

Р. Н. Яковлев, мл. науч. сотр., iakovlev.r@mail.ru, Ю. И. Рубцова, мл. науч. сотр., julia_rubik@mail.ru,

А. А. Ерашов, программист, quietIsaac@yandex.ru,

Федеральное государственное бюджетное учреждение науки

"Санкт-Петербургский Федеральный исследовательский центр Российской академии наук" (СПб ФИЦ РАН),

Санкт-Петербургский институт информатики и автоматизации Российской академии наук,

г. Санкт-Петербург, Россия

Сравнительная оценка подходов к определению точек захвата объектов робототехническим средством

Настоящее исследование посвящено сравнительной оценке современных методов определения точек захвата объектов сцены с использованием средств технического зрения. В работе исследуемые методы определения точек захвата объектов применяются совместно с методом построения карт глубины, основанном на нейросетевой модели ResNet-50, что позволяет отказаться от использования специализированных датчиков глубины в процессе проведения экспериментов. В исследовании представлены зависимости оценок вероятности успешного захвата от типов целевых объектов. Усредненные по типам объектов оценки вероятности успешности захвата для исследуемых методов GPD, 6-DOF GraspNet, VPG составили соответственно 0.690, 0.741, 0.613. Также были исследованы зависимости оценок вероятности успешного захвата от размеров целевых объектов, расстояний между объективом фиксирующей видеокамеры и целевыми объектами сцен, уровней освещенности сцены, а также от углов наблюдения сцены по вертикальной оси. Для рассмотренных методов GPD, 6-DOF GraspNet, VPG выявлены нелинейные возрастающие зависимости усредненных по типам объектов вероятностей успешного захвата от уровня освещенности сцены. Также было установлено, что зависимости усредненной вероятности успешного захвата объекта для всех остальных параметров носят нелинейный характер и являются немонокотонными. В работе определены диапазоны значений исследованных параметров сцены, на которых достигаются наибольшие значения вероятностей успешного захвата объектов для данных методов. По результатам проведенной экспериментальной оценки решение 6-DOF GraspNet продемонстрировало наилучшее качество работы на подавляющем большинстве комбинаций рассмотренных параметров сцены. Использование данного метода является предпочтительным для решения задачи определения точек захвата объектов при использовании подходов, обеспечивающих восстановление карты глубины без применения специализированных устройств.

Ключевые слова: точки захвата объекта, карта глубины, робототехнические средства, нейронные сети, GPD, 6-DOF GraspNet, VPG, ResNet-50

Введение

Современные запросы коллаборативной, ассистивной и промышленной робототехники требуют существенного повышения качества функционирования промышленных манипуляторов и сервисных роботов, а также значительного расширения их функциональных возможностей. Для выполнения производственных и бытовых задач робототехническими средствами (РС) необходимо обеспечить их взаимодействие с окружающим пространством и объектами различных типов, размеров, формы и материала: возможность захватывать, перемещать, позиционировать объекты в пространстве и т.д. Таким образом, разработка подходов к универсализации средств взаимодействия промышленных и сервисных роботов

с объектами различных форм и размеров является актуальной научной задачей, поскольку способствует расширению функциональности РС и возможностей автоматизации в области промышленности и сервиса.

Одной из центральных задач при разработке подходов к автоматизированному перемещению объектов является обеспечение захвата целевого объекта (ЦО) посредством РС [1, 2]. При захвате объектов любого типа важно обеспечить, с одной стороны, сохранение целостности объекта, с другой — надежное сцепление с ним, что предполагает учет геометрии контакта захватного механизма РС с ЦО, а также управление силой захвата для обеспечения его устойчивости. Смена типа объекта манипуляций, как правило, требует дополнительной программной настройки РС, что не

всегда представляется возможным. Универсализация подходов к оперированию ЦО различного типа, таким образом, предполагает снижение требований к учету свойств ЦО при реализации программных средств взаимодействия с ними, что касается в том числе и процесса захвата ЦО посредством РС как основы для дальнейших манипуляций.

В целом для обеспечения захвата ЦО рабочим органом РС (схватом манипулятора) необходимо решить следующий комплекс задач технического зрения: 1) детектирование ЦО; 2) локализация ЦО на сцене; 3) определение точек захвата ЦО; 4) отслеживание траектории движения исполнительного механизма; 5) определение соответствия точек рабочей поверхности исполнительного механизма точкам схвата на ЦО; 6) детектирование успешного захвата ЦО. Предметом настоящего исследования является анализ методов решения задачи определения точек захвата объекта (ТЗО), обеспечивающих возможность осуществления надежного захвата для последующего оперирования ЦО.

ТЗО представляют собой центр сжатия объекта посредством РС [3]. При реализации методов определения ТЗО основная сложность заключается в наличии некоторой функциональной зависимости между положением ТЗО и такими свойствами ЦО, как: размер, геометрические особенности, свойства материала, упругость, жесткость. Отсутствие априорной информации о данных свойствах ЦО, а также о виде и параметрах соответствующей зависимости накладывает значительные ограничения на перспективы разработки универсальных математических моделей, обеспечивающих определение ТЗО. Таким образом, при решении задачи определения ТЗО целесообразным представляется применение нейросетевых моделей и иных средств машинного обучения для обеспечения универсальности подходов к определению ТЗО.

Многие существующие нейросетевые методы определения ТЗО основаны на анализе информации о положении и пространственных характеристиках окружающих объектов, получаемой с помощью специальных датчиков глубины, таких как LIDAR [4] и Kinect [5]. Сканирующие лидары и RGB-D-камеры формируют карту глубины сцены, на основе которой и определяются положение и характеристики захватываемых объектов. Однако такая аппаратура обладает крайне высокой стоимостью в сравнении с бытовыми видео- и фотокамерами. Разработка подходов к определению ТЗО, не предполагающих применение специализи-

рованных устройств, требует использования альтернативных подходов к построению карт глубины сцены. Интеграция соответствующих решений в киберфизические системы, где РС выступают в качестве агентов [6–8], позволит расширить спектр задач, подлежащих автоматизации, и тем самым расширить функциональность систем данного класса без необходимости в установке дополнительного дорогостоящего оборудования.

Данная работа посвящена сравнительной оценке качества работы существующих методов определения ТЗО при использовании подходов, обеспечивающих восстановление карты глубины без применения специализированных устройств.

