letter code	Number	Binary code	Gray's code
Contact + on top	a	0	000	000
Contact + top right	b	1	001	001
Contact + on the right	c	2	010	011
Contact + bottom right	d	3	011	010
Contact + at the bottom	e	4	100	110
Contact + bottom left	f	5	101	111
Contact + on the left	g	6	110	101
Contact + top left	h	7	111	100

To describe the relations between primitives, we introduce their encoding. For encoding, we use Gray's code, an analog of binary coding, each value of which differs from the previous and from the next one by one bit. To describe the relations, eight encoded states need to be introduced, as shown in Table 1 and Fig. 3. Thus, much different relations will have a difference of 2 bits, and neighboring relations will have a difference of 1 bit.

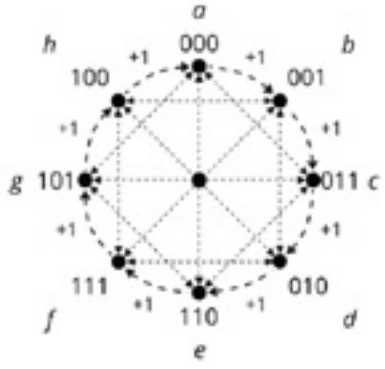


Fig. 7 – Graphical representation of relation encoding (+1 shows the direction of rotation when switching from one state to another)

Thus, the process of object recognition is reduced to the description of known objects and further search for matches in the scene. For example, there is a description of a «house» represented by primitives in Fig. 8, Formula 1.

$$House = \overset{e}{\frown}BA \tag{1}$$

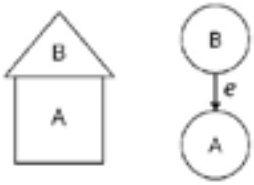


Fig. 8 – Description of the «house» represented as primitives and relations between them.

In such a description of the object, it is necessary to specify that changing the order of primitives when writing should lead to a change of the relation between them to the opposite, e.g., Formula 2.

$$BA = \overset{e}{\frown}\overset{a}{\frown}AB \tag{2}$$

Consider the scene with the objects shown in Figure 9a. The representation of the scene as primitives is as shown in Figure 9b.

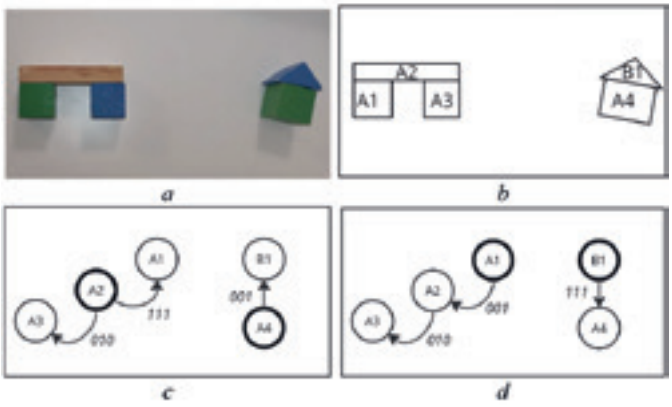


Fig. 9 – Example of a relation estimation between primitives (*a* – scene, *b* – scene representation as primitives, *c* and *d* – relation estimation with different start primitives)

Next, a start primitive has to be selected for recognition. From the start primitive along the chain, relations are estimated between primitives; an example is shown in Figures 9c and 9d. Considering situations when objects in the scene are rotated relative to the description, as, for example, in Figure 9a, it is necessary to check matches by rotating the object in the scene. To do this, a unit is added to or subtracted from all the object relations, and, if a match is found, the object is detected.

The example of recognition for Figure 9c looks as follows:

$$A1B1 \overset{-1}{\Rightarrow} \overset{b}{\frown}\overset{a}{\frown}A1B1 = House \tag{3}$$

By rotating the relations by one step, a match with the searched object is found.

Similarly, a recognition example is presented for Figure 5d. In this case, the primitive B1 is chosen as the

start element and a permutation of the primitives with a change of relations is also applied for recognition.

$$\overset{f}{\frown}\overset{e}{\frown}\overset{a}{\frown} \quad \overset{-1}{\Rightarrow} \quad B1A1 \Rightarrow B1A1 = A1B1 = House \tag{4}$$

As a result, the following steps of the algorithm can be defined (Figure 10).

