



УДК 004.932.2

© 2018 г. **В.А. Бобков**, д-р техн. наук,
В.П. Май, канд. техн. наук

(Институт автоматизации и процессов управления ДВО РАН, Владивосток)

О ПОВЫШЕНИИ ЭФФЕКТИВНОСТИ РЕШЕНИЯ 3D SLAM ЗАДАЧИ ПО СТЕРЕОИЗОБРАЖЕНИЯМ*

Предложен адаптивный метод визуальной одометрии с длительным прослеживанием особенностей на изображениях. Реализована схема многоступенчатой фильтрации данных в задаче расчета траектории автономного робота и построения 3D модели среды по стереоизображениям. Вычислительные эксперименты на модельных сценах проведены с использованием имитационного моделирующего комплекса.

Ключевые слова: 3D SLAM, автономный робот, стереоизображения, сопоставление особенностей, адаптивный метод, фильтрация, триангуляция.

DOI: 10.22250/isu.2018.56.14-23

Введение

Проблема восстановления структуры по движению, известная в компьютерном зрении как SfM (Structure from Motion), а в робототехнике как SLAM (Simultaneous Localization and Mapping) активно исследуется уже на протяжении длительного времени. За это время было предложено много подходов, основанных на обработке видеоинформации. Одним из наиболее востребованных является подход, основанный на методе визуальной одометрии. С обзором по различным алгоритмическим реализациям SLAM-алгоритмов можно ознакомиться в [1]. Поскольку метод визуальной одометрии базируется на интеграции последовательно вычисляемых относительных перемещений камеры, его основным недостатком является накопление погрешности расчета траектории камеры и координатного представления фиксируемой пространственной обстановки.

* Работа выполнена при поддержке РФФИ (проекты № 18-07-00165 и №16-07-00350), а также программы «Дальний Восток» (проект 18-5-014) и Программы Президиума РАН № 27.

В данной статье предлагается подход к повышению эффективности применения визуального подхода в решении 3D SLAM-проблемы, основанный на интеграции в единую вычислительную схему адаптивного метода визуальной навигации AP и метода длительного прослеживания особенностей на изображениях; многоуровневой фильтрации ошибок, связанных с обработкой изображений и построением 3D модели объектов.

Общий подход

В рассматриваемой постановке автономный робот (AP) оснащен стереокамерой, с помощью которой он выполняет съемку при движении по траектории в априори неизвестной среде. Задача заключается в точном вычислении параметров траектории AP с одновременным построением модели среды по получаемой последовательности кадров. Согласно методу визуальной одометрии, решение строится на последовательном вычислении локальных перемещений аппарата, описываемых соответствующими локальными матрицами геометрических преобразований. Объединение локальных матриц всех предшествующих позиций позволяет получить матрицу преобразования, связывающую мировую систему координат (СК) с локальной СК, привязанной к аппарату в каждой текущей позиции его траектории. Применяемый в настоящей работе метод вычисления матрицы локального преобразования следует уже сложившемуся подходу, основанному на сопоставлении 2D особенностей на изображениях и применении ICP (Iterative Closest Point) алгоритма, который оценивает расхождение двух 3D облаков точек, получаемых на базе стереопар. Вычислительную схему определения параметров текущей позиции АПР кратко можно представить следующим образом:

- 1) выделение общего множества особенностей на четырех изображениях двух стереопар;
- 2) генерация и фильтрация двух облаков 3D точек;
- 3) вычисление с помощью ICP-алгоритма локального геометрического преобразования, связывающего локальные СК двух соседних позиций. В качестве альтернативы применению MATLAB для вычисления локального преобразования может использоваться разработанный авторами алгоритм быстрого вычисления относительного смещения [2];
- 4) вычисление параметров (6DOF) текущей позиции AP в мировой СК путем объединения локальных преобразований предшествующих позиций.

Сопутствующие методу визуальной одометрии традиционные ошибки порождаются: а) ошибочным сопоставлением особенностей; б) накоплением ошибок вычисления относительных перемещений AP при вычислении абсолютных координат AP.

