DOI: 10.17587/mau.24.533-541

A. D. Voronkov, Postgraduate Student, a.voronkov.rtu@yandex.ru,
S. A. K. Diane, Cand. Tech. Sc., Assistant Professor, diane1990@yandex.ru,
MIREA — Russian Technological University, Moscow, 119454, Russian Federation

Corresponding author: **Voronkov Andrey**, Postgraduate Student, MIREA — Russian Technological University, Moscow, 119454, Russian Federation, e-mail: a.voronkov.rtu@yandex.ru

Accepted on July 3, 2023

Grasping of Unknown Objects with an Autonomous Manipulator: State of the Art, Problems and Prospects

Abstract

To fulfill the practical needs of modern robotics, it is necessary to develop approaches for grasping unknown objects, since in the real world the robot faces a large variety of them. Approaches that imply the availability of complete information about the objects of the working area (3D model, weight and size characteristics) are not practical and can only be used in controlled conditions, such as working on a conveyor with standard details. Therefore, the scientific community and a number of industries are interested in research methods that increase the robot's ability to adapt to new, unfamiliar conditions. This article presents main problems and research directions in the field of visual scene perception and grasping unknown objects by a manipulative robot. We discuss the differences in existing approaches according to various criteria, as well as advantages and disadvantages of existing solutions. The article may be useful to get acquainted with the subject area.

Keywords: *unknown objects grasping, collision avoidance, manipulative robot, machine learning, grasping objects of static scene, unknown objects perception*

For citation:

Voronkov A. D., Diane S. A. K. Grasping of Unknown Objects with an Autonomous Manipulator: State of the Art, Problems and Prospects, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2023, vol. 24, no. 10, pp. 533—541.

DOI: 10.17587/mau.24.533-541

УДК 004.021, 004.896

DOI: 10.17587/mau.24.533-541

А. Д. Воронков, аспирант, a.voronkov.rtu@yandex.ru,
С. А. К. Диане, канд. техн. наук, доц., diane1990@yandex.ru,
МИРЭА — Российский технологический университет, г. Москва

Захват неизвестных объектов с помощью автономного манипулятора: современное состояние, проблемы и перспективы

Для выполнения практических задач, стоящих перед современной робототехникой, требуется разработка подходов к захвату незнакомых объектов, поскольку в условиях реального мира робот сталкивается с большим их разнообразием. Подходы, подразумевающие наличие полной информации об объектах рабочей области (3D-модель, массогабаритные характеристики), не практичны и могут быть использованы только в контролируемых условиях, таких как работа на конвейере с типовыми деталями. Поэтому научное сообщество и ряд отраслей промышленности заинтересованы в исследовании методов, повышающих способность робота адаптироваться к новым, незнакомым условиям.

В данной статье приводится подборка основных направлений в задачах визуального анализа сцены и захвата неизвестных объектов манипуляционным роботом. Рассмотрены отличия существующих подходов по различным критериям, преимущества и недостатки имеющихся решений. Статья может быть полезна для ознакомления с предметной областью.

Ключевые слова: *захват незнакомых объектов, избегание столкновений, манипуляционный робот, машинное обучение, захват объектов статической сцены, восприятие незнакомых объектов*

Introduction

Grasping various objects is an integral part of applied tasks for industrial and service manipulative robots. There are a large number of works devoted to this problem in the scientific literature, but the problem of grasping unknown objects in heavily cluttered scenes with a success rate close to 100 % has not been solved yet.

The task of grasping objects of unknown shape by a manipulative robot can be considered in the form of several subtasks (Fig. 1): the task of perceiving the scene, implying receiving data from the onboard sensor subsystem, processing and analyzing this data; the task of synthesizing the optimal grasping configuration and the task of planning movement to the grasping configuration with collision avoidance.

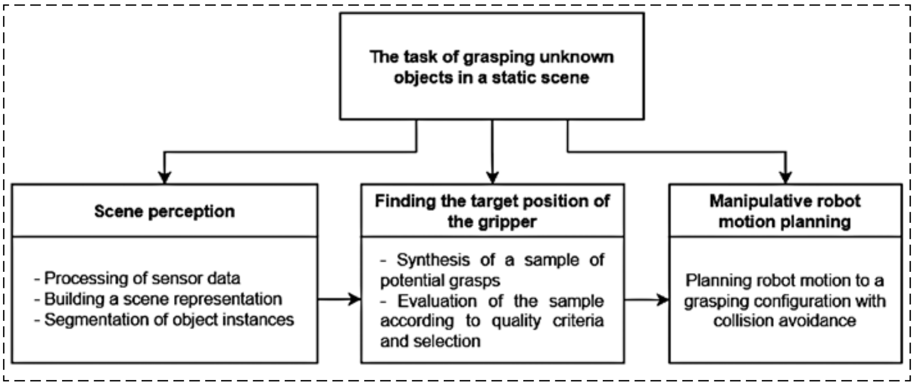


Fig. 1. Decomposition of the problem of capturing unknown objects of a static scene

The main difficulties in the grasping task are such aspects as: clutteriness of the scene, complicating perception due to overlaps between objects; imperfection of the onboard sensor subsystem, leading to noisy and distorted measurements; the presence of objects of unknown shape on the scene; the need of grasping in real time, imposing a strict limitation of search algorithms.

