УДК 004.93 **DOI** 10.15622/ia.22.3.2

П.А. Сливницин, Л.А. Мыльников

РАСПОЗНАВАНИЕ ОБЪЕКТОВ ПО СОСТАВЛЯЮЩИМ ИХ ПРИМИТИВАМ И ОТНОШЕНИЯМ МЕЖДУ НИМИ

Сливницин П.А., Мыльников Л.А. **Распознавание объектов по составляющим их примитивам и отношениям между ними.**

Аннотация. Целью работы является разработка способа и алгоритма распознавания объектов окружающего пространства, качество работы которого не будет зависеть от числа типов объектов реального мира, которые он может распознавать. Для этого поставлены и решены задачи распознавания множества элементарных геометрических объектов (признаков-примитивов), определения отношений между ними и поиска соответствий между найденными признаками-примитивами и отношениями и заданными шаблонами–описаниями сложносоставных и простых объектов реального мира. Для распознавания элементарных геометрических фигур применена нейронная сеть свёрточного типа. Для её обучения использовались искусственно сгенерированные изображения с элементарными геометрическими фигурами (3D примитивами), которые располагались на сцене случайным образом с различными свойствами их поверхностей и текстурами. В результате обучения была получена нейронная сеть, способная распознавать объекты примитивы. Сформировано множество отношений, необходимое для распознавания объектов, которые могут быть представлены как составные из признаков-примитивов. В предложенном способе распознавания количество классов для поиска ограничивается набором признаков-примитивов. Проверка на фотографиях реальных объектов показала способность распознавать объекты реального мира в независимости от их типа (в случаях, когда возможны их разные модели и модификации) и материала изготовления, а также способность успешно решать задачи поиска объектов в условиях частичного перекрытия объектов и их ограниченной видимости и частичной деформации. В работе рассмотрен пример с распознаванием светильника уличного освещения. Пример показывает способность алгоритма не только выявлять объект на изображении, но и определять ориентацию положения его составляющих. Предложенное решение может быть использовано в задачах манипуляции объектами внешнего мира робототехническими системами.

Ключевые слова: распознавание объектов, признаки-примитивы, отношения признаков, нейронная сеть, компьютерное зрение.

1. Введение. Распознавание объектов окружающего пространства является задачей компьютерного зрения, которая на данный момент нашла широкое применение в распознавании людей/положения их фигур/лиц, автомобильных номеров, транспортных средств и др. [1]. Алгоритмы распознавания нашли применение в промышленности, например, при контроле состояния оборудования, контроле ношения средств индивидуальной защиты, контроле качества сырья, полуфабрикатов и готовой продукции [1, 2], аутентификации персонала [3], а также в оцифровке текстов [4]. Решение задач распознавания объектов внешнего мира является одним из ключевых элементов в создании автономных

систем, таких как автомобили с автопилотом [5], автономные робототехнические системы [6, 7], системы дополненной реальности [8]. Делаются шаги по развитию существующих алгоритмов для идентификации состояний, действий, ситуаций [1, 9].

Задачи компьютерного зрения можно разделить на задачи:

– классификации (присвоение класса изображению в зависимости от того, что на нём изображено) [10];

– поиска объектов (локализация и классификация объектов из конечного набора классов на изображении) [11];

– сегментации объектов (отнесение каждого пикселя на изображении к некоторому классу, логика классификации пикселей может отличаться в зависимости от типа сегментации, среди которых семантическая, экземплярная, паноптическая) [12, 13].

Их различие заключается в получаемой из входных данных информации об объектах. Если прикладная задача требует более детальное представление, определение его частей, говорят о комплексных задачах распознавания. Например, поиск ключевых точек объекта [14], более детальная сегментация лица человека [15]. На основе комплексных задач распознавания решаются ещё более сложные задачи, связанные с оценкой положения объекта относительно сенсора, с которого поступают входные данные [16].

Задачи определения положения объектов на данный момент являются одними из основополагающих задач, которые необходимы для создания автономных робототехнических систем, которые находят применение в агропромышленном комплексе [17], обслуживании объектов городской инфраструктуры [6], обслуживании складских помещений [8].

Подходы к распознаванию, как правило, основываются на выделении признаков распознаваемых объектов для их поиска и классификации [18]. На основе выделенных признаков составляется прототип объекта, который представляет собой «усреднённое» представление объекта. В современных алгоритмах, основанных на глубоком обучении, признаки объекта выделяются автоматически на основе размеченных примеров из обучающей выборки.

Наиболее распространённым инструментом для распознавания объектов на данный момент являются нейронные сети свёрточного типа [19], которые вытеснили не нейросетевые алгоритмы распознавания (метод Виолы-Джонса [20], метод Далала-Триггса [21], DPM [22]). При этом говорить о преимуществе одних методов

над другими не представляется возможным, т.к. выбор метода и алгоритма (а также других параметров, таких как функция потерь, оптимизатор и т.д.) распознавания подбирается под конкретную, частную задачу [23] (при разном применении один и тот же метод может давать хорошие результаты и плохие). В связи с этим в литературе встречаются работы, связанные со сравнением и выбором различных методов (реализаций) для конкретной задачи распознавания [24]. Однако работ, сравнивающих эффективность использования разных концепций в рамках области или класса задач не наблюдается, поскольку нет применимых метрик, позволяющих провести сравнительную оценку разных концепций, таких как классификация, поиск объектов, сегментация. Разные алгоритмы могут учитывать специфику задачи и имеют разную эффективность в одинаковой прикладной задаче. Поэтому сравнение доступно только на качественном уровне. Пример некоторого количества алгоритмов распознавания на качественном уровне представлен в таблице 1. Точность распознавания для каждой задачи зависит от большого количества параметров, таких как выбранная метрика, обучающая выборка данных, нюансы самих распознаваемых объектов. Зачастую, чтобы подобрать подходящую выборку, проводится несколько итераций обучения и сравнения [24]. Также встречаются работы, в которых рассматривается влияние количества классов на точность распознавания объектов [25]. При значительном увеличении количества распознаваемых классов наблюдается снижение точности распознавания, что также подтверждается эмпирическими наблюдениями.

Следуя из этого, можно говорить об ограниченной сфере применения алгоритмов распознавания, где для перенастройки и адаптации при применении алгоритма, решающего конкретную задачу, для нового применения требуются дополнительные затраты. Более сложные задачи, основанные на распознавании объектов, таких как манипуляция объектами автономными роботизированными системами, требуют выявления ещё большей информации об объекте, что проблематично реализуется существующими алгоритмами и накладывает ещё больше ограничений при распознавании (например, такие свойства объекта, как прозрачность или отражающая способность, хрупкость, место и способ захвата).

Таблица 1. Качественное сравнение алгоритмов распознавания (оценки низкие/средние/высокие являются мнением авторов, полученным на основе практической реализации этих алгоритмов)

Метод	**Извлекаемая информация об изображении / объекте**	**Вычислительная сложность**	**Сложность реализации**	**Возможность добавления новых классов объектов**
1. Методы поиска ассоциативных правил	Класс изображения (бинарная/multi-class классификация)	Низкие требования к вычислительным ресурсам	Требует описания правил	Необходимо описание новых правил
2. Экспертные системы	Класс изображения (бинарная/multi-class классификация)	Низкие требования к вычислительным ресурсам	Требует описания правил	Необходимо описание новых правил
3. Теория предикатов	Класс изображения (бинарная/multi-class классификация)	Теория не предъявляет высоких требований к вычислительным ресурсам (зависит от реализации)	Требует небольшое количество обучающих данных	Необходимо описание новых объектов
4. Метод Виолы-Джонса (Haar Cascades)	Положение объекта класса на изображении (ограничивающая рамка)	Низкие требования к вычислительным ресурсам	Требует небольшое количество обучающих данных	Необходимо формирование новой обучающей выборки
5. Метод Далала-Триггса (HOG descriptor)	Положение объекта класса на изображении (ограничивающая рамка)	Средние требования к вычислительным ресурсам	Требует небольшое количество обучающих данных	Необходимо формирование новой обучающей выборки
6. Deformable Part Model detector (DPM)	Положение объекта класса и его частей на изображении (ограничивающая рамка)	Средние требования к вычислительным ресурсам	Требует небольшое количество обучающих данных	Необходимо формирование новой обучающей выборки
7. Нейронные сети	Класс изображения (бинарная/multi-class/multi-label классификация) / положение объекта класса на изображении (ограничивающая рамка) / положение объекта класса на изображении (контур объекта) / положение частей объекта класса на изображении (ключевые точки)	Значительные требования к вычислительным ресурсам, как при обучении алгоритма, так и при использовании	Требует большого количества обучающих данных и времени на обучение	Необходимо формирование новой обучающей выборки

Все существующие на данный момент методы распознавания работают либо с пикселями, либо с вокселями. Кроме этого, методы делятся на те, которые используют предварительную/дополнительную обработку данных и – нет. Последние ориентированы на поиск и распознавание единичных заранее заданных объектов, под которые разрабатываются и обучаются специализированные модели. Таких моделей разработано большое количество (например, YOLO [26], SSD [27]). Для оценки эффективности работы таких моделей используются такие метрики как среднее значение и дисперсия IoU [28].

