

МЕТОДОЛОГИЯ РЕШЕНИЯ ПРОБЛЕМЫ ОДНОВРЕМЕННОЙ НАВИГАЦИИ И ПОСТРОЕНИЯ КАРТЫ НА ОСНОВЕ КОМБИНИРОВАНИЯ ВИЗУАЛЬНЫХ И СЕМАНТИЧЕСКИХ ХАРАКТЕРИСТИК ОКРУЖАЮЩЕЙ СРЕДЫ

А. В. Вохминцев

Аннотация. Представлен обзор результатов исследования проблемы одновременной навигации и построения карты окружающей среды. Целью исследования является повышение эффективности процессов обработки информации в методах и алгоритмах для реконструкции трехмерных сцен окружающего пространства, визуализации трёхмерных моделей и навигации мобильных платформ в неизвестной среде на основе комбинирования информации о семантических характеристиках окружающей среды и доступной мультисенсорной информации, получаемой с различных типов датчиков. Для достижения цели решаются следующие задачи. Разработка методологии для точного решения задачи SLAM, основанная на комбинировании визуальных и семантических характеристик неизвестной окружающей среды. Разработка алгоритма поиска и сопоставления особых точек на основе двумерных изображений для обработки информации о цвете в кадре. Разработка точных комбинированных методов реконструкции контекстуально сложных крупномасштабных трехмерных сцен на основе итеративного алгоритма ближайших точек для группы аффинных преобразований и для группы ортогональных преобразований. Разработка метода решения задачи навигации и определения положения мобильной платформы на сцене, основанного на комбинировании визуальных и семантических характеристик неизвестной окружающей среды и расширенной калмановской фильтрации. Разработка быстрого рекурсивного алгоритма для вычисления дискретного косинусного преобразования (ДКП) для решения задач сжатия и восстановления изображений, получаемых со множества датчиков на мобильной платформе. Разработка подхода к оптимизации графа положений мобильной платформы и обнаружению «замыканий цикла» на основе метода корзина слов и комбинировании визуальных и семантических характеристик неизвестной окружающей среды. Апробация предложенной комбинированной визуальной системы SLAM для решения задачи спасения людей. Объектом исследования является система обработки визуальной и семантической информации о неизвестной окружающей среде для решения задачи реконструкции контекстуально сложных крупномасштабных трехмерных сцен и задачи определения положения мобильной платформы на сцене в каждый момент времени. Предметом исследования являются методы, алгоритмы для решения проблемы одновременной навигации и построения карты на основе комбинирования информации о семантических и визуальных характеристиках окружающей среды.

Ключевые слова: Simultaneous Localization and Mapping; регистрация данных; Iterative Closest Point; метод Хорна; сопоставление особых точек; проблема замыкания цикла, расширенный фильтр Калмана; EKF-SLAM; Visual SLAM.

ВВЕДЕНИЕ

Проблема одновременной навигации и построения карты (Simultaneous Localization and Mapping, SLAM) заключается в реконструкции двухмерной или трехмерной карты неизвестного окружающего пространства вокруг автономного интеллектуального объекта управления, например, мобильной платформы и определении своего положения на этой карте в каждый момент времени во время движения. Проблема SLAM является одной из ключевых проблем в современной робототехнике и машинном зрении, так как на ее решении основано создание автономных интеллектуальных робототехнических комплексов. Данная проблема

считается полностью решенной для построения двумерных карт и трехмерных карт для небольших статических сцен, однако для динамических, контекстуально сложных и крупномасштабных сцен является активной областью научных исследований. По оценкам аналитиков Агентства по перспективным оборонным научно-исследовательским разработкам (DARPA), существуют следующие актуальные проблемы в области SLAM: разработка быстрых методов для навигации в неизвестном динамическом пространстве; разработка интеллектуальных методов навигации на основе семантических свойств окружающей среды; создание методов для навигации на сценах, искаженных помехами окружающей среды различной природы; проблема позиционной точности методов построения трехмерной карты, связанная с принципиальной возможностью построения высоко точных моделей трехмерных объектов на основе визуальной информации, одометрической информации и информации о глубине. Данное исследование направлено на решение второй проблемы. Решение представленных проблем в области SLAM позволит перейти к созданию новых видов роботизированных коллективных систем, с помощью которых будет решаться широкий спектр задач, таких как распознавание и слежение за динамическими объектами в заданном пространстве, предсказание траектории их перемещения, согласованное и оптимизированное планирование движений мобильных платформ в заданном пространстве с учетом возможных статических и динамических препятствий. Важным направлением применения результатов исследования является коллаборативная робототехника, основным тренд в которой связан с созданием комплексов, оснащенных собственной системой машинного зрения и системой SLAM. Перспективным направлением применения результатов исследования является антропоморфная робототехника и связанные с ней области исследований.

СТЕПЕНЬ РАЗРАБОТАННОСТИ ТЕМЫ И ОБСУЖДЕНИЕ РЕШАЕМОЙ ЗАДАЧИ

Принято различать следующие методологии SLAM: визуальная (Visual SLAM), на основе частичного фильтра (Particle Filter SLAM, FAST SLAM), на основе расширенного фильтра Калмана (Extended Kalman Filter, EKF), на основе графов (Graph-Based SLAM). Известным методологиям свойственны следующие недостатки:

- методологии имеют ограничения по применению в окружающей среде, искаженной различными естественными помехами, в том числе классическая методология Visual SLAM показывает неудовлетворительные результаты для сцен с неравномерным освещением;
- методологии характеризуются высокой вычислительной и алгоритмической сложностью, поэтому их применение ограничено в приложениях, где требуется осуществлять одновременную навигацию и картографирование для контекстуально сложных крупномасштабных сцен в режиме реального времени;
- методология EKF-SLAM не учитывает при решении поставленной задачи визуальную информацию об окружающей среде. Оценка карты и определение положения навигационных ориентиров на сцене осуществляются без использования данных о форме реконструируемой поверхности объекта, поэтому точность получаемых оценок положения мобильной платформы на карте относительно навигационных ориентиров невысока;
- в методологии Visual SLAM информация о визуально связанных характеристиках и трехмерных облаках точек не используется для совместного (комбинированного) решения вариационной задачи ICP. Результат реконструкции сцены с точки зрения точности и сходимости обусловлен выбором начального приближения для решения вариационной задачи;
- визуальные методы для решения проблемы SLAM, которые основаны на итеративном алгоритме ICP для группы аффинных преобразований, не позволяют проводить точную регистрацию структурных элементов сцены;
- в методологии Visual SLAM не используется информация о семантических характеристиках окружающей среды, вследствие чего все особые точки в двумерных и трехмерных данных имеют одинаковую значимость;

– в методологии Visual SLAM информация о визуально связанных характеристиках окружающей среды используется для улучшения точности и сходимости всей вычислительной процедуры только на этапе сопоставления данных об особых точках на изображении.

Для преодоления указанных недостатков в работе предлагается разработать новую методологию решения проблемы SLAM на основе комбинирования визуальных и семантических характеристик окружающей среды. Предложенная методология основана на использовании трех типов данных: визуально связанные характеристики окружающей среды (изображения, данные о цвете), трехмерные облака точек (данные о глубине сцены), данные о семантических свойствах окружающей среды, при этом визуальная и семантическая информация о трехмерной сцене будет использоваться на всех ключевых шагах решения проблемы SLAM.

Проблемы создания теории, моделей, методов и алгоритмов для создания системы одновременной навигации и картографирования рассматриваются в следующих научных трудах: T. Rui, Q. Zhang, Y. Zhou, J. Xing, G. Dissanayake, P. Newman, S. Clark, H.-F. Durrant-Whyte, M. Csorba, S. K. Zhou, R. Chellappa, B. Moghaddam, M. Montemerlo, S. Thrun, D. Koller, V. Wegbreit, C. Estrada, J. Neira, J.D. Tardós, В. И. Кобера, Е. Н. Соснова, С. Л. Зенкевича и В. С. Лемпицкого. Были созданы ряд передовых технологий в области SLAM в неизвестной среде, среди которых можно выделить разработки Stanford Artificial Intelligence Lab, Stanford Racing Team, Volkswagen Electronics Research Laboratory, Massachusetts Institute of Technology Lincoln Laboratory, Pixon Imaging. Данные технологии нашли применение в военной сфере для организации наблюдения на поле боя или при разведывательных операциях, для обеспечения безопасности военных и стратегических промышленных объектов, при организации спасательных операций и для эвакуации раненых с поля боя. В работе рассмотрено применение предлагаемой комбинированной методологии SLAM для создания систем спасения людей. Работа связана с тематикой работ, осуществляемых в соответствии с Национальным проектом «Цифровая экономика». На основании сказанного выше сформулированы основная цель и задачи исследования.

Целью является повышение эффективности процессов обработки информации в методах и алгоритмах для реконструкции трехмерных сцен окружающего пространства, визуализации трёхмерных моделей и навигации мобильных платформ в неизвестной среде на основе комбинирования информации о семантических характеристиках окружающей среды и доступной мультисенсорной информации, получаемой с различных типов датчиков. Для достижения поставленной цели решаются следующие задачи:

1. Разработать методологию для точного решения задачи SLAM, основанную на комбинировании визуальных и семантических характеристик неизвестной окружающей среды.
2. Разработать алгоритм поиска и сопоставления особых точек на основе двумерных изображений для обработки информации о цвете в кадре.
3. Разработать точные комбинированные методы реконструкции контекстуально сложных крупномасштабных трехмерных сцен на основе итеративного алгоритма ближайших точек для группы аффинных преобразований и для группы ортогональных преобразований.
4. Разработать метод решения задачи навигации и определения положения мобильной платформы на сцене, основанный на комбинировании визуальных и семантических характеристик неизвестной окружающей среды и расширенной калмановской фильтрации.
5. Разработать быстрый рекурсивный алгоритм для вычисления дискретного косинусного преобразования (ДКП) для решения задач сжатия и восстановления изображений, получаемых со множества датчиков на мобильной платформе.
6. Предложить подход к оптимизации графа положений мобильной платформы и обнаружению «замыканий цикла» на основе метода корзина слов и комбинировании визуальных и семантических характеристик неизвестной окружающей среды.
7. Произвести апробацию предложенной комбинированной визуальной системы SLAM для решения задачи спасения людей.

