

**Московский Антон Дмитриевич**

**МЕТОДЫ И АЛГОРИТМЫ РАСПОЗНАВАНИЯ СЦЕН ДЛЯ ЗАДАЧ  
ГЛОБАЛЬНОЙ ЛОКАЛИЗАЦИИ МОБИЛЬНЫХ РОБОТОВ**

Специальность 2.3.5 –

Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Автореферат

диссертации на соискание ученой степени кандидата технических наук

Москва – 2026

Работа выполнена в Федеральном государственном бюджетном учреждении  
Национальном исследовательском центре «Курчатовский институт»

**Научный руководитель** доктор технических наук, доцент Карпов Валерий  
Эдуардович.

**Официальные оппоненты:**

\_\_\_\_\_ (фамилия, имя, отчество - при наличии, ученая степень, ученое звание, организация/ место работы, должность)

\_\_\_\_\_ (фамилия, имя, отчество - при наличии, ученая степень, ученое звание, организация/ место работы, должность)

\_\_\_\_\_ (фамилия, имя, отчество - при наличии, ученая степень, ученое звание, организация/ место работы, должность)

**Ведущая организация** \_\_\_\_\_ (название организации, подготовившей отзыв)

Защита состоится «    » \_\_\_\_\_ 2026 г. в \_\_\_\_\_ на заседании диссертационного  
совета 24.1.107.03 при Федеральном государственном бюджетном учреждении  
науки Институте проблем управления им. В.А. Трапезникова Российской академии  
наук по адресу: 117997, Москва, Профсоюзная ул., 65, ИПУ РАН.

С диссертацией можно ознакомиться в научно-технической библиотеке и на сайте  
ИПУ РАН <https://www.ipu.ru>

Автореферат разослан \_\_\_\_\_ (дата)

Ученый секретарь  
диссертационного совета,  
доктор технических наук, доцент

Е.А. Барабанова

## Общая характеристика работы

**Актуальность темы.** Задача локализации (определения положения в пространстве) является краеугольной для большинства практических задач в области мобильной робототехники, ведь для автономного движения (навигации) роботу обычно требуется знать свои координаты в каждый момент времени. Развитие методов одновременного картирования и локализации (SLAM, simultaneous localization and mapping), позволяющих строить карту неизвестного пространства и проводить на ней локализацию, значительно расширило применение мобильных роботов. Подобные карты обычно представляют собой набор геометрических примитивов или облака точек, получающихся посредством «сканирования» пространства дальномерами. Помимо задачи SLAM отдельно выделяют задачу глобальной локализации (global localization) – т.е. определение положения уже на построенной карте. Определение положения по таким картам обычно происходит путём поиска лучшего «наложения» текущих данных с сенсоров робота на имеющуюся карту. Если во время задачи SLAM поиск такого наложения происходит в локальной области карты, то в случае глобальной локализации область задаётся размерами карты, что требует многочисленных наложений. Это же создаёт сложности, связанные с симметрией и однородностью пространства – в местах, где пространственные сенсоры одинаково хорошо накладываются на несколько точек карты.

Одним из способов решения описанных проблем является добавление размеченных на карте объектов (называемых семантической картой) определенных классов. Семантика позволяет понизить «размерность» пространства поиска, а также может добавить «особенности» к однообразной геометрии. С другой стороны, семантические объекты характеризуются тем, что зачастую представлены в среде во многих экземплярах, что затрудняет процедуру сопоставления видимых объектов с картой. Традиционно данная проблема решается с помощью методов распознавания связанных групп объектов - *методов распознавания сцен*. Их задача – выделение геометрических особенностей во взаимоположении объектов на наблюдаемой сцене и поиск этих же особенностей на семантической карте. Типичными проблемами при решении этих задач являются ошибки в системах распознавания ввиду неточности детектора или перекрытия другими объектами, а также ошибки в определении их положения в пространстве. Также следует учитывать изменчивость среды со временем. Хотя семантический уровень считается в целом более разнообразным, чем геометрический, всё же встречаются объекты, расположенные в некотором свойственном для них порядке и практически не отличающиеся внешне друг от друга. Все перечисленные проблемы приводят к тому, что системы распознавания сцен должны иметь возможность предоставить несколько гипотез, а также дополнительные средства отсеивания ошибочных вариантов.

Стоит отметить, что решение задачи распознавания сцены является первым этапом в более общей задаче визуальной локализации, и лишь определяет соответствие между видимыми ориентирами и ориентирами на карте. Далее требуется по этим данным определить положение робота. Но работа с визуальными ориентирами обычно сопровождается большими погрешностями, особенно в части

определения расстояний до них, а иногда во входных данных содержатся противоречивости, что приводит к весьма низкой точности определения положения. В тоже время, более точные методы локализации по данным дальномеров в постановке локализации по видимым ориентирам практически не используются, хотя локально могут обеспечить более высокую точность. Следовательно, требуется разработать специальные методы, позволяющие в таких условиях определить некую локальную область на карте, а затем произвести уточнение местоположения на ней геометрическими методами на основе облаков точек.

Таким образом, научная проблема создания эффективных способов сопоставления наблюдаемых объектов с семантической картой и дальнейшего определения положения робота по этим данным является **актуальной**. Речь идет о создании моделей и методов, позволяющих решать задачу глобальной локализации по визуальным данным. Актуальность определяется также и практическим аспектом работы. В ходе исследования была разработана комплексная технология глобальной локализации по визуальным ориентирам, позволяющая повысить автономность робототехнических средств за счёт решения задачи определения стартового положения робота на больших территориях и в сложных условиях, связанных с симметричностью и повторяемостью пространства. Разработкой методов распознавания сцен, в применении к локализации мобильных роботов, а также средствами верификации их результатов занимается ряд научных коллективов, что также говорит об актуальности темы.

**Сегодня** исследования в этой области семантической локализации охватывают много различных поднаправлений. Этим направлением в целом занимается ряд отечественных и зарубежных учёных, например: Ющенко А.С., Кирильченко А.А., Романов А.М., Платонов А.К., Яковлев К.С., Юдин Д.А., Колубин С.А., Фомин И.С., Burgard W., How J.P., Lusk P.C., Ankenbauer J. Liu Y., Matsuzaki S. и др. Много работ посвящено именно задачам семантического SLAM – построению карты с размеченными объектами, которые в дальнейшем используются в других робототехнических применениях.

Анализ работ вышеупомянутых отечественных и зарубежных исследователей позволяет сделать вывод, что создание устойчивых к перечисленным проблемам систем распознавания сцены и уточнения своего положения в применении к задачам глобальной локализации является актуальным.

Опираясь на вышесказанное, были поставлены цель и задачи.

**Целью диссертационного исследования** является повышение качества процедуры глобальной локализации мобильных робототехнических систем на основе применения методов распознавания сцен с учетом визуального сходства объектов и использования недоопределённых вычислений.

Для достижения заданной цели были определены следующие **задачи**:

1. Провести анализ существующих подходов в области глобальной локализации мобильных роботов, в частности по визуальным ориентирам и дальномерным данным, с целью выявления ограничений, влияющих на качество решения.

2. Разработать методы распознавания сцен, позволяющие найти соответствие между видимым набором объектов и семантической картой местности, которые

способны работать в условиях высокой погрешности определения расстояния и т.н. выбросов (отсутствия нужных и наличия лишних объектов), а также определять сразу несколько гипотез и снабжать их численной степенью уверенности.

3. Разработать метод определения положения робота по данным визуальных ориентиров в условиях больших погрешностей, противоречивости данных и множественных гипотез, результат которого может быть использован классическими геометрическими методами определения положения для дальнейшего уточнения.

4. Объединить разработанные методы и модели в комплексную технологию глобальной локализации, дополнив её модификациями классических методов локализации с использованием облаков точек, которые способны работать в ограничениях.

5. Подготовить на основе открытых наборов данных систему тестирования, включающую в себя как данные, так и метрики оценивания решений задачи распознавания сцен и задач локализации по видимым ориентирам.

6. Разработать комплекс программных и технических решений, реализующий созданные модели и методы для решения прикладных задач в области локализации мобильных роботов, провести оценку работоспособности и эффективности предложенных решений.

**Объектом исследования** является процедура глобальной локализации мобильных роботов.

**Предметом исследования** являются методы, алгоритмы глобальной локализации на основе распознавания сцен.

**Научная новизна** работы заключается в следующем:

1. Предложен новый метод распознавания сцен на основе алгоритмов поиска изоморфного подграфа, позволяющий по видеоданным с бортовых сенсоров сопоставлять видимые объекты с семантической картой, учитывая визуальное сходство объектов методами на основе машинного обучения (пункты 4 и 7 паспорта специальности 2.3.5).

2. Предложен новый метод N-локализации на основе недоопределённых моделей, позволяющий получить решение задачи локализации по ориентирам в интервальных терминах (ограничениях), что делает возможным использование классических методов локализации по дальномерным данным (пункт 7 паспорта специальности 2.3.5.).

3. Предложена архитектура комплекса глобальной локализации, включающей в себя программные средства распознавания сцен, недоопределённой локализации и классических методов локализации, а также порядок их взаимодействия (пункт 3 паспорта специальности 2.3.5).

4. Разработаны метрики оценки качества и требования к данным для тестирования программных реализаций алгоритмов распознавания сцен (пункт 10 паспорта специальности 2.3.5).