Step 1. Extraction of a set of primitives from an image (e.g., as in [17]);
Step 2. Estimation of relations between primitives in the scene;
Step 3. Partitioning of the set of primitives into isolated groups. Further steps are performed for each group separately;
Step 4. Checking sufficiency of primitives in the group to contain known objects. If there are enough primitives, go to step 5, if not, go to the next group;
Step 5. Choosing a start primitive, relative to which the chain of relations will be estimated;
Step 6. Constructing a relation graph sequentially from each primitive to all adjacent primitives;
Step 7. Matching search by combinations of primitives and relations, taking into account possible object rotation. If a match is found, add the object to the set of detected objects, remove primitives from the set of primitives. If a match is not found, go to the next combination.
Step 8. If not all the groups are checked, go to step 4, otherwise output the set of detected objects.

Fig. 10 – Component-based object recognition algorithm

IMPLEMENTATION AND APPLICATION OF COMPONENT-BASED OBJECT RECOGNITION ALGORITHM

JSON language has been used for the description of objects and relations between them (listing 1). Consider the scene shown in Figure 11a. The found primitives and the relations between them are presented in listing 2. The recognition result is shown in Fig. 11a.

As can be seen from the above recognition example, the proposed component-based object recognition algorithm can be used to detect objects in an image.

Listing 1. Description example of the object «house» (v1 – description variant, nodes – primitive set, edges - relations)

```
"house": { "v1": {
      "nodes": {
      "A1": {"name": "square_prism"},
      "B1":{"name":"triangular_prism"}},
      "edges": {
      "a1": {"from": "A1", "to": "B1",
            "data": null}
}}}
```

Listing 2. Object recognition example (possible_objects - hypothesis about objects in the image, boxes - coordinates of bounding boxes of detected primitives, relations - relations between detected primitives, version - version of hypothesis description; detected_objects - verified hypotheses about the presence of an object in the image, box - bounding box of the object, class - a certain class of object based on primitives and relations between them).

```
{'arc_0': {'boxes': {'A0': [516, 416, 109, 102],
        'A1': [605, 348, 282, 60],
        'A2': [700, 417, 99, 103]},
        'relations': {('A0', '011', 'A2'),
              ('A1', '010', 'A2')},
        'version': 'v1'},
 'house_1':{'boxes':{'A3':[1167,424,152,115],
        'B4':[1179,352,160,82]},
        'relations':{('A3','000','B4')},
        'version': 'v1'}}
detected_objects:
{'arc_0': {'box': [462, 318, 287, 150],
       'class': 'arc'},
 'house_1': {'box': [1091, 311, 168, 170],
        'class': 'house'}}
```

As can be seen from the above recognition example, the proposed component-based object recognition algorithm can be used to detect objects in an image.

CONCLUSION

The paper presents an approach to object recognition based on the hypothesis that objects can be recognized using primitives and relations between them. An approach to primitive recognition, object description, and relation encoding for two-dimensional image space is presented; object recognition examples in an image are demonstrated. The proposed recognition approach allows to recognize new objects without the need to retrain the algorithm on a new training dataset. To expand the list of recognized objects, it is necessary to expand the database of known objects with the description of a new object or a class of objects. The implementation of the approach will reduce the risks of recognition quality deterioration when the number of recognized object classes increases and reduce labor costs for adapting recognition algorithms for new objects.

The research results can be applied in algorithms for recognizing traffic participants as well as traffic signaling objects.

By comparing the behavior of an identified traffic participant with the detected traffic signaling, its condition can be corrected and the traffic situation managed.

With an appropriate mobile communication network for data transmission and distance monitoring of traffic participants, their speed and distance from each other can be corrected and safe traffic flow can be organized.