The ability of the manipulative robot to grasp objects regardless of their visual and physical properties, such as shape, color, texture, as well as in conditions of severe clutter and changing lighting opens up wide opportunities for such applied tasks as cleaning [1], sorting [2, 3], working in warehouses [4] and emergency rescue operations [5].

The task of perceiving a static scene

Obtaining scene representation. At the stage of scene perception, a generalized representation of the working environment of the manipulative robot is formed, which will then serve as the basis for the methods of grasp synthesis. Usually, an RGBD-camera or a depth-camera is used at this stage. The use of an RGBD image is preferable to the use of an RGB image, since it contains more information and serves as a source of features allocated by the neural network model. In existing studies, different tactics are used to obtain initial data:

- obtaining an RGBD image or a depth image from one camera position (Fig. 2, *a,* see the second side of the cover). In this case, the camera position is fixed on the end link of the robot [6, 7] or above the stage [8—12]. Approaches involving a fixed camera position are applicable in controlled environments and have the advantage that the robot does not spend time moving the camera over the scene to scan before grasping the object;
- merging of individual RGBD or depth images into a single point cloud of the scene. In this case, the manipulative robot moves the camera in several positions, set, for example, by spherical coordinates, while taking pictures of the scene [13];
- continuous integration of RGBD or depth images into a point cloud while following the scanning trajectory [13, 14].

In [14], point cloud building strategies are classified into an active — movement of a wrist-mounted camera along a complex trajectory with a constant direction to the center of the scene, and a passive — merging of images from two cameras with known positions mounted on the robot body. Thus, moving along the scanning trajectory allows to get significantly more information, while with a passive strategy, part of the scene is not perceived. Exploratory movements of the robot in order to obtain depth images from different camera positions, for building a more detailed point cloud, may not be possible if the robot works in a limited space, since cluttered scenes significantly limit the robot's working area [15].

The possibility of using depth images in grasp synthesis methods is also due to the fact that a training sample with realistic images can be obtained by means of virtual modeling. At the same time, a neural network model trained on an artificially created training set can be applied in the real world without additional model changes [7, 13]. In addition to point clouds, during the operation of the grasp synthesis method, a representation of the scene in the form of a voxel grid can be used to evaluate the collision of the gripper with scene objects [6] and a truncated signed distance field (TSDF) representation as input data of a neural network [13] (Fig. 2, *b*, see the second side of the cover).

In the template library-based approach [17], a convex hull consisting of polygons is constructed around an object previously segmented on the plane of the table. Vectors of normals to the centers of polygons of the resulting shell are used to create a discrete sample of potential positions of the gripper. A similar approach can be found in [18], where a limiting 3D bounding box is used, the selection of points in the faces of which allows to get a set of potential positions of the gripper. Thus, the representation of scene objects in the form of polygonal surfaces is applicable.

Scene segmentation. An important role for the application of grasp synthesis methods in practice is played by their ability to segment a set of objects of a complex static scene into separate instances. Classical computer vision algorithms such as morphological operations, Canny boundary detector, superpixel extraction [19], watershed algorithm, etc. are applicable only as an intermediate stage of image processing, since they can lead to over-segmentation [8] and are able to provide information about object instances only in the simplest or controlled cases. Therefore, the most qualitative solution to the problem of segmentation of unfamiliar objects is performed using neural network technologies [20—24].

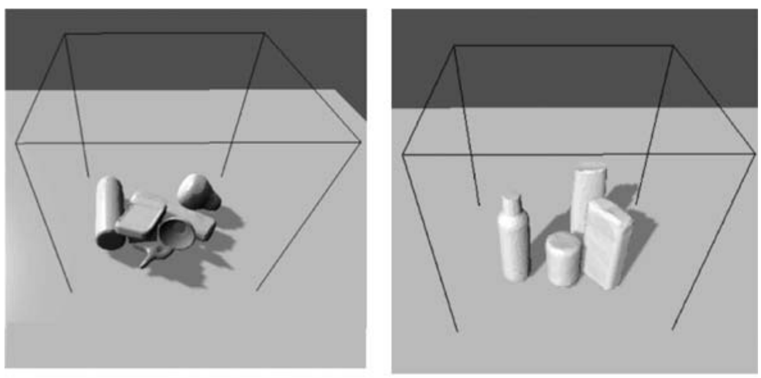
Segmentation of scene instances avoids the situation when the robot cannot determine the boundaries of objects and grasps two objects simultaneously [14]. At the same time, such a disadvantage is insignificant if the applied task is to remove all objects from the scene. It is worth noting that in undemanding tasks, many grasp generation methods resort to selecting a flat surface using the RANSAC model parameter estimation algorithm and thus obtain information about point clouds of loosely spaced objects or a single scene object [15].