Объектом исследования является система обработки визуальной и семантической информации о неизвестной окружающей среде для решения задачи реконструкции контекстуально сложных крупномасштабных трехмерных сцен и задачи определения положения мобильной платформы на сцене в каждый момент времени. Предмет исследования – методы, алгоритмы для решения проблемы одновременной навигации и построения карты на основе комбинирования информации о семантических и визуальных характеристиках окружающей среды.

Повышение степени автономности управления роботизированным комплексом – основная цель, преследуемая исследователями при разработке современных робототехнических систем. Достижение этой цели возможно за счет интеграции теории автоматического управления, теории искусственного интеллекта, компьютерного когнитивного моделирования и создания на основе такой интеграции систем одновременной навигации и построения карты нового поколения. В связи с этим исследовано современное состояние проблемы, рассмотрены известные методологии решения проблемы: визуальная (Visual SLAM), на основе частичного фильтра (FAST SLAM), на основе расширенного фильтра Калмана (EKF SLAM), на основе графов (Graph-Based SLAM). Предложенная комбинированная методология наследует некоторые свойства визуальной методологии SLAM (Visual SLAM), которая, в свою очередь, развивает методологию на основе расширенного фильтра Калмана (EKF SLAM). Определены основные проблемы известных методологий применительно к решению задачи SLAM для контекстуально сложных крупномасштабных сцен, представлена информация об основных конкурентах в данной области, проведен патентный поиск по базе ЦИТИС по данной тематике. Исследованы известные способы решения задачи автоматической трехмерной реконструкции сцен и задачи планирования траектории применительно к решению поставленной практической задачи спасения людей, рассмотрены некоторые аспекты решения задачи некооперативного распознавания личности по лицу на расстоянии. Для решения задачи трехмерной реконструкции используется итеративный алгоритм ближайших точек (Iterative Closest Point, ICP) либо его модификации. Полученные результаты в разрозненном виде приведены в работах [1–11]. В данной работе представлена целостная картина исследования.

МЕТОДОЛОГИЯ РЕШЕНИЯ ПРОБЛЕМЫ SLAM

Предложена методология для решения проблемы SLAM, основанная на комбинировании визуальных и семантических характеристик неизвестной окружающей среды (рисунок 1) [1].

Методология основана на использовании трех типов информации: визуально связанные характеристики окружающей среды (данные о цвете), трехмерные облака точек (данные о глубине сцены), данные о семантических свойствах окружающей среды. В данной методологии входной информацией являются последовательность RGB-D кадров и критерии останова для алгоритмов. Входные данные для компьютерного моделирования и вычислительных экспериментов получены со множества камер, расположенных на мобильной платформе (камеры видимого диапазона, камеры глубины, RGB-D камеры). Выходной информацией являются реконструкция трехмерной сцены и граф положений мобильной платформы, связанный с координатами трехмерной карты в различные моменты времени движения мобильной платформы. Предлагаемая методология в отличие от известных использует мультисенсорную и семантическую информацию на всех этапах решения проблемы SLAM для повышения его точности и состоит из следующей повторяющейся последовательности шагов:

1. Получение данных с визуальных камер.
2. Процедура предварительной обработки данных (сжатие изображений и восстановление изображений от помех различной природы).
3. Снижение размерности двумерных и трехмерных данных в RGB-D кадре с целью уменьшения количества точек в анализируемых данных.
4. Извлечение информации об особых точках на изображении с использованием предложенного в работе двумерного дескриптора на основе локальных гистограмм ориентированных градиентов.

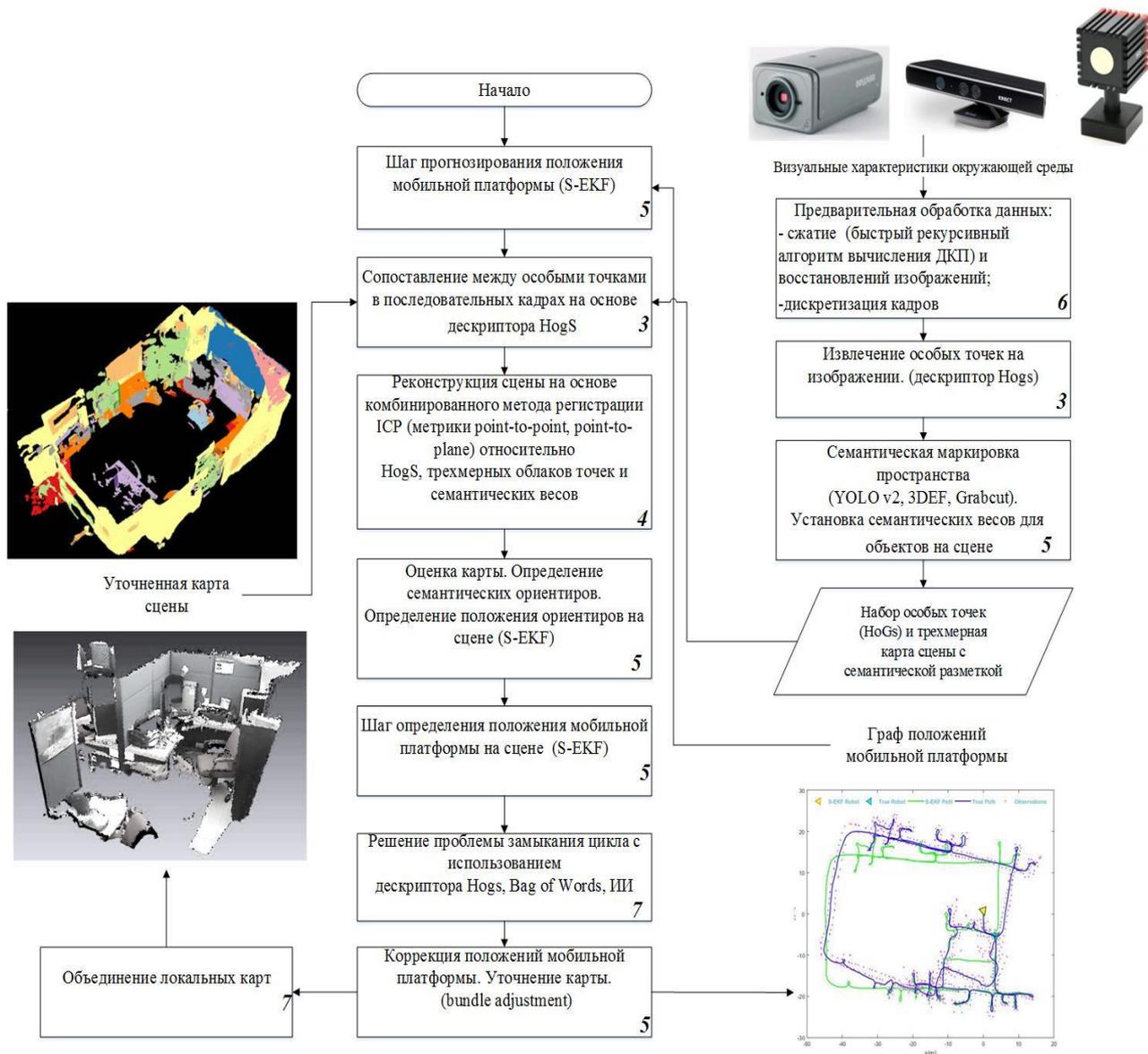


Рис. 1 Комбинированная методология SLAM.

5. Семантическая маркировка 3D пространства на основе 2D изображений и вычисление семантических весов для объектов на сцене.

6. Прогнозирование положения мобильной платформы на основе расширенного фильтра Калмана (EKF).

7. Сопоставление особых точек на двумерных и трехмерных данных в последовательности кадров.

8. Реконструкция сцены на основе комбинированного метода ICP с метрикой «точка – точка» и метрикой «точка – плоскость» для группы аффинных и группы ортогональных преобразований соответственно с использованием информации о семантических весах объектов.

9. Оценка карты и определение положения навигационных ориентиров на сцене.

10. Вычисление положения мобильной платформы на сцене с использованием EKF: вычисление сдвига мобильной платформы на трехмерной сцене и сопоставление данных о трехмерной сцене, полученных на текущем шаге, с данными, которые были получены на предыдущем шаге.

11. Решение проблемы «замыкания цикла» с использованием предложенного дескриптора, известного подхода корзина слов и предложенного метода для поиска визуального подобия на основе инвертированного индекса.

12. Коррекция положения мобильной платформы на трехмерной сцене и уточнение карты. Оптимизация графа положений платформы.

13. Объединение локальных карт. Изменение трехмерной карты окружающей среды, используя результаты, полученные на шагах 11–12.

Визуально связанные характеристики сопоставляются с использованием предложенного в работе алгоритма сопоставления данных на основе дескриптора ДЛ ГНГ. Трехмерные данные сопоставляются с использованием метода ближайших соседей. Информация о семантических характеристиках окружающей среды получена в результате использования известных методов семантической маркировки пространства (YOLO и 3DEF). Ключевым шагом предлагаемой методологии является регистрация данных [2]. Данный шаг связан с комбинированным (совместным) решением вариационной задачи алгоритма ICP относительно трехмерных облаков точек, визуально связанных и семантических характеристик окружающей среды. Задача регистрации данных заключается в построении модели объекта по множеству его проекций в окружающем пространстве. Данные о характерных признаках изображений были применены на двух основных этапах метода регистрации ICP: для задания начальных значений и при решении вариационной задачи. Семантические маркеры были связаны со значениями весов слагаемых в рассматриваемом функционале для улучшения сходимости метода регистрации при решении вариационной задачи. В работе предложен точный комбинированный метод решения вариационной задачи для группы аффинных и группы ортогональных преобразований без требования начальной инициализации с использованием информации о семантических характеристиках окружающей среды. Также предложен метод навигации и определения положения мобильной платформы S-EKF SLAM, основанный на расширенном фильтре Калмана, использовании мультисенсорной динамической информации, пространственном согласовании последовательных кадров данных и локализации семантических навигационных ориентиров. В работе предложен подход к решению проблемы «замыкания цикла», а также коррекции движения мобильной платформы на основе известного алгоритма «корзина слов».