Состояние исследований

Рассмотрим современные методы решения задачи определения ТЗО с использованием нейронных сетей (НС). В работе [9] представлена модель НС GPD для определения ТЗО. Процесс определения ТЗО с использованием данного метода включает в себя несколько основных этапов. На первом этапе происходит преобразование входного RGB-D-изображения в облако точек, а также вокселизация и оптимизация полученного облака точек в целях устранения шумов и статистических выбросов. Далее на основе обработанного облака точек идентифицируется область интереса (ROI), на которой необходимо провести захват ЦО. Далее определяются положения потенциальных ТЗО (кандидатов) на идентифицированной области интереса. На третьем этапе полученные данные о так называемых кандидатах поступают на вход четырехслойной сверточной нейронной сети (CNN). Данная НС решает задачу бинарной классификации и на основе полученной оценки принимает решение о возможности захвата ЦО для каждого из кандидатов. На заключительном шаге осуществляется выбор оптимального кандидата и выполняется непосредственно сам захват объекта посредством РС. В данной работе была проведена экспериментальная оценка полученного решения на 16 объектах цилиндрической формы и 29 объектах кубической формы. Вероятность успешного захвата ЦО посредством РС с использованием данного метода в среднем составляет 84 %.

В работе [10] предлагается модель НС 6-DOF GraspNet, предназначенная для определения ТЗО для различного класса объектов посредством РС. В соответствии с предложенной ав-

торами моделью на первом этапе из полученной на вход карты глубины извлекается облако точек ЦО, затем разработанная авторами НС Grasp Sampler Network принимает на вход это облако точек и определяет расположение потенциальных ТЗО. На заключительном этапе осуществляется оптимизация полученных результатов с использованием стохастического градиентного спуска. В данной работе апробация модели осуществлялась с использованием коллаборативного робота Franka Panda со встроенной RGB-D-камерой. В качестве ЦО выступали объекты различных форм и размеров: коробки, цилиндры, глубокие миски и кружки. Согласно результатам экспериментов успешный захват ЦО с использованием данного метода был осуществлен в среднем в 88 % случаев.

В работе [11] предлагается метод обучения роботизированных агентов Visual Pushing and Grasping (VPG) на основе сверточной нейронной сети Fully Convolutional Network (FCN) [12], предназначенный для обеспечения захвата объектов и дальнейшего манипулирования ими с использованием РС. Авторами используется метод Q-обучения, он позволяет роботизированному агенту формировать стратегию собственного поведения в среде, опираясь на опыт предшествующих взаимодействий с ЦО. В данной работе была проведена экспериментальная оценка полученного решения на коллаборативном манипуляторе UR5 с захватным устройством RG2. В качестве ЦО использовались игрушечные блоки небольшого размера. Результаты экспериментов показали, что вероятность успешного захвата ЦО с использованием данного метода в среднем составляет 83,3 %, однако стоит отметить, что данное решение было протестировано авторами только на объектах, имеющих небольшой вес и простую геометрическую форму.

В работе [13] представлена модель НС PointNetGPD, которая предназначена для решения задачи локализации ТЗО и принимает на вход необработанное облако точек сцены. Данная модель основана на НС PointNet [14] и обучена на наборе данных, представленном в работе [15], включающем объекты различных форм, размеров и текстур. В данной работе были проведены эксперименты по захвату ЦО с помощью коллаборативного манипулятора UR5 и роботизированного захватного устройства 3-Finger Adaptive Robot Gripper. В качестве ЦО использовались различные предметы, включая: коробки, жестяные банки, муляжи фруктов и овощей. В результате эксперимен-

тальной оценки было выявлено, что вероятность успешного захвата ЦО с использованием данного метода в среднем составляет 82 %.

Авторами работы [16] предложена модель НС, позволяющая определять ТЗО с использованием РС. Данная модель основана на остаточной нейронной сети ResNet [17]. Подаваемая на вход карта глубины сцены делится на фрагменты размером 168x168 пикселей. Сеть Region Prediction Networks [16] принимает на вход полученные фрагменты и выдает карту вероятности успеха захвата. Результаты проведенных экспериментов показали, что в 82,17 % случаев захват ЦО манипулятором был успешным, однако стоит также отметить, что предложенная модель НС тестировалась только на объектах, имеющих простую цилиндрическую форму, а также небольшой вес и маленький размер.

По результатам проведенного анализа методов и подходов к определению ТЗО для проведения дальнейшей сравнительной оценки были выбраны решения, представленные в работах [9—11], характеризующиеся наибольшими вероятностями успешного захвата ЦО. Рассмотрим далее существующие подходы, обеспечивающие восстановление карты глубины без применения специализированных устройств.

В общем случае методы построения карт глубины условно можно разделить на две группы: аппаратные [4, 5] и программные методы, в частности, основанные на использовании нейросетевых моделей [18—27]. Последние не требуют применения специального оборудования и позволяют осуществить восстановление карты глубины сцены по 2D-изображению, полученному с обыкновенной монокулярной камеры. В частности, авторами работы [18] предложен метод, в соответствии с которым проблема восстановления карты глубины формулируется как задача дискретно-непрерывной оптимизации, где непрерывные переменные кодируют глубину суперпикселей во входном 2D-изображении, а дискретные переменные представляют собой отношение между соседними суперпикселями. Авторы используют многочастичный алгоритм распространения доверия для вывода графической модели. В рамках экспериментальной оценки данного метода авторами было получено значение среднеквадратичной ошибки (СКО) восстановления карты глубины, равное 1,06...1,08. В более позднем исследовании достигнуто значение СКО, равное 0,824 [19].

В работе [20] представлен метод построения карты глубины посредством обработки RGB-изображения с использованием глубокой НС.

Процесс построения карты глубины с использованием данного метода можно разделить на два этапа: на первом этапе происходит оценка глубины на основании всего изображения, а на втором этапе полученная карта глубины оптимизируется в целях устранения статистических выбросов, что позволяет на выходе НС получить карту глубины с более четкими границами объектов. СКО предсказания глубины с применением данного метода составляет 0,871.

В работе [21] предлагается метод построения карты глубины посредством обработки трехканального изображения глубокой НС. Данная НС основана на архитектуре сверточной нейронной сети VGG [22] и обучена на наборе данных NYU Depth с применением метода Transfer Learning [21]. В рамках экспериментальной оценки данного метода авторами было получено значение СКО восстановления карты глубины, равное 0,833.

В работе [23] предложена глубокая сверточная нейронная сеть, предназначенная для выполнения ряда задач: прогнозирования глубины, оценки поверхности и семантической маркировки. Данная модель изначально основана на архитектуре, представленной в работе [24], но с некоторыми улучшениями: увеличено число сверточных слоев, а также повышено разрешение третьего слоя свертки. По данным, представленным в исследовании, СКО восстановления глубины с использованием данного метода составляет 0,641.