Методы с дополнительной или предварительной обработкой данных являются более перспективными и включают в себя самые разные подходы. Примерами реализации таких подходов могут быть методы селективного поиска для чего выделяется специальный слой нейронной сети для выявления области интереса (области содержащей объект с высокой вероятностью), с последующей классификацией каждого региона на принадлежность к искомым классам и уточнение местоположения ограничивающих рамок с помощью регрессора (Faster R-CNN [29]). Распространение получают методы использующие специальные структуры для идентификации объектов и определения их положения [30].

Развиваются методы реконструкции пространства (методы группы SLAM++) [31 – 33] которые помимо составления облака точек для исключения столкновения при перемещении в пространстве ориентированы на распознавание отдельных объектов, что дает дополнительную информацию о локации в которой находится робототехническая система. Развивается направление распознавания объектов по частям [34], распознавания множества предопределённых в базе знаний объектов [35], а также поиска особых взаимосвязей между пикселями или вокселями которые будут выступать ключевыми признаками в процессе распознавания [36]. Для оценки качества работы таких методов необходимо учитывать возможность распознавания множества объектов на рассматриваемой сцене, что приводит к ситуации, когда точность работы зависит от алгоритмов используемых для предварительной обработки данных (например, распознавания частей), а фактором показывающим качество работы всего алгоритма становится не метрика работы отдельного алгоритма, а сам факт распознавания ключевого объекта/локации/элемента в разных условиях или их множеств. Тогда к алгоритмам распознавания становятся применимы метрики, используемые в задачах классификации (правильное распознавание объекта/класса).

Упомянутые выше методы распознавания элементов сцен или частей объектов используют в своей работе все множество возможных модификаций объектов для идентификации, например, спинок/ножек стульев, тем самым полагаясь на то, что используемый метод (как правило, нейронная сеть) сам выделит необходимые уникальные признаки. Таким образом, в ситуациях с большим количеством вариантов конструкции качество работы методов будет снижаться как у методов, не использующих дополнительные этапы/предобработку. Решением может быть выделение некоторых универсальных признаков/объектов/черт комбинация которых делает объект или его часть уникальными.

Подходы, основанные на комбинации алгоритмов, как правило, работают дольше, однако имеют более высокую точность распознавания, это связано с тем, что последующие алгоритмы уточняют предсказание и уменьшают ошибку предыдущих. В связи с этим точность многоэтапных подходов не может быть оценена обобщением качества работы промежуточных алгоритмов. Точность алгоритма при распознавании простых объектов, которые могут быть описаны одним признаком-примитивом, является точностью исходной модели поиска. Точность классификации объектов как комбинации признаков и отношений между ними является предметом дальнейших исследований, поскольку требует подготовки набора данных для оценки точности.

Исходя из этого, внедрение систем распознавания и автономных роботизированных систем хоть и происходит, но не носит масштабный характер. Решение этой проблемы лежит в разработке универсальных методов распознавания, которые позволят расширить способы их применения.

И. Бидерман показал, что человек для распознавания окружающих его объектов использует множество компонент и учитывает их расположение относительно друг друга [37]. При этом существуют некоторые границы, после превышения которых в ориентации и расстоянии между компонентами человек перестает воспринимать объект (набор компонент) как единое целое, что стало обобщением данных экспериментов (например, иллюзия Тэтчер [38]). Способность распознавать также в зависит от набора компонент с помощью которых происходит формирование объекта [39]. В рамках своей теории распознавания по компонентам И. Бидерман рассматривал вопросы когнитивной психологии, связанные с процессом распознавания объектов человеком. Она строится на предположении, что каждый предмет может быть представлен

совокупностью геометрических фигур – геонов. Каждый геон может быть описан совокупностью неслучайных свойств, которые неизменны при изменении угла зрения [18, 37]. Развитие подходов к распознаванию основанных на таком представлении объектов можно проследить в [30, 40]. Задача распознавания в таком случае сводится к определению необходимого набора элементов для идентификации объектов мира или только необходимых для рассматриваемой области деятельности, выбору и обучению модели для их распознавания и выработки правил на основании соответствия которым объект будет однозначно идентифицирован.

Некоторые авторы показывают, что число распознаваемых типов компонентов может быть, в некоторых случаях, сведено даже до одного – двух элементов. Так работает вычислительная теория восприятия Д. Марра [41], которая предполагает, что для распознавания необходима многоэтапность с возрастающей детализацией объектов. Сначала обрабатывается информация о контурах, краях и пятнах, затем о глубине и ориентации видимых поверхностей, после чего генерируется трёхмерная модель распознаваемого объекта. Модели, согласно этой теории, состоят из канонических форм (например, цилиндров).

2. Методология. ***Определение 1.*** *Признаками-примитивами будем называть множество пространственных геометрических фигур и их свойств, составляющих множество объектов распознавания (общие элементы во всем множестве целевых объектов распознавания).*

Определение 2. *Отношениями признаков-примитивов будем называть общий для всех целевых объектов набор отношений между признаками-примитивами для описания их взаимодействия в рамках каждого целевого объекта.*

Определение 3. *Сложным признаком будем называть множество признаков-примитивов и других сложных признаков отношения, между которыми описываются без использования отношения «без отношений». Сложный признак, как и простые признаки-примитивы, может состоять в отношениях с другими признаками-примитивами и/или сложными признаками.*

Среди алгоритмов распознавания существуют примеры использования только отношений компонентов объектов. Такой подход работает при распознавании поз и жестов. Однако информации только об отношениях, как правило, не достаточно для распознавания объектов [42].

Набор признаков-примитивов может зависеть от предметной области распознаваемых объектов и учитывать особенности, характерные для этой области, например, в области обслуживания наружного освещения [6] (ограниченное множество креплений, форм, текстур и т.д.). Однако даже при отсутствии ограничений, накладываемых предметной областью число отношений и признаков примитивов, является конечным (в отличие от форм объектов окружающего мира).

Примерами признаков-примитивов могут быть признаки формы (простые и сложные трёхмерные фигуры: призма, сфера, тор и т.д.), признаки цвета, текстуры, материала и др.

Множество известных объектов обозначим через $O = \{o_1, o_2, o_3, ..., o_i, ..., o_n\}$, каждому из элементов которого может быть сопоставлена пара множеств признаков-примитивов и отношений $o_i = (s^{o_i}, q^{o_i})$, где $s^{o_i} \in S, q^{o_i} \in Q$. При этом $\forall s_j^{o_i} \in s^{o_i}$ можно определить множество отношений из $q_j^{o_i} \in q^{o_i}$, которые описывают его взаимодействие с $\forall s_k^{o_i} \in s^{o_i}$ для $j \neq k$.

Предположение 1 (о представлении объекта). *Любой объект может быть представлен как конечное множество признаков-примитивов $s \in S$, связанных между собой конечным множеством отношений $q \in Q$.*

Предположение 2 (о распознавании объекта). *Распознавание объекта o из множества известных объектов O может осуществляться как поиск совокупности признаков объекта $s \in S$, связанных между собой совокупностью отношений $q \in Q$.*

Предположение 3 (о сложном признаке). *Множество признаков* ***S*** *может включать в себя наряду с признаками-примитивами сложные признаки.*

Множество сложных признаков вводится для упрощения распознавания сложносоставных объектов.

Предположение 4 (о распознавании сложносоставных объектов). *Распознавание сложносоставного объекта o из множества известных объектов O может осуществляться как поиск совокупности сложных признаков и признаков примитивов $s \in S$, связанных между собой совокупностью отношений $q \in Q$.*

Множество отношений Q описывает взаимодействие объектов O в трехмерном пространстве. При работе с реальными объектами мы

можем видеть их с разных точек зрения, использовать различные проекции.

Для оперирования объектами и отношениями между ними необходимо введение точки отсчёта направления осей, относительно которых будем описывать объекты и отношения. В этом случае, зная нахождение точки зрения, можно выполнить необходимые преобразования в пространстве и преобразовать фигуры и отношения, использовать для распознавания объектов соответствующую точке зрения модель.

Для реализации описанного подхода для введённого множества примитивов необходимо определить положения, например, с помощью алгоритма Direct Linear Transform (DLT) [43], по ключевым точкам (рисунок 1).

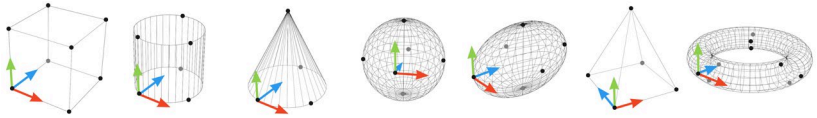


Рис. 1. Примеры геометрических примитивов с нанесёнными ключевыми точками и направляющими осями

Для описания сцены введём следующее множество пространственных отношений между объектами, формула (1).

$$Q = \{\text{над, под, справа от, слева от, за,}$$
$$\text{перед, выше, ниже, правее, левее,}$$
$$\text{дальше, ближе, наклон, поворот, .} \quad (1)$$
$$\text{удалённость, размер, соприкосновение,}$$
$$\text{стоит на, вставлен в}\}$$

Для идентификации необходимо несколько снимков со смещениями для определения удалённости.