**Теоретическая значимость работы** заключается в создании новых методов распознавания сцен, способных учитывать как геометрические особенности наблюдаемой сцены, так и визуальное сходство объектов. Применение подхода недоопределённых вычислений может использоваться для отсеивания гипотез и

определения положения робота по визуальным ориентирам в условиях больших погрешностей.

**Практическая значимость** полученных результатов заключается в повышении эффективности решения задачи глобальной локализации роботов, оперирующих на больших/симметричных/повторяемых территориях, что расширяет их автономность и функциональность.

**Методы исследования.** В диссертационной работе применяются методы теории графов, недоопределённых вычислений, теории вероятности, оптимизации, машинного обучения, имитационного моделирования, технического зрения.

**Положения, выносимые на защиту:**

1. Новый метод распознавания сцен на основе поиска изоморфного подграфа с учетом визуального сходства и его реализация, повышающие точность получаемого решения в диапазоне 13-49% в сравнении с современными аналогами.

2. Метод и алгоритмы недоопределённой локализации, позволяющие проводить как глобальную, так и пошаговую (непрерывную) локализацию в ограничениях (интервалах), что позволяет уточнять положение классическими методами локализации, тем самым повышая до порядка общую точность и/или скорость локализации.

3. Комплексная технология глобальной локализации, включающая разработанные алгоритмы распознавания сцен и недоопределённой локализации, а также модификации классических алгоритмов локализации с использованием облаков точек, позволяющая повысить точность (от 5 до 48% по метрике попадания положения робота в доверительный интервал) и скорость процедуры (до 4х раз в сравнении с геометрическими методами) глобальной локализации по визуальным ориентирам и дальномерным данным.

4. Алгоритм формирования набора данных на основе KITTI-360, а также метрики, позволяющие проводить оценку качества алгоритмов распознавания сцен.

**Достоверность** результатов подтверждается численными экспериментами, проведенными на открытом наборе данных, а также в средах симуляции и на реальных роботах. Все реализованные алгоритмы подробно описаны, и их реализация частично выложена в открытый доступ, что позволяет проверить полученные результаты.

**Апробация результатов** проходила на ряде конференций и семинаров, основными из которых являются: V Всероссийский научно-практический семинар «Беспилотные транспортные средства с элементами искусственного интеллекта» (БТС-ИИ-2019), 22-24 мая 2019 г., Санкт-Петербург; 4-я практическая конференция «ROS Meetup» (удалённый формат), 5 декабря 2020 г., Москва; XIX Национальная конференция по искусственному интеллекту с международным участием (КИИ-2021), 11-16 октября 2021 г., Таганрог; Семинар «Теория управления организационными системами» (ТУОС), 3 февраля 2022 г., ИПУ РАН, Москва (удалённый формат); 34-я международная научно-техническая конференция «Экстремальная робототехника» (ЭР-2023), 23-24 ноября 2023 г., Санкт-Петербург; Семинар «Проблемы управления автономными робототехническими комплексами», 15 февраля 2024 г., ИПУ РАН, Москва; 7-я практическая конференция «ROS Meetup», 27-28 апреля 2024 г., Москва; XII Международная научно-практическая

конференция «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (ИММВ-2024), 14-17 мая 2024 г., Коломна; 35-я международная научно-техническая конференция «Экстремальная робототехника» (ЭР-2024), 29-30 октября 2024 г., Санкт-Петербург; Объединенный семинар "Интеллектуальные системы" и "Системное программирование", 13 февраля 2025 г., ИСИ СО РАН, Новосибирск; Донецкий международный научный круглый стол «Искусственный интеллект: теоретические аспекты, практическое применение», 28 мая 2025 г., ФГБНУ ИПИИ, Донецк (удалённый формат).

**Внедрение результатов.** Результаты исследований были использованы при:

1. Проведении НИР «Разработка экспериментального образца аппаратно-программного комплекса сбора и обработки данных для контроля характеристик и калибровки целевой аппаратуры КА ДЗЗ в части создания макета подвижной платформы для проведения подспутниковых наземных радиометрических измерений» (шифр «Мониторинг-СГ-1.3.1.1.1-МПП») (акт о внедрении, полезная модель).

2. В учебном процессе кафедры Информатики и вычислительных сетей Физтех-школы природоподобных, плазменных и ядерных технологий им. И.В. Курчатова (КНТ МФТИ) (акт о внедрении).

3. При проведении НИР «Исполнительные механизмы и системы электродвижения» по теме «Роботизированная многофункциональная коляска для людей с ограниченными возможностями» (2019-2022гг.) и НИР «Разработка робототехнических комплексов, в том числе энергетически автономных» (тематический план НИЦ «Курчатовский институт», тематика 1п.1.2, приказ №86 от 20.01.2023 г. «О проведении фундаментальной научно-исследовательской работы по комплексной теме "Прикладные междисциплинарные исследования в области создания природоподобных технологий"») (акт о внедрении).

**Личный вклад.** Все результаты, приведенные в данной работе, включая разработку алгоритмов, их программную реализацию и проведение экспериментов, получены автором лично. В работе [5] автору принадлежит обзор методов локализации, методов навигации без карты, а также участие в обсуждении и формировании тенденций и открытых проблем области. Остальные работы выполнены без соавторов.

**Соответствие паспорту специальности.** Диссертация выполнена в соответствии с паспортом научной специальности 2.3.5 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» по следующим пунктам:

– п.3. Модели, методы, архитектуры, алгоритмы, языки и программные инструменты организации взаимодействия программ и программных систем (создание архитектуры комплексной технологии локализации, включающей в себя программные средства методов распознавания сцен, N-локализации и классических подходов).

– п.4. Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов (создание гибридных методов распознавания сцен с использованием технологий машинного обучения, позволяющих учитывать визуальное сходство объектов).

– п.7. Модели, методы, архитектуры, алгоритмы, форматы, протоколы и программные средства человеко-машинных интерфейсов, компьютерной графики, визуализации, обработки изображений и видеоданных, систем виртуальной реальности, многомодального взаимодействия в социокиберфизических системах (создание систем распознавания сцен и локализации по ориентирам, опирающихся на обработку визуальных данных).

– п.10. Оценка качества, стандартизация и сопровождение программных систем (создание набора данных и разработка метрик оценки качества решений задачи распознавания сцен).

**Публикации.** Основные результаты изложены в девяти работах [1-9], две из которых входят в список научных изданий *ВАК* по специальности 2.3.5, три в журналах, индексируемых в *Scopus*, две в журналах, индексируемых в *Web of Science*. Получен патент на полезную модель, в которой частично использовались разработанные алгоритмы и подходы к локализации [10]. По результатам создания программного обеспечения, реализующего разработанные методы, было получено два свидетельства о регистрации РИД (программы для ЭВМ) [11, 12].

## Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, формулируется цель, ставятся задачи работы, излагаются научная новизна и практическая значимость представляемой работы.

В **первой главе** приводится краткая ретроспектива зарождения области распознавания объектов и локализации в робототехнике. Далее дается подробный обзор современных работ на эту тему. Показывается, что добавление семантики в карты робототехнических устройств преследует не только цель улучшения локализации, но также даёт доступ к созданию на их основе высокоуровневых систем планирования и средств человеко-машинного взаимодействия, в том числе и на естественном языке. Рисунок 1 показывает тенденцию развития методов распознавания сцен с разбиением на подгруппы. На основе приведенного обзора делается вывод, что разработка методов, позволяющих опираться на оба критерия (поиск групп объектов и учёт визуального сходства), является актуальной задачей и потенциально позволит улучшить качество распознавания сцен. Помимо решения задачи распознавания сцен, требуется определить положение робота по этим данным, поэтому приводится обзор основных подходов в этой области. Подавляющее большинство этих методов решает задачу локализации по ориентирам и не рассматривает облака точек как источник данных. Заключается, что разработка методов, способных работать с облаками точек, используя результат задачи распознавания сцен, может принести новое качество. Для этого предлагается обратиться к подходам, оперирующим интервальными величинами. Полученные интервальные ограничения на положение робота могут быть использованы, чтобы сократить область поиска методами, работающим с облаками точек.