В работе [25] архитектура предложенной авторами НС построена на основе остаточной нейронной сети ResNet-50 [26] и дополняет ее Upsampling слоями, которые повышают точность восстановления карт глубины. Авторы работы [25] провели экспериментальную оценку качества работы данной архитектуры в сравнении с моделями AlexNet [27] и VGG [22]. Данные нейросетевые модели были протестированы на обучающих наборах данных NYU Depth и Make3D. Результаты сравнения моделей показали, что предложенная архитектура значительно превосходит аналогичные решения в задаче построения карты глубины по входному RGB-изображению. В рамках экспериментальной оценки авторами было получено значение СКО восстановления карты глубины с использованием разработанной модели, равное 0,573.

На основе проведенного анализа методов построения карты глубины с использованием глубоких нейронных сетей была выбрана модель НС, представленная в работе [25], поскольку она обладает меньшим числом обучающих

параметров и, следовательно, требует меньшего объема данных для обучения. Кроме того, данная модель характеризуется наименьшим значением СКО, а значит, обладает наибольшей точностью предсказания выходной карты глубины в сравнении с другими рассмотренными решениями.

Далее перейдем к проведению сравнительной оценки выбранных подходов к определению ТЗО: GPD [9], 6-DOF GraspNet [10], VPG [11]. В качестве входных данных будем использовать карту глубины исследуемой сцены, полученную с использованием НС [25].

Экспериментальная оценка подходов к осуществлению захвата различных классов объектов с использованием робототехнических средств

В рамках настоящего исследования сравнительная оценка выбранных подходов к определению ТЗО [9–11] проводилась на основе тестового набора данных, включающего в себя 33 000 изображений различных сцен, смоделированных в программной среде gazebo [28]. Каждому изображению I_i из тестового набора данных соответствует набор значений параметров P_i :

1. Тип ЦО T . Включает в себя 12 классов объектов: прямоугольные коробки (1), банки цилиндрической формы (2), кружки цилиндрической формы (3), чашки сферической формы (4), миски (5), глубокие миски (6), кувшины (7), вазы сложной формы (8), муляжи фруктов (яблоко (9), банан (10)) и овощей (морковь (11), перец (12)).

2. Размер ЦО l (см). В качестве данного параметра используется среднее значение длин проекций объекта на каждую из осей координат. Каждый целевой объект был отнесен к одной из следующих категорий исходя из значения l : 1 — [2, 4], 2 — (4, 5.5], 3 — (5.5, 7], 4 — (7, 9], 5 — (9, 12].

3. Расстояние между объективом фиксирующей видеокамеры и ЦО сцены d (м). Для тестового набора данных изображения снимались при дискретных значениях d в промежутке от 0,3 м до 1,2 м с шагом в 0,05 м.

4. Уровень освещенности сцены lum . Тестовый набор данных включает в себя изображения, сформированные при различных уровнях освещенности сцены: от 100 до 50 % с шагом в 5 %, где за 100 % взят нормативный уровень освещенности для офисных помещений [29].

5. Угол наблюдения сцены по вертикальной оси α . Данный угол определяется как угол отклонения оптической оси видеокамеры от вертикального направления. При формировании тестового набора данных изображения снимали при дискретных значениях α в промежутке от 0 до 60° с шагом в 3°.

Для всех тестовых сцен в качестве стандартного был выбран набор значений параметров $\mathbf{P}_{base} = \{l_{base} = 3, d_{base} = 0,75, lum_{base} = 100 \%, \alpha_{base} = 36^\circ\}$, данный набор значений параметров не зависит от типа ЦО. Базовая тестовая сцена представляет собой плоскую горизонтальную поверхность, на которой размещается шестизвенный манипулятор [30], а также ЦО размера l_{base} . На рабочий орган РС устанавливается видеокамера, посредством которой получается соответствующее тестовое изображение исследуемого участка сцены. ЦО при этом размещается по центру поля зрения видеокамеры таким образом, чтобы дистанция между объективом видеокамеры и ЦО составляла d_{base} , а угол наблюдения сцены по вертикальной оси был равен α_{base} . Освещенность моделируемой сцены в стандартных условиях составляет 100 %. В процессе формирования тестового набора данных **Img** для каждого отдельного ЦО осуществлялось последовательное варьирование значений параметров $\mathbf{P}_i$. Таким образом было получено множество изображений, отличающихся как типами, так и размерами ЦО, дистанциями между объективом фиксирующей видеокамеры и ЦО сцены, уровнями освещенности сцены и углами наблюдения сцены по вертикальной оси. Пример тестового изображения, полученного видеокамерой

при стандартных значениях параметров сцены ($\mathbf{P}_{base}$), представлен на рис. 1, а.

С помощью метода построения карт глубины [25] на основе изображений **Img** был получен результирующий набор карт глубины **Depth**. Применение рассматриваемых подходов к определению ТЗО [9–11] к данному набору карт глубины позволило сформировать оценки качества работы соответствующих решений в зависимости от параметров сцены и характеристик ЦО. Для оценки качества работы данных решений использовалась вероятность успешного захвата S — доля успешных попыток захвата ЦО при использовании предлагаемых данными методами ТЗО в рамках моделирования соответствующих процессов.

Рассмотрим полученные оценки качества работы данных методов в зависимости от типа ЦО T . В рамках данного эксперимента для проведения оценки использовались лишь те карты глубины **Depth_i**, которые были получены из изображений сцен **Img_i** со стандартными значениями параметров ($\mathbf{P}_i = \mathbf{P}_{base}$). Полученные оценки S для каждого из рассматриваемых методов представлены на рис. 2.

Из данных, представленных на рис. 2, можно заключить, что наилучшие результаты для всех типов ЦО демонстрирует подход 6-DOF GraspNet [10]. Усредненная по типам объектов оценка вероятности успешности захвата для данного решения составила 0,741. Несколько худшее качество показывает метод GPD [9], полученная усредненная оценка в данном случае составила 0,690. Наихудшие результаты в рамках данного эксперимента продемонстрировал подход VPG [11], усредненная оценка S