Из приведённых предположений можно сделать вывод, что комбинации признаков и отношения между ними определяют распознаваемый объект. Последовательность распознавания объектов может быть представлена следующим образом:

1) Идентифицируем признаки-примитивы на изображении и определяем их положение относительно точки наблюдения (далее при проведении экспериментов использовалась свёрточная нейронная сеть YOLACT [44, 45] для поиска признаков-примитивов и ансамбли

регрессионных деревьев [46] вместе с алгоритмом Direct Linear Transform (DLT) [43] для определения положения признаков-примитивов).

2) Выбираем базовый признак-примитив, точку отсчёта. В качестве базового примитива может выбираться любой признак-примитив (его выбор необходим для определения отношений между признаками-примитивами).

3) Вычисляем положение точки наблюдения относительно системы координат базового признака-примитива.

4) Формируем множество всех признаков-примитивов и сложных признаков (пространственных отношений), присутствующих на анализируемом изображении $S_{img} \in S$, и множество отношений между ними $Q_{img} \in Q$.

5) Проверяем достаточность выделенных признаков для определения объектов на изображении, для этого формируем множество объектов (Q_{img}), для которых выполняется $S^{(O)} \in S_{img}$, $\forall O$.

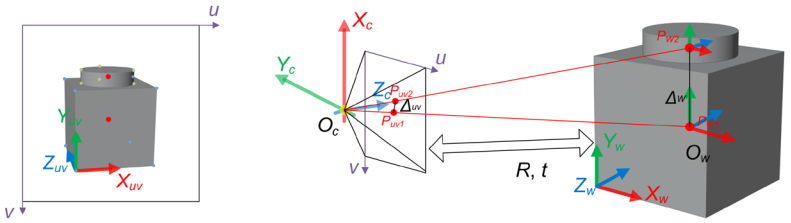
6) Если $O_{img} = \emptyset$, то объектов на изображении нет. Переходим на шаг 9.

7) Для каждого элемента (i) из множества Q_{img} проверяем между соответствующими ему признаками $S_{img}^{(O_{img}^{(i)})} \in S_{img}$ наличие необходимых отношений $Q_{img}^{(O_{img}^{(i)})} \in Q_{img}$. Из элементов, прошедших проверку, формируем множество $O_{img}^{*} \in O_{img}$ и исключаем соответствующие им признаки из множества признаков $S_{img} = S_{img} \Big/ S_{img}^{(O_{img}^{*})}$.

8) Добавляем распознанные объекты O_{img}^{*} в множество признаков $S_{img} = S_{img} + O_{img}^{*}$ и переходим на **шаг 2.**

9) Выводим множество O_{img}^{*}, которое представляет собой множество распознанных объектов.

Рассмотрим случай, когда общая для рассматриваемой сцены система координат ориентирована так, что ось 0X направлена вправо, ось 0Y вверх, а ось 0Z от наблюдателя (рисунок 2).



а б

Рис. 2. Пример проекции отношения на плоскость изображения: а) проекция объекта на плоскость изображения; б) трёхмерный вид соотношения объекта и камеры, где X_W, Y_W, Z_W, O_W – мировая система координат; X_C, Y_C, Z_C, O_C – система координат камеры; u, v – система координат изображения; R – матрица поворота камеры; t – вектор перемещения камеры)

На основе ограничивающих рамок и ключевых точек, отношения между примитивами A и B могут быть описаны следующими выражениями:

1) строгие отношения положения:

A над B: $\min(y^{(A)}) \geq \max(y^{(B)})$;

B под A: $\max(y^{(B)}) \leq \min(y^{(A)})$;

A справа от B: $\min(x^{(A)}) \geq \max(x^{(B)})$;

A слева от B: $\max(x^{(B)}) \leq \min(x^{(A)})$;

A за B: $\min(z^{(A)}) \geq \max(z^{(B)})$;

A перед B: $\max(z^{(A)}) \leq \min(z^{(B)})$;

2) «мягкие» отношения положения (примитивы имеют пересечение):

A выше B: $\min(y^{(A)}) < \max(y^{(B)}) \wedge \max(y^{(A)}) > \max(y^{(B)})$;

B ниже A: $\max(y^{(B)}) > \min(y^{(A)}) \wedge \min(y^{(B)}) > \min(y^{(A)})$;

A правее B: $\min(x^{(A)}) < \max(x^{(B)}) \wedge \max(x^{(A)}) > \max(x^{(B)})$;

A левее B: $\min(x^{(A)}) < \min(x^{(B)}) \wedge \max(x^{(A)}) > \min(x^{(B)})$;

A дальше B: $\min(z^{(A)}) < \max(z^{(B)}) \wedge \max(z^{(A)}) > \max(z^{(B)})$;

A ближе B: $\min(z^{(A)}) < \min(z^{(B)}) \wedge \max(z^{(A)}) > \min(z^{(B)})$;

3) отношения положения и взаимодействия:

Наклон A относительно B представлен формулой 2:

$$OY^{(A)} \angle OX^{(B)} \in [c_1, c_2] \ \vee \ OZ^{(A)} \angle OZ^{(B)} \in [c_1, c_2], \qquad (2)$$

где $OX^{(A)}$ – прямая оси **0*X*** для примитива **A** (остальные аналогично), $[c_1, c_2]$ – интервал значений наклона, характерного для пары примитивов при описании объекта.

Поворот A относительно B представлен формулой 3**:**

$$OY^{(A)} \angle OY^{(B)} \in [c_1, c_2], \tag{3}$$

где $OX^{(A)}$, $OY^{(B)}$ – прямые оси **0*Y*** для примитивов **A** и **B** соответственно, $[c_1, c_2]$ – интервал значений поворота, характерного для пары примитивов при описании объекта.

Удалённость A от B представлен формулой 4:

$$\Delta = \sqrt{\left(x_c^{(A)} - x_c^{(B)}\right)^2 + \left(y_c^{(A)} - y_c^{(B)}\right)^2 + \left(z_c^{(A)} - z_c^{(B)}\right)^2}, \tag{4}$$

где $x_c^{(A)}$, $x_c^{(A)}$, $y_c^{(A)}$, $y_c^{(B)}$, $z_c^{(A)}$, $z_c^{(B)}$ – координаты центра примитива, рассчет для $x_c^{(A)}$ представлен формулой (5) остальные рассчитываются аналогично.

$$x_c^{(A)} = \frac{\max(x^{(A)}) - \min(x^{(A)})}{2}. \tag{5}$$

Размер A относительно B представлен формулой 6:

$$V^{(A)} : V^{(B)}, \tag{6}$$

где $V^{(A)}$, $V^{(B)}$ – габаритные размеры примитива, рассчет для $V^{(A)}$ представлен формулой 7 остальные рассчитываются аналогично.

$$V^{(A)} = \left(\max(x^{(A)}) - \min(x^{(A)})\right) \cdot \left(\max(y^{(A)}) - \min(y^{(A)})\right) \cdot \cdot \left(\max(z^{(A)}) - \min(z^{(A)})\right) \tag{7}$$

Отношение площадей сечения A и B представлено формулой (8):

$$\begin{aligned} &S^{(A)}_{x\_\min} : S^{(B)}_{x\_\min} \in [c_1, c_2] \vee S^{(A)}_{x\_\max} : S^{(B)}_{x\_\max} \in [c_1, c_2] \vee \\ &\vee S^{(A)}_{y\_\min} : S^{(B)}_{y\_\min} \in [c_1, c_2] \vee S^{(A)}_{y\_\max} : S^{(B)}_{y\_\max} \in [c_1, c_2] \vee , \\ &\vee S^{(A)}_{z\_\min} : S^{(B)}_{z\_\min} \in [c_1, c_2] \vee S^{(A)}_{z\_\max} : S^{(B)}_{z\_\max} \in [c_1, c_2] \end{aligned} \quad (8)$$

где $S^{(A)}_{x\_\min}$ – площадь сечения $\min(x^{(A)})$ примитива $\boldsymbol{A}$, рассчет для $S^{(A)}_{x\_\min}$ представлен формулой (9) остальные рассчитываются аналогично.

$$S^{(A)}_{x\_\min} = \left(\max(y^{(A)}) - \min(y^{(A)})\right) \cdot \left(\max(z^{(A)}) - \min(z^{(A)})\right). \quad (9)$$

Сопрокосновение A с B представлено формулой (10):

$$\begin{aligned} &\max(x^{(A)}) = \min(x^{(B)}) \vee \max(y^{(A)}) = \min(y^{(B)}) \vee \\ &\vee \max(z^{(A)}) = \min(z^{(B)}) \vee \max(x^{(B)}) = \min(x^{(A)}) \vee . \\ &\vee \max(y^{(B)}) = \min(y^{(A)}) \vee \max(z^{(B)}) = \min(z^{(A)}) \end{aligned} \quad (10)$$

Выражение **A стоит на B** представлено формулой (11):

$$\begin{aligned} &\max(y^{(A)}) = \min(y^{(B)}) \wedge \\ &\wedge(\min(x^{(A)}) < \max(x^{(B)}) \wedge \max(y^{(A)}) < \min(y^{(B)})) \wedge . \\ &\wedge(\min(z^{(A)}) < \max(z^{(B)}) \wedge \max(z^{(A)}) < \min(z^{(B)})) \end{aligned} \quad (11)$$

Выражение **A вставлен в B** представлено формулой (12):

$$\begin{aligned} &V^{(A)} < V^{(B)} \wedge \min(y^{(A)}) < \max(y^{(B)}) \wedge \\ &\wedge(\min(x^{(A)}) < \max(x^{(B)}) \wedge \max(x^{(A)}) > \min(x^{(B)})) \wedge . \\ &\wedge(\min(z^{(A)}) < \max(z^{(B)}) \wedge \max(z^{(A)}) > \min(z^{(B)})) \end{aligned} \quad (12)$$

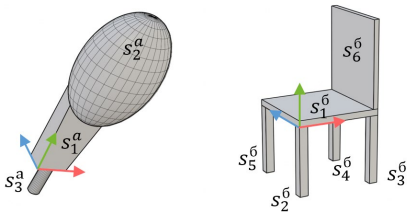
Работа алгоритма опирается на множество известных признаков S , отношений между признаками Q и множество известных объектов с предзаданной структурой O. На выходе алгоритма мы получаем множество, состоящее из распознанных объектов, сложных признаков и признаков-примитивов. Такой подход даёт нам возможность анализировать полученные результаты

и вносить изменения в множества примитивов, отношений и объектов, улучшая качество работы предлагаемого алгоритма.