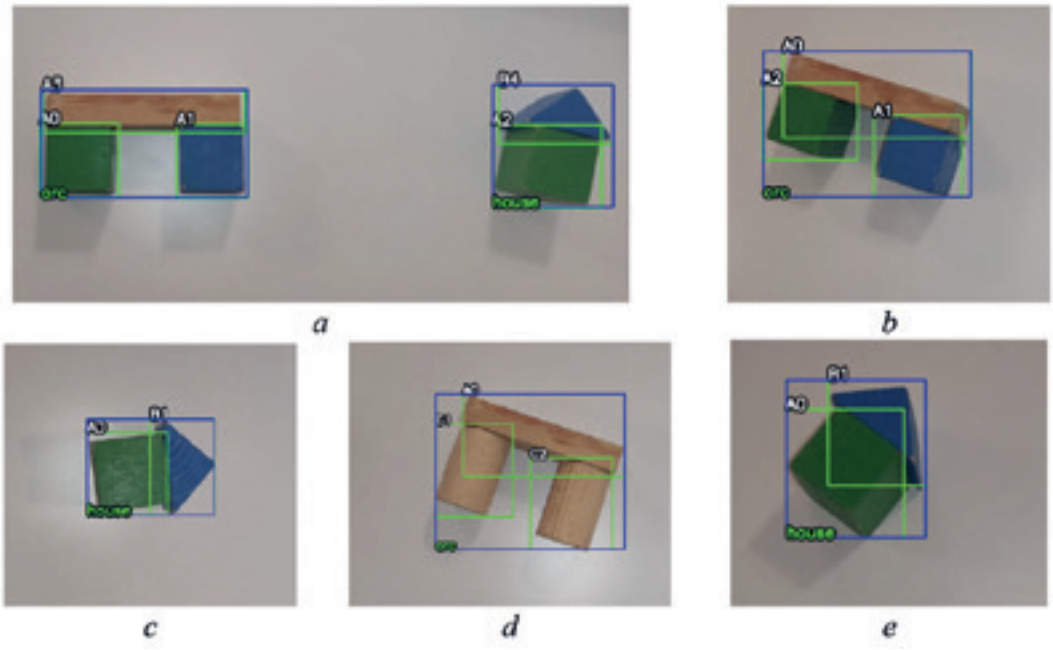


Fig. 11 – Object recognition example using component-based object recognition algorithm

REFERENCES

[1] V. Meel, 'The 87 Most Popular Computer Vision Applications for 2023', 2022. [Online]. Available: https://viso.ai/applications/computer-vision-applications/. [Accessed: 23-Nov-2022].

[2] A. Urbonas, V. Raudonis, R. Maskeliunas, and R. Damaševičius, 'Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning', *Appl. Sci.*, vol. 9, no. 22, 2019.

[3] L. Bureš, I. Gruber, P. Neduchal, M. Hlaváč, and M. Hrúz, 'Semantic text segmentation from synthetic images of full-text documents', *SPIIRAS Proc.*, vol. 18, no. 6, pp. 1380–1405, 2019.

[4] A. N. Oreshin and I. Y. Lisanov, 'A new method for automation of the personnel authentication process using a video stream', *SPIIRAS Proc.*, vol. 5, no. 54, pp. 35–56, 2017.

[5] L. Mylnikov, P. Slivnitsin, and A. Mylnikova, 'Robotic System Operation Specification on the Example of Object Manipulation', *Proc. Int. Conf. Appl. Innov. IT*, vol. 10, no. 1, pp. 51–59, 2022.

[6] I. Biederman, 'Recognition-by-Components: A Theory of Human Image Understanding', *Psychol. Rev.*, vol. 94, no. 2, pp. 115–147, 1987.

[7] P. H. Winston, *Artificial intelligence*, 3rd ed. Addison-Wesley Longman Publishing Co., Inc., 75 Arlington Street, Suite 300 Boston, MA, United States, 1992.

[8] H. M. Gomes, 'Model learning in iconic vision', 2002.

[9] P. Slivnitsin, A. Kniazev, L. Mylnikov, S. Schlechtweg, and A. Kokoulin, 'Influence of Synthetic Image Datasets on the Result of Neural Networks for Object Detection', *Proc. Int. Conf. Appl. Innov. IT*, vol. 9, no. 1, pp. 55–60, 2021.

[10] F. Abramovich and M. Pensky, 'Classification with many classes: Challenges and pluses', *J. Multivar. Anal.*, vol. 174, pp. 1–25, 2019.

[11] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, 'OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks', *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, Dec. 2013.

[12] P. Viola and M. Jones, 'Rapid Object Detection using a Boosted Cascade of Simple Features', in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, pp. 511–518.

[13] N. Dalal and B. Triggs, 'Histograms of oriented gradients for human detection', *Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005*, vol. I, no. 16, pp. 886–893, 2005.

[14] P. Felzenszwalb, D. McAllester, and D. Ramanan, 'A discriminatively trained, multiscale, deformable part model', in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, vol. 330, no. 6, pp. 1–8.

[15] D. H. Wolpert and W. G. Macready, 'No free lunch theorems for optimization', *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.

[16] P. Slivnitsin, A. Bachurin, and L. Mylnikov, 'Robotic system position control algorithm based on target object recognition', in *Proceedings of International Conference on Applied Innovation in IT*, 2020, vol. 8, no. 1, pp. 87–94.

[17] P. Slivnitsin and L. Mylnikov, 'Object Recognition by Components and Relations between Them', *Informatics Autom.*, vol. 22, no. 3, pp. 511–540, May 2023.