МЕТОД ПОИСКА И СОПОСТАВЛЕНИЯ ОСОБЫХ ТОЧЕК НА ИЗОБРАЖЕНИЯХ НА ОСНОВЕ ДВУМЕРНЫХ ДЕСКРИПТОРОВ

Точность и сходимость итерационных методов при решении вариационной задачи, связанной с минимизацией функции ошибки, зависит от выбора начальных значений параметров метода ICP. Как известно, итеративный алгоритм ближайших точек инициализируется начальными значениями параметров геометрического преобразования, такими как поворот, параллельный перенос и изменение масштаба. Выбор параметров инициализации метода ICP существенно влияет на сходимость данного алгоритма. Для преодоления данного недостатка в работе предлагается использовать визуально связанные характеристики RGB-D кадра, которые позволяют совмещать кадры без требования начальной инициализации. Известные методы сопоставления изображений на основе дескрипторов (SIFT, SURF, ORB) имеют высокую алгоритмическую сложность, поэтому их нельзя использовать для решения поставленной задачи – одновременной навигации и построения карты для крупномасштабных контекстуально сложных сцен. Для обработки визуальных характеристик сцены предложен алгоритм сопоставления изображений на основе рекурсивного вычисления гистограмм направленных градиентов (ГНГ) по нескольким круглым скользящим окнам и пирамидальному разложению изображения (ДЛ ГНГ) [3]. Пусть $W_i = \{(x, y) \in R^2\}, i = 1, \dots, M$ – набор медианных фильтров (МФ) в виде круглых скользящих окон:

$$W_i = \{(x - x_i)^2 + (y - y_i)^2 \leq r_i\}, \varphi(x, y) = \left[\frac{\varrho}{360} \arctan\left(\frac{gy}{gx}\right) + \frac{1}{2} \right], \quad (1)$$

где $\{x_i, y_i\}$ – координаты центра медианного фильтра и r_i – радиус i -го диска, соответствующего медианному фильтру, Q – дискретный уровень. В каждом положении i -го МФ из набора вычислим значение градиентов с использованием оператора Собеля. Далее, используя значения модулей градиентов $\{\text{Mag}_i(x, y): (x, y) \in W_i\}$ и величины углов направлений $\{\varphi_i(x, y): (x, y) \in W_i\}$, вычислим ГНГ в каждом положении i -го МФ:

$$\text{HOG}_i(\alpha) = \begin{cases} \sum_{(x,y) \in W_i} \delta(\alpha - \varphi_i(x, y)), & \text{if } \text{Mag}_i(x, y) \geq \text{Med}, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

где $\alpha = \{0, \dots, Q - 1\}$ – порядковый номер дискретного уровня ГНГ, Med – медиана модулей градиентов МФ, $\delta(z) = 1$, если $z = 0$; иначе 0, и представляет из себя символ Кронекера. В исследовании используется корреляционный оператор, при помощи которого осуществляется процедура сравнения объектов на анализируемой двумерной сцене с соответствующей ей ГНГ эталонного изображения. Данный оператор позволяет обеспечить инвариантность дескриптора к преобразованию «поворот». Далее введем формулу для определения нормализованной центрированной ГНГ эталонного изображения

$$\overline{\text{HOG}}_i^R(\alpha) = \frac{\text{HOG}_i^R(\alpha) - \text{Mean}^R}{\sqrt{\text{Var}^R}}, \quad (3)$$

где Mean^R – среднее значение ГНГ, Var^R – дисперсия ГНГ. Тогда для каждого i -го МФ в позиции k можно определить корреляционную функцию

$$C_i^k(\alpha) = \text{IFT} \left[\frac{HS_i^k(\omega) HR_i^*(\omega)}{\sqrt{Q \sum_{q=0}^{Q-1} (\text{HOG}_i^k(q))^2 - (HS_i^k(0))^2}} \right], \quad (4)$$

где $HS_i^k(\omega)$ – преобразование Фурье ГНГ внутри i -го МФ входной сцены, $HR_i(\omega)$ – преобразование Фурье $\overline{\text{HOG}}_i^R(\alpha)$, * для i -го преобразования Фурье обозначает «комплексное сопряжение». Для определения подобия двух ГНГ применяется корреляционный пик $P_i^k = \max_{\alpha} \{C_i^k(\alpha)\}$. Предложенный двумерный дескриптор обладает лучшими характеристиками по сравнению с известными дескрипторами при малых поворотах вне области сцены.

РЕКОНСТРУКЦИЯ ТРЕХМЕРНЫХ СЦЕН НА ОСНОВЕ ТОЧНЫХ РЕШЕНИЙ ВАРИАЦИОННОЙ ЗАДАЧИ РЕГИСТРАЦИИ МЕТОДА ИСР

Во многих приложениях робототехники для построения трехмерных моделей по картам глубины используется алгоритм ИСР. Главный этап алгоритма регистрации ИСР связан с поиском соответствующего геометрического преобразования (ортогонального или аффинного), которое наилучшим образом совмещает два облака точек в разных кадрах для выбранной метрики (вариационная задача алгоритма). Наиболее распространенными вариантами ИСР являются алгоритмы с метриками «точка – точка» и «точка – плоскость». Точность реконструкции трехмерной сцены существенно зависит от выбора метрики для оценки геометрического преобразования и метода решения вариационной задачи. Результат применения итерационных методов для решения задачи минимизации выбранного функционала зависит от правильности выбора начального приближения параметров геометрического преобразования. Использование решений вариационной задачи в замкнутой форме позволяет избежать этих проблем. Выбор класса геометрических преобразований оказывает значительное влияние на результат реконструкции трехмерной сцены. Предложены методы решения вариационной подзадачи ИСР на основе комбинации данных о визуально связанных характеристиках, трехмерных облаках точек и данных о семантических свойствах окружающей среды.

Рассмотрим вариационную задачу метода регистрации ИСР для аффинных преобразований [4]. Давайте обозначим через $X = \{x_1, \dots, x_n\}$ трехмерные данные об исходном RGB-D кадре и через $Y = \{y, \dots, y_m\}$ данные о целевом RGB-D кадре в \mathbb{R}^3 . Пусть отношения между точками в кадрах X и Y такие, что для каждой точки в x_i можно вычислить соответствующую точку в y_i . Тогда итеративный алгоритм ближайших точек (ICP) можно рассмотреть как геометрическое преобразование ригидных объектов из X в Y следующего вида

$$R x_i + T, \quad (5)$$

где R матрица поворота, T вектор параллельного переноса, $i = 1, \dots, n$. Пусть $J(R, T)$ можно представить как функцию следующего вида

$$J(R, T) = \sum_{i=1}^n \| R x_i + T - y_i \|^2, \quad (6)$$

$$\text{где } R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}, T = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}, x_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \end{pmatrix}, y_i = \begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \end{pmatrix}. \quad (7)$$

Положим, что $\sum_{i=1}^n x_{1i}^2 \neq 0, \sum_{i=1}^n x_{2i}^2 \neq 0, \sum_{i=1}^n x_{3i}^2 \neq 0$, тогда мы можем решить вариационную задачу относительно матрицы поворота R :

$$\frac{\partial J(R)}{\partial r_{1k}} = \sum_{i=1}^n 2 (r_{11} x_{1i} + r_{12} x_{2i} + r_{13} x_{3i} + t_1 - y_{1i}) x_{ki} = 0, k = 1, 2, 3, \quad (8)$$

$$r_{1k} = - \frac{\sum_{i=1}^n (r_{1m} x_{mi} + r_{1n} x_{ni} - y_{1i}) x_{ki}}{\sum_{i=1}^n x_{ki}^2}. \quad (9)$$

Введем следующие обозначения для упрощения вида функционала $J(R)$:

$$G_{mi} = x_{mi} - x_{ki} \frac{\sum_{j=1}^n x_{mj} x_{kj}}{\sum_{j=1}^n x_{kj}^2}, \quad (10)$$

$$G_{pi} = x_{ni} - x_{ki} \frac{\sum_{j=1}^n x_{nj} x_{kj}}{\sum_{j=1}^n x_{kj}^2}, G_{ki} = y_{1i} - x_{ki} \frac{\sum_{j=1}^n y_{1j} x_{kj}}{\sum_{j=1}^n x_{kj}^2}.$$

Тогда с учетом обозначений в (10) представим функционал $J(R)$ как

$$J(R) = \sum_{i=1}^n (r_{1m} G_{mi} + r_{1n} G_{pi} - G_{ki})^2 + (r_{21} x_{1i} + r_{22} x_{2i} + r_{23} x_{3i} - y_{2i})^2 + (r_{31} x_{1i} + r_{32} x_{2i} + r_{33} x_{3i} - y_{3i})^2. \quad (11)$$

Тогда определим частную производную $J(R)$ по r_{1m}

$$r_{1m} = - \frac{r_{1n} \sum_{i=1}^n G_{mi} G_{pi} - \sum_{i=1}^n G_{mi} G_{ki}}{\sum_{i=1}^n G_{mi}^2}. \quad (12)$$

Введем следующие обозначения:

$$q_1 = G_{pi} - \frac{G_{mi} \sum_{j=1}^n G_{mj} G_{pj}}{\sum_{j=1}^n G_{mj}^2}, \quad q_2 = G_{ki} - \frac{G_{mi} \sum_{j=1}^n G_{mj} G_{kj}}{\sum_{j=1}^n G_{mj}^2}. \quad (13)$$

Теперь определим частную производную $J(R)$ по r_{1n} :

$$\frac{\partial J(R)}{\partial r_{1n}} = 2 \sum_{j=1}^n (r_{1n} q_1 - q_2) q_1 = 0, \quad r_{1n} = \frac{\sum_{k=1}^n q_1 q_2}{\sum_{k=1}^n q_1^2}. \quad (14)$$

Таким образом определены все элементы первой строки матрицы поворота R : r_{1m}, r_{1k}, r_{1n} . Аналогичным образом получим вторую и третью строку матрицы поворота. Рассмотрим комбинированный метод решения вариационной задачи «точка – точка» в замкнутой форме для аффинных преобразований

$$J(RV, RD) = \alpha \frac{1}{W} \frac{1}{|A_f|} \sum_{i \in A_f}^m w_i \|M(RVF_x^i) - M(F_y^i)\|^2 + (1 - \alpha) \frac{1}{W} \frac{1}{|A_d|} \sum_{j \in A_d}^n w_j \|RDx_j + T - y_j\|^2, \quad (15)$$

где RV, RD – матрицы аффинного преобразования для визуально связанных характеристик сцены и для данных о глубине сцены соответственно, α, W – набор параметров для нормировки данных, подбираемый эмпирическим путем, w_i, w_j – весовые характеристики данных, связанные с семантической маркировкой пространства, A_f – подмножество содержит связи между особыми точками в двух кадрах, A_d – содержит связи между соответствующими точками x_j и y_j в трехмерных облаках точек в двух кадрах данных, F_x^i, F_y^i – визуально связанные характеристики сцены. В общем случае, $RV \neq RD$, в данной работе находится совместное решение вариационной задачи для частного случая, когда $RV = RD = RT$. Пусть T_{km} – начальная оценка для ИСП с использованием кинематической модели движения камеры, k_{\max}, ε – пороги алгоритма ИСП по числу шагов и по величине ошибки E соответственно, RT^* – лучшее преобразование на i -м шаге метода, δ – порог для точек инлайнеров. Комбинированный метод состоит из шагов:

1. Вычисление гистограмм направленных градиентов.
2. Сопоставление между особыми точками F_x^i и F_y^i для выбранных подмножеств. Решение вариационной задачи регистрации данных для визуально связанных характеристик изображения. Получим RT^* и A_f .
3. Проверка: Если $|A_f| < \delta$, то $RT^* = T_{km}$ и $A_f = \emptyset$. Положим $i = 1$.
4. Определение соответствующих точек A_d в исходном и целевом облаке точек с использованием метода ближайших соседей. Определение весовых коэффициентов для каждой полученной пары A_d .
5. Решение комбинированной вариационной задачи относительно RT^*, A_f и A_d .
6. Проверка: $(E(RT^*_i) - E(RT^*_{i+1})) < \varepsilon$ или (Номер итерации $> k_{\max}$). Если условие не верно, то возврат на Шаг 4 и $i = i + 1$, иначе получено искомое преобразование RT^* .