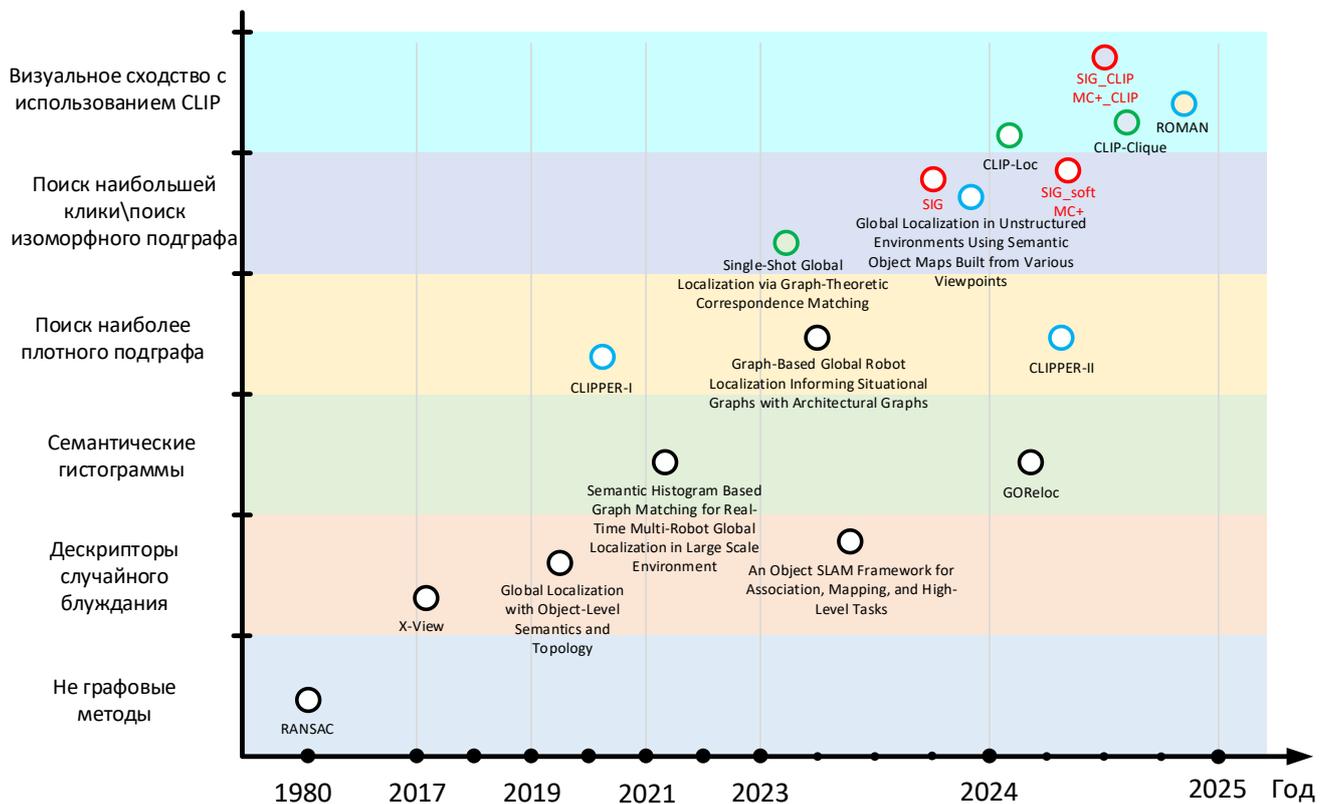


Рисунок 1. Некоторые работы по тематике распознавания сцен в задачах локализации мобильных роботов. Цветом кроме чёрного отмечены работы пересекающихся коллективов, в т. ч. автора диссертации (красным)

Исследование работ по теме распознавания сцен также показало, что для данной области отсутствует устоявшийся набор данных для тестирования методов. Авторам приходится использовать наборы данных из смежных задач локализации, таких как SLAM, и вносить в них свои изменения. Также в обзорной главе подчеркивается, что ввиду специфики области семантической локализации, всегда используются дополнительные способы верификации полученных решений задачи распознавания сцен, разработка которых также является актуальной задачей.

**Вторая глава** посвящена теоретической базе предлагаемых методов.

**Раздел 2.1** начинается с формальной постановки задачи распознавания сцен. Даны множества объектов  $M$  (карта) и  $S$  (сцена), где каждый объект  $o$  этих множеств характеризуется следующими элементами: класс (семантическая метка)  $c$  (идентификатор из заранее определённого множества), положение в пространстве  $p$ , визуальный образ  $I$ . При этом положения в пространстве  $p$  для каждого из множеств задано в отдельной системе координат, преобразование между которыми не известно. Однако будучи отражением реального мира, карта и сцена содержат набор одних и тех же объектов этого мира. Это значит, что для этого набора объектов сохраняются классы объектов  $c$ , их визуальные образы  $I$  и геометрические отношения, рассчитанные на основе положений объектов. Цель – найти такое сопоставление из множества  $S$  в множество  $M$ , которое бы сохраняло как геометрические отношения между объектами, так и их визуальное сходство, при этом учитывая класс объектов.

Для решения поставленной задачи был предложен **метод распознавания сцен на основе алгоритмов поиска изоморфного подграфа** семейства VF, которые

позволяют учитывать разметку графов. Разметка позволяет эффективно искать изоморфные подграфы данными алгоритмами, если предварительно подготовить графы и ввести дополнительные критерии сходства вершин и рёбер. Для каждого из множеств  $M$  и  $S$  строится свой граф, где вершины соответствуют объектам, а их разметка - классам объектов. Таким образом, дополнительным критерием сходства вершин графов выступает совпадение их разметки, т.е. семантических классов объектов. Для определения и добавления на граф рёбер требуется расчёт расстояний между объектами. Для этого был предложен способ оценки таких расстояний и их погрешностей для множества объектов сцены по данным сенсоров робота. На основе расстояний между объектами на карте и в сцене, а также погрешности их определения вводится следующая мера сравнения  $P$ :

$$p(D_1, D_2, \sigma D) = e^{-\frac{1}{2} \left( \frac{D_1 - D_2}{\sigma D} \right)^2}, \quad (1)$$

где  $D_1$  и  $D_2$  – расстояния на карте и сцене соответственно, а  $\sigma D$  – среднеквадратичная ошибка определения расстояния на сцене. Такая мера нормирована на отрезок  $[0, 1]$ . Графы сцены  $G_S$  и карты  $G_M$  таким образом становятся полностью связными, где каждому ребру соответствует расстояние между объектами. Однако для графа карты можно «проредить» рёбра, оставив только те, расстояния которых не превышают удвоенной рабочей дистанции сенсора глубины. Далее можно сформировать дополнительный критерий сходства рёбер на графах как удовлетворение некоторому заданному порогу  $\tau$  по введённому критерию по расстоянию  $P$  (1):

$$e_i \in E_M, e_j \in E_S: P(D^{e_i}, D^{e_j}, \sigma D^{e_j}) \geq \tau \rightarrow e_i \simeq e_j, \quad (2)$$

где  $e_i$  и  $e_j$  – это рёбра из множества рёбер  $E$  графов карты и сцены. После формирования данных критериев можно воспользоваться алгоритмами семейства VF. Алгоритм на выходе даёт набор соответствий  $\{f\}$  вершин графов  $G_S$  в  $G_M$ , что автоматически является искомым соответствием  $S \rightarrow M$ . Для каждого такого соответствия  $f$  предлагается рассчитать следующую степень уверенности  $dc$ :

$$dc(f) = \frac{1}{|E_S|} \sum_{(o_i^S, o_j^S) \in E_S} \left[ u \left( I^{o_i^S}, I^{o_i^M} \right) \cdot u \left( I^{o_j^S}, I^{o_j^M} \right) \right]. \quad (3)$$

$$P \left( D \left( p^{o_i^M}, p^{o_j^M} \right), D \left( p^{o_i^S}, p^{o_j^S} \right), \sigma D \left( p^{o_i^S}, p^{o_j^S} \right) \right), o_i^M = f(o_i^S),$$

где  $u$  – это функция извлечения и сравнения векторных представлений визуальных образов объектов, которые получаются с помощью нейросетевой архитектуры CLIP или аналогов. Сравнение векторных представлений в  $u$  организовано как косинусное сходство. Мера сравнения (3) нормирована на отрезок  $[0, 1]$  и учитывает как геометрическое, так и визуальное сходство объектов, позволяя отсортировать полученные решения по качеству.

Также для данного метода было разработано мягкое расширение, позволяющее работать в условиях выбросов (объектов не нанесенных на карту, но распознанных). Такие объекты нарушают изоморфизм полученных графов и могут приводить к тому, что решение не будет найдено. Для этого на граф карты добавляются особые вершины, помеченные как «фиктивные», по  $k$  элементов для каждого класса объектов на сцене. Такие вершины соединяются с другими

вершинами «фиктивными» рёбрами. Критерий сходства рёбер (3) в таком случае всегда будет считаться выполненным, если хотя бы один его аргумент ложный. Однако функции  $P$  и  $u$  в расчёте коэффициента уверенности  $dc$  (3) выбираются равными нулю, если хотя бы одним аргументом является ложный элемент. Такое расширение позволяет получать решения в условиях лишних объектов, а также в случае высоких погрешностей определения расстояний, но увеличивает время расчёта. Алгоритм предложенного метода представлен на схеме ниже (Рисунок 2).