Для решения задачи построения 3D модели объектов окружающей обста-

новки применяется метод реконструкции по изображениям, предложенный в [3]. Результатом является триангуляционная модель среды (так называемая splat-модель), предназначенная для быстрой визуализации. Входной информацией для метода реконструкции служат триангуляционные сетки отдельных видов, которые интегрируются в единую модель. Процессу генерации триангуляционных сеток сопутствуют ошибки, приводящие к появлению ложных треугольников.

Поэтому рассматриваются два направления повышения эффективности визуального подхода к решению проблемы SLAM (рис.1).

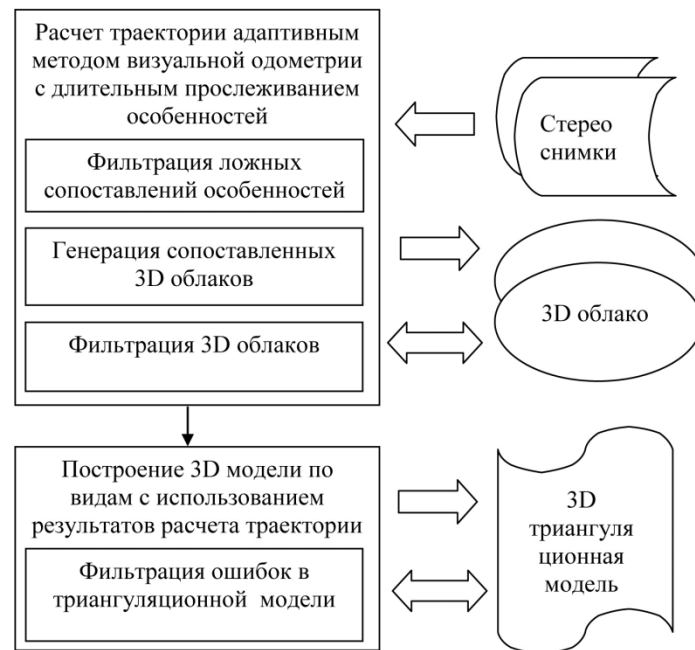


Рис. 1. Схема фильтрации ошибок при построении 3D модели среды по стереоизображениям с использованием адаптивного метода визуальной навигации.

Во-первых, это объединение ранее разработанных авторами адаптивного метода визуальной навигации AP [4] и метода длительного прослеживания особенностей на изображениях [5] в рамках единой вычислительной схемы. Во-вторых, это фильтрация ошибок на всех уровнях обработки данных в методе визуальной навигации и в методе построения 3D модели среды, включая:

фильтрацию ошибок при выделении и сопоставлении особенностей на изображениях;

фильтрацию ложных сопоставлений точек в 3D облаках;

устранение ошибок при генерации триангуляционной сетки, представляющей 3D модель объектов обстановки, фиксируемых камерой.

Фильтрация ложных сопоставлений особенностей

Сопоставление точечных особенностей выполняется на четырех изображениях двух стереопар (1-2 и 3-4), соответствующих двум последовательным позициям AP на траектории. Результат сопоставления проиллюстрирован на рис. 2.