3. Результаты. Рассмотрим сцену, приведённую на рисунке 3. Множества S и Q для неё примут следующие значения:

$$S = \{\textit{призма}, \textit{цилиндр}, \textit{конус}, \textit{эллипсоид}, \textit{пирамида}, \textit{тор}\}, \quad (13)$$

$$Q = \{\textit{над}, \textit{под}, \textit{справа от}, \textit{слева от}, \textit{за}, \textit{перед}, \textit{выше}, \textit{ниже}, \textit{правее}, \textit{левее}, \textit{дальше}, \textit{ближе}, \textit{наклон}, \textit{поворот}, \textit{удалённость}, \textit{размер}, \textit{соприкосновение}, \textit{стоит на}, \textit{вставлен в}, \textit{отношение сторон}\}. \quad (14)$$



а б

Рис. 3. Представление объектов окружающего мира с использованием признаков-примитивов

Для распознавания объектов необходимо ввести эталонные объекты и отношения, которые будут однозначно идентифицировать рассматриваемые объекты.

Для фонаря, приведённого на рисунке 3(а), такие эталонные описания могут быть представлены с использованием множеств $S_a = \{s_1^a, s_2^a, s_3^a\}$, где $s_1^a = \textit{призма}$, $s_2^a = \textit{эллипсоид}$, $s_3^a = \textit{цилиндр}$ и $Q_a = \{\textit{наклон}, \textit{стоит на}, \textit{выше}, \textit{ниже}, \textit{над}, \textit{под}, \textit{размер}, \textit{удалённость}\}$ в форме представленной формулой (15):

$$\begin{gathered} Q_a^S = \{s_2^a \ \textit{стоит на}\ s_1^a \ \textit{И}\ s_1^a \ \textit{размер}\ s_2^a \ \textit{И} \\ \textit{И} \ \ s_1^a \ \textit{наклон}\ s_2^a \ \textit{И}\ s_1^a \ \textit{ниже}\ s_2^a \ \textit{И} \\ \textit{И}\ s_1^a \ \textit{удаленность}\ s_2^a \ \textit{И}\ s_1^a \ \textit{размер}\ s_3^a \ \textit{И} \\ \textit{И} \ \ s_1^a \ \textit{наклон}\ s_3^a \ \textit{И}\ s_1^a \ \textit{над}\ s_3^a \ \textit{И} \\ \textit{И}\ s_1^a \ \textit{удаленность}\ s_3^a \ \}\ \textit{ИЛИ} \\ \textit{ИЛИ}\ \{s_2^a \ \textit{стоит на}\ s_1^a \ \textit{И}\ s_1^a \ \textit{размер}\ s_2^a \ \textit{И} \\ \textit{И}\ s_1^a \ \textit{ниже}\ s_2^a \ \textit{И}\ s_1^a \ \textit{удаленность}\ s_2^a \}. \end{gathered} \quad (15)$$

Для стула, приведённого на рисунке 3(б), эталонные описания могут быть представлены с использованием множеств $S_{б} = \{s_1^{б}, s_2^{б}, s_3^{б}, s_4^{б}, s_5^{б}, s_6^{б}\}$, где $s_1^{б} = призма$, $s_2^{б} = призма$, $s_3^{б} = призма$, $s_4^{б} = призма$, $s_5^{б} = призма$, $s_6^{б} = призма$ и $Q_{б}$, представленная формулой (16). Для рисунка 3(б) одним из вариантов описания объекта может быть представлено формулой (17).

$$Q_{б} = \{наклон, стоит\ на, левее, правее, выше, ниже, \\ размер, отношение\ площадей\}, \tag{16}$$

$$\begin{gathered} Q_{б}^{S} = \{s_1^{б}\ стоит\ на\ s_2^{б}\ \ И\ s_1^{б}\ над\ s_2^{б}\ И \\ И\ \ s_1^{б}\ размер\ s_2^{б}\ И\ s_1^{б}\ правее\ s_2^{б}\ И \\ И\ s_1^{б}\ дальше\ s_2^{б}\ \ И\ s_1^{б}\ удаленность\ s_2^{б}\ И \\ И\ \ s_1^{б}\ отношение сторон\ s_2^{б}\ И \\ И\ s_1^{б}\ стоит\ на\ s_3^{б}\ \ И\ s_1^{б}\ над\ s_3^{б}\ И \\ И\ \ s_1^{б}\ размер\ s_3^{б}\ И\ s_1^{б}\ правее\ s_3^{б}\ И \\ И\ s_1^{б}\ дальше\ s_3^{б}\ \ И\ s_1^{б}\ удаленность\ s_3^{б}\ И \\ И\ \ s_1^{б}\ отношение сторон\ s_3^{б}\ И \\ И\ s_1^{б}\ стоит\ на\ s_4^{б}\ \ И\ s_1^{б}\ над\ s_4^{б}\ И \\ И\ \ s_1^{б}\ размер\ s_4^{б}\ И\ s_1^{б}\ левее\ s_4^{б}\ И \\ И\ s_1^{б}\ ближе\ s_4^{б}\ \ И\ s_1^{б}\ удаленность\ s_4^{б}\ И \\ И\ \ s_1^{б}\ отношение сторон\ s_4^{б}\ И \\ И\ s_1^{б}\ стоит\ на\ s_5^{б}\ \ И\ s_1^{б}\ над\ s_5^{б}\ И \\ И\ \ s_1^{б}\ размер\ s_5^{б}\ И\ s_1^{б}\ правее\ s_5^{б}\ И \\ И\ s_1^{б}\ ближе\ s_5^{б}\ \ И\ s_1^{б}\ удаленность\ s_5^{б}\ И \\ И\ \ s_1^{б}\ отношение сторон\ s_5^{б}\ И \\ И\ s_6^{б}\ стоит\ на\ s_1^{б}\ \ И\ s_6^{б}\ над\ s_1^{б}\ И \\ И\ \ s_1^{б}\ размер\ s_6^{б}\ И\ s_1^{б}\ левее\ s_6^{б}\ И \\ И\ s_1^{б}\ удаленность\ s_6^{б}\ И\ s_1^{б}\ отношение сторон\ s_6^{б}\}. \end{gathered} \tag{17}$$

Для идентификации признаков-примитивов при распознавании объектов предложенным методом будем использовать свёрточную нейронную сеть для экземплярной сегментации объектов YOLACT

[44, 45]. Выбор этой архитектуры нейронной сети был основан на сравнении скорости и точности известных архитектур нейронных сетей приведённом в статье [44] (выбранная архитектура нейронной сети согласно данным приведённой статьи обладает наибольшей производительностью и входит в 1/3 лучших по точности, что крайне важно для автономных робототехнических систем, которые и манипулируют объектами внешнего мира). Сравнение сегментационных моделей производилось на наборе данных Common Object in Context (COCO) [47]. В качестве признаков будем использовать трёхмерные фигуры – призма, цилиндр, эллипсоид, пирамида. Для идентификации признаков-примитивов сеть была переобучена на искусственно сформированном наборе данных.

В качестве обучающей выборки используем набор данных, сгенерированный с помощью программной платформы BlenderProc [48]. Она позволяет сформировать размеченную выборку данных для экземплярной сегментации в формате COCO на основе 3D сцены в Blender 3D [49]. Для генерации использовалось 4 признака-примитива (призма, цилиндр, эллипсоид, сфера). Для их разметки были сформированы сцены в Blender 3D, на которых признаки-примитивы располагались случайным образом (их положение и поворот) в ограниченных пределах координат (заданных размером сцены). При разметке определялись границы признака-примитива на сгенерированном изображении путём их автоматического проецирования на плоскость изображения камеры (объекта Blender 3D).

В процессе генерации будем выделять признаки-примитивы контуром и ограничивающей рамкой, а также случайным образом задавать освещение сцены, положение камеры, накладывать случайные текстуры, выбирать фон для сцены, задавать положение и ориентацию признаков-примитивов (рисунок 4).