Для проведения компьютерного моделирования были выбраны четыре набора данных из базы данных NYU Depth Dataset V2, содержащие фрагменты крупномасштабных сцен помещений: Classrooms, Living Rooms (1/4), Offices (1/2), Offices (2/2). В результате серии тестов установлены зависимости скорости сходимости методов от выбора ошибки метрики и условий проведения экспериментов (рисунок 2). Было установлено, что в контролируемых условиях предложенный метод регистрации данных имеет сходимость, близкую к методу, который использует метрику «точка – плоскость» (рисунок 2, а), в неконтролируемых условиях (при неравномерном освещении) предложенный метод показывает лучшую сходимость, чем указанные выше методы регистрации данных (рисунок 2, б). Был проведен сравнительный анализ зависимости скорости сходимости методов регистрации от выбора типа дескриптора и условий проведения экспериментов (рисунок 3).

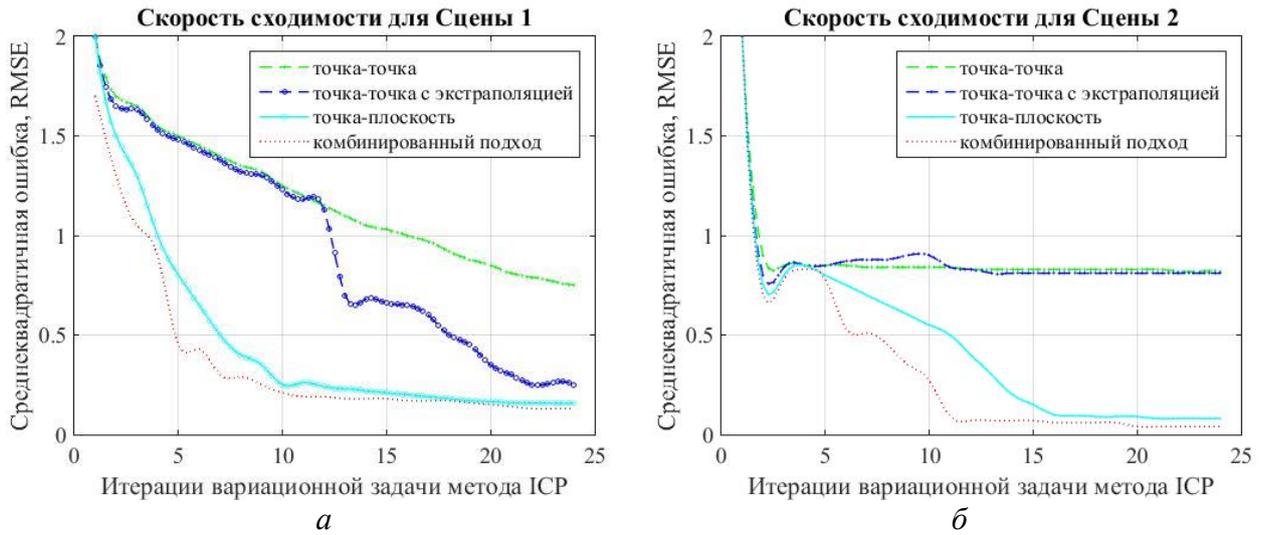


Рис. 2 Сравнение скорости сходимости в зависимости от ошибки метрики и условий наблюдения:
a – сходимость в контролируемых условиях; *б* – сходимость в неконтролируемых условиях.

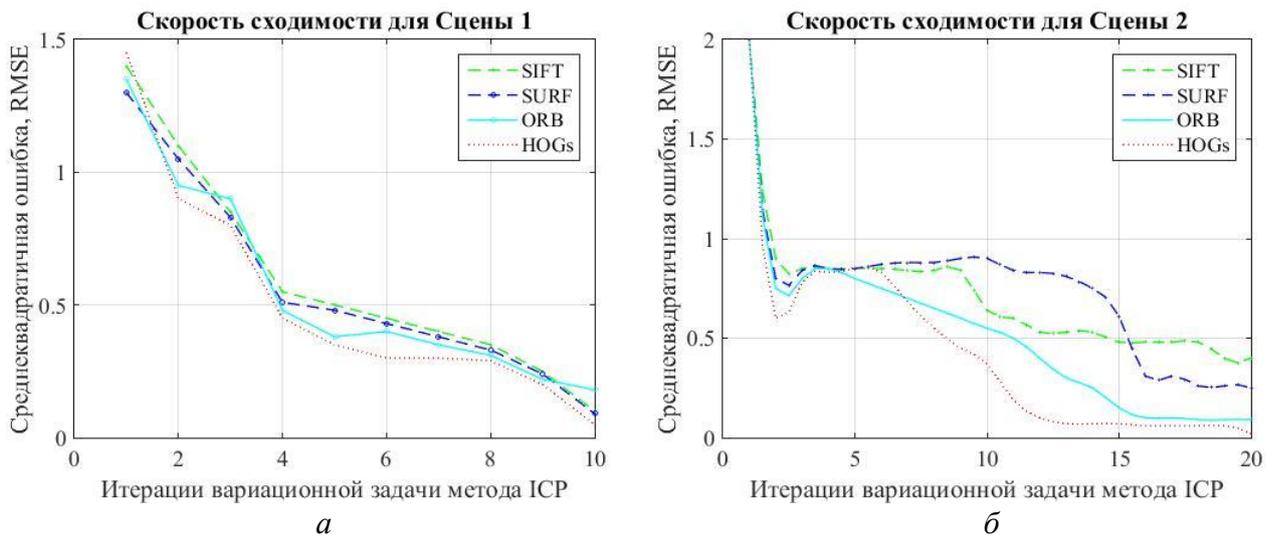


Рис. 3 Сравнение скорости сходимости в зависимости от типа дескриптора и условий наблюдения:
a – сходимость в контролируемых условиях; *б* – сходимость в неконтролируемых условиях.

Установлено, что в контролируемых условиях тип используемого дескриптора имеет ограниченное влияние на сходимость метода регистрации данных (рисунок 3, *a*), в неконтролируемых условиях использование дескриптора ДЛ ГНГ позволяет получить лучшую сходимость в сравнении с другими дескрипторными методами: предложенный метод регистрации сходится уже после 11 итераций, тогда как при использовании дескриптора ORB метод сходится только после 16 итераций (рисунок 3, *б*). Известно, что метрика «точка – плоскость» превосходит метрику «точка – точка» по точности и скорости сходимости. Решение вариационной задачи минимизации функционала в замкнутой форме как условной вариационной задачи с ограничением на принадлежность решения классу ортогональных матриц приводит к необходимости работать с существенно нелинейным подмножеством ортогональных матриц многообразия матриц линейных преобразований [5]

$$J(R) = \sum_{i=1}^n (\langle Rx^i - y^i n^i \rangle)^2, \quad (16)$$

где \langle , \rangle есть скалярное произведение, R – матрица аффинного преобразования в однородных координатах, x^i есть точка в облаке точек X , y^i есть точка в облаке точек Y , n^i есть унитарная нормаль к касательной плоскости $T(y^i)$.

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad x^i = \begin{pmatrix} x_1^i \\ x_2^i \\ x_3^i \\ 1 \end{pmatrix}, \quad y^i = \begin{pmatrix} y_1^i \\ y_2^i \\ y_3^i \\ 1 \end{pmatrix}, \quad n^i = \begin{pmatrix} n_1^i \\ n_2^i \\ n_3^i \\ 1 \end{pmatrix}. \quad (17)$$

Рассмотрим функцию $J(R)$

$$\begin{aligned} J(R) &= \sum_{i=1}^n (\langle Rx^i - y^i, n^i \rangle)^2 = \sum_{i=1}^n (\langle Rx^i, n^i \rangle - \langle y^i, n^i \rangle)^2 = \\ &= \sum_{i=1}^n (\langle Rx^i, n^i \rangle)^2 - 2\langle Rx^i, n^i \rangle \langle y^i, n^i \rangle + (\langle y^i, n^i \rangle)^2. \end{aligned} \quad (18)$$

Слагаемое $\langle y^i, n^i \rangle$ является постоянным относительно R , следовательно, вариационная задача (18) может быть представлена в следующем виде

$$\sum_{i=1}^n (\langle Rx^i, n^i \rangle)^2 - 2\langle Rx^i, n^i \rangle \langle y^i, n^i \rangle + (\langle y^i, n^i \rangle)^2 \rightarrow \min_R. \quad (19)$$

Рассмотрим компоненты градиента относительно переменной t

$$\frac{\partial J}{\partial t_j} = \sum_{i=1}^n 2\langle Rx^i, n^i \rangle n_j^i - 2n_j^i \langle y^i, n^i \rangle = 0, \quad i, j = 1, \dots, 3. \quad (20)$$

Аналогично найдем частные производны $J(R)$ относительно r_{ij}

$$\frac{\partial J}{\partial r_{ij}} = \sum_{k=1}^n 2\langle Rx^k, n^k \rangle x_j^k n_i^k - 2x_j^k n_i^k \langle y^k, n^k \rangle = 0. \quad (21)$$