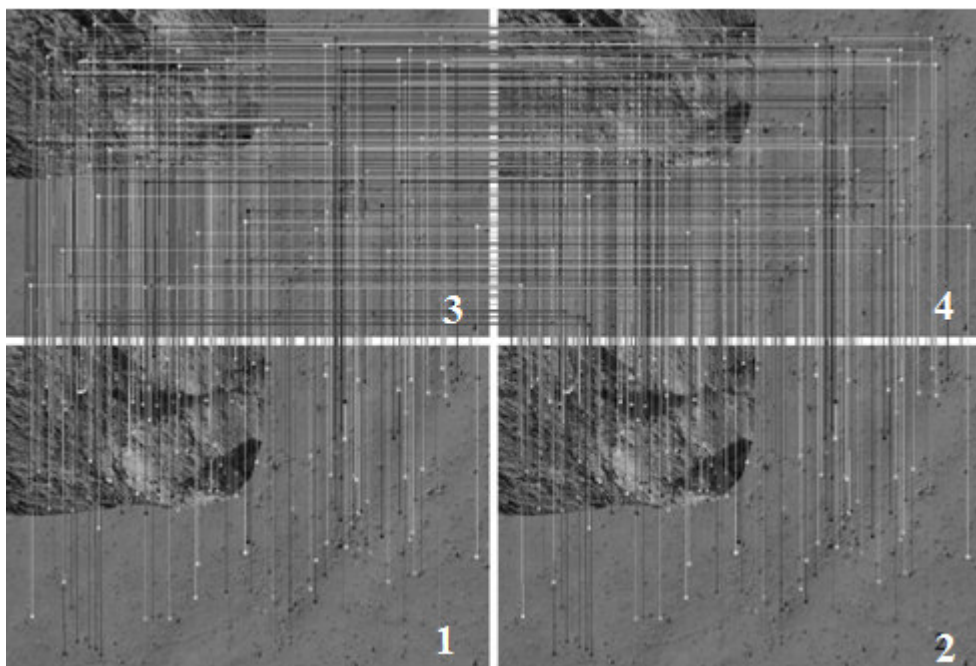


Рис. 2. Выделение единого множества сопоставленных особенностей на четырех изображениях двух стереопар.

Были рассмотрены и проверены несколько схем получения единого множества особенностей для четверки изображений двух стереопар с альтернативным использованием известных детекторов SURF, FREAK, FAST, а также алгоритма вычисления диспаратности для калиброванных очищенных пар. При программной реализации использовалась библиотека OpenCV.

Выбор был сделан в пользу двух схем:

а) последовательное сопоставление пар изображений 1-3, 3-4, 4-2 с выделением общего множества особенностей. На всех трех этапах работает SURF. Для каждой пары изображений выполняется сопоставление слева направо и справа налево (cross – checking). Для исключения ложных сопоставлений применяется эпиполярный фильтр для пары 3-4 и для пары 1-2;

б) вначале выполняется сопоставление пары 1-3 детектором SURF, затем применяется алгоритм определения диспаратности для калиброванных пар 1-2 и 3-4.

На изображениях 1, 2, 3, 4 было сгенерировано более 1000 особенностей. На каждом этапе 1-3, 3-4, 4-2 было сопоставлено несколько сот точек. Полученное в итоге множество особенностей, сопоставленных для всех четырех изображений, позволяет далее построить два сопоставленных облака 3D точек, отвечающих двум стереопарам (двум позициям AP).

Ошибки при сопоставлении возникают из-за наличия шумов (различие в освещенности изображений, слабая текстурированность или регулярность текстуры), что приводит к построению ошибочных точек в 3D облаках и, в конечном счете, к неправильному вычислению матриц локального перемещения. Ошибки

можно разделить на два типа. *Ошибка 1-го типа* появляется, если одна из сопоставленных точек двух облаков ложная (рис. 3а, рис. 3б).

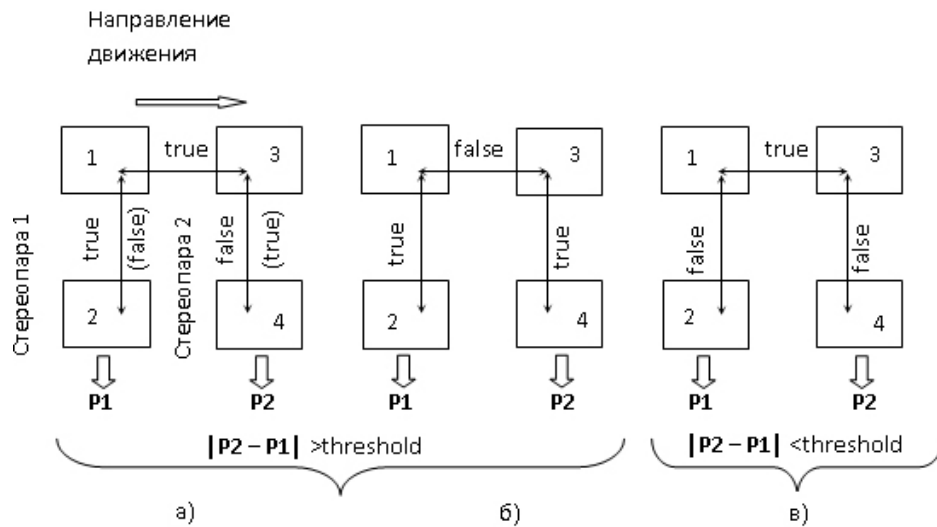


Рис. 3. Возможные комбинации ложных сопоставлений особенностей:

- а) ложное сопоставление между левым и правым снимком стереопары 1 или стереопары 2;
- б) ложное сопоставление между снимками стереопары 1 или стереопары 2;
- в) ложное сопоставление между снимками первой и второй стереопар.