Existing methods for image segmentation based on machine learning show good results. In [20], the modified architecture of the Mask R-CNN neural network is investigated for segmentation of objects without defining their category. Authors of [21] explore segmentation using a feature map obtaining via neural network. Clusters of points with the same features are considered as one object, and thus segmentation of instances of unknown objects is carried out (Fig. 3, *a*, see the second side of the cover). An approach presented in [22] uses a two-stage process with two autoencoders for regression and refinement of segments of unknown objects. In [23] the authors use the hierarchical structure of the neural network to perform segmentation of the scene and receive information not only about visible segments, but also about the expected segments of parts of objects invisible on the frame (Fig. 3, *b*, see the second side of the cover).

The neural network architecture PointNet [24] and its improved version PointNet++ [25] allow to extract features directly from three-dimensional data, such as point clouds. This architecture was originally developed for use in classification and segmentation tasks, so it is being used in recent studies to extract the properties of the original point cloud and generate a set of potential grasps.

Types of static scenes. A number of studies considers two types of cluttered scenes that affect the distribution of potential grasps. In the works [6, 9, 13], cluttered scenes with a random arrangement of objects and scenes consisting of objects arranged in order are considered (Fig. 4).

A feature of scenes with randomly arranged objects is the random position and orientation of all objects. An example of such scenes in the real world can be a blockage, a set of unsorted objects. In this kind of scenes, the potential poses of the gripper are oriented mostly vertically. Cluttered scenes with an ordered arrangement of objects contain objects stably located on the surface. An example of such a scene is a set of tightly arranged objects on a shelf. In this kind of scene, the potential poses of the gripper can be distributed both vertically and horizontally. In scenes with an ordered arrangement of objects, an unintentional collision of a gripper with an object leads to much more noticeable consequences than in random scenes: an object standing vertically can fall, touch other objects, greatly changing the scene. In applied tasks, this may be extremely undesirable [6]. Thus, an important role in evaluating the effectiveness of the method is associated with its verification on both types of scenes reflecting real-world conditions.



a) *b)*

Fig. 4. Static scenes [13]:
a — with a random objects arrangement; *b* — with an ordered objects arrangement

Comparison of approaches to scene perception in existing studies

<table>
<tr><td>Parameter</td><td colspan="30">Realization</td></tr>
<tr><td rowspan="2">Sensor type</td><td colspan="15">RGBD-camera</td><td colspan="15">Depth-camera</td></tr>
<tr><td colspan="15">[6][7][8][9][10][11][14][16]</td><td colspan="15">[12][13][17]</td></tr>
<tr><td rowspan="2">Sensor location</td><td colspan="15">Fixed statically</td><td colspan="15">Fixed on the robot link</td></tr>
<tr><td colspan="15">[8][9][10][11][12][16][17]</td><td colspan="15">[6][7][13][14]</td></tr>
<tr><td rowspan="2">Scene representation obtaining strategy</td><td colspan="15">One-position image</td><td colspan="15">Merging images from multiple positions</td></tr>
<tr><td colspan="15">[6][7][8][9][10][11][12][16][17]</td><td colspan="15">[13][14]</td></tr>
<tr><td rowspan="2">Scene representation</td><td colspan="6">Point cloud</td><td colspan="6">Voxelized</td><td colspan="6">TSDF</td><td colspan="6">N-channel image</td><td colspan="6">Polygonal</td></tr>
<tr><td colspan="6">[6][7][8][9][12][14][16]</td><td colspan="6">[6]</td><td colspan="6">[13]</td><td colspan="6">[10][11]</td><td colspan="6">[17]</td></tr>
<tr><td rowspan="2">Segmentation method (if applied)</td><td colspan="10">Classical methods of image processing</td><td colspan="10">Clustering</td><td colspan="10">Neural networks</td></tr>
<tr><td colspan="10">[8]</td><td colspan="10">[7][16][17]</td><td colspan="10">[6][9][10]</td></tr>
</table>

Summarizing the analysis of existing approaches to the perception of the scene, we can group them into Table 1.

Manipulative robot motion planning task

The task of grasping in number of studies is considered as the task of synthesizing such a position of the gripper that would allow the manipulative robot to reliably grasp the object, avoiding collisions with other objects of the scene [6, 7, 9, 13, 15, 26]. Thus, the grasping-based task of cleaning the workspace from objects [10] implies the sequential removal of objects from the workspace, simulating the applied tasks of analyzing the blockage and cleaning the room. The scientific literature presents many approaches that solve this problem, which differ in the complexity of implementation; the necessary equipment for scene perception and data processing; computational complexity. They will be discussed in more detail.