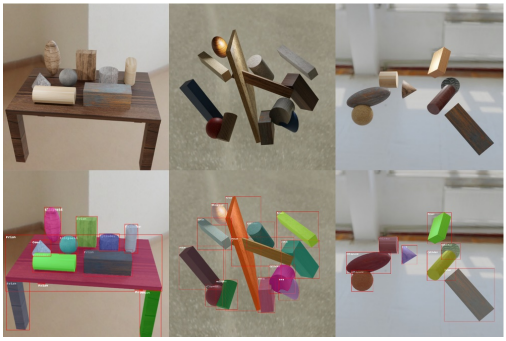


Рис. 4. Примеры сгенерированных для обучающей выборки сцен

Сгенерированные для обучения изображения содержат, в том числе, пересечения различных признаков-примитивов друг с другом, что вызывает ситуации, когда часть признака-примитива не видна. Такие примеры позволяют учесть ситуацию, когда при разделении объекта на признаки-примитивы, части признака формы скрыты другим признаком, что довольно распространено.

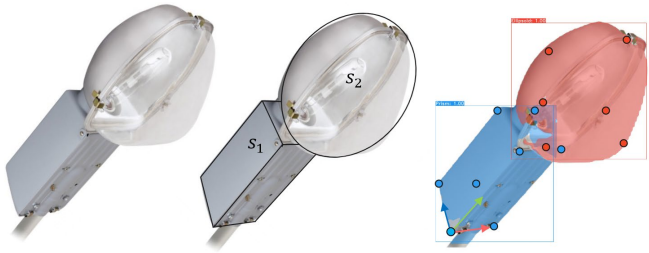
Для обучения был сгенерирован набор данных, состоящий из 2500 размеченных изображений для обучения и 1000 изображений для оценки качества модели. Обучение продолжалось в течение 480 эпох (для обучения использовался GPU Nvidia GeForce RTX 2060).

Для поиска точек, требуемых для определения положения примитива, используем ансамбли регрессионных деревьев. Пример использования ансамбля регрессионных деревьев для поиска ключевых точек представлен на рисунке 5 [46, 50].

Рассмотрим пример идентификации объектов реального мира на примере светильников уличного освещения. Пример работы алгоритма для распознавания светильника наружного освещения приведён на рисунке 5.

Исходя из изображений, представленных на рисунке 5, сформировать описание объекта $O_{св}$ можно из множеств $S_{св} = \{s_1^{св}, s_2^{св}\}$, где $s_1^{св} = призма$; $s_2^{св} = эллипсоид \vee призме \vee цилиндр$; $Q_{св} = \{наклон, стоит\ на, выше, ниже, размер, удалённость\}$. $Q_{св}^S$ представлена формулой (18).

$$Q_{св}^S = (s_2^{св}\ стоит\ на\ s_1^{св}\ \ И\ s_1^{св}\ размер\ s_2^{св}\ \ И \\ И\ \ s_1^{св}\ ниже\ s_2^{св}\ И\ s_1^{св}\ удаленность\ s_2^{св}\ ), \quad (18)$$



а б в

Рис. 5. Иллюстрация работы алгоритма распознавания на простом примере распознавания светильников уличного освещения: а) распознаваемый объект, б) выделенные примитивы из объекта, в) поиск примитивов и выделение ключевых точек)

При распознавании объекта на изображении выделены следующие признаки и определено их положение: $S_{img} = \{s_1^{img}, s_2^{img}\}$, где $s_1^{img} = призма$; $s_2^{img} = эллипсоид$.

Среди выделенных признаков базовым был выбран s_1^{img}. Относительно его положения далее описываются отношения между примитивами. Сформировано множество отношений на изображении Q_{img}^S, представленное формулой (19).

$$Q_{img}^S = (s_2^{img}\ стоит\ на\ s_1^{img}\ \ И\ s_1^{img}\ размер\ s_2^{img}\ И \\ И\ \ s_1^{img}\ ниже\ s_2^{img}\ И\ \ s_2^{img}\ выше\ s_1^{img}\ И \\ И\ s_1^{св}\ удаленность\ s_2^{св}\ ) \quad . \qquad (19)$$

Выделенного множества признаков и отношений между ними достаточно для идентификации единственного искомого объекта (светильника). После проверки соответствия наличия необходимых отношений между всеми признаками объекта, множества O_{img}, S_{img}, Q_{img} примут следующий вид: $O_{img} = \{O_{св}\}$, $S_{img} = \{O_{св}\}$, $Q = \emptyset$ – что говорит о наличии объекта класса «светильник» на изображении. Представленный пример, опирается на сегментацию признаков-примитивов на входном изображении свёрточной нейронной сетью. В качестве обучающих данных были использованы автоматически сгенерированные изображения, полученные с помощью программной платформы BlenderProc [48] на основе 3D сцен с объектами в Blender 3D. Данная программная платформа позволяет формировать изображения с признаками примитивами учитывая разные изменяемые параметры: освещение, точку обзора, фон, положение признаков-примитивов, текстуры признаков-примитивов. Для обучения сети, описанной в статье, были использованы три сцены с заданным набором признаков-примитивов, набор из 20 текстур (среди которых есть текстуры с составным рисунком, например, кирпичная кладка) и 25 фоновых изображений, что позволяет сделать относительно разнообразную выборку, но не охватывающую все возможные варианты, которые могут встречаться в реальном мире. Приведённый пример позволяет подтвердить гипотезу о том, что из объектов можно выделить признаки-примитивы существующими методами распознавания.

4. Дискуссия. Описание всех известных объектов и текущей сцены необходимо хранить в памяти во время процесса распознавания.

В зависимости от задачи, возможно большое количество распознаваемых на сцене отношений и признаков-примитивов, что выдвигает требования к способу хранения информации о них (формированию базы отношений и примитивов, описывающих объекты). В логику алгоритма хорошо вписываются графовые базы данных [18], что позволяет в явном виде хранить информацию. Если рассмотреть такой способ хранения информации об уличном светильнике, то получим структуру, где признаки-примитивы – это узлы графа, а отношения – это направленные рёбра графа (рисунок 6).

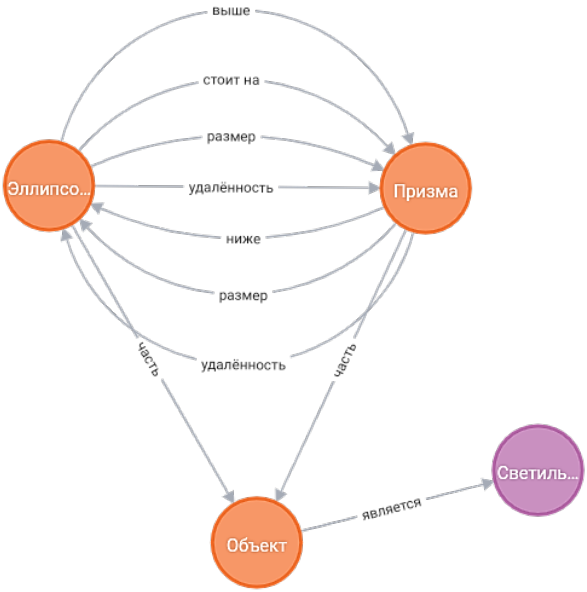


Рис. 6. Пример описания светильника наружного освещения в виде графа

Качество распознавания признаков-примитивов напрямую зависит от набора данных, однако процесс подготовки большого размеченного набора реальных данных является очень трудоёмким. На основе анализа работы текущего обученного алгоритма сформулированы следующие возможные способы повышения качества распознавания при использовании полностью искусственного набора данных:

– увеличение выборки, в том числе увеличение разнообразия сцен с объектами, собранными из признаков-примитивов (например, стол (рисунок 5));

– добавление сцен с объектами, состоящими из признаков-примитивов с плавными переходами между признаками-примитивами;

– возможность использования одинаковой текстуры для всех признаков-примитивов на сцене;

– использование разных текстур для разных граней одного признака-примитива.

Разработанный алгоритм распознавания, опирающийся на поиск признаков-примитивов объектов и определение отношений между ними, позволяет распознавать разнообразные объекты без необходимости переобучения алгоритма с предварительной подготовкой большого количества обучающих данных. Для дополнения списка распознаваемых объектов достаточно дополнить базу распознаваемых объектов описанием нового объекта или класса объектов. Помимо этого, множество признаков примитивов, детализация описания объекта, алгоритм поиска могут быть изменены в зависимости от требований к качеству распознавания и требований к вычислительным ресурсам. Распознанные признаки-примитивы и отношения между ними могут быть использованы для описания новых, неизвестных объектов в полуавтоматическом режиме. Множества признаков-примитивов и отношений между ними были сформулированы, исходя из опыта авторов.

Недостатком является необходимость описания различных конфигураций объектов одного класса, что может быть затруднительно. Однако такой процесс может быть автоматизирован, например, с помощью самого алгоритма, или, например, с помощью дополнительного ПО, которое позволит собирать объекты из блоков (признаков-примитивов) и автоматически генерировать описание для них.

Из экспериментов, проведённых авторами, можно ожидать, что могут быть ложноположительные срабатывания алгоритма на конструкции, похожие на распознаваемые объекты, но не являющиеся ими. Поэтому одним из направлений дальнейшего развития описанного подхода может быть обработка исключений при распознавании. Исключения могут формулироваться при описании объекта (например, требования к цвету, материалу, текстуре, положению, анализ областей соединения признаков-примитивов).