Такая ошибка может быть выявлена оценкой расхождения точек в пространстве системы координат одной из камер после вычисления матрицы геометрического преобразования между системами координат камер двух соседних позиций. Для оценки расхождения используется пороговый критерий, с предварительно заданным порогом. В качестве порога используется характерная для этой дальности ошибка измерения. Ошибка измерения вычисляется по формуле, учитывающей заданную величину ошибки сопоставления особенностей на 2D снимках стереопары (например, расстояние, соответствующее 1 пикселю снимка), а также учитывающую другие параметры, характеризующие геометрию этой сцены (база стереопары, угол обзора камеры, разрешение снимка). *Ошибка 2-го типа* возникает, если обе сопоставляемые точки двух облаков ложные, но при этом расхождения между ними нет (рис. 3в). Такая ситуация может возникнуть, когда детектор одинаковым образом ложно сопоставил особенности в каждой из стереопар. Ошибка этого типа не выявляется пороговым критерием при совмещении двух облаков как в случае ошибки 1-го типа. Здесь следует отметить, что порождаемый этой ошибкой ложный пик в видимой поверхности может быть как большим, так и маленьким. Эксперименты на модельных сценах подтвердили наличие ошибок такого типа. Таким образом, ошибки 1-го типа могут быть устранены пороговым фильтром после вычисления матрицы локального преобразования. Но ошибки 2-го типа при этом остаются. Для устранения ошибок 2-го типа предлагается фильтрация каждого из двух 3D облаков на основе критерия связности соседних точек. Критерий связности можно построить на вычислении градиента,

который указывает на наличие/отсутствие пика (выброса). Поскольку фильтрация с произвольным порогом может удалять наряду с артефактами и «хорошие точки» (пик на поверхности действительно имеет место), то целесообразно фильтровать только большие пики. Устранение при такой фильтрации наряду с outliers небольшого числа «хороших» точек не является недостатком, поскольку оставшегося количества «хороших» точек в облаке достаточно для точного вычисления локального преобразования. Чтобы этот критерий мог применяться, необходима плотная карта глубин, по которой можно построить непрерывные фрагменты видимых поверхностей объектов сцены.

Адаптивность с длительным прослеживанием особенностей на снимках

Для навигации AP, как дополнение к используемому штатному навигационному оборудованию, был разработан метод навигации, основанный на визуальной одометрии с использованием стереокамеры. Предлагаемый метод визуальной навигации отличают от классической реализации метода визуальной одометрии два алгоритмических решения, направленные на повышение точности локализации и снижение вычислительных затрат. Во-первых, это адаптивная методика, минимизирующая число расчетных позиций траектории [4]. Во-вторых, это алгоритм с длительным прослеживанием особенностей на снимках участка траектории, предшествующего обрабатываемой позиции [5].

Дополнительным преимуществом, позволяющим снижать накапливаемую погрешность локализации при длительных перемещениях AP, является учет измерений сенсоров штатного навигационного оборудования в вычислительной схеме визуального метода навигации. При описании метода визуальной навигации будем использовать следующую терминологию. Траектория AP состоит из позиций – точек траектории, в которых выполняется съемка камерой (съемка ведется с заданной частотой). Позиции, в которых выполняется обработка снимков, будем называть узловыми позициями. Шаг – это расстояние между узловыми позициями, которое можно измерять числом снятых кадров (оно является эквивалентом времени, поскольку известна частота съемки). Фрагмент – это участок траектории, состоящий из нескольких шагов, или из некоторого числа позиций.

Схема расчета относительного перемещения камеры/AP показана на рис. 4.