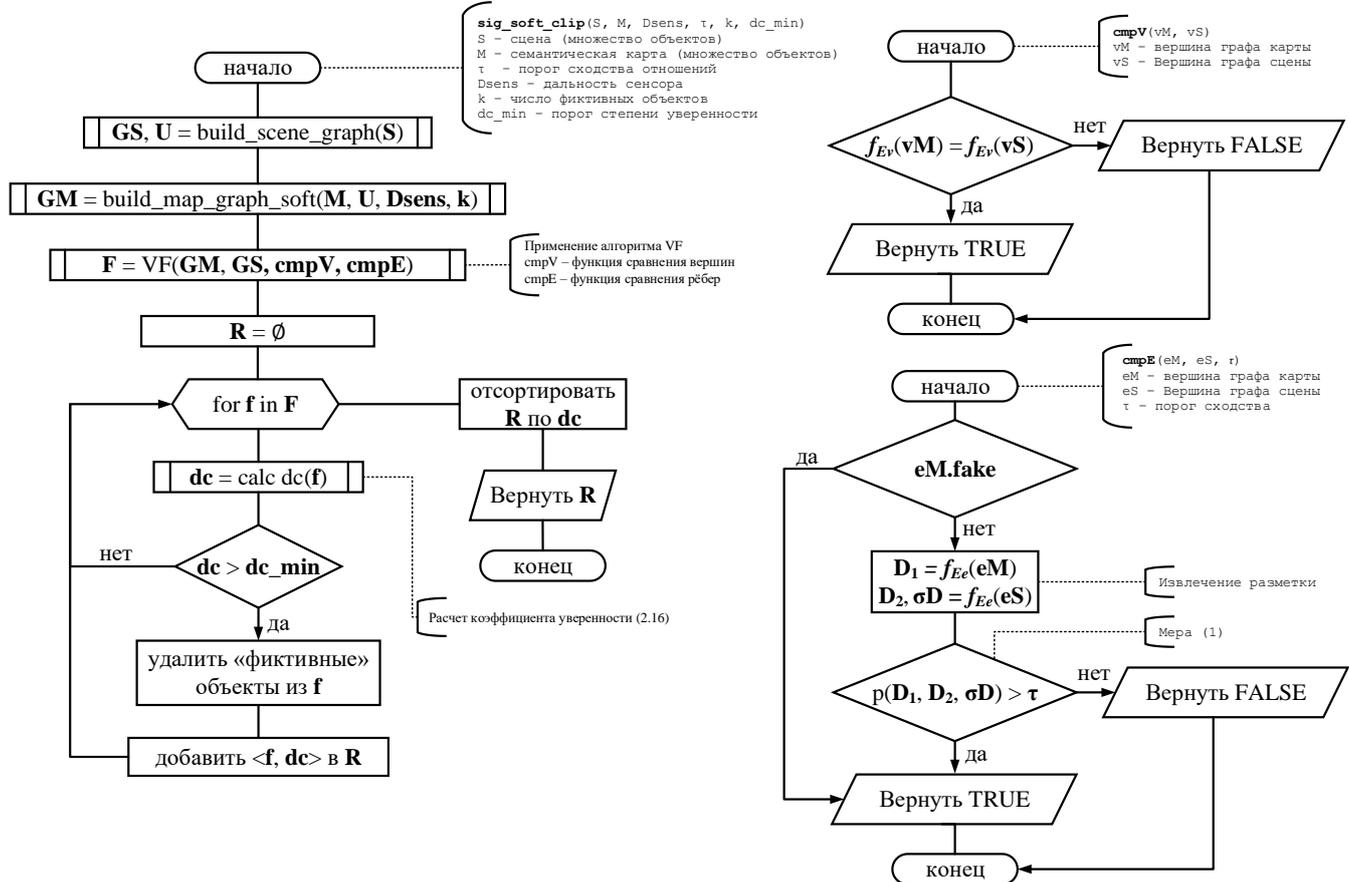


Рисунок 2. Схема алгоритма, реализующего метод распознавания сцен на основе поиска изоморфного подграфа

Алгоритм **sig\_soft\_clip** (Рисунок 2) принимает на вход множество объектов карты  $M$  и множество объектов сцены  $S$ , максимальную рабочую дальность сенсора  $D$ , порог сходства отношений  $\tau$ , число ложных объектов  $k$ , порог степени уверенности  $dc_{min}$ . Функции **build\_scene\_graph** и **build\_map\_graph** строят графы сцены и карты, функция **p** выполняет расчет согласно (1). Функция **VF** выполняет поиск изоморфного подграфа одноименным алгоритмом с указанием критериев сходства вершин и рёбер, функция **calc\_dc** рассчитывает коэффициент уверенности согласно (3). Алгоритм возвращает отсортированный массив  $R$  соответствий  $f:S \rightarrow M$  с указанием их коэффициентов уверенности  $dc$ . Общая схема разработанного метода с визуализацией процесса и данных представлена ниже (Рисунок 3).

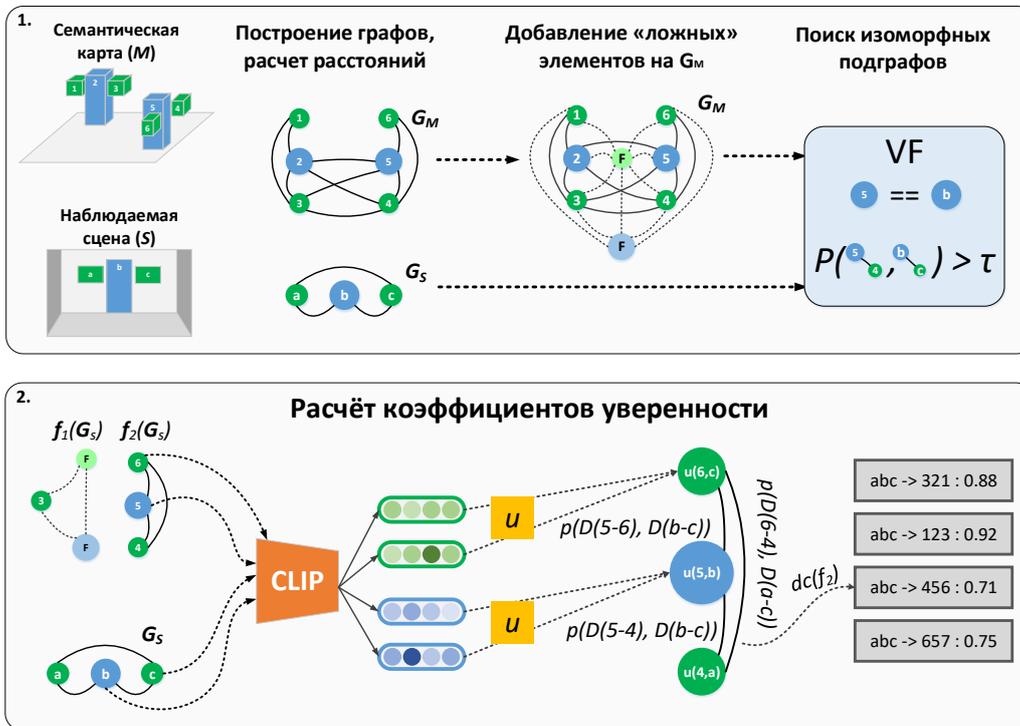


Рисунок 3. Общая схема метода распознавания сцен на основе поиска изоморфного подграфа

Далее, в разделе 2.2, предлагается метод **Н-локализации**. Ввиду погрешностей в системах распознавания и локализации объектов, результат разработанного метода распознавания сцен (Рисунок 2) может содержать ошибки, а также выдавать несколько близких друг к другу вариантов ответа. Для работы с такими данными был создан метод локализации на основе механизма недоопределённых моделей (далее Н-моделей), предложенного А. С. Нариньяни. Оригинальный механизм Н-моделей предусматривает итерационное уточнение так называемых недоопределённых переменных (Н-переменных), которые представлены в виде некоего множества значений при помощи функций присваивания, которые отражают зависимость между Н-переменными. Также особенностью данного аппарата является наличие функций проверки корректности, которые позволяют определять противоречия во входных данных, возникающие в процессе решения задачи распознавания сцен. Таким образом Н-модель задаётся четвёркой:

$$M = \langle V, R, W, C \rangle, \quad (4)$$

где  $V$  – множество Н-переменных,  $R$  – ограничения на их значения, которые должны быть функционально интерпретируемы,  $W$  – функции интерпретации,  $C$  – функции проверки корректности. Н-переменные могут быть заданы по-разному, однако для задачи локализации целесообразным является представление в виде мульти-интервала, т.е. набора упорядоченных отрезков:

$$^*a = \{[a_i, a_{i+1}]\}, a_i \leq a_{i+1}, \quad (5)$$

здесь и далее знак \* слева от имени переменной обозначает, что она является недоопределённой.

Для адаптации данного подхода (формализации модели) для задачи локализации по визуальным ориентирам требуется: описать положение робота в терминах  $N$ -переменных в виде мульти-интервалов (5); реализовать  $N$ -операции над мульти-интервалами, позволяющие работать с ними как с обычными переменными; определить функции интерпретации, связывающие положение робота с данными о позициях и измерении ориентиров (определение расстояний и углов на них); определить функции проверки корректности на основании входных данных; реализовать критерии остановки вычислительной процедуры.