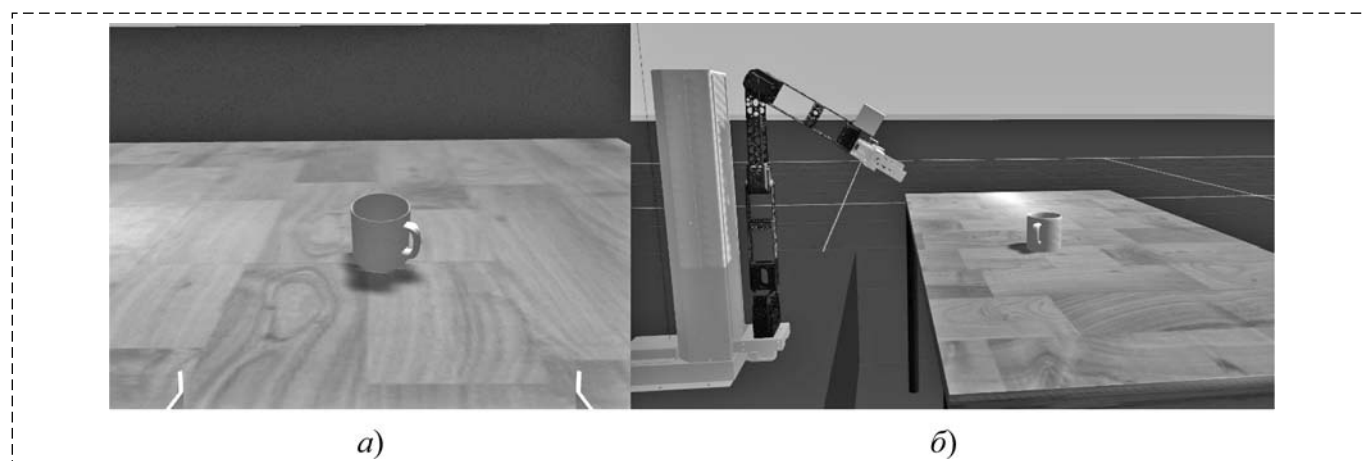


Рис. 1. Пример смоделированной сцены:

а — тестовое изображение, полученное видеокамерой; б — общий вид смоделированной сцены

Fig. 1. Examples of the simulated environment:

а — test image, obtained with camera; б — general view of the simulated environment

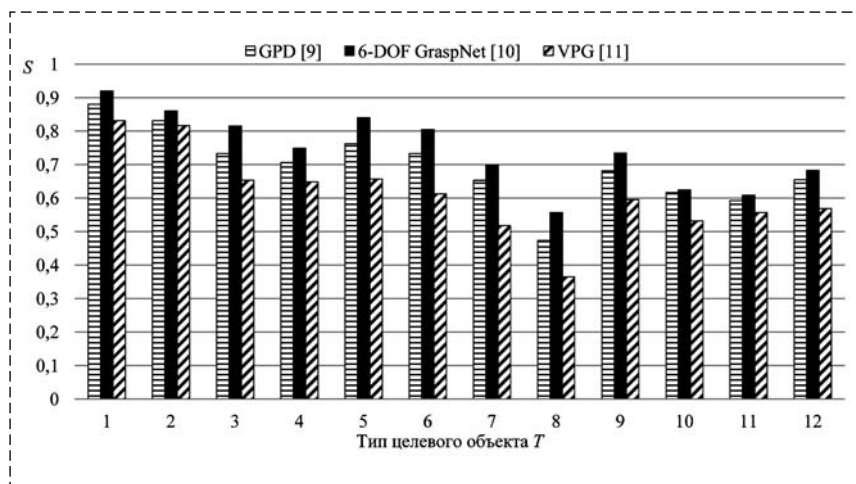


Рис. 2. Оценка качества работы исследуемых методов в зависимости от типа ЦО T , в стандартных условиях эксперимента ($\forall P_i = P_{base}$)

Fig. 2. Performance quality assessment of the methods under consideration depending on the type of target object T in default experimental setup ($\forall P_i = P_{base}$)

для данного метода составила 0,613. Следует отметить, что ни одно из представленных решений в данном эксперименте не достигло тех уровней показателей качества работы, которые в процессе апробации были получены авторами соответствующих исследований. Данные результаты могут быть объяснены следующими факторами: для восстановления карт глубины из тестовых изображений в данной работе использована нейросетевая модель из работы [25], которая обладает меньшей точностью в сравнении с аппаратными методами восстановления карт глубины; в работе рассматривается более широкий набор типов ЦО, включающий классы объектов, обладающие сложной нестандартной формой. Важно отметить, что для всех рассматриваемых подходов усредненная вероятность успешного захвата превысила значение 0,8 для таких классов объектов, как коробки прямоугольной формы (1) и жестяные банки цилиндрической формы (2). Данные типы объектов характеризуются простой геометрической формой и использовались авторами соответствующих подходов в процессе их апробации. Полученные для данных классов объектов значения показателей успешного захвата в целом соответствуют значениям, полученным авторами соответствующих подходов в исследованиях [9–11]. В то же время для таких нестандартных классов, как кувшины (7) и вазы сложной формы (8), все рассматриваемые подходы продемонстрировали значительно более низкое качество работы. В частности, для типа объектов (8) полученные оценки составили 0,474; 0,557 и 0,365 для методов из работ [9], [10] и [11] соот-

ветственно. Таким образом, можно заключить, что снижение качества работы рассматриваемых подходов в данном эксперименте обуславливается в первую очередь именно многообразием рассматриваемых типов ЦО, а не использованием нейросетевой модели, представленной в работе [25], для восстановления карт глубины.

Далее перейдем к оценке влияния размеров ЦО l на качество работы рассматриваемых решений. В процессе формирования соответствующего набора тестовых изображений на исследуемых сценах осуществлялось варьирование размеров ЦО каждого типа, остальные параметры принимали стандартные

значения из набора P_{base} . Результаты проведенной серии экспериментов для каждой категории размера ЦО представлены на рис. 3

Из графиков, представленных на рис. 3, можно заметить, что для всех категорий размера ЦО наилучшие результаты по-прежнему в среднем демонстрирует подход 6-DOF GraspNet [10], чуть худшие значения S демонстрирует подход GPD [9], а самые низкие — метод VPG [11]. Наблюдаемые зависимости вероятности успешного захвата от типа ЦО имеют схожий вид для всех рассмотренных категорий размера ЦО. Во всех случаях наилучшие результаты характерны для объектов классов (1) и (2), а наихудшие — для объектов класса (8). Следует отметить, что наибольшие усредненные оценки успешности захвата наблюдаются для категории размера объектов $l = 3$ и равны 0,69; 0,741; 0,613 для методов [9–11] соответственно. При переходе как к меньшим, так и к большим размерам ЦО наблюдается снижение данных усредненных оценок, при этом минимальные значения усредненных оценок вероятности успешности захвата были получены для категории $l = 1$ и составляют соответственно 0,63; 0,682; 0,515. Данное снижение при применении рассматриваемых методов к ЦО более мелких размеров может быть объяснено снижением точности выделения сегмента объекта на изображениях в процессе восстановления карты глубины при уменьшении размера ЦО. Увеличение размеров ЦО до определенного предела не оказывает столь значимого влияния на качество восстановления карты глубины, в данном случае снижение качества работы методов, представленных в работах [9–11], является не столь