Рассмотрим комбинированный метод решения вариационной задачи «точка – плоскость» в замкнутой форме для ортогональных преобразований с использованием визуально связанных характеристик [6]

$$\begin{aligned} J(RV, RD) &= \alpha \frac{1}{W} \frac{1}{|A_f|} \sum_{i=1}^n w_i (\langle RV * F_x^i - F_y^i n^i \rangle)^2 + \\ &+ (1 - \alpha) \frac{1}{W} \frac{1}{|A_d|} \sum_{i=1}^n w_i (\langle RD * x^i - y^i n^i \rangle)^2. \end{aligned} \quad (22)$$

Обозначим $\eta_1 = \alpha \frac{1}{W} \frac{1}{|A_f|}$ и $\eta_2 = (1 - \alpha) \frac{1}{W} \frac{1}{|A_d|}$. Представим функцию $J(RV, RD)$ в следующем виде:

$$\begin{aligned} \eta_1 \sum_{i=1}^n (\langle RV F_x^i, n^i \rangle)^2 - 2\langle RV F_x^i, n^i \rangle \langle F_y^i, n^i \rangle + \eta_2 \sum_{i=1}^n (\langle RD x^i, n^i \rangle)^2 - \\ - 2\langle RD x^i, n^i \rangle \langle y^i, n^i \rangle \rightarrow \min_{RV, RD}. \end{aligned} \quad (23)$$

Рассмотрим компоненты градиента относительно переменных t и b :

$$\frac{\partial J}{\partial t_j} = \sum_{i=1}^n 2 \langle RD x^i, n^i \rangle n_j^i - 2 n_j^i \langle y^i, n^i \rangle = 0, \quad j = 1, 2, 3, \quad (24)$$

$$\frac{\partial J}{\partial b_j} = \sum_{i=1}^n 2 \langle RV F_x^i, n^i \rangle n_j^i - 2 n_j^i \langle F_y^i, n^i \rangle = 0, \quad j = 1, 2, 3. \quad (25)$$

Используем свойство трассировки матрицы, тогда получим:

$$\sum_{i=1}^n Tr \left(RD \cdot (n_j XN)^i \right) = Tr \left(RD \left(\sum_{i=1}^n (n_j XN)^i \right) \right) = \alpha_j, \quad (26)$$

$$\sum_{i=1}^n Tr \left(RV \cdot (n_j FXN)^i \right) = Tr \left(RV \left(\sum_{i=1}^n (n_j FXN)^i \right) \right) = \beta_j. \quad (27)$$

Обозначим через SV^j и SD^j , $j = 1, 2, 3$ следующие матрицы:

$$\sum_{i=1}^n (n_j XN)^i = SD^j, \quad \sum_{i=1}^n (n_j FXN)^i = SV^j. \quad (28)$$

Найдем частные производные функционала $J(RD)$ относительно r_{ij} , f_{ij} :

$$\frac{\partial J}{\partial r_{ij}} = \sum_{k=1}^n 2 \langle RD x^k, n^k \rangle x_j^k n_i^k - 2 x_j^k n_i^k \langle y^k, n^k \rangle = 0, \quad (29)$$

$$\frac{\partial J}{\partial f_{ij}} = \sum_{k=1}^n 2 \langle RV F_x^k, n^k \rangle F_{jx}^k n_i^k - 2 F_{jx}^k n_i^k \langle F_y^k, n^k \rangle = 0. \quad (30)$$

Частная сумма $\sum_{k=1}^n x_j^k n_i^k \langle y^k, n^k \rangle$ есть константа относительно RD , а частная сумма $\sum_{i=1}^n F_{jx}^k n_i^k \langle F_y^k, n^k \rangle$ – относительно RV . Пусть $\sum_{k=1}^n x_j^k n_i^k \langle y^k, n^k \rangle = \alpha_{ij}$ и $\sum_{k=1}^n F_{jx}^k n_i^k \langle F_y^k, n^k \rangle = \beta_{ij}$. Представим решение комбинированной вариационной задачи для ортогональных преобразований в матричной форме:

$$\begin{cases} Tr(RD \cdot SD^j) = \alpha_j, \\ Tr(RV \cdot SV^j) = \beta_j, \\ \sum_{k=1}^n Tr(RD \cdot SD^{kj}) = \alpha_{ij} \Rightarrow Z \cdot \Omega = \theta, \\ \sum_{k=1}^n Tr(RV \cdot SV^{kj}) = \beta_{ij}, \end{cases} \quad (31)$$

где матрица Z имеет размер 18×24 и состоит из элементов SD^{ij} и SV^{ij} ; θ есть матрица из 24 столбцов α_j , β_j , α_{ij} и β_{ij} . Комбинированный метод для ортогональных преобразований состоит из шагов:

1. Вычисление ГНГ на изображениях.
2. Сопоставление между особыми точками F_x^i и F_y^i для выбранных подмножеств. Решение вариационной задачи регистрации данных для визуально связанных характеристик изображения. Получим RT^* и A_f .
3. Проверка: Если $|A_f| < \delta$, то $RT^* = T_{km}$ и $A_f = \emptyset$. Положим $i = 1$.
4. Определение соответствующих точек A_d в исходном и целевом облаке точек с использованием метода ближайших соседей. Определение весовых коэффициентов для каждой пары A_d .

5. Рассмотрим компоненты градиента относительно переменных t и b (векторы переноса) и r_{ij}, f_{ij} (матрицы поворота) соответственно. Получим проекцию элемента многообразия матриц линейных преобразований на подмногообразии ортогональных матриц $F_{jx}^k n_i^k FXN^k$, $x_j^k n_i^k XN^k$ и $n_j^i XN$, $n_j^i FXN$.

6. Решение комбинированной вариационной задачи относительно RT^* , A_f и A_d .

7. Проверка: $(E(RT^*_i) - E(RT^*_{i+1}) < \varepsilon)$ или (Номер итерации $> k_{\max}$). Если условие не верно, то возврат на Шаг 4 и $i = i + 1$, иначе получено искомое преобразование RT^* .

В результате серии тестов установлены зависимости скорости сходимости методов от выбора ошибки метрики и условий проведения экспериментов (рисунок 4). Был проведен сравнительный анализ зависимости скорости сходимости методов регистрации для ортогональных преобразований от выбора типа дескриптора и условий проведения экспериментов (рисунок 5).

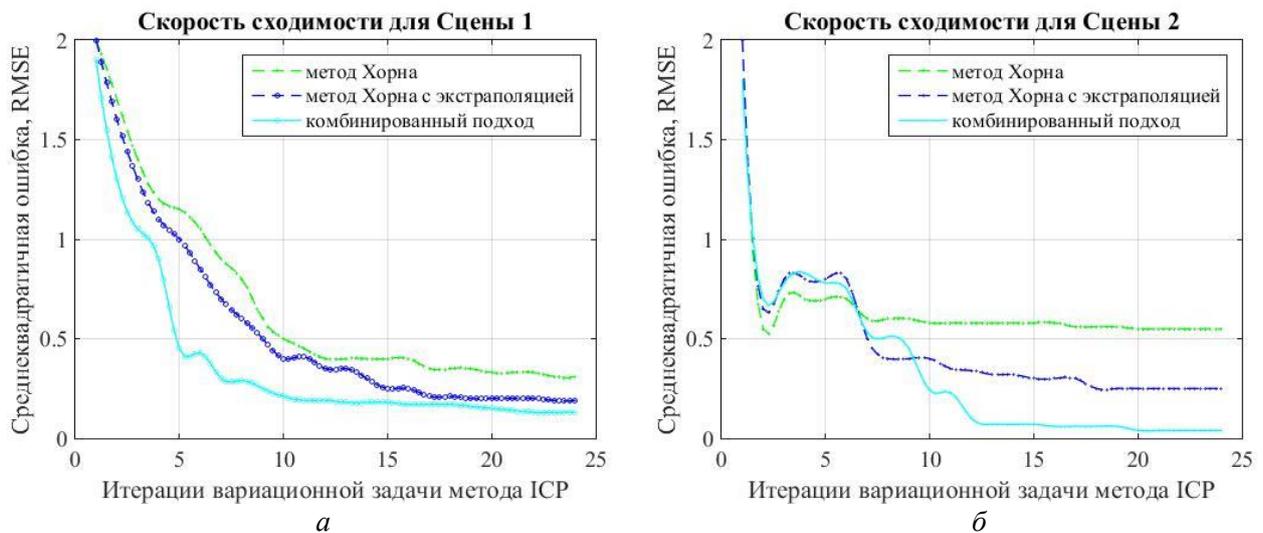


Рис. 4 Исследование сходимости в зависимости от ошибки метрики и условий наблюдения:

a – скорость сходимости в контролируемых условиях;

$б$ – скорость сходимости в неконтролируемых условиях.

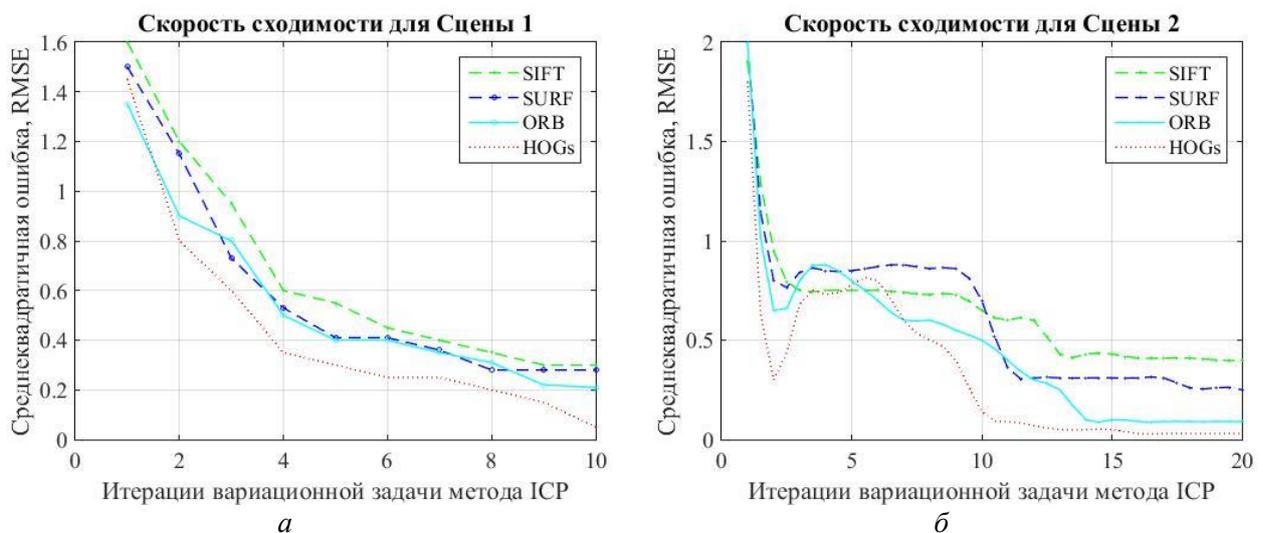


Рис. 5 Исследование сходимости в зависимости

от выбора типа дескриптора и условий наблюдения:

a – скорость сходимости в контролируемых условиях;

$б$ – скорость сходимости в неконтролируемых условиях.