Положение робота, которое требуется найти, будет выражено в виде трёх  $N$ -переменных (для двумерного случая) – положения по двум координатам и углу – таким образом множество  $V$  (4) принимает вид:

$$V = \langle *X, *Y, *\theta \rangle. \quad (6)$$

Далее требуется связать данные  $N$ -переменные с известной роботу информацией об окружающей среде, которая позволит определить его положение. Такой информацией выступает положение ориентиров в пространстве, а также измеренное до них расстояние и угол. Данную связь легко можно получить из геометрических соотношений и выразить через введённые  $N$ -переменные, например:

$$\begin{aligned} (Y_l - *Y)^2 + (X_l - *X)^2 &= (r + *\Delta r)^2, \\ *\theta &= \arctan\left(\frac{Y_l - *Y}{X_l - *X}\right) - (\alpha + *\Delta\alpha), \end{aligned} \quad (7)$$

где  $X_l, Y_l$  – положение ориентира,  $r$  и  $\alpha$  – расстояние и угол до него соответственно. Формулы (7) содержат погрешности определения измеряемых величин (расстояния  $*\Delta r$  и угла  $*\Delta\alpha$ ), также выраженные как  $N$ -переменные по правилу « $N$  сигм», где  $*\Delta\alpha = [-N\sigma_\alpha, N\sigma_\alpha]$ . Выразив из формул (7) (и их аналогов) каждую из искомым  $N$ -переменных (6), получим функции интерпретации  $R$  (4). Однако для того, чтобы воспользоваться ими, требуется реализовать  $N$ -операции, позволяющие производить с  $N$ -переменными все требуемые действия, от простых алгебраических операций до тригонометрических функций. Такие операции работают с границами элементов мульти-интервалов (5) и могут быть выражены следующим образом:

$$w(*a) = [\inf *_a(w), \sup *_a(w)], w \in W. \quad (8)$$

Когда получено новое значение некоторой  $N$ -переменной, требуется, используя функцию присваивания  $W$ , обновить старое значение, для этого используется пересечение полученных мульти-интервалов. Такой подход позволяет гарантировать, что процесс не будет расходящимся. А функции проверки корректности  $C$  (4) позволят определить, что пересечение не пустое или же прервать процедуру ввиду противоречивости входных данных. В работе также предлагаются дополнительные функции интерпретации и проверки корректности, построенные на других зависимостях: взаиморасположении робота и двух ориентиров, порядке появления ориентиров по значению угла в поле зрения робота, знаниях об их ориентации и др. Дополнительные функции проверки корректности  $C'$  применяются после исполнения основного вычислительного процесса с элементами декартового

произведения полученных мульти-интервалов. Алгоритм проверки результатов методом распознавания сцен представлен ниже (Рисунок 4).

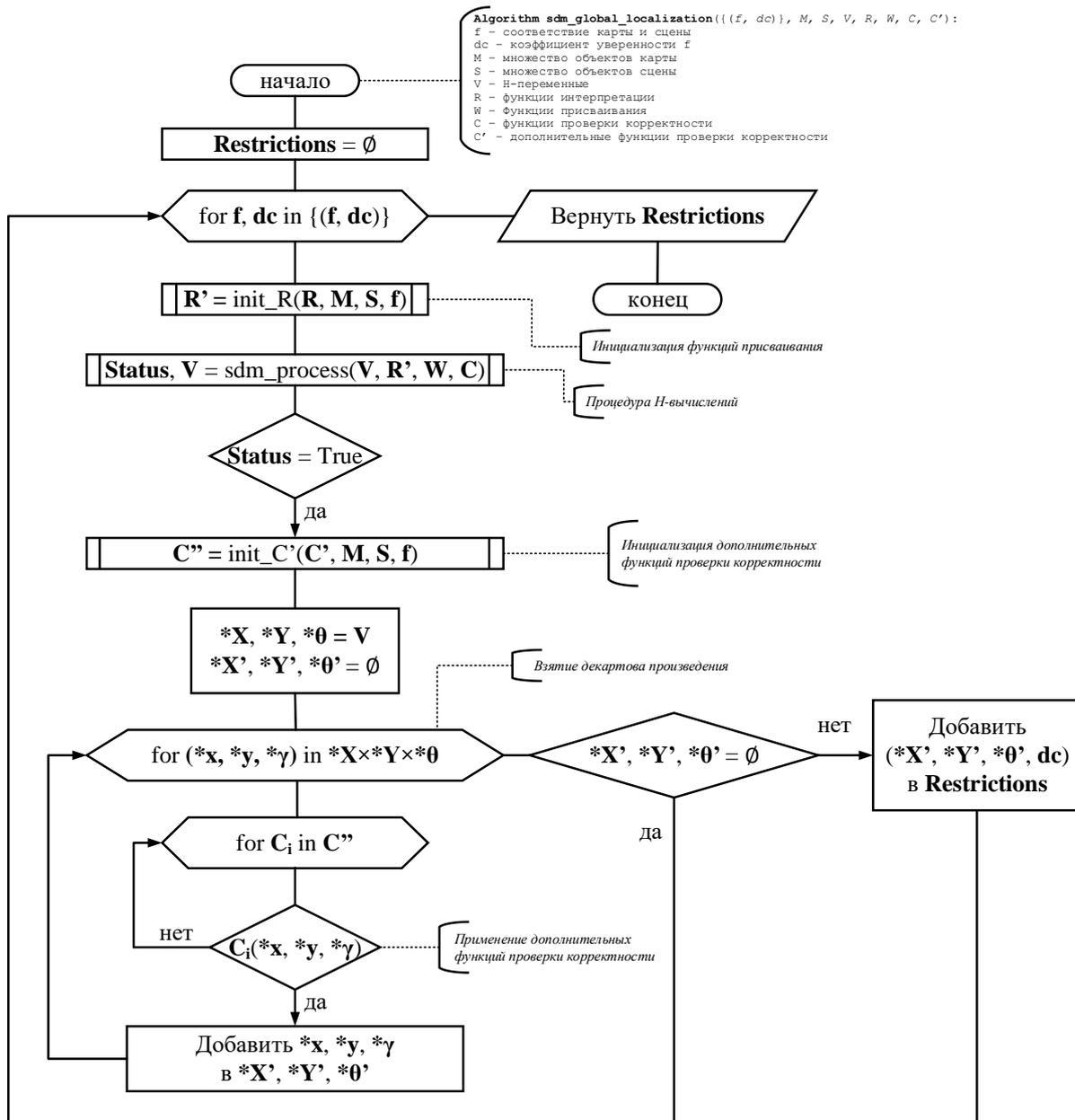


Рисунок 4. Схема алгоритма глобальной локализации с помощью  $N$ -моделей ( $N$ -локализации)

Алгоритм **sdm\_global\_localization** принимает на вход результат методов распознавания сцен  $\{f, dc\}$ , карту  $M$  и наблюдаемую сцену  $S$ , элементы  $N$ -модели  $V, R, W, C$ , а также дополнительные функции проверки корректности  $C'$ . Функции **init\_R** и **init\_C'** инициализируют параметры функций интерпретации и проверки корректности исходя из соответствия  $f$ . Функция **sdm\_process** отражает процедуру  $N$ -вычислений и выдаёт набор ограничений на положение робота, который проходит дополнительную проверку и возвращается алгоритмом.

В разделе 2.3 на основе метода распознавания сцен и метода  $N$ -локализации предлагается архитектура комплексной технологии глобальной локализации по визуальным ориентирам, которая позволяет также использовать точные методы локализации по дальномерным данным (Рисунок 5).

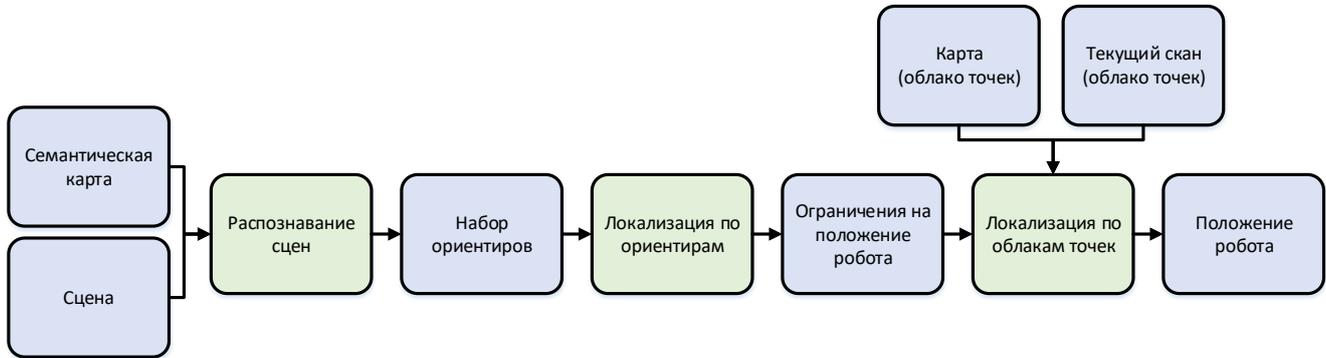


Рисунок 5. Предлагаемая архитектура комплексной технологии глобальной локализации по семантическим данным

Предложен алгоритм обрезки облака точек карты в соответствии с полученными от N-локализации ограничениями, который позволяет далее применить классические алгоритмы поиска положения по двум облакам точек. Также предложены алгоритмы распределения набора частиц в полученных ограничениях, что позволит решать задачу локализации при помощи фильтра частиц.

Далее в главе (раздел 2.4) формализуются и предлагаются новые метрики оценки решения задачи распознавания сцен, в том числе и для оценивания всего набора гипотез, который выдаётся методом. На основании верного сопоставления между картой и сценой  $\bar{f}$  и отсортированного по коэффициенту уверенности (3) набора гипотез  $\tilde{f} = \{f_i\}$  предлагается метрика *Score*:

$$Score(\tilde{f}, \bar{f}) = \begin{cases} \frac{1}{i}, f_i \in \tilde{f}, f_i = \bar{f} \\ 0, \bar{f} \notin \tilde{f} \end{cases} \quad (9)$$

На основании (9) была введена также метрика *Degree* показывающая тип полученного набора ответов  $\tilde{f}$ :

$$Degree(\tilde{f}, \bar{f}) = \begin{cases} \text{"ИСТИНА"}, Score(\tilde{f}, \bar{f}) = 1, \\ \text{"ПРАВДОПОДОБИЕ"}, \bar{f} \in \tilde{f}, Score(\tilde{f}, \bar{f}) \neq 1, \\ \text{"ЛОЖЬ"}, Score(\tilde{f}, \bar{f}) = 0, \tilde{f} \neq \emptyset, \\ \text{"ПУСТО"}, \tilde{f} = \emptyset, \end{cases} \quad (10)$$

где *ИСТИНА* соответствует полностью правильно найденному ответу, *ПРАВДОПОДОБИЕ* тому что ответ был найден среди гипотез, *ЛОЖЬ* что верный ответ отсутствует и *ПУСТО* показывает что решения не было найдено. Метрика *Degree* (10) применяется к наборам сцен формируя доли указанных типов, показывая общую возможность методов находить верное решение.