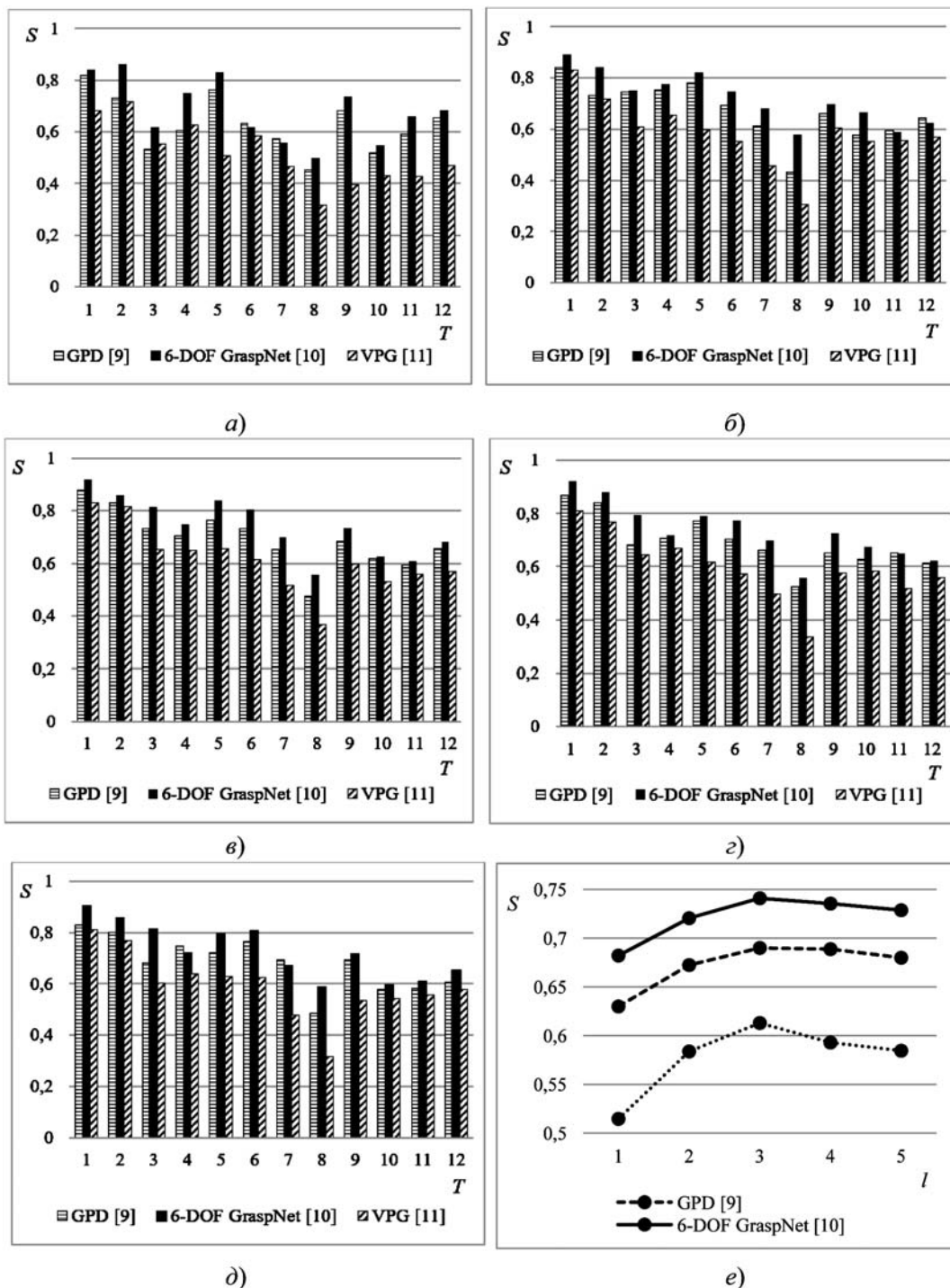


Рис. 3. Оценка влияния размеров ЦО l на качество работы исследуемых методов:

$a - l = 1$; $b - l = 2$; $в - l = 3$; $г - l = 4$; $д - l = 5$; $e -$ усредненная по T оценка вероятности успешного захвата ЦО в зависимости от l

Fig. 3. Assessment of influence of target object dimensions l on the performance of the methods under consideration:

$a - l = 1$; $b - l = 2$; $в - l = 3$; $г - l = 4$; $д - l = 5$; $e - T$ -averaged probability estimate of successful grip of target object depending on l

сильным и объясняется повышением сложности процесса поиска оптимальной ТЗО в связи с увеличением числа потенциальных ТЗО. Значения усредненной оценки вероятности успешности захвата для $l = 5$ составили соответственно 0,679; 0,728; 0,584 для методов [9–11].

Для проведения оценки влияния остальных описанных ранее параметров на качество работы рассматриваемых решений был выбран тестовый набор данных, где тип ЦО $T = (1), (2), (7)$ и (8) . Данные классы были выбраны, поскольку на них по результатам проведен-

ных выше экспериментов исследуемые методы демонстрируют наилучшее для $T = (1), (2)$ и наихудшее для $T = (7), (8)$ качество работы соответственно. На рис. 4 (см. третью сторону обложки) представлены полученные зависимости вероятности успешного захвата в зависимости от d , lum и α для классов объектов (1) и (8): а) вероятность успешного захвата S для типов ЦО (1), (8) в зависимости от d ; б) усредненные по T оценки вероятности успешного захвата ЦО в зависимости от d ; в) вероятность успешного захвата S для типов ЦО (1), (8) в зависимости от lum ; г) усредненные по T оценки вероятности успешного захвата ЦО в зависимости от lum ; д) вероятность успешного захвата S для типов ЦО (1), (8) в зависимости от α ; е) усредненные по T оценки вероятности успешного захвата ЦО в зависимости от α . В приведенных экспериментах усреднение по T проводилось для классов {1, 2, 7, 8}. Отметим, что при варьировании каждого параметра остальным параметрам были присвоены стандартные значения из набора $\mathbf{P}_{base}$.