Полученный алгоритм позволяет расширить существующие подходы сбора и хранения информации об окружающем пространстве и объектах в нём, построении карт пространств (SLAM методы [51]). Описанный подход позволяет устранить главный недостаток SLAM алгоритмов, связанный с отсутствием информации об объектах в

пространстве, с которыми могут производиться манипуляции робототехническими системами.

В настоящее время задачи манипуляции объектами решаются индивидуально для каждого случая. Например, в статье [16] рассматривается решение задачи захвата нужного объекта и складывания его в корзину в рамках конкурса Amazon Picking Challenge. Описанный подход основывается на нейронной сети для распознавания и сегментации ограниченного количества объектов на изображении, которые используются в конкурсе, и опирается на их 3D модели для определения положения. Для адаптации подхода к новым объектам необходима трудоёмкая подготовка новых обучающих данных и 3D моделей объектов.

В статье [52] рассматривается задача бросания роботом предметов на основе обучения с подкреплением. На основе разработанного авторами подхода робот успешно совершает броски объектов, однако при захвате не производится классификация объектов, а захват осуществляется только на основе данных о геометрии объектов.

В статье [53] рассматривается задача распознавания и оценки положения объекта на основе одного изображения. Распознавание основывается на построении унифицированного представления множества экземпляров категории объектов. Для создания унифицированного представления используется вариационный автоэнкодер, который формирует независимое от положения представление объекта для каждой из категорий. Для оценки положения новых объектов они сравниваются с этим унифицированным представлением. Такой подход к распознаванию объектов позволяет довольно точно распознавать объекты и определять их положение на основе одного изображения, однако требует большого количества размеченных данных для переобучения и адаптации алгоритма к распознаванию новых объектов, а также ограничивает распознавание только классовой принадлежностью объекта, не позволяя классифицировать объекты внутри одного класса.

Разработанный алгоритм отличается тем, что:

– не требует для дополнения списка распознаваемых объектов подготовки новых обучающих данных и переобучение модели, это заменяется на дополнение базы описаний объектов описанием нового объекта;

– алгоритм не предъявляет требований к выбору способа реализации поиска признаков-примитивов (выбор может

осуществляться в зависимости от требований к качеству распознавания и требований к вычислительным ресурсам);

– распознанные признаки-примитивы и отношения между ними могут быть использованы для описания новых объектов в автоматическом/полуавтоматическом режиме (подход позволяет автоматизировать процесс формирования описания объектов (формирование графов взаимосвязей, рисунок 6)).

5. Заключение. В работе рассмотрен алгоритм идентификации объектов реального мира на фотографическом изображении, основанный на гипотезе о возможности распознавания объектов окружающего мира опираясь на ограниченное число признаков-примитивов и отношений между ними. Это позволяет не снижать качество распознавания при увеличении числа распознаваемых объектов.

В результате проведённого исследования удалось выяснить, что задача идентификации объектов внешнего мира может быть решена на синтетических данных и экспертных знаниях об устройстве/конфигурации объектов внешнего мира, что решает проблему создания качественного набора данных для обучения нейронной сети.

Предложенный алгоритм позволяет уйти от сбора большого количества примеров объекта и трудоёмкой разметки наборов данных для обучения, позволяя расширить количество распознаваемых объектов, добавив только их описание. Результаты, описанные в статье, могут быть адаптированы для идентификации объектов в других формах представления. Например, при получении облака точек (получаемого в результате 3D сканирования) для реконструкции окружающего мира (получения 3D модели). Информация об окружающем пространстве в виде облака точек включает в себя информацию об удалённости объектов от сканера в явном виде (в отличие от плоского снимка), что позволяет наиболее точно определить геометрию и пространственные отношения объектов для их идентификации, но при этом несет и дополнительные особенности, которые связанны с возможной деформацией полигональной сетки, наличием множества моделей объектов [54]. Построение 3D сцен, в свою очередь, открывает возможности применения в робототехнических системах в задачах ориентации в пространстве и манипуляции объектами реального мира. При этом можно ожидать, что ориентация на выявление примитивов (крупных форм) позволит показать лучшие результаты работы с помехами, связанными с освещением, осадками (туман, дождь, снег), перекрытиями

прозрачными и частично прозрачными объектами (перекрытия сетками, деревьями и т.п.), при деформации объектов (при сильном ветре, получении вмятин и трещин).

Кроме того, разделение объектов на примитивы позволяет задуматься о возможности для использования в генеративных моделях и научно-техническом творчестве на основе методов морфологического анализа и синтеза применительно к дизайну исследуемых объектов.

Литература

1. Meel V. The 87 Most Popular Computer Vision Applications for 2023. 2022. Available at: https://viso.ai/applications/computer-vision-applications/ (accessed: 23.11.2022).
2. Urbonas A., Raudonis V., Maskeliūnas R., Damaševičius R. Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning // Appl. Sci. 2019. vol. 9(22). pp. 4898.
3. Орешин А.Н., Лысанов И.Ю. Новый метод автоматизации процессов аутентификации персонала с использованием видеопотока // Труды СПИИРАН. 2017. Т. 5. № 54. С. 35–56.
4. Bureš L., Gruber I, Neduchal P., Hlaváč M., Hruz M. Semantic text segmentation from synthetic images of full-text documents // SPIIRAS Proc. 2019. vol. 18(6). pp. 1380–1405.
5. Yu F., Chen H, Wang X., Xian W., Chen Y., Liu F., Madhavan V., Darrell T. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2020. pp. 2633–2642.
6. Slivnitsin P., Bachurin A., Mylnikov L. Robotic system position control algorithm based on target object recognition // Proceedings of International Conference on Applied Innovation in IT. Anhalt University of Applied Sciences. 2020. vol. 8(1). pp. 87–94.
7. Чиров Д.С., Чертова О.Г., Потапчук Т.Н. Методика обоснования требований к системе технического зрения робототехнического комплекса // Труды СПИИРАН. 2017. Т. 2. № 51. С. 152–176.
8. Delfanti A., Frey B. Humanly Extended Automation or the Future of Work Seen through Amazon Patents // Sci. Technol. Hum. Values. 2021. vol. 46. no. 3. pp. 655–682.
9. Al-Azzo F., Taqi A.M., Milanova M. Human related-health actions detection using Android Camera based on TensorFlow Object Detection API // Int. J. Adv. Comput. Sci. Appl. 2018. vol. 9. no. 10. pp. 9–23.
10. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Zh., Karpathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge // Int. J. Comput. Vis. 2015. vol. 115. no. 3. pp. 211–252.
11. Zou Z., Chen K., Shi Zh., Guo Yu., Ye J. Object Detection in 20 Years: A Survey // arXiv. 2019. pp. 1–39.
12. He K., Gkioxari G., Dollár P., Girshick R.. Mask R-CNN // IEEE Trans. Pattern Anal. Mach. Intell. 2020. vol. 42. no. 2. pp. 386–397.
13. Kirillov A., He K., Girshick R., Rother C., Dollar P. Panoptic segmentation // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019. vol. 2019-June. pp. 9396–9405.

14. Bazarevsky V, Grishchenko I, Raveendran K., Zhu T., Zhang F., Grundmann M. BlazePose: On-device real-time body pose tracking // arXiv. 2020.
15. Khan K., Ahmad N., Ullah K., Din I. Multiclass semantic segmentation of faces using CRFs // Turkish J. Electr. Eng. Comput. Sci. 2017. vol. 25. no. 4. pp. 3164–3174.
16. Zeng A, Yu K.-T., Song S., Suo D., Walker Jr.E., Rodriguez A., Xiao J. Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge // 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017. pp. 1386–1383.
17. Yaguchi H., Nagahama K., Hasegawa T., Inaba M. Development of an autonomous tomato harvesting robot with rotational plucking gripper // IEEE Int. Conf. Intell. Robot. Syst. 2016. vol. 2016-Novem. pp. 652–657.
18. Mylnikov L., Slivnitsin P., Mylnikova A. Robotic System Operation Specification on the Example of Object Manipulation // Proc. Int. Conf. Appl. Innov. IT. 2022. vol. 10. no. 1. pp. 51–59.
19. Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks // 2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc. 2013. 16 p.
20. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features // Proceedings IEEE Conf. on Computer Vision and Pattern Recognition. 2001. pp. 511–518.
21. Dalal N., Triggs B. Histograms of oriented gradients for human detection // Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005. 2005. vol. 1(16). pp. 886–893.
22. Felzenszwalb P., McAllester D., Ramanan D. A discriminatively trained, multiscale, deformable part model // 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008. vol. 330. no. 6. pp. 1–8.
23. Wolpert D.H., Macready W.G. No free lunch theorems for optimization // IEEE Trans. Evol. Comput. 1997. vol. 1(1). pp. 67–82.
24. Slivnitsin P., Kniazev A., Mylnikov L., Schlechtweg S., Kokoulin A. Influence of Synthetic Image Datasets on the Result of Neural Networks for Object Detection // Proc. Int. Conf. Appl. Innov. IT. 2021. vol. 9(1). pp. 55–60.
25. Abramovich F., Pensky M. Classification with many classes: Challenges and pluses // J. Multivar. Anal. 2019. vol. 174. pp. 1–25.
26. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016. vol. 2016-Decem. pp. 779–788.
27. Liu W., Anguelov D., Erhan D. Szegedy C., Reed S., Fu C.-Y., Berg A.C SSD: Single Shot MultiBox Detector // Eccv / (Eds.: Leibe B.). Cham: Springer International Publishing, 2016. vol. 9905. pp. 398–413.
28. Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2019. vol. 2019-June. pp. 658–666.
29. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // IEEE Trans. Pattern Anal. Mach. Intell. 2017. vol. 39(6). pp. 1137–1149.
30. Gomes H.M. Model learning in iconic vision // PQDT – UK & Ireland. 2002. 212 p.
31. Salas-Moreno R.F., Newcombe R.A., Strasdat H., Kelly P.H.J., Davison A.J. SLAM++: Simultaneous localisation and mapping at the level of objects // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2013. pp. 1352–1359.
32. Dai A., Nießner M. 3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2018. vol. 11214 LNCS. pp. 458–474.