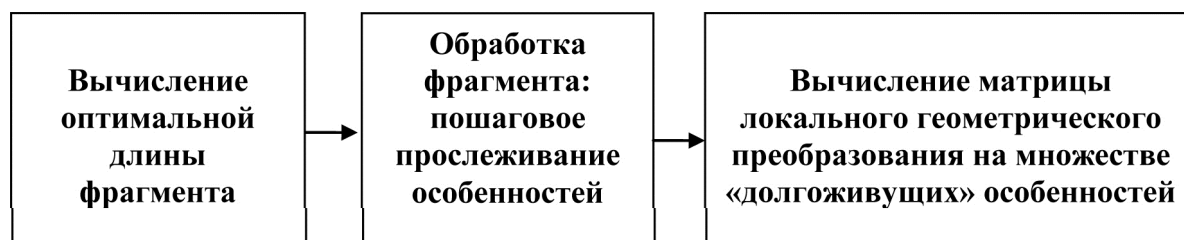


Рис. 4. Адаптивный метод визуальной навигации с длительным прослеживанием особенностей.

Сначала вычисляется максимально возможная длина подлежащего обработке фрагмента траектории при условии сохранения заданной степени перекрытия зон видимости камеры в начальной и конечной позициях фрагмента. На втором этапе выполняется пошаговое прослеживание особенностей на снимках узловых позиций фрагмента с последующим вычислением матрицы геометрического преобразования для относительного перемещения камеры/АР.

Для вычисления максимально возможной длины фрагмента на очередном шаге применяется разработанная авторами адаптивная методика [4]. Расчет основывается на согласовании параметров, определяющих относительное перемещение камеры при условии сохранения требуемой величины общей зоны видимости для начальной и конечной позиций (камера направлена вниз). К определяющим параметрам относятся: заданная величина перекрытия зон видимости (в %); угол обзора камеры; частота съемки; высота над дном и скорость перемещения АР, получаемые экстраполяцией значений предыдущего шага. Далее фрагмент с вычисленной длиной (в кадрах) обрабатывается алгоритмом «длительного прослеживания особенностей». Алгоритм основывается на выявлении «долгоживущих» особенностей на снимках фрагмента. Как показали предварительные эксперименты [5, 6], такие особенности позволяют получить в конечном счете более высокую точность локализации, вычисляемую методом визуальной одометрии. Фрагмент рассматривается как состоящий из нескольких шагов (размер шага предварительно фиксируется в алгоритме). Прослеживание особенностей осуществляется пошагово (рис. 5), для прослеживания на очередном шаге фрагмента используется множество особенностей, прослеженное на предыдущем шаге.

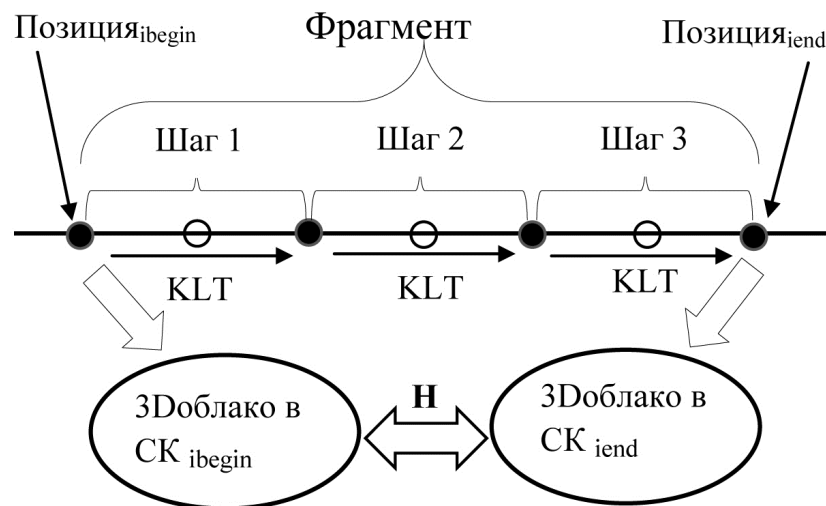


Рис. 5. Пошаговое прослеживание особенностей трекером KLT и вычисление локальной матрицы геометрического преобразования H для фрагмента, состоящего из трех шагов (шаг = 2), где узловые позиции обозначены кружочками.

Сопоставление на снимках стереопары выполняется детектором SURF, а прослеживание на снимках узловых позиций – за счет трекера KLT. Если на оче-

редном шаге прослеживания количество особенностей становится меньше заданного порога, то процесс прослеживания завершается, что эквивалентно уменьшению размера фрагмента. Полученное в результате такого многошагового прослеживания множество особенностей в конце фрагмента сопоставляется с исходным множеством (проекция на начало фрагмента).