**Третья глава** затрагивает вопросы реализации и проверки работоспособности полученных методов.

В разделе 3.1 для проведения исследований с методом распознавания сцен была предложена модификация широко применяющегося метода на основе алгоритмов поиска максимальной клики. В оригинальный алгоритм был добавлен вероятностный учет отношений на основе (1) и учет визуального сходства моделью

CLIP, в итоге получив на выходе алгоритма численную оценку решения подобную (3). Также был предложен алгоритм разделения семантической карты на подкарты с пересечением, чтобы ускорить процедуры распознавания сцен, в основе которых лежит решение NP-полных задач (поиска изоморфного подграфа и поиска максимальной клики), и при этом не потерять решения вблизи границы разрезов карты. Была описана программная реализация методов распознавания сцен а также произведена проверка работоспособности и сравнение с аналогами в среде имитационного моделирования по предложенной метрике *Degree* (10) в трех различных сценариях: *clear* – без выбросов, *holes* – с выбросами, *mix* – равномерная смесь предыдущих сценариев (Рисунок 6).

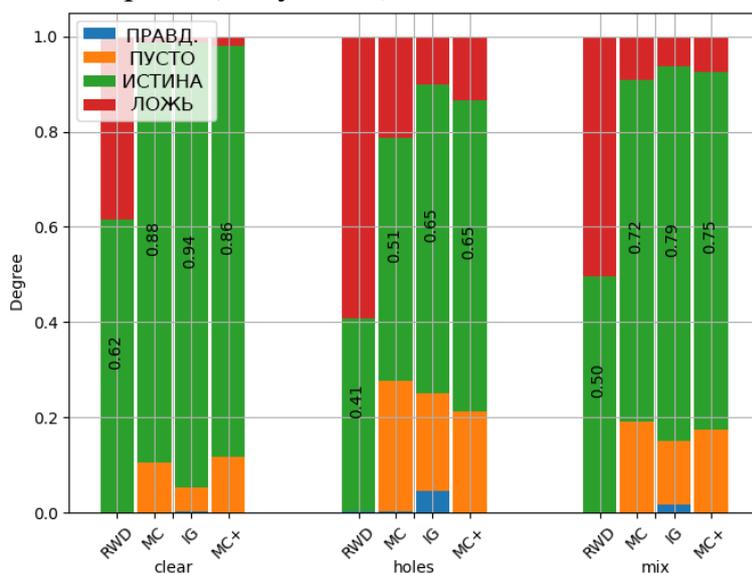


Рисунок 6. Проверка работоспособности и сравнение с аналогами метода распознавания сцен на основе поиска изоморфного подграфа (IG)

Проведённые эксперименты (Рисунок 6) показали преимущество предложенного метода с аналогами (*RWD* – дескрипторы случайного блуждания, *MC* – поиск максимальной клики, *MC+* – предложенная модификация *MC*) во всех сценариях.

**Раздел 3.2** посвящен реализации метода N-локализации, а также их аналогов, с которыми в работе проводится сравнение. Для оценки метода N-локализации как способа сокращения исследуемого пространства поиска было произведено сравнение с подходами на основе гистограммного фильтра (*hf*) и фильтра частиц (*amcl*). Тестировались разные сценарии локализации по ориентирам: по камере (*Camera*, измеряя угол и расстояние в узком секторе), по акустическому сигналу (*LBL*, измеряя только расстояние по всем направлениям) и по пеленгу (*Bearing*, измеряя только угол). Сравнились распространённые подходы в локализации на основе гистограммного фильтра, фильтра частиц и их расширения с помощью N-локализации, которая предварительно сужала область для фильтров. Предложенные модификации рассматриваемых подходов (*sdl\_amcl\_lite* и *sdl\_hf*) показали как преимущества по точности, так и по времени работы (Рисунок 7).

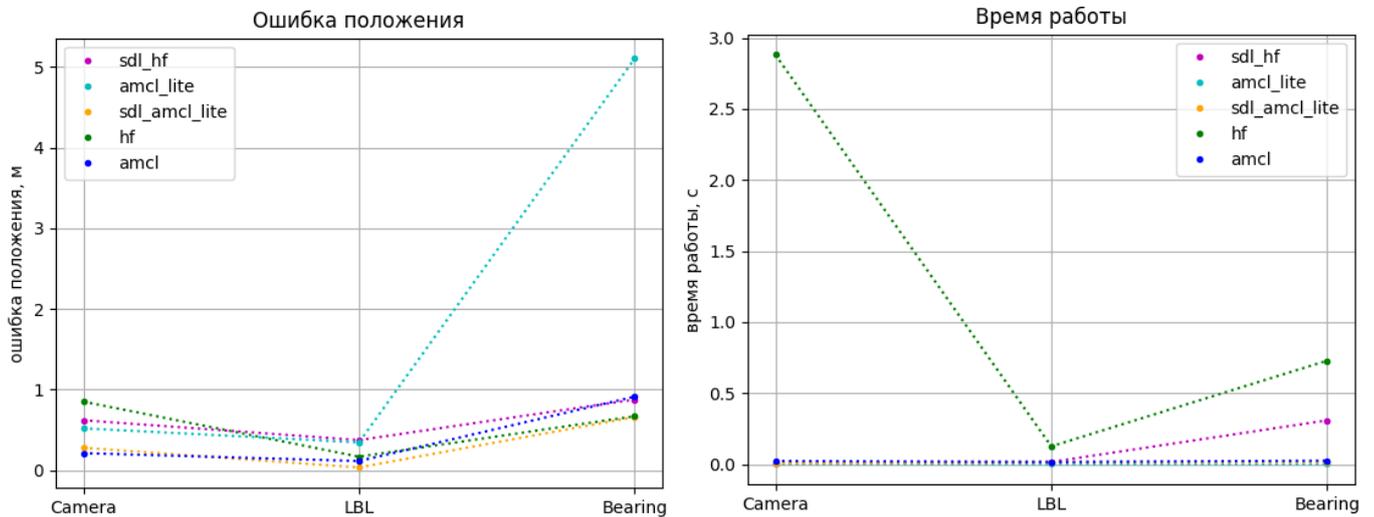


Рисунок 7. Сравнительные графики для исследованных методов (слева) ошибки положения (справа) времени работы

В разделе 3.3 представлена реализация и эксперименты с комплексной технологией глобальной локализации. Поскольку первым технология базируется на описании сцены, то представлена система распознавания объектов для робототехнических задач Extended Object Detection, реализующее мульти-атрибутное распознавание объектов и распознавание сложных (составных) объектов. Разработанная система активно применялась в последующих экспериментах в средах физического моделирования и на реальных роботах.

Исследование предлагаемой комплексной технологии, в виде комбинации средств распознавания сцен и подхода N-локализации вкупе с определением своего положения на основе данных 2D-лидара проводилось в среде физического моделирования Gazebo. В модели отражалось симметричное помещение с точки зрения геометрии. В одном случае глобальная локализация производилась исключительно методом AMCL (Augmented Monte-Carlo Localization), в другом предлагаемой технологией. Предлагаемый подход показал улучшение точности первоначального положения с 6.37 м до 2.88 м, с использованием метода распознавания сцен без учета визуального сходства (метод *sig* на Рисунок 9). Добавление метода с визуальным сходством, дополнительно увеличит точность.

Для тестирования методов с использованием визуального сходства требуются реальные размеченные данные. В разделе 3.4 описана модификация открытого набора данных KITTI-360, которая восстанавливает соответствие идентификаторов семантической двумерной и трёхмерной разметки, отсутствующее в оригинальном варианте (Рисунок 8). Полученный набор для девяти маршрутов содержит семантическую карту и набор сцен, полученных с камер и лидара автомобиля. Каждый объект характеризуется своим классом, положением, а также вектором признаков, определённым архитектурой CLIP. Такой набор, позволяющий проверять качество задачи распознавания сцен, выложен в открытый доступ<sup>1</sup>.