33. Dai A., Chang A.X., Savva M., Halber M., Funkhouser T., Nießner M. ScanNet: Richly-annotated 3D reconstructions of indoor scenes // Proc. – 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017. vol. 2017-Janua. pp. 2432–2443.
34. Le T., Duan Y. PointGrid: A Deep Network for 3D Shape Understanding // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2018. pp. 9204–9214.
35. Su H., Maji S., Kalogerakis E., Learned-Miller E. Multi-view Convolutional Neural Networks for 3D Shape Recognition // 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015. vol. 32(1). pp. 945–953.
36. Choy C., Park J., Koltun V. Fully convolutional geometric features // Proc. IEEE Int. Conf. Comput. Vis. 2019. vol. 2019-Octob. pp. 8957–8965.
37. Biederman I. Recognition-by-Components: A Theory of Human Image Understanding // Psychol. Rev. 1987. vol. 94(2). pp. 115–147.
38. Thompson P. Margaret Thatcher: A New Illusion // Perception. 1980. vol. 9(4). pp. 483–484.
39. Biederman I. Visual object rocognition // An Invitation to Cognitive Science. (Eds.: Kosslyn S.M., Osherson D.N.) Cambridge: MIT Press, 1995. pp. 121–165.
40. Winston P.H. Artificial intelligence. Addison-Wesley Longman Publishing Co., Inc. Boston, MA: Addison-Wesley Publishing Company, 1992. 737 p.
41. Marr D., Poggio T. A computational theory of human stereo vision // Proc. R. Soc. London - Biol. Sci. 1979. vol. 204. no. 1156. pp. 301–328.
42. Marr D., Nishihara H.K. Representation and recognition of the spatial organization of three-dimensional shapes // Proc. R. Soc. London. Ser. B. Biol. Sci. 1978. vol. 200. no. 1140. pp. 269–294.
43. Abdel-Aziz Y.I., Karara H.M. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry // Photogramm. Eng. Remote Sensing. 2015. vol. 81(2). pp. 103–107.
44. Bolya D, Zhou C., Xiao F., Lee Y.J. YOLACT++ Better Real-Time Instance Segmentation // IEEE Trans. Pattern Anal. Mach. Intell. 2022. vol. 44. no. 2. pp. 1108–1121.
45. Bolya D, et al. You Only Look At CoefficienTs. 2020. Available at: https://github.com/dbolya/yolact (accessed: 11.11.2022).
46. Kazemi V., Sullivan J. One Millisecond Face Alignment with an Ensemble of Regression Trees // Rev. Anthropol. 1992. vol. 21(2). pp. 147–157.
47. Lin T.Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick C.L., Dollar P. Microsoft COCO: Common objects in context // Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2014. vol. 8693 LNCS(5). pp. 740–755.
48. Denninger M., Sundermeyer M., Winkelbauer D., Zidan Y., Olefir D., Elbadrawy M., Lodhi A., Katam H.T. BlenderProc. 2019. 7 p. DOI: 10.48550/arXiv.1911.01911.
49. Blender 3D. Available at: https://www.blender.org/ (accessed: 22.11.2022).
50. Slivnitsin P. Position estimation for robotic system positioning using the example of outdoor luminaire replacement: master thesis. Koethen: HS Anhalt, 2021. 54 p.
51. Vershinin D., Mylnikov L. A review and comparison of mapping and trajectory selection algorithms // Proc. Int. Conf. Appl. Innov. IT. 2021. vol. 9(1). pp. 85–92.
52. Zeng A. и др. TossingBot: Learning to Throw Arbitrary Objects With Residual Physics // IEEE Trans. Robot. 2020. vol. 36(4). pp. 1307–1319.
53. Chen D. и др. Learning Canonical Shape Space for Category-Level 6D Object Pose and Size Estimation // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020. pp. 11970–11979.
54. Koch S. и др. ABC: A big cad model dataset for geometric deep learning // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019. vol. 2019-June. pp. 9593–9603.

Сливницин Павел Александрович — аспирант, кафедра информационных технологий и автоматизированных систем электротехнического факультета, ФГАОУ ВО «Пермский национальный исследовательский политехнический университет». Область научных интересов: информационные системы и процессы, кибернетика, data science, искусственный интеллект, компьютерное зрение, распознавание образов. Число научных публикаций — 11. slivnitsin.pavel@gmail.com; улица профессора Поздеева, 7, 614013, Пермь, Россия; р.т.: +7(342)239-1354.

Мыльников Леонид Александрович — доцент, кафедра микропроцессорных средств автоматизации электротехнического факультета, кафедра информационные технологии и автоматизированные системы электротехнического факультета, ФГАОУ ВО «Пермский национальный исследовательский политехнический университет»; заведующий лабораторией, научно-учебная лаборатория междисциплинарных эмпирических исследований, НИУ ВШЭ. Область научных интересов: информационные системы и процессы, кибернетика, системная инженерия, управление в организационных системах. Число научных публикаций — 110. lamylnikov@hse.ru; улица Студенческая, 38, 614070, Пермь, Россия; р.т.: +7(342)200-9555.

Поддержка исследований. Работа выполнена при финансовой поддержке Правительства Пермского края в рамках научного проекта № С-26/692.

DOI 10.15622/ia.22.3.2

P. SLIVNITSIN, L. MYLNIKOV

OBJECT RECOGNITION BY COMPONENTS AND RELATIONS BETWEEN THEM

Slivnitsin P., Mylnikov L. **Object Recognition by Components and Relations between Them.**

Abstract. The paper's goal is to develop a methodology and algorithm for the recognition of objects in the environment, keeping the quality with an increasing number of objects. For this purpose, the following problems were solved: recognition of the shape features, estimation of relations between features, and matching between the found features and relations and the defined templates (descriptions of complex and simple objects of the real world). A convolutional neural network is used for the shape feature recognition. In order to train it we used artificially generated images with shape features (3D primitive objects) that were randomly placed on the scene with different properties of their surfaces. The set of relations necessary to recognize objects, which can be represented as a combination of shape features, is formed. Testing on photos of real-world objects showed the ability to recognize real-world objects regardless of their type (in cases where different models and modifications are possible). This paper considers an example of outdoor luminaire recognition. The example shows the algorithm's ability not only to detect an object in the image but also to estimate the position of its components. This solution makes it possible to use the algorithm in the task of object manipulation performed by robotic systems.

Keywords: object recognition, shape features, shape feature relation, computer vision, neural network.

References

1. Meel V. The 87 Most Popular Computer Vision Applications for 2023. 2022. Available at: https://viso.ai/applications/computer-vision-applications/ (accessed: 23.11.2022).
2. Urbonas A., Raudonis V., Maskeliūnas R., Damaševičius R. Automated identification of wood veneer surface defects using faster region-based convolutional neural network with data augmentation and transfer learning. Appl. Sci. 2019. vol. 9(22). pp. 4898.
3. Oreshin A.N., Lisanov I.Y. [A new method for automation of the personnel authentication process using a video stream]. Trudy SPIIRAN – SPIIRAS Proc. 2017. vol. 5(54). pp. 35–56. (In Russ.).
4. Bureš L., Gruber I, Neduchal P., Hlaváč M., Hruz M. Semantic text segmentation from synthetic images of full-text documents. SPIIRAS Proc. 2019. vol. 18(6). pp. 1380–1405.
5. Yu F., Chen H, Wang X., Xian W., Chen Y., Liu F., Madhavan V., Darrell T. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2020. pp. 2633–2642.
6. Slivnitsin P., Bachurin A., Mylnikov L. Robotic system position control algorithm based on target object recognition. Proceedings of International Conference on Applied Innovation in IT. Anhalt University of Applied Sciences. 2020. vol. 8(1). pp. 87–94.