Было установлено, что в контролируемых условиях тип используемого дескриптора имеет ограниченное влияние на сходимость метода регистрации данных (рисунок 5, а), в неконтролируемых условиях использование дескриптора ДЛ ГНГ позволяет получить лучшую сходимость и точность в сравнении с другими дескрипторными методами: предложенный метод регистрации сходится уже после 10 итераций, тогда как при использовании дескриптора ORB предложенный метод регистрации сходится только после 14 итераций (рисунок 5, б). Проведенные вычислительные эксперименты показали значительное повышение точности предложенного метода ICP для разного класса сцен.

МЕТОД ОПРЕДЕЛЕНИЯ ДИНАМИЧЕСКОГО ПОЛОЖЕНИЯ МОБИЛЬНОЙ ПЛАТФОРМЫ

Предложен метод определения динамического положения мобильной платформы (S-EKF SLAM) на основе расширенной калмановской фильтрации (EKF), истории положений мобильной платформы, семантических ориентиров и совмещении полученных трехмерных карт [7]. Входной информацией для каждой итерации метода S-EKF SLAM является множество неточных измерений положений платформы на сцене для некоторого дискретного момента времени k , выходной информацией для каждой итерации является последовательность оценок положений мобильной платформы на сцене.

В исследовании используются детектор объектов (YOLO) и семантический подход сегментации, основанный на 3D-запутанных лесах (3-D Entangled Forests 3DEF) для определения принадлежности объектов к классам предметной области (семантической маркировки окружающего пространства). Метод 3DEF позволяет точно сегментировать многие структурные элементы сцены и другие объекты, но плохо классифицирует плоские классы объектов (различные малые формы интерьера). Для преодоления данного недостатка в работе используется детектор объектов YOLO. Для интеграции детектора объектов YOLO и метода 3DEF необходимо: выделить объект в кадре при помощи детектора; внутри ограничительной рамки детектора провести процедуру сегментации объекта вдоль его контуров на основе метода Grabcut. Объединение результатов семантической сегментации метода 3DEF и комбинации методов YOLOv3 + Grabcut осуществляется с использованием слияния на основе байсевского подхода.

Для каждого шага метода S-EKF SLAM (рисунок 6) определяется положение мобильной платформы и вычисляется оценка ее следующей позиции. Каждый кадр содержит информацию о местоположении особых точек в пространстве, относительно которых можно вычислить положение мобильной платформы на сцене. Далее определяется значение дельта-функции, которая является разностью полученных оценок положения платформы (реальных и прогнозируемых), затем на основе значения дельта-функции обновляются веса особых точек в RGB-D кадре, производится коррекция положения и траектории движения мобильной платформы на сцене. 3D карта представляет собой вектор состояния большой размерности положений мобильной платформы и навигационных ориентиров: $x = [R \ M]^T$, где R – положение платформы и $M = \{L_1, \dots, L_n\}$ – множество навигационных ориентиров, n – текущее количество отображаемых ориентиров. В EKF SLAM данное отображение моделируется гауссовой переменной с использованием матрицы средних значений и матрицы ковариации вектора состояний и обозначается соответственно как \bar{x} и C :

$$\bar{x} = \begin{bmatrix} \bar{R} \\ \bar{M} \end{bmatrix} = \begin{bmatrix} \bar{R} \\ \bar{L}_1 \\ \vdots \\ \bar{L}_N \end{bmatrix}, \quad C = \begin{bmatrix} C_{R|R} & C_{R|M} \end{bmatrix} = \begin{bmatrix} C_{R|R} & C_{R|L_1} & \dots & C_{R|L_N} \\ C_{L_1|R} & C_{L_1|L_1} & \dots & C_{L_1|L_n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{L_N|R} & C_{L_n|L_1} & \dots & C_{L_n|L_n} \end{bmatrix}. \quad (32)$$

В предложенном методе S-EKF SLAM большая часть трехмерной карты инвариантна относительно движения платформы, поэтому можно представить разреженные матрицы Якоби со следующей структурой:

$$F_x = \begin{bmatrix} \frac{dF_R}{dR} & 0 \\ 0 & I \end{bmatrix}, F_y = \begin{bmatrix} \frac{dF_R}{dn} \\ 0 \end{bmatrix}. \quad (33)$$

Получим разреженные уравнения для прогнозирования траектории движения платформы в пространстве с помощью расширенного фильтра Калмана:

$$\begin{aligned} \bar{R} &\leftarrow f_R(\bar{R}, u, 0), C_{R|R} \leftarrow \frac{dF_R}{dR} C_{R|R} \frac{dF_R^T}{dR} + \frac{dF_R}{dn} N \frac{dF_R^T}{dn}, \\ C_{R|M} &\leftarrow \frac{dF_R}{dR} C_{R|M}, C_{M|R} \leftarrow C_{R|M}. \end{aligned} \quad (34)$$

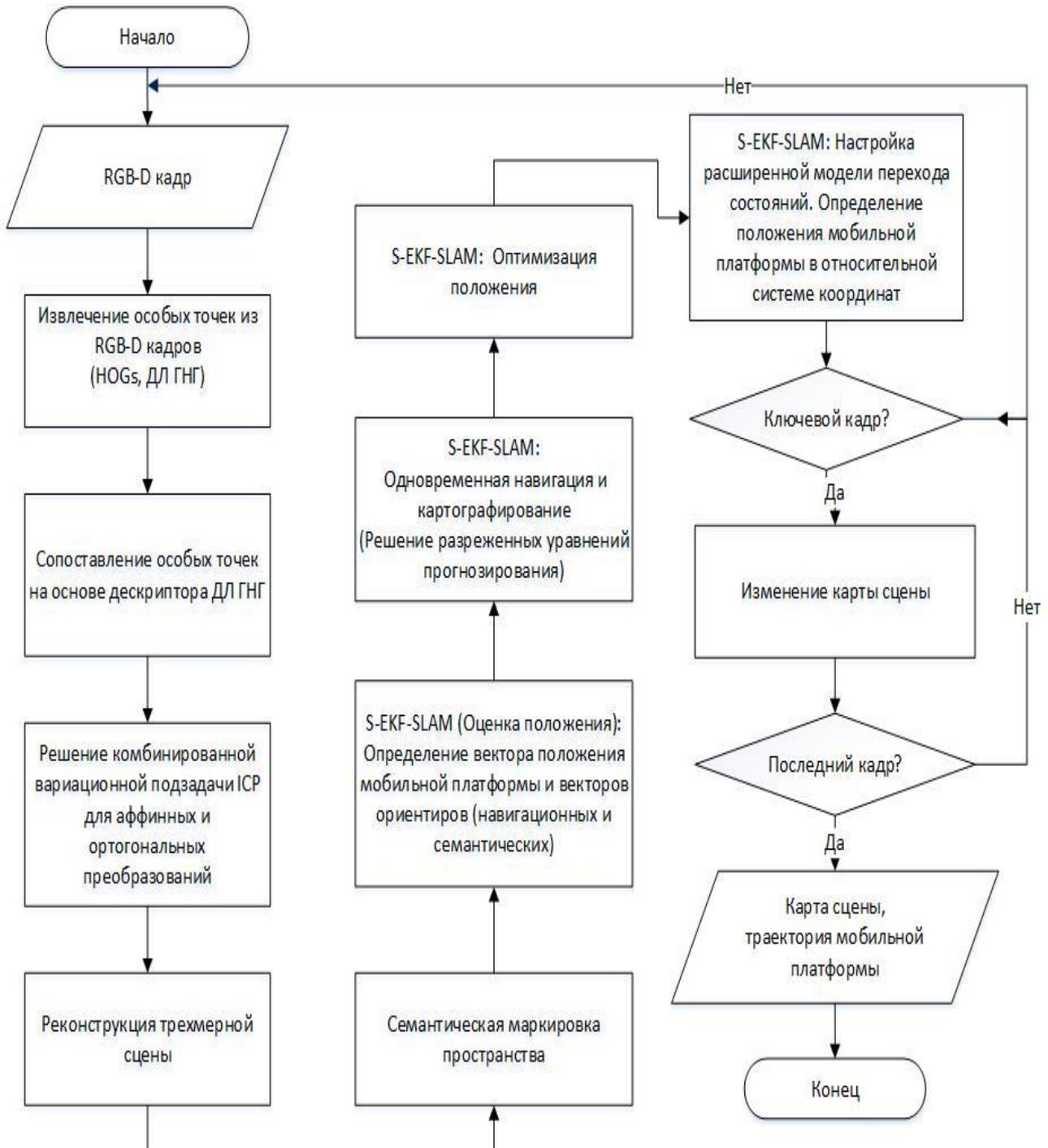


Рис. 6 Общая блок-схема метода S-EKF SLAM.