Как можно заметить, все представленные на рис. 4, а (см. третью сторону обложки) зависимости обладают схожей структурой независимо от используемого метода определения ТЗО и типа ЦО T : при значительном отклонении от стандартного значения параметра d как в большую, так и в меньшую сторону наблюдается устойчивое снижение качества работы всех рассматриваемых методов на всех типах ЦО. Рассматривая усредненные по T оценки S , представленные на рис. 4, б, можно заметить, что оптимальные S для всех методов достигаются при значениях параметра d в диапазоне от 0,6 до 0,8 м. При дальнейшем увеличении d у всех методов наблюдается плавное устойчивое снижение вероятности успешного захвата ЦО. Данный эффект может быть объяснен тем, что при удалении ЦО в процессе построения карты глубины из исследуемого изображения сцены удастся извлечь меньше информации о характеристиках ЦО, что негативным образом сказывается на качестве работы рассматриваемых решений. В то же время при уменьшении d до значений менее 0,5 м также наблюдается значительный рост темпа падения качества работы рассматриваемых методов [9–11]. Это может объясняться тем, что при слишком малых значениях d объект занимает слишком большую область изображения, что не позволяет осуществить корректное построение карты глубины исследуемой сцены и, соответственно, приводит к значительным погрешностям в работе рассматриваемых методов.

На всем рассматриваемом диапазоне d метод 6-DOF GraspNet [10] демонстрирует наилучшие усредненные показатели качества работы, результирующая оценка S_{Td} вероятности успешного захвата усредненная по T и d для данного метода составила 0,685, несколько более низкие результаты демонстрирует решение GPD [9] $S_{Td} = 0,629$, а самые слабые показатели вновь характерны для решения VPG [11] $S_{Td} = 0,552$.

По результатам исследования зависимостей вероятности успешного захвата от уровня освещенности сцены lum для типов ЦО (1) и (8) (рис. 4, в), а также аналогичных зависимостей, усредненных в разрезе каждого метода по типам ЦО T (рис. 4, г), можно сделать вывод о схожести данных зависимостей независимо от используемого метода определения ТЗО и типов ЦО. При снижении уровня освещенности, начиная с некоторого момента, оценки вероятностей успешного захвата начинают существенно снижаться, в частности, при понижении уровня освещенности с 60 до 55 % усредненные по T оценки вероятности успешного захвата на всех методах дополнительно снижаются более чем на 35 %. Более того, для ЦО типа (8) оценки вероятности успешного захвата для всех методов при уровне освещенности уже в 50 % обращаются в ноль. Данный эффект может быть объяснен тем, что значительное снижение уровня освещенности оказывает критическое влияние на качество работы используемого метода построения карты глубины сцены, а низкое качество полученных карт глубины, в свою очередь, значительно снижает точность определения потенциальных ТЗО или же вовсе не позволяет методам успешно реализовать данный процесс.

На основе полученных экспериментальных оценок можно сделать вывод о том, что на всем исследованном диапазоне уровней освещенности сцен метод 6-DOF GraspNet [10] демонстрирует наилучшие усредненные показатели качества работы в сравнении с другими рассматриваемыми решениями [9, 11]. Результирующие оценки S_{Tlum} вероятности успешного захвата усредненные по T и lum составили 0,547; 0,597; 0,476 для методов из работ [9–11] соответственно.

Аналогичным образом была исследована зависимость оценки вероятности успешного захвата ЦО от α . Полученные значения данных оценок для типов ЦО (1) и (8) в разрезе каждого рассматриваемого метода представлены на рис. 4, д (см. третью сторону обложки), а соответствующие усредненные по T зависимости

представлены на рис. 4, *е*. Следует отметить, что полученные зависимости характеризуются несколько более высокой вариабельностью результатов при снижении и увеличении α . Тем не менее, во всех представленных случаях наблюдаются четкие тренды снижения оценок S при отклонении α от стандартного значения как в большую, так и в меньшую сторону. Полученные зависимости объясняются тем, что при изменении α на изображениях сцены фиксируются разные участки объекта: при малых α на изображении фиксируется лишь одна сторона ЦО, что негативно сказывается на качестве определения структуры ЦО и не позволяет рассматриваемым методам выявить значительную долю потенциальных ТЗО. При больших α на изображениях сцены фиксируется лишь поверхность, на которой непосредственно установлен ЦО, и сам ЦО захвата, что может привести к снижению точности построения карт глубины на основе таких изображений исследуемых сцен. Это и является предполагаемой причиной снижения оценок вероятности успешного захвата ЦО в данном случае. На всем рассматриваемом диапазоне α метод 6-DOF GraspNet [10], как правило, демонстрирует наилучшие усредненные показатели качества работы, несколько более низкие результаты демонстрирует решение GPD [9], а самые низкие показатели вновь характерны для решения VPG [11]. Результирующие оценки $S_{T\alpha}$ вероятности успешного захвата, усредненные по T и α , составили 0,579; 0,631; 0,506 для методов [9–11] соответственно.

Таким образом, можно сделать вывод о том, что представленные в работе стандартные значения параметров сцены $P_{base} = \{l_{base} = 3, d_{base} = 0,75, lum_{base} = 100 \%, \alpha_{base} = 36^\circ\}$ являются одними из оптимальных с точки зрения максимизации вероятности успешного захвата ЦО каждым из рассмотренных методов [9–11]. По результатам проведенной экспериментальной оценки решение 6-DOF GraspNet [10] демонстрирует наилучшие показатели качества работы для подавляющего большинства комбинаций рассмотренных параметров сцены, и его использование является предпочтительным для решения задачи определения ТЗО при использовании подходов, обеспечивающих восстановление карты глубины без применения специализированных устройств.

Заключение

Проведенные исследования показывают, что рассмотренные методы GPD [9], 6-DOF GraspNet

[10], VPG [11] достигают наибольших значений показателей качества своей работы на объектах простой геометрической формы, например, на прямоугольных коробках и банках цилиндрической формы. Анализ зависимости оценок вероятности успешного захвата от размера ЦО и параметров сцены позволил определить факторы, негативным образом сказывающиеся на качестве работы рассматриваемых решений и определить набор параметров, при которых достигаются наилучшие результаты. Модель HC 6-DOF GraspNet [10] демонстрирует наилучшие показатели качества работы для подавляющего большинства комбинаций рассмотренных параметров сцены. Ее использование является предпочтительным для решения задачи определения ТЗО при использовании подходов, обеспечивающих восстановление карты глубины без применения специализированных устройств.