7. Chirov D.S., Chertova O.G., Potapchuk T.N. [Methods of study requirements for the complex robotic vision system]. Trudy SPIIRAN – SPIIRAS Proc. 2017. vol. 2(51). pp. 152–176. (In Russ.)
8. Delfanti A., Frey B. Humanly Extended Automation or the Future of Work Seen through Amazon Patents. Sci. Technol. Hum. Values. 2021. vol. 46. no. 3. pp. 655–682.
9. Al-Azzo F., Taqi A.M., Milanova M. Human related-health actions detection using Android Camera based on TensorFlow Object Detection API. Int. J. Adv. Comput. Sci. Appl. 2018. vol. 9. no. 10. pp. 9–23.
10. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Zh., Karpathy A., Khosla A., Bernstein M., Berg A.C., Fei-Fei L. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. Vis. 2015. vol. 115. no. 3. pp. 211–252.
11. Zou Z., Chen K., Shi Zh., Guo Yu., Ye J. Object Detection in 20 Years: A Survey. arXiv. 2019. pp. 1–39.
12. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN. IEEE Trans. Pattern Anal. Mach. Intell. 2020. vol. 42. no. 2. pp. 386–397.
13. Kirillov A., He K., Girshick R., Rother C., Dollar P. Panoptic segmentation. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019. vol. 2019-June. pp. 9396–9405.
14. Bazarevsky V, Grishchenko I, Raveendran K., Zhu T., Zhang F., Grundmann M. BlazePose: On-device real-time body pose tracking. arXiv. 2020.
15. Khan K., Ahmad N., Ullah K., Din I. Multiclass semantic segmentation of faces using CRFs. Turkish J. Electr. Eng. Comput. Sci. 2017. vol. 25. no. 4. pp. 3164–3174.
16. Zeng A, Yu K.-T., Song S., Suo D., Walker Jr.E., Rodriguez A., Xiao J. Multi-view self-supervised deep learning for 6D pose estimation in the Amazon Picking Challenge. 2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017. pp. 1386–1383.
17. Yaguchi H., Nagahama K., Hasegawa T., Inaba M. Development of an autonomous tomato harvesting robot with rotational plucking gripper. IEEE Int. Conf. Intell. Robot. Syst. 2016. vol. 2016-Novem. pp. 652–657.
18. Mylnikov L., Slivnitsin P., Mylnikova A. Robotic System Operation Specification on the Example of Object Manipulation. Proc. Int. Conf. Appl. Innov. IT. 2022. vol. 10. no. 1. pp. 51–59.
19. Sermanet P., Eigen D., Zhang X., Mathieu M., Fergus R., LeCun Y. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks. 2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc. 2013. 16 p.
20. Viola P., Jones M. Rapid Object Detection using a Boosted Cascade of Simple Features. Proceedings IEEE Conf. on Computer Vision and Pattern Recognition. 2001. pp. 511–518.
21. Dalal N., Triggs B. Histograms of oriented gradients for human detection. Proc. - 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, CVPR 2005. 2005. vol. 1(16). pp. 886–893.
22. Felzenszwalb P., McAllester D., Ramanan D. A discriminatively trained, multiscale, deformable part model. 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008. vol. 330. no. 6. pp. 1–8.
23. Wolpert D.H., Macready W.G. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1997. vol. 1(1). pp. 67–82.
24. Slivnitsin P., Kniazev A., Mylnikov L., Schlechtweg S., Kokoulin A. Influence of Synthetic Image Datasets on the Result of Neural Networks for Object Detection. Proc. Int. Conf. Appl. Innov. IT. 2021. vol. 9(1). pp. 55–60.
25. Abramovich F., Pensky M. Classification with many classes: Challenges and pluses. J. Multivar. Anal. 2019. vol. 174. pp. 1–25.

26. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016. vol. 2016-Decem. pp. 779–788.
27. Liu W., Anguelov D., Erhan D. Szegedy C., Reed S., Fu C.-Y., Berg A.C SSD: Single Shot MultiBox Detector. Eccv (Eds.: Leibe B.). Cham: Springer International Publishing, 2016. vol. 9905. pp. 398–413.
28. Rezatofighi H., Tsoi N., Gwak J., Sadeghian A., Reid I., Savarese S. Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2019. vol. 2019-June. pp. 658–666.
29. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017. vol. 39(6). pp. 1137–1149.
30. Gomes H.M. Model learning in iconic vision. PQDT – UK & Ireland. 2002. 212 p.
31. Salas-Moreno R.F., Newcombe R.A., Strasdat H., Kelly P.H.J., Davison A.J. SLAM++: Simultaneous localisation and mapping at the level of objects. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2013. pp. 1352–1359.
32. Dai A., Nießner M. 3DMV: Joint 3D-multi-view prediction for 3D semantic scene segmentation. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2018. vol. 11214 LNCS. pp. 458–474.
33. Dai A., Chang A.X., Savva M., Halber M., Funkhouser T., Nießner M. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. Proc. – 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017. 2017. vol. 2017-Janua. pp. 2432–2443.
34. Le T., Duan Y. PointGrid: A Deep Network for 3D Shape Understanding. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2018. pp. 9204–9214.
35. Su H., Maji S., Kalogerakis E., Learned-Miller E. Multi-view Convolutional Neural Networks for 3D Shape Recognition. 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2015. vol. 32(1). pp. 945–953.
36. Choy C., Park J., Koltun V. Fully convolutional geometric features. Proc. IEEE Int. Conf. Comput. Vis. 2019. vol. 2019-Octob. pp. 8957–8965.
37. Biederman I. Recognition-by-Components: A Theory of Human Image Understanding. Psychol. Rev. 1987. vol. 94(2). pp. 115–147.
38. Thompson P. Margaret Thatcher: A New Illusion. Perception. 1980. vol. 9(4). pp. 483–484.
39. Biederman I. Visual object rocognition. An Invitation to Cognitive Science. (Eds.: Kosslyn S.M., Osherson D.N.) Cambridge: MIT Press, 1995. pp. 121–165.
40. Winston P.H. Artificial intelligence. Addison-Wesley Longman Publishing Co., Inc. Boston, MA: Addison-Wesley Publishing Company, 1992. 737 p.
41. Marr D., Poggio T. A computational theory of human stereo vision. Proc. R. Soc. London - Biol. Sci. 1979. vol. 204. no. 1156. pp. 301–328.
42. Marr D., Nishihara H.K. Representation and recognition of the spatial organization of three-dimensional shapes. Proc. R. Soc. London. Ser. B. Biol. Sci. 1978. vol. 200. no. 1140. pp. 269–294.
43. Abdel-Aziz Y.I., Karara H.M. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. Photogramm. Eng. Remote Sensing. 2015. vol. 81(2). pp. 103–107.
44. Bolya D, Zhou C., Xiao F., Lee Y.J. YOLACT++ Better Real-Time Instance Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2022. vol. 44. no. 2. pp. 1108–1121.
45. Bolya D, et al. You Only Look At CoefficienTs. 2020. Available at: https://github.com/dbolya/yolact (accessed: 11.11.2022).

46. Kazemi V., Sullivan J. One Millisecond Face Alignment with an Ensemble of Regression Trees. Rev. Anthropol. 1992. vol. 21(2). pp. 147–157.
47. Lin T.Y., Maire M., Belongie S., Bourdev L., Girshick R., Hays J., Perona P., Ramanan D., Zitnick C.L., Dollar P. Microsoft COCO: Common objects in context. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 2014. vol. 8693 LNCS(5). pp. 740–755.
48. Denninger M., Sundermeyer M., Winkelbauer D., Zidan Y., Olefir D., Elbadrawy M., Lodhi A., Katam H.T. BlenderProc. 2019. 7 p. DOI: 10.48550/arXiv.1911.01911.
49. Blender 3D. Available at: https://www.blender.org/ (accessed: 22.11.2022).
50. Slivnitsin P. Position estimation for robotic system positioning using the example of outdoor luminaire replacement: master thesis. Koethen: HS Anhalt, 2021. 54 p.
51. Vershinin D., Mylnikov L. A review and comparison of mapping and trajectory selection algorithms. Proc. Int. Conf. Appl. Innov. IT. 2021. vol. 9(1). pp. 85–92.
52. Zeng A. и др. TossingBot: Learning to Throw Arbitrary Objects with Residual Physics. IEEE Trans. Robot. 2020. vol. 36(4). pp. 1307–1319.
53. Chen D. и др. Learning Canonical Shape Space for Category-Level 6D Object Pose and Size Estimation. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020. pp. 11970–11979.
54. Koch S. и др. ABC: A big cad model dataset for geometric deep learning. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019. vol. 2019-June. pp. 9593–9603.

Slivnitsin Pavel — Graduate student, Department of information technologies and automation system, electrical engendering faculty, Perm National Research Polytechnic University. Research interests: data science, cybernetics, data science, artificial intelligence, computer vision, object recognition. The number of publications — 11. slivnitsin.pavel@gmail.com; 7, Professora Pozdeyeva St., 614013, Perm, Russia; office phone: +7(342)239-1354.

Mylnikov Leonid — Associate professor, microprocessor automation means Department, electrical engendering faculty, information technologies and automation system Department, electrical engendering faculty, Perm National Research Polytechnic University; Head of laboratory, Interdisciplinary empirical studies laboratory, HSE University. Research interests: design science, cybernetics, system science, management. The number of publications — 110. lamylnikov@hse.ru; 38, Student St., 614070, Perm, Russia; office phone: +7(342)200-9555.

Acknowledgements. The study was supported by the Government of Perm Region, project no. C-26/692.