<sup>1</sup> [https://github.com/MoscowskyAnton/scene\\_recognition\\_kitti\\_360](https://github.com/MoscowskyAnton/scene_recognition_kitti_360)

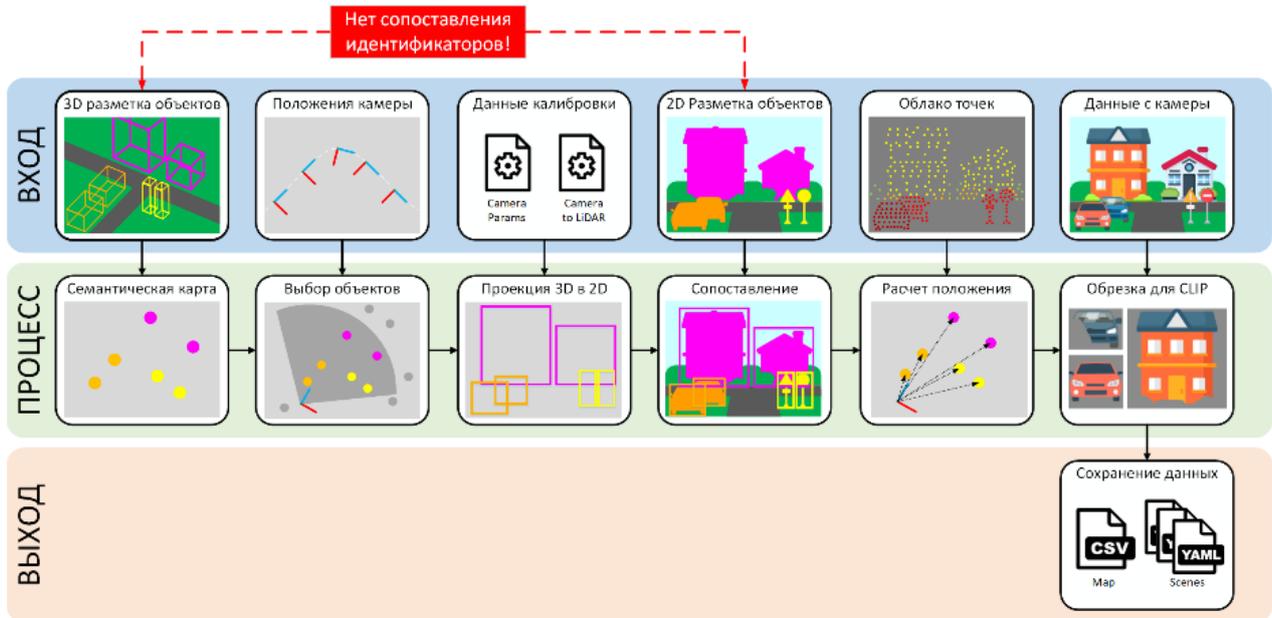


Рисунок 8. Модификация набора данных KITTI-360 для задачи распознавания сцен

**Четвёртая глава** посвящена экспериментальным исследованиям и апробации разработанных методов и моделей.

В разделе 4.1 на полученном наборе данных было проведено исследование нескольких разработанных и улучшенных методов, а также сторонних подходов. В исследовании замерялись величина *Precision* и введённая в работе величина *Degree* (10) (Рисунок 9).

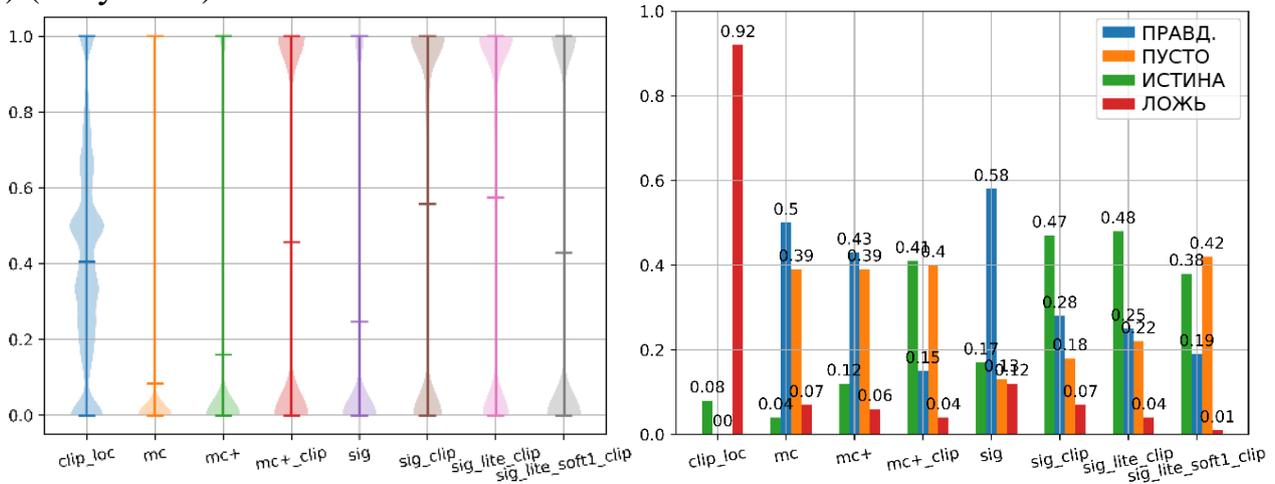


Рисунок 9. Значения *Precision* (слева) и *Degree* (справа) исследуемых методов распознавания сцен

Исследование показало, что методы, которые используют как геометрические особенности, так и визуальное сходство (*mc+\_clip*, *sig\_clip*, *sig\_lite\_clip*, *sig\_lite\_soft1\_clip*), значительно превосходят по точности методы, основанные только на геометрических особенностях (*mc*, *mc+*, *sig*), и метод, основанный исключительно на визуальном сходстве (*clip\_loc*). Общие выводы из проведённого исследования показывают превосходство предложенного подхода, основанного на комбинации геометрических особенностей и визуального сходства.

**Раздел 4.2** посвящен экспериментам с методом N-локализации по камере в среде физического моделирования Gazebo, а в дальнейшем и на роботах YARP8, YARP13 и MENTOR-1 лаборатории НИЦ «Курчатовский институт». В качестве

ориентиров были использованы распространенные ArUco-метки. Эксперименты в среде симуляции и натурные испытания показали, что данный подход, взаимодействуя вместе с фильтром частиц, показывает результаты, при которых робот способен успешно проводить навигацию в рабочей среде

В разделе 4.3 приводится эксперимент с комплексной технологией целиком (Рисунок 5) на полученном наборе данных на основе KITTI-360. Метод N-локализации применялся к лучшим 10 отобранным результатам метода распознавания сцен. Были замерены возможности метода N-локализации сужать поисковое пространство без потери верного решения. Общий размер полученных ограничений для корректного ответа метода распознавания сцен в среднем составлял 3.83% от размера всей карты при средней точности в 99.9%. А при использовании алгоритмов уточнения N-переменных размер ограничений составил 0.32% при точности 97.6%. К полученным ограничениям применялись методы локализации с использованием облаков точек на основе фильтра частиц (SDM+PF) и с классическим подходом на основе методов RANSAC и ICP (SDM+RANSAC+ICP). Также исследовался этот подход отдельно и два варианта метода локализации ориентиров на основе сингулярного разложения – SDV и этот же метод с последующей коррекцией с помощью ICP (SVD+ICP). Из полученных результатов (Рисунок 10) были сделаны выводы, что предложенная комплексная технология улучшает качество задачи глобальной локализации для сложных сценариев, когда метод распознавания сцен не находит верный ответ, но имеет его среди гипотез.

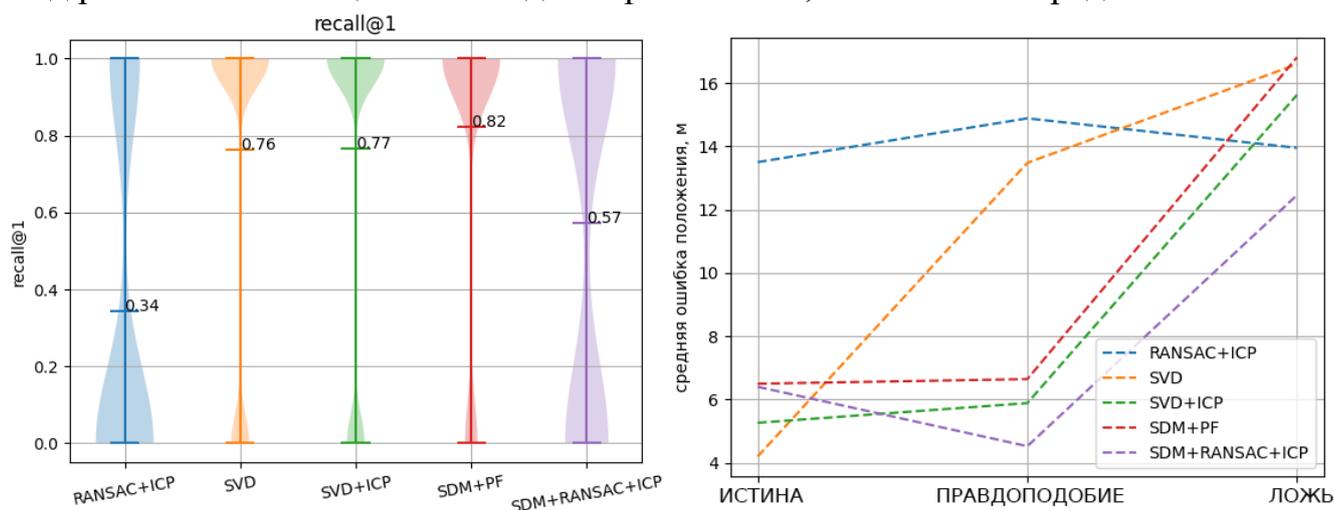


Рисунок 10. Метрика Recall@1 и средняя ошибка положения с порогом в 20м для исследованных методов

Также предлагаемая комплексная технология была апробирована целиком на работе RItRover (Рисунок 11) лаборатории робототехники НИЦ «Курчатовский институт».