Список литературы

1. Cutkosky M. R., Howe R. D. Human grasp choice and robotic grasp analysis // *Dextrous robot hands*. Springer, New York, NY. 1990. P. 5–31.
2. Morales A., Asfour T., Azad P., Knoop S., Dillmann R. Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands // 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE. 2006. P. 5663–5668.
3. Лесков А. Г., Илларионов В. В., Калеватых И. А., Морозкин С. Д., Бажинова К. В., Феоктистова Е. В. Аппаратно-программный комплекс для решения задач автоматического захвата объекта манипуляторами // *Инженерный журнал: наука и инновации*. 2015. Т. 1, № 37.
4. Fernald F. G. Analysis of atmospheric lidar observations: some comments // *Applied optics*. 1984. Vol. 23, N. 5. P. 652–653.
5. Zhang Z. Microsoft kinect sensor and its effect // *IEEE multimedia*. 2012. Vol. 19, N. 2. P. 4–10.
6. Ronzhin A., Saveliev A., Basov O., Solonyj S. Conceptual model of cyberphysical environment based on collaborative work of distributed means and mobile robots // *International Conference on Interactive Collaborative Robotics*. Springer, Cham. 2016. P. 32–39.
7. Яковлев Р. Н., Ватаманюк И. В. Алгоритмическая модель распределенной системы корпоративного информирования в рамках киберфизической системы организации // *МОИТ*. 2019. Т. 7, № 4. С. 32–33.
8. Левоневский Д. Архитектура облачной системы распределения контента в киберфизических системах // *МОИТ*. 2019. Т. 7, № 4. С. 16–17.
9. ten Pas A., Gualtieri M., Saenko K., Platt R. Grasp pose detection in point clouds // *The International Journal of Robotics Research*. 2017. Vol. 36, N. 13–14. P. 1455–1473.
10. Mousavian A., Eppner C., Fox D. 6-dof graspnet: Variational grasp generation for object manipulation // *Proceedings of the IEEE International Conference on Computer Vision*. 2019.
11. Zeng A., Song S., Welker S., Lee J., Rodriguez A. Learning synergies between pushing and grasping with self-supervised deep reinforcement learning // 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE. 2018. P. 4238–4245.
12. Li B. 3d fully convolutional network for vehicle detection in point cloud // 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017. P. 1513–1518.
13. Liang H., Ma X., Li S., Görner M., Tang S., Fang B., Sun F., Zhang J. Pointnetgpd: Detecting grasp configurations

from point sets // 2019 International Conference on Robotics and Automation (ICRA). IEEE. 2019. P. 3629–3635.

14. **Qi C. R., Su H., Mo K., Guibas L. J.** Pointnet: Deep learning on point sets for 3d classification and segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. P. 652–660.

15. **Calli B., Singh A., Walsman A., Srinivasa S., Abbeel P., Dollar A. M.** The ycb object and model set: Towards common benchmarks for manipulation research // 2015 international conference on advanced robotics (ICAR). IEEE. 2015. P. 510–517.

16. **Shao Q., Hu J., Wang W., Fang Y., Liu W., Qi J., Ma J.** Suction Grasp Region Prediction using Self-supervised Learning for Object Picking in Dense Clutter // 2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR). IEEE. 2019. P. 7–12.

17. **Szegedy C., Ioffe S., Vanhoucke V., Alemi A. A.** Inception-v4, inception-resnet and the impact of residual connections on learning // Thirty-first AAAI conference on artificial intelligence. 2017.

18. **Liu M., Salzmann M., He X.** Discrete-continuous depth estimation from a single image // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 716–723.

19. **Liu F., Shen C., Lin G.** Deep convolutional neural fields for depth estimation from a single image // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. P. 5162–5170.

20. **Eigen D., Puhrsch C., Fergus R.** Depth map prediction from a single image using a multi-scale deep network // Advances in neural information processing systems. 2014. P. 2366–2374.

21. **Zhu J., Ma R.** Real-time depth estimation from 2D images. 2016.

22. **Simonyan K., Zisserman A.** Very deep convolutional networks for large-scale image recognition // arXiv preprint arXiv:1409.1556. 2014.

23. **Eigen D., Fergus R.** Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture // Proceedings of the IEEE international conference on computer vision. 2015. P. 2650–2658.

24. **Geiger A., Lenz P., Stiller C., Urtasun R.** Vision meets robotics: The kitti dataset // The International Journal of Robotics Research. 2013. Vol. 32, N. 11. P. 1231–1237.

25. **Laina I., Rupprecht C., Belagiannis V., Tombari F., Navab N.** Deeper depth prediction with fully convolutional residual networks // 2016 Fourth international conference on 3D vision (3DV). IEEE. 2016. P. 239–248.

26. **He K., Zhang X., Ren S., Sun J.** Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. P. 770–778.

27. **Iandola F. N., Han S., Moskewicz M. W., Ashraf K., Dally W. J., Keutzer K.** SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size // arXiv preprint arXiv:1602.07360. 2016.

28. **Koenig N., Howard A.** Design and use paradigms for gazebo, an open-source multi-robot simulator // 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (IEEE Cat. No. 04CH37566). IEEE. 2004. Vol. 3. P. 2149–2154.

29. **ГОСТ Р. 55710-2013** Освещение рабочих мест внутри зданий // Нормы и методы измерений. 2013.

30. **Funabashi H., Horie M., Kubota T., Takeda Y.** Development of spatial parallel manipulators with six degrees of freedom // JSME international journal. Ser. 3, Vibration, control engineering, engineering for industry. 1991. Vol. 34, N. 3. P. 382–387.

Comparative Evaluation of Approaches for Determination of Grasp Points on Objects, Manipulated by Robotic Systems

R. N. Iakovlev, iakovlev.r@mail.ru, **J. I. Rubtsova**, julia_rubik@mail.ru, **A. A. Erashov**, quietisaac@yandex.ru,
St. Petersburg Federal Research Center of the Russian Academy of Sciences (SPC RAS),
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences,
St. Petersburg, 199178, Russian Federation

Corresponding author: Iakovlev R. N., Junior Researcher, St. Petersburg Federal Research Center of the Russian Academy of Sciences, Saint-Petersburg, 199178, Russian Federation, e-mail: iakovlev.r@mail.ru

Accepted on October 21, 2020

Abstract

This paper considers comparative evaluation of recent methods for grip point determination for manipulations with objects in the scene. This research is aimed to compare and evaluate the modern approaches of grip point determination, when this process is aided by computer vision. The methods of object gripping, considered in this paper, are employed in connection with depth map composition, backed by neural network model ResNet-50, which allowed to omit application of specific depth sensors in the course of experiments. This research shows dependencies of successful grip probability from the things being manipulated. Probability scores, averaged over different types of objects for the methods GPD, 6-DOF GraspNet, VPG, were, accordingly, 0.690, 0.741, 0.613. The paper also considers dependencies of successful grip probability from object sizes, distances from the capturing camera and target objects in the scene, luminosity levels, as well from the angles of scene inspection along the vertical axis. In terms of the considered methods GPD, 6-DOF GraspNet, VPG, non-linear increasing dependencies are revealed for object type-averaged probabilities of successful grip from luminosity level of the scene. It was also discovered, that the dependencies of successful grip for all the other parameters are non-linear and non-monotonic. The ranges of the values for scene parameters under consideration are defined in this paper, which ensure the highest probability values for object grip in these approaches. Upon the results of the performed experimental evaluation, the 6-DOF GraspNet solution showed the best performance for the vast majority of the considered parameters of the scene. The approach, presented in this paper, is the preferable way for solution of grip point problem, in context of methods, which assume depth map reconstruction without specific equipment.