Состояние исследуемой системы в каждый момент времени состоит из данных о положении и ориентации платформы, а также данных о положении всех семантических ориентиров относительно платформы. Состояние платформы на временном шаге k обозначается $x_v(k)$. Движение платформы моделируется обычным линейным уравнением перехода дискретно-временного состояния или моделью процесса вида $p_i(k+1) = p_i(k) = p_i$. Функцию перехода состояния для i -го семантического ориентира можно представить как $s_i(k+1) = s_i(k) = s_i$. Пусть вектор всех N семантических ориентиров есть $s = [s_1^T \dots s_N^T]^T$, тогда расширенный вектор состояния представим в виде $x(k) = [x_v^T(k) \ p_1^T \dots \ p_N^T]^T$. Для получения оценки вектора состояния при движении платформы по серии зашумленных измерений и семантических ориентиров модель данного процесса представлена в виде матричного уравнения:

$$\begin{bmatrix} x_v(k+1) \\ p_1 \\ \vdots \\ p_N \end{bmatrix} = \begin{bmatrix} JT_v(k+1) & 0 & \dots & 0 \\ 0 & L_{sp_1} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & L_{sp_N} \end{bmatrix} \text{dist} \left(\begin{bmatrix} x_v(k) \\ p_1 \\ \vdots \\ p_M \end{bmatrix}, \begin{bmatrix} x_v(k) \\ s_1 \\ \vdots \\ s_N \end{bmatrix} \right) + \begin{bmatrix} U_v(k+1) \\ O_{p_1} \\ \vdots \\ O_{p_N} \end{bmatrix} + \begin{bmatrix} V_v(k+1) \\ O_{p_1} \\ \vdots \\ O_{p_N} \end{bmatrix}, \quad (35)$$

где L_{spi} есть $\dim(p_i) \times \dim(p_i)$ – единичная матрица и O_{pi} есть $\dim(p_i)$ – нуль-вектор, dist – функция, которая формирует новый вектор навигационных ориентиров p размерностью N из вектора визуальных ориентиров p размерностью M и вектора семантических ориентиров s . Обновление истории перемещений мобильной платформы происходит не на каждом шаге, а если положение платформы существенно изменилось. В EKF SLAM вектор состояния имеет большую размерность, что затрудняет использование методологии для крупномасштабных сцен. Использование семантических ориентиров позволяет сократить размерность вектора состояния при решении проблемы локализации. В работе были проведены компьютерное моделирование для эталонных баз данных и вычислительные эксперименты [10, 11]. Оценка точности предложенного метода S-EKF SLAM для навигации мобильной платформы проводилась в серии экспериментов: первый эксперимент – RGBD-D камера переносилась человеком в направлении движения; второй эксперимент – камера Kinect 2.0 и камера MESA SR4000 были установлены на мобильной платформе Husky UGV. Платформы с аналогами систем для тестирования были взяты с сайта openslam.org и были протестированы в различных условиях. На рисунке 7 представлены результаты второго вычислительного эксперимента.

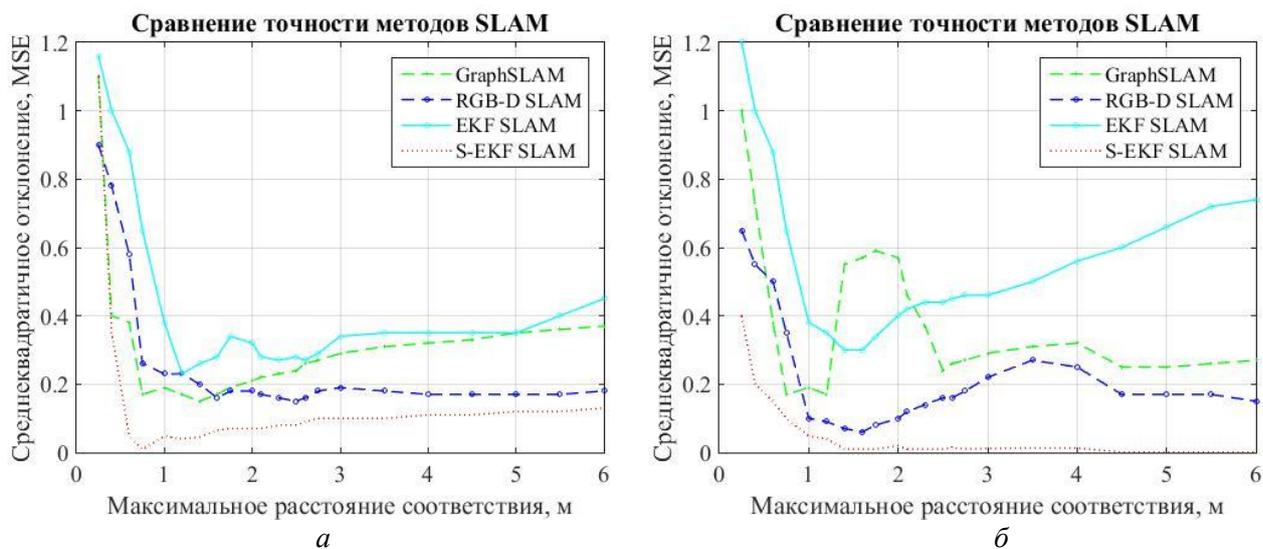


Рис. 7 Результаты компьютерного моделирования методов SLAM в контролируемых (а) и в неконтролируемых (б) условиях.

Было установлено, что при отсутствии различных шумов и хорошем освещении предложенный метод S-EKF SLAM показывает похожие по точности результаты с классическим

Visual SLAM, но при этом предложенный метод имеет лучшую сходимость (рисунок 7, а). В неконтролируемых условиях (рисунок 7, б) метод показывает лучшие по точности результаты, чем известные подходы.

ОПТИМИЗАЦИЯ ГРАФА ПОЛОЖЕНИЙ МОБИЛЬНОЙ ПЛАТФОРМЫ

Для повышения сходимости предложенной методологии SLAM необходимо исключить из полученного решения (траектории движения) участки с «замыканием цикла». Предложен подход к оптимизации графа положений мобильной платформы и обнаружению «замыканий цикла» с использованием метода корзина слов и предложенного двумерного дескриптора на основе ГНГ [8]. В данном решении для каждого кадра устанавливается в соответствие предложенный двумерный дескриптор. Для определения «замыкания цикла» на каждом шаге сохраняется история RGB-D кадров и проверяется, совпадает ли текущий кадр с некоторыми из предыдущих кадров. Если текущий кадр является визуально подобным [9, 10] с другим кадром, замеченным в прошлом, то вычисляется преобразование между этими кадрами, и в граф положений добавляется новое ограничение из данного кадра. С помощью новых полученных ограничений положений кумулятивная ошибка графа положений может быть значительно уменьшена. Оптимизация графа положений приводит к глобально выровненному набору кадров. Метод уравнивания связей (bundle adjustment) одновременно оптимизирует граф положений и трехмерную плотную карту. Использование процедуры «замыкания циклов», в общем случае, позволяет существенно улучшить точность локализации мобильной платформы на сцене для длинных циклов (таблицы 1 и 2), для коротких циклов применение процедуры имеет ограниченное влияние, что подтверждается результатами экспериментов.

Таблица 1

**Точность локализации методов SLAM
без использования процедуры «замыкания цикла» (для длинных циклов)**

Название метода	Среднее отклонение координат, мм	Средне-квадратичное отклонение координат, мм	Среднее отклонение ориентации, град	Средне-квадратичное отклонение ориентации, град
EKF SLAM	215,15	296,98	2,2	1,5
Graph SLAM	629,13	670,33	2,87	1,74
S-EKF SLAM	120,84	198,08	1,71	1,04

Таблица 2

**Точность локализации методов SLAM
с использованием процедуры «замыкания цикла» (для длинных циклов)**

Название метода	Среднее отклонение координат, мм	Средне-квадратичное отклонение координат, мм	Среднее отклонение ориентации, град	Средне-квадратичное отклонение ориентации, град
EKF SLAM	14,35	7,1	0,3	0,12
Graph SLAM	45,81	30,44	0,65	0,28
S-EKF SLAM	4,05	1,81	0,14	0,09

Предложенный метод S-EKF SLAM на основе комбинирования визуальных и семантических характеристик окружающей среды имеет преимущества с точки зрения точности локализации перед известными методами SLAM на основе EKF. Для повышения производительности подхода в работе используется метод поиска изображений по визуальному подобию с применением инвертированных индексов (ИИ) цветовых гистограмм.

ПРАКТИЧЕСКИЕ ПРИЛОЖЕНИЯ МЕТОДОЛОГИИ

Укажем практические приложения предложенной методологии для решения проблемы SLAM, рассмотрено применение полученных результатов в системах спасения людей. Система спасения людей состоит из следующих компонент: система одновременной навигации и картографирования; система некооперативного распознавания личности по лицу; система планирования траектории мобильной платформы и позиционирования в неизвестном окружающем пространстве. Предложены новые методы и подходы: метод дистанционного распознавания личности по лицу с использованием автокодировщика; метод обратной индексации для поиска кандидатов по эталону для распознавания личности по лицу.

ЗАКЛЮЧЕНИЕ

Проведенные исследования позволили сформулировать следующие основные выводы и получить практические результаты:

1. Методология для решения задачи одновременной навигации и построения карты, основанная на комбинировании визуальных и семантических характеристик неизвестной окружающей среды, позволяет решать проблему SLAM для контекстуально сложных крупномасштабных сцен в реальном масштабе времени. Предлагаемая методология в отличие от известных использует мультисенсорную визуальную и семантическую информацию на всех этапах решения проблемы SLAM для повышения точности и сходимости получаемого решения. Реализация задач выполнена посредством прорывных научных исследований и разработок мирового уровня в области информационных технологий, предложенная комбинированная методология может быть использована для создания систем навигации и машинного зрения в современных автономных интеллектуальных робототехнических комплексах и системах.

2. Алгоритм представления окрестности особых точек в виде вектора признаков на основе рекурсивного вычисления гистограмм направленных градиентов по нескольким круглым скользящим окнам и пирамидальному разложению изображения позволяет выполнять быстрое сопоставление изображений, инвариантное к повороту и небольшому изменению масштаба для крупномасштабных сцен.

3. Комбинированные методы решения вариационной задачи с метриками «точка – точка» и «точка – плоскость» на основе метода ICP для аффинных и ортогональных преобразований позволяют решить важную проблему цифровой обработки информации и робототехники, связанную с реконструкцией точных трехмерных моделей окружающего пространства во время движения автономного интеллектуального объекта управления. Данные о визуально связанных характеристиках были использованы при решении вариационной задачи и позволяют решить проблему зависимости результата решения вариационной задачи от правильности выбора начальных значений. Предложенный метод регистрации данных для аффинных преобразований создает основу для применения метода Хорна на случай с не ригидными объектами на сцене. Предложены решения вариационной задачи «точка – точка» для вырожденных случаев, позволяющие проводить точную регистрацию структурных элементов сцены. Некоторые результаты, касающиеся использования информации о характерных точках для решения вариационной задачи «точка – плоскость», были известны и ранее, однако явные формулы для совместного решения вариационной задачи в замкнутой форме в поставленной постановке проблемы для решения проблемы SLAM были получены впервые. Комбинированные методы регистрации используются для регистрации облаков точек с произвольным пространственным разрешением и масштабом относительно друг друга. Применение визуально связанных характеристик для решения вариационной задачи «точка – точка» и «точка – плоскость» позволяет улучшить точность и сходимость методов регистрации данных в контролируемых и неконтролируемых условиях.