Такая схема обработки в отличие от традиционной вычислительной схемы, во-первых, повышает достоверность сопоставления особенностей и, соответственно, уменьшает ошибку локализации, а во-вторых, сокращает время вычислений в целом за счет увеличения длины обрабатываемых участков (фрагментов) траектории. По полученным сопоставленным множествам особенностей согласно методу визуальной одометрии строятся два множества 3D точек (3D облака). По сопоставленным 3D облакам стандартным образом (методом минимизации квадратичных расхождений, сопоставленных 3D точек в одной системе координат) вычисляется матрица геометрического преобразования \mathbf{H} между системами координат начальной и конечной позиций фрагмента. С помощью этой матрицы определяется относительное изменение положения и ориентации камеры/робота. Тогда абсолютные координаты камеры/АПР в конечной позиции фрагмента вычисляются как $\mathbf{P}_{iend} = \mathbf{P}_{ibegin} \cdot \mathbf{H}$.

Фильтрация 3D облаков. Для исключения outliers из 3D облаков реализован итерационный фильтр, основанный на учете структурной когерентности облаков (сохраняющееся взаиморасположение точек в двух облаках). Алгоритмическая реализация аналогична описанной в работе [7]. Следует отметить, что такого рода фильтрация удаляет в том числе и особенности, связанные с динамическими объектами.

Фильтрация триангуляционной сетки

Вычислительные эксперименты по реконструкции сцен показали, что основной источник ошибок – невидимые участки поверхности, приводящие к «дырам» на 3D модели, и ложные поверхности, возникающие в результате ошибок в локальных триангуляционных сетках, подаваемых на вход алгоритму объединения. Ошибки проявляются в виде ложных треугольников, которые возникают на этапе генерации 2.5D триангуляционной сетки. Ложные треугольники возникают по двум причинам – разреженность множества точек, образующих сетку, и разрывность глубин на соседних точках. Разрывность глубин появляется на границе между видимой и невидимой частью одной поверхности для данного ракурса или на границе двух наблюдаемых поверхностей. Механизм появления ложного треугольника в обоих случаях один – три точки, образующие треугольник, принадлежат двум разным поверхностям в сцене и порождают, таким образом, ложную поверхность. При этом ложный треугольник закрывает видимый участок реаль-

ной поверхности на данном виде или на соседних видах.

Для устранения этого дефекта предлагается предварительная фильтрация подаваемых на вход триангуляционных сеток, направленная на удаление ложных поверхностей. Ниже дается описание ряда реализованных фильтров, применение которых позволило эффективно бороться с указанным дефектом.

Удаление «длинных» треугольников. Критерий – длина стороны треугольника больше пороговой величины.

Удаление треугольников с превышением пороговой величины по глубине (координата Z).

Оба указанных фильтра удаляют треугольники, где потенциально возможно нарушение связности поверхности.

Удаление «наклонных» треугольников. Удаляются треугольники, у которых угол между нормалью и лучом из точки наблюдения близок к прямому. Предполагается, что такие треугольники вычисляются с высокой степенью неопределенности.

Удаление вершин «наклонных» треугольников. Фильтр работает так же, как и предыдущий, только удаляются вершины, а не треугольник. В этом случае требуется повторная триангуляция после удаления вершин.

Фильтр с кросс-проверкой. Карта глубин, полученная для вида i , корректируется с учетом вида $(i + 1)$ следующим образом. Каждая 3D точка вида i проецируется на изображение вида $(i + 1)$. Для полученной 2D точки алгоритмом «дальномер» вычисляется ее 3D образ. Если расстояние между 3D точкой и вычисленным 3D образом больше порогового, то данная точка исключается из карты глубин вида i .

Следует заметить, что хотя применение перечисленных фильтров и порождает «дырки» на локальных поверхностях, объединение локальных поверхностей их устраняет, поскольку объединяются виды с разных ракурсов.

Заключение

Вычислительные эксперименты проводились на модельных сценах с использованием моделирующего комплекса [8] применительно к подводному AP (рис. 6, 7).