Formalized description of the position and orientation of the gripper. A formalized description of the position and orientation of the gripper plays an important role for the effectiveness of the grasp synthesis methods. Usually, when synthesizing the grasping configurations of a manipulative robot, 2 ways of representing the pose of the gripper are considered, differing in the number of degrees of freedom:

1) representation of a pose with 3 degrees of freedom [11, 27—29]. In this case, the pose of the gripper is described by 3 independent parameters relative to the image plane: (x, y, θ), where (x, y) are the coordinates of the center of the gripper device in the image, θ is the angle of rotation of the gripper relative to the horizontal axis. The magnitude and direction of the longitudinal displacement of the gripper is determined from the coordinates of the grasping center point in the image and the normal vector to the surface at this point [11]. The use of such a representation greatly limits the variety of grasps considered by those whose approach direction is perpendicular to the image plane. As a result, grasping an object may be kinematically impossible;

2) pose representation with 6 degrees of freedom [6, 7, 9]. In this case, the pose of the gripper is defined by 6 independent parameters: $(x, y, z, \alpha, \beta, \gamma)$, where (α, β, γ) are 3 Euler angles and (x, y, z) are 3 projections of the translation vector. The advantage in such approaches is the ability to grasp objects from any side. The disadvantage is the fact that position regression in three-dimensional Euclidean space is more complicated than regression in two-dimensional space [9];

3) apart from the above mentioned parametrization methods extended approach is used for the case of a multifingered grippers. Depending on the complexity of the gripper kinematics, a bigger number of parameters is required to describe its pose and orientation [26]. Nonetheless many researchers simplify the pose estimation task and reduce the optimization problem to 3 or 6 parameters.

In [9], the authors rely on the observation that in most cases one of the contact points of a parallel gripper with an object lies on the visible part of the surface. This observation allowed them to use the dimension of the representation of the gripper equal to 4 in order to optimize regression. Thus, the learning process of the neural network was facilitated and, moreover, the accuracy of estimating the position of the gripper increased.

Target object awareness for the grasping algorithm. In addition to the grasp representation, the existing approaches for the presence of the target object of manipulation can be classified into the following:

1) approaches using only an array of source data without selection of a target object [12—14]. According to these approaches, collision avoidance grasps are synthesized for the entire scene as a whole, without dividing it into separate objects. This can lead to the grasping of two objects at the same time, since there is no functional for their differentiation (Fig. 5, *a*, see the second side of the cover). This type of approach can be used in applied tasks that do not require the grasping of the target object. The advantage of these methods is a simplified interpretation of the grasping task and less complexity due to skipping the segmentation step;

2) target oriented approaches — in this group of approaches, there is a matching between synthesized grasps and specific target objects [6, 7, 30] (Fig. 5, *b*, see the second side of the cover). Target oriented approaches have an advantage over approaches without specifying a target object, since they are more applicable in practical usage.

Grasp planning based on pattern matching. Approaches involving the use of known objects present in the robot's workspace, such as [31, 32], are not always compatible with the requirement to work in uncontrolled working environments. However, due to the availability of this information, broad generalizing abilities are not required from the perception system, as a result of which such approaches benefit in speed and quality of grasping. In the real world, a manipulative robot can encounter a large number of objects of different categories and with different physical parameters, such as shape, weight and texture, so one way to generalize information is to use a template library.

In [17], the authors present an algorithm for grasping unknown objects based on matching with a template library. The authors rely on the hypothesis that objects similar in shape can be grasped in a same way. Initially, the database with templates is filled by kinesthetic training of the robot on some training sample. The templates selected using the descriptor developed by the authors are then compared with the data from the depth camera during the robot's operation and, in case of successful matching, grasps from the library are used. During the operation of the algorithm, the library increases its efficiency due to feedback. During the experimental evaluation of the algorithm, it showed from 62 to 87 % success rate of grasps. The calculation of the ranked list of grasps for the on-board computer of the robot PR2 took from 5 to 30 seconds. The advantage of the algorithm is the use of descriptors to generalize templates to different objects.

The category of approaches involving object models includes methods related to the use of a family of superquadrics. So, in [16], the selected and processed point cloud of the object is approximated and replaced by a superquadric in order to then search for the position of the gripper. The advantage of superquadrics is the possibility of defining them by 11 parameters and the simplicity of obtaining a point cloud of a superquadric. In order to ensure the reliability of the grasp, the criteria of the distance to the centroid of the superquadric and the curvature of the contact areas are used.

In addition to explicitly specifying models of workspace objects, there are studies aimed at reconstructing the shape of an incomplete point cloud. In this way, in [33], the shape of an object is reconstructed using a random forest model along a part of the surface. However, since the reconstruction process is computationally expensive, the most effective approaches use machine learning to output grasps in one pass of a network, in the hidden space of which possible grasps are mapped to an incomplete cloud of object points.