4. Метод решения задачи навигации и определения положения мобильной платформы на сцене, основанный на комбинировании визуальных и семантических характеристик неизвестной окружающей среды и расширенной калмановской фильтрации (ЕКФ), имеет хорошую сходимос ть и может использоваться в приложениях, работающих в реальном масштабе времени в контролируемых и неконтролируемых условиях. Получены формулы для оценки вектора состояния при движении мобильной платформы по серии зашумленных измерений и семантических навигационных ориентиров. В предлагаемом методе комбинирование информации происходит на уровне выделенных семантических и мультимодальных характеристик, сходимос ть предложенного метода улучшается за счет использования семантических ориентиров в структуре расширенного фильтра Калмана.

5. Быстрый рекурсивный алгоритм для вычисления ДКП дискретного сигнала в скользящем окне с произвольным шагом основан на рекурсивном соотношении между тремя последовательными локальными спектрами ДКП, которые вычисляются в равноудаленных положениях окна. В работе приведены граничные значения для шага скользящего окна, при которых предлагаемый алгоритм является более эффективным, чем быстрые алгоритмы ДКП. Данный алгоритм используется для локальной пространственно-неоднородной обработки сигналов в спектральной области с произвольным пространственным разрешением.

6. Подход к оптимизации графа положений мобильной платформы и обнаружению «закрывающихся цикла» на основе метода корзина слов и двумерного дескриптора на основе гистограммы ориентированных градиентов позволяет улучшить сходимос ть алгоритма обнаружения «закрывающихся цикла» для длинных циклов; для коротких циклов применение подхода имеет ограниченное влияние. Предложенные методы, алгоритмы и программно-технические решения позволили создать прототип системы одновременной навигации и построения карты лучшего качества, чем известные системы на основе визуальной методологии SLAM и методологии на основе ЕКФ, работающие в аналогичных условиях. Разработанный прототип системы апробирован при решении задачи спасения людей и может быть использован для решения широкого спектра задач: распознавание и слежение за динамическими объектами в заданном пространстве, предсказание траектории их перемещения, согласованное и оптимизированное планирование движений мобильных платформ в заданном пространстве с учетом возможных статических и динамических препятствий.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Vokhmintsev A., Yakovlev K. "A real-time algorithm for mobile robot mapping based on rotation-invariant descriptors and ICP" // *Communications in Computer and Information Science*. 2016. Vol. 661. Pp. 338-350.
2. Vokhmintsev A., Makovetskii A., Kober V., Sochenkov I., Kuznetsov V. "A fusion algorithm for building three-dimensional maps" // In: *Proceedings of the XXXVIII SPIE's Annual Meeting: Applications of Digital Image Processing (San Diego, USA, August 2015)*. 2015. Vol. 9599. Pp. 959929-1.
3. Vokhmintsev A. V., Sochenkov I. V., Kuznetsov V. V., Tikhonkikh D. V. "Face recognition based on matching algorithm with recursive calculation of local oriented gradient histogram" // *Doklady Mathematics*. 2016. Vol. 466 (3). Pp. 453-459.
4. Вохминцев А. В., Мельников А. В., Пачганов С. В. Комбинированные методы навигации и составления карты на основе решения вариационной задачи точка-плоскость ICP для аффинных преобразований в трехмерном пространстве // *Информатика и ее применения*. 2020. Т. 14 (1). С. 101-112. [[Vokhmintsev A. V., Melnikov A. V., Pachganov S. V. "Combined methods of navigation and mapping based on the solution of the variational point-plane problem ICP for affine transformations in three-dimensional space" // *Informatics and its Applications*. 2020. Vol. 14 (1), pp. 101-112. (In Russian).]]
5. Vokhmintsev A. V., Melnikov A. V., Mironov K. V., Burlutskiy V. V. "Reconstruction of three-dimensional map based on closed form solution of variational problem of multi-sensor data registration" // *Doklady Mathematics*. 2019. Vol. 99 (1). Pp. 1-5.
6. Вохминцев А. В. Решение вариационной задачи точка-плоскость ICP на основе комбинирования визуальных и семантических характеристик трехмерной сцены // *Труды Института системного анализа РАН*. 2020. Т. 70 (1). С. 3-14. [[Vokhmintsev A. V. "Solution of the variational point-plane problem ICP based on the combination of visual and semantic characteristics of a three-dimensional scene" // in: *Proceedings of the Institute of System Analysis of the Russian Academy of Sciences*. 2020. V. 70 (1), pp. 3-14. (In Russian).]]
7. Vokhmintsev A., Botova T., Sochenkov I., Sochenkova A., Makovetskii A. "Robot mapping algorithm based on Kalman filtering and symbolic tags" // *Proceedings SPIE of the XL Applications of Digital Image Processing (San Diego, USA, August 2017)*. Vol. 10396. Pp. 103962I-7.

8. Vokhmintsev A., Timchenko M., Karyakina A. "Real-time visual loop-closure detection using fused iterative close point algorithm and extended Kalman filter" // IEEE Proceedings of the 3rd International Conference on Industrial Engineering, Applications and Manufacturing, ICIEAM, (Saint-Petersburg, Russia, May 2017). 2017. Pp. 1-6.

9. Соченков И. В., Вохминцев А. В., Кузнецов В. В., Хромов Н. А., Григорьева К. С. Поиск изображений по визуальному подобию с применением инвертированных индексов цветовых гистограмм // Информационные технологии и вычислительные системы. 2015. № 4. С. 86-94. [[Sochenkov I. V., Vokhmintsev A. V., Kuznetsov V. V., Khromov N. A., Grigoryeva K. S. "Image search by visual similarity using inverted color histogram indices" // Information Technologies and Computing Systems. 2015. No. 4, pp. 86-94. (In Russian).]]

10. Соченков И., Соченкова А., Вохминцев А. Метод обратной индексации для поиска лиц по эталону // Искусственный интеллект и принятие решений. 2017. № 1. С. 50-58. [[Sochenkov I., Sochenkova A., Vokhmintsev A. "Reverse indexing method for face search by reference" // Artificial Intelligence and Decision Making. 2017. No. 1, pp. 50-58. (In Russian).]]

11. Vokhmintsev A., Timchenko M., Melnikov A., Kozko A., Makovetskii A. "Robot path planning algorithm based on symbolic tags in dynamic environment" // Proceedings of the XL SPIE Applications of Digital Image Processing (San Diego, USA, August 2017). 2017. Vol. 10396. Pp. 103962E1-9.

Поступила в редакцию 28 августа 2023 г.

МЕТАДААННЫЕ / METADATA

Title: Methodology for solving the problem of simultaneous navigation and mapping based on a combination of visual and semantic characteristics of the environment.

Abstract: A review of the results of the study of the problem of simultaneous navigation and mapping of the environment is presented. The aim of the study is to improve the efficiency of information processing processes in methods and algorithms for the reconstruction of three-dimensional scenes of the surrounding space, visualization of three-dimensional models and navigation of mobile platforms in an unknown environment based on combining information about the semantic characteristics of the environment and available multi-sensor information obtained from various types of sensors. To achieve the goal, the following tasks are solved. Development of a methodology for the exact solution of the SLAM problem, based on a combination of visual and semantic characteristics of an unknown environment. Development of an algorithm for searching and matching singular points based on two-dimensional images for processing color information in a frame. Development of accurate combined methods for the reconstruction of contextually complex large-scale three-dimensional scenes based on an iterative algorithm of nearest points for a group of affine transformations and for a group of orthogonal transformations. Development of a method for solving the problem of navigation and determining the position of a mobile platform on the stage, based on a combination of visual and semantic characteristics of an unknown environment and extended Kalman filtering. Development of a fast recursive algorithm for calculating the discrete cosine transform (DCT) for solving the problems of compressing and restoring images obtained from multiple sensors on a mobile platform. Development of an approach to optimizing the position graph of a mobile platform and detecting "loop closures" based on the word basket method and combining the visual and semantic characteristics of an unknown environment. Approbation of the proposed combined visual system SLAM for solving the problem of saving people. The object of research is a system for processing visual and semantic information about an unknown environment to solve the problem of reconstructing contextually complex large-scale three-dimensional scenes and the problem of determining the position of a mobile platform on the stage at any given time. The subject of the study are methods, algorithms for solving the problem of simultaneous navigation and map building based on a combination of information about the semantic and visual characteristics of the environment.

Key words: Simultaneous Localization and Mapping; data registration; Iterative Closest Point; Horn method; matching singular points; loop closure problem, extended Kalman filter; EKF-SLAM; Visual SLAM.

Язык статьи / Language: русский / Russian.

Об авторе / About the author:

ВОХМИНЦЕВ Александр Владиславович

ФГБОУ ВО «Челябинский государственный университет»,
ФГБОУ ВО «Югорский государственный университет»,
Россия.

Зав. науч.-иссл. лабораторией интеллектуальных информационных технологий и систем, проф. каф. информационных технологий и экономической информатики. Профессор инженерной школы цифровых технологий. Дипл. инж. по вычисл. машинам, комплексам, системам и сетям (Ю.-Уральск. гос. ун-т, 2000). Д-р техн. наук по сист. анализу, управлению и обр-ке информации (Уфимск. гос. авиац. техн. ун-т, 2020). Иссл. в обл. цифр. обр. информации, сопоставления и регистрации изображений, одновр. навигации и составления карты в неизв. пространстве, распознаван. личности по лицу.

E-mail: vav2000@inbox.ru

ORCID: <https://orcid.org/0000-0002-2402-2963>

URL: https://elibrary.ru/author_profile.asp?authorid=291824

VOKHMINTCEV Aleksandr Vladislavovich

Chelyabinsk State University,
Ugra State University,
Russia.

Head Research Laboratory of Intelligent Information Technologies and Systems, Professor of the Department of Information Technologies, and Economic Informatics. Professor at the Engineering School of Digital Technologies. Dipl. Eng. on computers, complexes, systems, and networks (South Ural State University, 2000). Dr. Tech. Sciences in system analysis, management, and information processing (Ufa State Aviation Technical University, 2020). Research in the field of digital information processing, comparison and registration of images, simultaneous navigation and mapping in an unknown space, personality recognition.

E-mail: vav2000@inbox.ru

ORCID: <https://orcid.org/0000-0002-2402-2963>

URL: https://elibrary.ru/author_profile.asp?authorid=291824