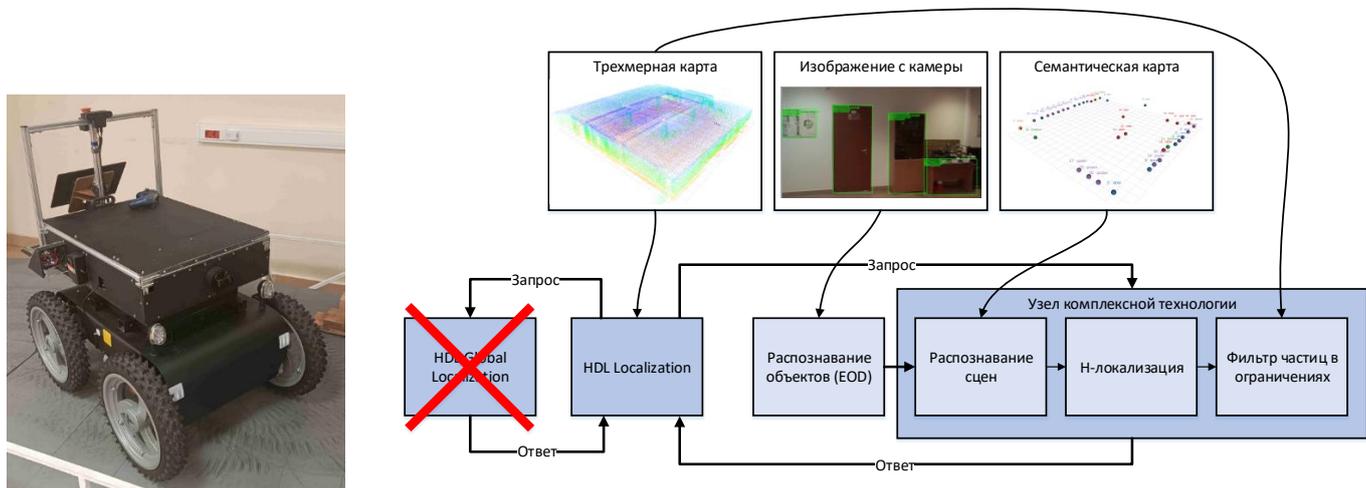


Рисунок 11. Робот RiT rover и вариант комплексной технологии глобальной локализации для него

Предлагаемая технология показала свою практическую применимость в поставленных условиях в симметричном помещении.

**В заключении** обобщаются основные результаты диссертационной работы, обсуждаются проблемы и перспективы развития данного направления исследований.

**В приложениях** приведены акты внедрения результатов исследований и свидетельства о регистрации РИД.

Таким образом, в ходе диссертационного исследования, была решена задача глобальной локализации (без начальных данных о положении) мобильного робота по визуальным ориентирам и дальномерным данным. Решение задачи включает в себя методы распознавания сцены, определяющих соответствие между наблюдением робота и семантической картой, N-локализации, определяющей положение робота по визуальным ориентирам в ограничениях. И классических подходов локализации по дальномерным данным, уточняющих положение робота в полученных ранее ограничениях.

Основные научные результаты диссертации заключаются в следующем:

1. Проведен анализ современных подходов к задаче распознавания сцен. Сделаны выводы, что подходы развиваются в сторону применения формальных подходов теории графов, позволяющих определять геометрические особенности взаиморасположения объектов на сцене, а также в сторону использования базисных моделей для определения визуального сходства.

2. Разработаны метод и алгоритмы распознавания сцен на основе алгоритмов поиска изоморфного подграфа, до этого не применявшихся в исследуемой области. Метод позволяет учитывать как геометрические особенности взаиморасположения объектов, так и визуальное сходство объектов. Мягкое расширение метода также позволяет работать с выбросами во входных данных. Метод позволяет получить ряд гипотез, снабженных численной оценкой. Полученный метод показал наивысшую точность (0.48) по метрике *Degree* (10) и по метрике *Precision* (0.57) среди исследовавшихся аналогов.

3. Было предложено использование модели CLIP для определения визуального сходства наблюдаемых объектов, которое повысило качество решения задачи

распознавания сцен по метрике *Degree* (10) на 30% по числу полностью верно найденных ответов. Данный подход был предложен одновременно с рядом других научных коллективов, занимающихся задачами распознавания сцен в применении к локализации.

4. Разработан ряд метрик, позволяющих оценивать качество решения задачи распознавания сцен в виде множества гипотез.

5. Реализован алгоритм формирования набора данных для тестирования задач распознавания сцен на основе открытого набора данных KITTI-360. Реализация алгоритма и полученные данные находятся в открытом доступе.

6. Разработан метод N-локализации посредством формализации N-модели к текущей области и разработки соответствующих алгоритмов, позволяющих проводить локализацию робота по визуальным ориентирам и получать интервальные оценки положения робота (ограничения на положение), а также бороться с противоречиями во входных данных и учитывать погрешности в определении расстояний и угла на объекты. N-локализация может применяться совместно с вероятностными методами локализации, увеличивая как скорость их работы (до 10 раз для гистограммного фильтра) так и точность полученных решений (до 9 раз в случае фильтра частиц).

7. Разработана система распознавания объектов для робототехнических применений на основе мультиатрибутной парадигмы, позволяющей комбинировать в рамках одного объекта ряд различных методов технического зрения. Реализация данной системы выложена в открытый доступ.

8. Предложена архитектура комплексной технологии глобальной локализации по визуальным ориентирам и дальномерным данным с использованием методов распознавания сцен и подхода N-локализации. Проведённые эксперименты показали увеличение точности определения положения от 5 до 52% в сравнении с существующими методами по метрике *Recall@1*.

9. Разработанные алгоритмы были апробированы на ряде реальных роботов, разного назначения и обладающих различными сенсорными и вычислительными возможностями, а также внедрены в ряд организаций.

## **Основные публикации автора по теме диссертации**

### В списке ВАК по специальности 2.3.5:

1. Московский А.Д. Распознавание сцен в задаче глобальной локализации мобильного робота с использованием моделей векторных представлений изображений и графовых подходов // Управление большими системами. 2025. № 114. С. 307–344.

2. Московский А.Д. Применение недоопределённых моделей в задаче глобальной локализации мобильного робота // Проблемы управления. 2025. № 4. С. 64–78.

### В изданиях WoS\Scopus:

3. Moscovsky A. Subdefinite Computations for Reducing the Search Space in Mobile Robot Localization Task // Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) /

ed. Kovalev S.M., Kuznetsov S.O., Panov A.I. Cham: Springer International Publishing, 2021. Vol. 12948 LNAI. P. 180–196. [WoS, Scopus]

4. Moscovsky A.D. Extended Object Detection: Flexible Object Description System for Detection in Robotic Tasks // Smart Electromechanical Systems: Recognition, Identification and Modeling. 2022. Vol. 419. P. 27–43. [Scopus]

5. Moscovsky A.D., Rovbo M.A., Sorokoumov P.S., Moskovskaya. E.D. Review of Methods for Autonomous Rough Terrain Traversal of Mobile Robots // Automation and Remote Control. – 2025. [WoS, Scopus]

В сборниках трудов конференций, индексируемых РИНЦ:

6. Московский А.Д. Недоопределённые модели в задаче локализации мобильного робота // Пятый Всероссийский научно-практический семинар “Беспилотные транспортные средства с элементами искусственного интеллекта” (БТС-ИИ-2019). Санкт-Петербург: Общероссийская общественная организация «Российская ассоциация искусственного интеллекта», 2019. P. 9–18.

7. Московский А.Д. Распознавание сцен для задачи глобальной локализации робота // Труды 34-й Международной научно-технической конференции «Экстремальная робототехника». СПб., 2023. P. 267–274.

8. Московский А.Д. Распознавание сцен для задач локализации мобильного робота в условиях неопределённости // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник научных трудов XII Международной научно\_практической конференции (ИММВ-2024, Коломна, 14-17 мая 2024 г.). В 2-х томах. Коломна: Универсум, 2024. Vol. 2. P. 255–266.

9. Московский А.Д. Применение архитектуры CLIP в задаче распознавания сцен для локализации мобильного робота // Сборник тезисов 35-й Международной научно-технической конференции «Экстремальная робототехника». 2024. P. 259–262.

Патенты на полезную модель:

10. Карпов В.В., Малышев А.А., **Московский А.Д.**, Ревин С.А., Ровбо М.А., Скрыльников Р.А., Тихонычев В.В. Мобильная платформа для проведения подспутниковых измерений. Патент на полезную модель 178972 от 24.04.2018 // 2018.

Свидетельства о государственной регистрации программ для ЭВМ:

11. Московский А.Д. Система автоматического формирования семантического слоя карты / Программа для ЭВМ, №2020619217 от 13.09.2020 // Свидетельство о государственной регистрации программы для ЭВМ. 2020.

12. Московский А.Д. Система локализации мобильного робота на основе недоопределённых моделей / Программа для ЭВМ, №2019667433 от 11.12.2019 // Свидетельство о государственной регистрации программы для ЭВМ. РФ: НИЦ КИ, 2019.