Grasp planning using machine learning. The most advanced results in the grasping task are achieved by using machine learning [6, 7, 9, 10—13, 27, 28, 34], because neural networks are a powerful tool for extracting features from input data. Most training-based approaches use an artificially created sample of training data and are thus transferred to real-world conditions without changes. When using machine learning in the task of grasping unknown objects, 3 types of approaches are used:

- reinforcement learning, where the robot interacts with the environment, receiving responses [35, 36]. The disadvantage of this approach is the impossibility of its fast deployment in a new environment;
- supervised learning with an annotated training data [6, 7, 13]. To create a training data, most methods of grasp synthesis involve modeling complexes;
- forming of a motion plan based on maximizing the objective function. This approach can use evolutionary algorithms [37], decision trees [38].

Approaches based on reinforcement learning [35, 36] consider the grasping task as learning a sequence of actions, as a result of which the target object will be grasped. To do this, it may be necessary to remove or move the blocking object, so 2 possible actions are considered in the process of completing the task:

- grasping — positioning the gripper near the target object in order to grasp it;
- pushing — shifting of an object in order to provide free space near another object.

Methods generating an initial sample of grasps based on the assessment of surface normals do not cope with objects containing thin parts and incomplete point clouds [14]. Neural network approaches are devoid of this disadvantage usually. However, in [9] it is noted that in existing approaches using complex algorithms consisting of several stages, there are disadvantages, expressed in the presence of several potential points of failure and in a relatively long execution time, which does not allow working in real time. Thus, such software-algorithmic complexes that use the smallest number of components are more preferable. So, in [9], a regression of a set of 4 parameters describing the pose of the gripper is carried out based on a cloud of scene points obtained using an RGBD-camera. To do this, the authors train a neural network with the PointNet++ architecture. The set of potential grasping configurations generated by the neural network is analyzed, and grasps with a high score are selected. The advantage of the method is that it does not require precise segmentation of the target object, uses a reduced dimension of the output space and has a high speed compared to other methods.

In practical applications robots often encounter scenes consisting of several objects, so there is a need to take into account possible undesirable collisions. So, in [7], an autoencoder based on PointNet generates a set of grasps for an isolated object. The authors solve the problem of collision avoidance by evaluating grasps taking into account the point cloud of the scene using a second autoencoder and using an algorithm for iterative refinement of grasps. Iterative adjustment of the found grasps plays an important role, since in some cases only a small change in its position or orientation is necessary to ensure the success of the grasp. Algorithms that work with scenes containing several objects are more applicable in practice [11], but at the same time they need to avoid collisions.

In a special group, it is worth highlighting the methods of grasp detection [6, 7, 13, 14, 30]. In this group of approaches, the task of finding the target poses of the gripper is considered similarly to the problem of detection in computer vision. Most grasp detection algorithms use some method to generate potential grasps based on the scene representation obtained at the perception stage. This generation of a variety of grasps facilitates the search for kinematically feasible and high-quality grasps. Usually the grasp detection algorithm consists of 2 main stages: generating a sample of grasps; evaluating the generated grasps and selecting the optimal ones according to one or more criteria (Fig. 6, see the second side of the cover).

Summarizing the analysis of existing approaches for grasping unknown objects, they can be grouped into Table 2.

Motion planning with collision avoidance. Motion planning of a manipulative robot with obstacle avoidance is crucial for ensuring the safe, efficient and long-term operation of robotic systems. It is also essential to avoid collisions with the surrounding environment for maximization of grasping success rate. There are several algorithms available for this task, including:

Table 2

Existing approaches for grasping unknown objects

Parameter	Realization				
Selection of target object	Yes		No		Single object
	[6][7][9][10][11][16] [30][31][35]		[8][12][13][36]		[17] [26][27][28][29][37][38]
The method for finding gripper pose	Analytical	Neural networks	Templates library	Decision trees	Genetic algorithm
	[8][16][26][31]	[6][7][9][10][11][12][13][27][28] [29] [30][35] [36]	[17]	[38]	[37]
Auxiliary actions	Yes		No		
	[10][35][36]		[6][7][8][9][11][12][13][16][17][26][27][28][30][31][37][38]		
Collision avoidance	Yes		No		
	[6][7][8][9][10][11][12][13][26][35][36][38]		[16][17][27][28][29][30][31][37]		

- artificial potential field algorithm (APF) — the algorithm assigns attractive forces toward the goal and repulsive forces from obstacles [39];
- rapidly exploring random tree (RRT) — where the basic idea is exploring high-dimension space in the form of the tree [40];
- genetic algorithm (GA) — the algorithm is able to achieve target position without collisions and singularity with specific objective functions [41].