Результаты экспериментов показали, что за счет адаптивной методики вычислительные затраты могут быть снижены в несколько раз в зависимости от характера сцены. Применение же фильтрации на всех указанных уровнях позволяет повысить точность расчета траектории AP на 30-50% и, соответственно, повысить качество генерируемой 3D модели среды.

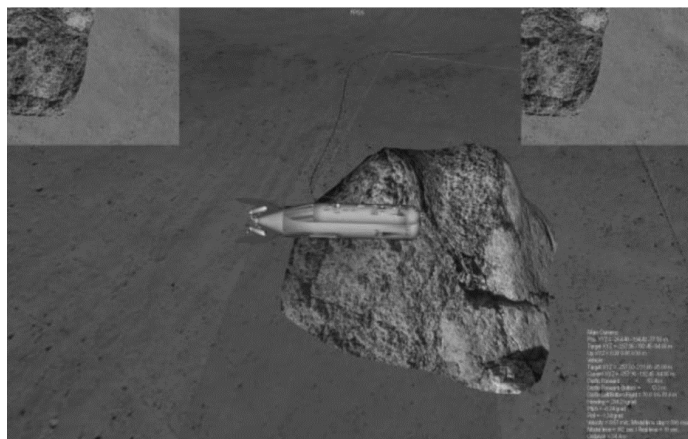


Рис. 6. Навигация АПР по оптической информации.

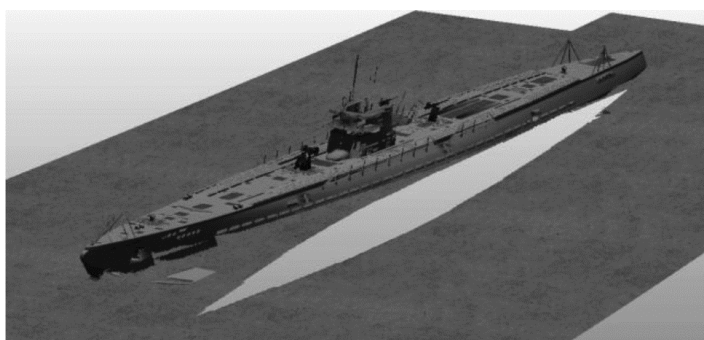


Рис. 7. 3D модель объекта подводной сцены.

ЛИТЕРАТУРА

1. Takafumi Taketomi, Hideaki Uchiyama, Sei Ikeda. Visual SLAM algorithms: a survey from 2010 to 2016 // IPSJ Transactions on Computer Vision and Applications, 2017 (<https://doi.org/10.1186/s41074-017-0027-2>).
2. Bobkov V.A., Mel'man S.V., Kudryashov A.P. Fast computation of local displacement by stereo pairs // Pattern Recognition and Image Analysis. – 2017. – Vol. 27, №. 3. – P. 458-465.
3. Бобков В.А., Кудряшов А.П. Реконструкция и визуализация пространственной сцены с использованием глобальной точечной модели // Информатика и системы управления – 2017. – №2(52). – С. 3-11.
4. Бобков В.А., Машенцев В.Ю. Навигация подводного робота по стереоизображениям // Мехатроника, автоматизация, управление. – 2016. – Т. 17, №2. – С. 101-109.
5. Бобков В.А. Визуальная одометрия с длительным прослеживанием особенностей // Информатика и системы управления. – 2017. – №3(53). – С. 17-23.
6. Hernan Badino, Akihiro Yamamoto, Takeo Kanade. Visual Odometry by Multi-frame Feature Integration // International Workshop on Computer Vision for Autonomous Driving @ ICCV. – December 2013. – P. 222-229.
7. Saez J.M., Hogue A., Escolano F., Jenkin M. Underwater 3D SLAM through Entropy Minimization // Proceedings 2006 IEEE International Conference on Robotics and Automation, ICRA, 2006. – P. 3562-3567.
8. Бобков В.А., Мельман С.В., Морозов М.А., Тарасов Г.В. Моделирующий комплекс для подводного робота на распределенной архитектуре с использованием кластера // Информатика и системы управления. – 2017. – №4 (54). – С. 32-42.

Статья представлена к публикации членом редколлегии А.И. Абакумовым.

E-mail:

Бобков Валерий Александрович – bobkov@iacr.dvo.ru;

Май Владимир Павлович – may@iacr.dvo.ru.