Keywords: object grip points, depth map, neural networks, GPD, 6-DOF GraspNet, VPG, ResNet-50, robotic systems

For citation:

Iakovlev R. N., Rubtsova J. I., Erashov A. A. Comparative Evaluation of Approaches for Determination of Grasp Points on Objects, Manipulated by Robotic Systems, *Mekhatronika, Avtomatizatsiya, Upravlenie*, 2021, vol. 22, no. 2, pp. 83–93.

DOI: 10.17587/mau.22.83-93

References

1. **Cutkosky M. R., Howe R. D.** Human grasp choice and robotic grasp analysis, *Dextrous Robot Hands*, Springer, New York, NY, 1990, pp. 5–31.
2. **Morales A., Asfour T., Azad P., Knoop S., Dillmann R.** Integrated grasp planning and visual object localization for a humanoid robot with five-fingered hands, *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2006, pp. 5663–5668.
3. **Leskov A. G., Illarionov V. V., Kalevatykh I. A., Moroshkin S. D., Bazhinova K. V., Feoktistova E. V.** Hardware-software complex for solving the task of automatic capture of the object with manipulators, *Inzhenernyy Zhurnal: Nauka i Innovacii*, 2015, vol. 1, no. 37 (in Russian).
4. **Fernald F. G.** Analysis of atmospheric lidar observations: some comments, *Applied optics*, 1984, vol. 23, no. 5, pp. 652–653.
5. **Zhang Z.** Microsoft kinect sensor and its effect, *IEEE multimedia*, 2012, vol. 19, no. 2, pp. 4–10.
6. **Ronzhin A., Saveliev A., Basov O., Solyonyj S.** Conceptual model of cyberphysical environment based on collaborative work of distributed means and mobile robots, *International Conference on Interactive Collaborative Robotics*, Springer, Cham, 2016, pp. 32–39 (in Russian).
7. **Vatamaniuk I. V., Iakovlev R. N.** Algorithmic model of a distributed corporate notification system in context of a corporate cyber-physical system, *MOIT*, vol. 7, no. 4, pp. 32–33 (in Russian).
8. **Levonevskiy D. K.** Architecture of a cloud system for distributing multimedia content in cyber-physical systems, *MOIT*, vol. 7, no. 4, pp. 16–17 (in Russian).
9. **ten Pas A., Gualtieri M., Saenko K., Platt R.** Grasp pose detection in point clouds, *The International Journal of Robotics Research*, 2017, vol. 36, no. 13–14, pp. 1455–1473.
10. **Mousavian A., Eppner C., Fox D.** 6-dof graspnet: Variational grasp generation for object manipulation, *Proceedings of the IEEE International Conference on Computer Vision*, 2019.
11. **Zeng A., Song S., Welker S., Lee J., Rodriguez A.** Learning synergies between pushing and grasping with self-supervised deep reinforcement learning, *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 4238–4245.
12. **Li B.** 3d fully convolutional network for vehicle detection in point cloud, *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017, pp. 1513–1518.
13. **Liang H., Ma X., Li S., Görner M., Tang S., Fang B., Sun F., Zhang J.** Pointnetgpd: Detecting grasp configurations from point sets, *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 3629–3635.
14. **Qi C. R., Su H., Mo K., Guibas L. J.** Pointnet: Deep learning on point sets for 3d classification and segmentation, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
15. **Calli B., Singh A., Walsman A., Srinivasa S., Abbeel P., Dollar A. M.** The ycb object and model set: Towards common benchmarks for manipulation research, *2015 international conference on advanced robotics (ICAR)*, IEEE, 2015, pp. 510–517.
16. **Shao Q., Hu J., Wang W., Fang Y., Liu W., Qi J., Ma J.** Suction Grasp Region Prediction using Self-supervised Learning for Object Picking in Dense Clutter, *2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR)*, IEEE, 2019, pp. 7–12.
17. **Szegedy C., Ioffe S., Vanhoucke V., Alemi A. A.** Inception-v4, inception-resnet and the impact of residual connections on learning, *Thirty-first AAAI conference on artificial intelligence*, 2017.
18. **Liu M., Salzmann M., He X.** Discrete-continuous depth estimation from a single image, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 716–723.
19. **Liu F., Shen C., Lin G.** Deep convolutional neural fields for depth estimation from a single image, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 5162–5170.
20. **Eigen D., Puhrsch C., Fergus R.** Depth map prediction from a single image using a multi-scale deep network, *Advances in Neural Information Processing Systems*, 2014, pp. 2366–2374.
21. **Zhu J., Ma R.** Real-time depth estimation from 2D images, 2016.
22. **Simonyan K., Zisserman A.** Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556, 2014.
23. **Eigen D., Fergus R.** Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2650–2658.
24. **Geiger A., Lenz P., Stiller C., Urtasun R.** Vision meets robotics: The kitti dataset, *The International Journal of Robotics Research*, 2013, vol. 32, no. 11, pp. 1231–1237.
25. **Laina I., Rupprecht C., Belagiannis V., Tombari F., Navab N.** Deeper depth prediction with fully convolutional residual networks, *2016 Fourth international conference on 3D vision (3DV)*, IEEE, 2016, pp. 239–248.
26. **He K., Zhang X., Ren S., Sun J.** Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
27. **Iandola F. N., Han S., Moskewicz M. W., Ashraf K., Dally W. J., Keutzer K.** SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size, arXiv preprint arXiv:1602.07360, 2016.
28. **Koenig N., Howard A.** Design and use paradigms for gazebo, an open-source multi-robot simulator, *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE Cat. No. 04CH37566), IEEE, 2004, vol. 3, pp. 2149–2154.
29. **GOST R. 55710-2013** Lighting of workplaces inside buildings, Norms and methods of measurements, 2013 (in Russian).
30. **Funabashi H., Horie M., Kubota T., Takeda Y.** Development of spatial parallel manipulators with six degrees of freedom, *JSME International Journal, Ser. 3, Vibration, Control Engineering, Engineering for Industry*, 1991, vol. 34, no. 3, pp. 382–387.