It should also be noted that employing these techniques requires either extensive equipment of the manipulator with obstacle sensors, or application of environment mapping algorithms, based on video analysis.

Assessment of grasp synthesis algorithms

Algorithm efficiency criteria. Evaluating of effectiveness of methods for the grasp synthesis plays important role since it allows to judge how reliable potential grasps are and how they are distributed. To evaluate the performance of the method for the grasp synthesis, various metrics can be used, including those based on heuristics, but the most common are the following:

- success rate — the percentage of successful grasps from the total number of grasps determined experimentally. In the case of grasping unknown objects, this metric reflects the generalizing abilities of the system, as well as the reliability of the generated grasps [7, 13];
- coverage — the proportion of coincidence of synthesized grasps with ground truth grasps. The high value of this metric reflects the fact that the synthesized grasps are distributed in the same way as the grasps of the training data, which, usually, are arranged so that the object can be grasped in different ways. Such a need arises in highly cluttered scenes, where several alternative ways of accessing the object are required [7, 13];
- recall at high precision — a metric introduced in [14] that evaluates coverage at a certain threshold value of the acceptability of the grasp classifier. This metric can be used in grasp detection methods, where a set of grasps is first synthesized, which is then evaluated using a classifier;
- planning time — the time between getting a view of the scene and getting a list of feasible grasps [13].

It is worth noting that the key feature of recall at high precision is that it evaluates the grasp detection system as a whole — simultaneously reflecting both the quality of the grasp synthesis subsystem and the accuracy of the classifier.

Quality criteria for evaluating synthesized potential gripper poses. For the selection of executable grasps, their ranking is required in accordance with robustness. The paper [42] provides methods for calculating criteria that determine the qualitative indicator of potential grasp. The authors identify several groups of analytical criteria based on:

- algebraic properties of the grasping matrix — the analysis of the grasping matrix, that is, the matrix connecting the contact forces acting on the object from the gripper with the general effect of fingers on the object, allows to assess the possibility of this grasping configuration to resist external disturbances;
- geometric properties of the grasp — when calculating these criteria, the geometric relationships between the contact points are taken into account. Thus, the degree of stability of the grasp is reflected, i.e. its ability to resist contact slippage;
- contact forces limitations — this set of criteria reflects the ability of the grasp to resist disturbing influences, taking into account the limitations of the forces applied to the object of manipulation by the gripper fingers;
- gripper configuration — this set of criteria is calculated based on the Jacobi grasping matrix, i.e. the matrix linking the gripper space and the manipulation object space. Thus, the degree of closeness of the gripper to the singular configuration is estimated.

The paper notes that the grasp assessment should be carried out in accordance with several criteria included in the overall assessment with weighting coefficients.

The ability of the grasp to resist disturbances and immobilize the object is determined by the properties of form closure and force closure [43]. The property of form closure is provided by the arrangement of contact areas in such a way as to restrict the movement of the object in space. The force closure reflects the ability of the forces applied to the object of manipulation by the fingers of the gripper to resist slipping and movement of the object. In the study [14], the search for antipodal frictionless grasps[1], is carried out, since with a non-zero coefficient of friction they are grips closed in force.

[1]Antipodal frictionless grasp is a grasp in which the vectors of the normal to the contact points are directed opposite to the direction of movement of the corresponding griper fingers and are colinear [44].

Conclusion

It can be concluded that the problem of grasping arbitrary objects by a manipulative robot is relevant and is currently widely discussed in the scientific community. Of great interest are the methods of detecting potential gripper poses and synthesis of grasping movements presented in recent studies, which allow for reliable grasp of an arbitrary object with obstacle avoidance. These methods are mainly based on the use of machine learning. However, to date, a sufficiently universal and reliable method has not been developed yet, the effectiveness of which would be comparable to the grasps performed by a person.

The most applicable and universal method in practice seems to be one that would provide a short solution output time, separation of the initial representation of a static scene into object instances with matching grasps to each object, as well as collision avoidance. Despite the fact that usually in recent works one of the possible strategies for constructing a point cloud is used, the possibility of using both passive and active point cloud construction strategy, depending on the limitations of the workspace, would be an advantage.

References

1. **Liu J., Balatti P., Ellis K., Hadjivelichkov D., Stoyanov D., Ajoudani A., Kanoulas D.** Garbage collection and sorting with a mobile manipulator using deep learning and whole-body control, *2020 IEEE-RAS 20th International Conference on Humanoid Robots (Humanoids),* Munich, Germany, 2021, pp. 408—414.

2. **Zhang H., Liang H., Ni T., Huang L.** Research on multi-object sorting system based on deep learning, *Sensors,* 2021, vol. 21, no. 18, 21 p.

3. **Guerin J., Stephane T., Nyiri E., Gibaru O.** Unsupervised robotic sorting: towards autonomous decision making robots, *International Journal of Artificial Intelligence & Applications,* 2018, vol. 9, pp. 81—98.

4. **Bormann R., Brito B., Lindermayr J., Omainska M., Patel M.** Towards automated order picking robots for warehouses and retail, *Computer Vision Systems. ICVS 2019. Lecture Notes in Computer Science,* 2019, vol. 11754, Springer, Cham.

5. **Habibian S., Dadvar M., Peykari B.** Design and implementation of a maxi-sized mobile robot (Karo) for rescue missions, *Robomech*, 2021, vol. 8, no. 1, 33 p.

6. **Murali A., Mousavian A., Eppner C., Paxton C., Fox D.** 6-DOF grasping for target-driven object manipulation in clutter, *ICRA*, 2020 (virtual), 2020, 7 p.

7. **Mousavian A., Eppner C., Fox D.** 6-DOF GraspNet: variational grasp generation for object manipulation, *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 2901—2910.

8. **Sauvet B., Levesque F., Park S., Cardou P.** Model-based grasping of unknown objects from a random pile, *Robotics*, 2019, vol. 8, no. 3, 18 p.

9. **Sundermeyer M., Mousavian A., Triebel R., Fox D.** Contact-GraspNet: efficient 6-DoF grasp generation in cluttered scenes, *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xian, China, 2021, pp. 13438—13444.

10. **Huang B., Han S., Boularias A., Yu J.** DIPN: Deep interaction prediction network with application to clutter removal, *Conference: 2021 IEEE International Conference on Robotics and Automation (ICRA),* Xian, China — 2021, pp. 4694—4701.

11. **Zhang H., Lan X., Bai S., Zhou X., Tian Z., Zheng N.** ROI-based robotic grasp detection for object overlapping scenes, *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS),* Macau, China — 2019, pp. 4768—4775.

12. **Qin Y., Chen R., Zhu H., Song M., Xu J., Su H.** S4G: amodal single-view single-shot SE(3) grasp detection in cluttered scenes, *3rd Conference on Robot Learning (CoRL 2019),* Osaka, Japan, 2019, pp. 53—65.

13. **Breyer M., Chung J., Ott L., Siegwart R., Nieto J.** Volumetric grasping network: real-time 6 DOF grasp detection in clutter, *4th Conference on Robot Learning (CoRL 2020)*, 2020, pp. 1602—1611.

14. **Pas A., Gualtieri M., Saenko K., Platt R.** Grasp pose detection in point clouds, *The International Journal of Robotics Research*, 2017, vol. 36, pp. 1455—1473.

15. **Lei Q., Meijer J., Wisse M.** A survey of unknown object grasping and our fast grasping algorithm-C shape grasping, *2017 3rd International Conference on Control, Automation and Robotics (ICCAR),* Nagoya, Japan, 2017, pp. 150—157.

16. **Abhijit M., Federico T., Perez-Gracia A.** Grasping unknown objects in clutter by superquadric representation, *2018 Second IEEE International Conference on Robotic Computing (IRC),* California, USA, 2018, pp. 292—299.

17. **Herzog A., Pastor P., Kalakrishnan M., Righetti L., Bohg J., Asfour T., Schaal S.** Learning of grasp selection based on shape-templates, *Autonomous Robots*, 2014, vol. 36, pp. 51—65.

18. **Kappler D., Bohg J., Schaal S.** Leveraging big data for grasp planning, *IEEE International Conference on Robotics and Automation (ICRA*), 2015, pp. 4304—4311.

19. **Savvin S. V., Sirota A. A.** Superpixel segmentation methods and their application for image analysis with heterogeneous texture, *Vestnik VGU, seriya: sistemnyj analiz i informacionnye tekhnologii,* 2016, no. 4, pp. 165—173 (in Russian).

20. **Danielczuk M., Matl M., Gupta S., Li A., Lee A., Mahler J., Goldberg K.** Segmenting unknown 3D objects from real depth images using Mask R-CNN trained on synthetic data, *2019 International Conference on Robotics and Automation (ICRA),* 2019, pp. 7283—7290.

21. **Xiang Y., Xie C., Mousavian A., Fox D.** Learning RGB-D feature embeddings for unseen object instance segmentation, *4th Conference on Robot Learning (CoRL 2020),* 2020, pp. 461—470.

22. **Xie C., Xiang Y., Mousavian A., Fox D.** Unseen Object Instance Segmentation for Robotic Environments, *IEEE Transactions on Robotics*, vol. 37, no. 5, pp. 1343—1359.

23. **Back S., Lee J., Kim T., Noh S., Kang R., Bak S., Lee K.** Unseen object amodal instance segmentation via hierarchical occlusion modeling, *2022 International Conference on Robotics and Automation (ICRA),* Philadelphia, PA, US, 2022, pp. 5085—5092.

24. **Charles R., Su H., Kaichun M.,Guibas L.** PointNet: Deep learning on point sets for 3D classification and segmentation, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, USA, 2017, pp. 77—85.

25. **Charles R., Yi L., Su H., Leonidas J.** PointNet++: deep hierarchical feature learning on point sets in a metric space, *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17),* Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 5105—5114.

26. **Leskov A. G., Seliverstova E. V.** Algorithm for planning and selecting a method for grasping a deformable object with a multi-fingered gripper of a manipulative robot, *Mekhatronika, Avtomatizaciya, Upravlenie*, 2017, vol. 18, no. 11, pp. 739—744 (in Russian).

27. **Mais Zh.** Application of convolution neural networks to train a manipulative robot to capture objects, *Politekhnicheskij molodezhnyj zhurnal*, 2020, no. 1, 5 p. (in Russian).

28. **Lenz I., Lee H., Saxena A.** Deep learning for detecting robotic grasps, *The International Journal of Robotics Research,* 2013, no. 34, pp. 705—724.

29. **Redmon J., Angelova A.** Real-time grasp detection using convolutional neural networks, *Proceedings — IEEE International Conference on Robotics and Automation*, 2015, pp. 1316—1322.

30. **Li T., Wang F., Ru C., Jiang Y., Li J.** Keypoint-based robotic grasp detection scheme in multi-object scenes, *Sensors*, 2021, vol. 21, no. 6, 15 p.

31. **Abdullov T. S., Muhametov A. N.** Robotic grasping of objects of known shape using a monocular camera, *Fiziko-matematicheskie, estestvenno-nauchnye i social'nye aspekty sovremennogo razvitiya nauki, tekhniki i obshchestva. Materialy I gorodskoj molodyozhnoj nauchnoj konferencii Fiziko-matematicheskie, estestvenno-nauchnye i social'nye aspekty sovremennogo razvitiya nauki, tekhniki i obshchestva. Materialy I gorodskoj molodyozhnoj nauchnoj konferencii,* Kazan — 2021, pp. 3—6 (in Russian).

32. **Agboh W., Dogar M.** Real-time online re-planning for grasping under clutter and uncertainty, *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, 2018, 8 p.

33. **Firman M., Aodha O., Julier S., Gabriel J.** Structured prediction of unobserved voxels from a single depth image, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5431—5440.

34. **Rawan A., Al Akkad M. A.** Machine learning in robotic grasping tasks: a survey, *Priborostroenie v XXI veke — 2020. Integraciya nauki, obrazovaniya i proizvodstva. Cbornik materialov XVI Vserossijskoj nauchno-tekhnicheskoj konferencii,* Izhevsk, 2021, pp. 164—170.

35. **Boularias A., Bagnell J. A. Stentz A.** Learning to manipulate unknown objects in clutter by reinforcement, *AAAI'15: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 1336—1342.

36. **Zeng A., Song S., Welker S., Lee J., Rodriguez A., Funkhouser T.** Learning synergies between pushing and grasping with self-supervised deep reinforcement learning, *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4238—4245.

37. **Zhang Z., Gu J.** Grasp planning of 3D objects using genetic algorithm, *2012 IEEE International Conference on Automation and Logistics,* Zhengzhou, China, 2012, pp. 646—651.

38. **Gabellieri C., Angelini F., Arapi V., Palleschi A., Catalano M., Grioli G., Pallottino L., Bicchi A., Bianchi M., Garabini M.** Grasp it like a pro: grasp of unknown objects with robotic hands based on skilled human expertise, *IEEE Robotics and Automation Letters*, 2020, vol. 5, no. 2, pp. 2808—2815.

39. **Palmieri G., Scoccia C.** Motion Planning and Control of Redundant Manipulators for Dynamical Obstacle Avoidance, *Machines*, 2021, vol. 9, no. 6—121.

40. **Wei K., Ren B.** A Method on Dynamic Path Planning for Robotic Manipulator Autonomous Obstacle Avoidance Based on an Improved RRT Algorithm, *Sensors*, 2018, vol. 18, no. 2—571.

41. **Ibraheem B., Ibrahim A., Talib A.** Motion Planning for a Robot Arm by Using Genetic Algorithm, *Jordan Journal of Mechanical and Industrial Engineering,* 2008, vol. 2, no. 3, pp. 131—136;

42. **Leskov A. G., Bazhinova K. V., Seliverstova E. V.** Methods for determining the quality of object grasping, *Vestnik MGTU im. N. E. Baumana. Ser. Mashinostroenie*, 2017, no. 3 (114), pp. 122—139 (in Russian).

43. **Bicchi A.** On the closure properties of robotic grasping, *The International Journal of Robotics Research,* 1995, vol. 27, no. 14, pp. 319—334.

44. **Gualtieri M., Pas A., Saenko K., Platt R.** High precision grasp pose detection in dense clutter, *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS*), Daejeon, Korea (South), 2016, pp. 598—605.