

# ПРОСТРАНСТВЕННАЯ РЕКОНСТРУКЦИЯ ОБЪЕКТОВ В ДИНАМИЧЕСКОЙ СЦЕНЕ ПО ВИДЕОПОТОКУ

**В.А. Бобков, А.П. Кудряшов**

Федеральное государственное бюджетное учреждение науки  
Институт автоматизации и процессов управления ДВО РАН<sup>1</sup>

Представлен метод 3D реконструкции объектов в динамической сцене в рамках общего подхода к решению проблемы восстановления структуры и движения по изображениям. Подход в целом основывается на точечном представлении объектов и применении метода визуальной одометрии для расчета движения динамических объектов по стереовидеопотоку. На этапе подготовки входных данных для метода реконструкции выполняются идентификация динамических объектов и вычисление траекторий их движения. Согласно предлагаемому методу 3D модель динамического объекта строится посредством объединения его видов, получаемых из разных позиций камеры. Проведенные эксперименты с виртуальными сценами подтвердили работоспособность предложенных алгоритмов.

## ВВЕДЕНИЕ

Задача 3D реконструкции объектов по изображениям актуальна для многих практических приложений, включая и приложения с использованием автономных подводных аппаратов для инспекции объектов подводной промышленной инфраструктуры, поисковых и спасательных работ.

Принципиальная сложность задачи реконструкции динамических сцен заключается в том, что, в отличие от статических сцен, здесь требуется восстановление движения динамических объектов (ДО). Такая постановка задачи подразумевает построение геометрических моделей движущихся объектов по отдельным видам, фиксируемым движущейся камерой (установленной на борту робота).

На сегодня известны различные методы решения задачи реконструкции объектов в статических и динамических сценах по видеоинформации. Они основываются на использовании силуэтов, цвета/стерео [1–7], освещенности и движения [8–10], извлекаемых из видео, чтобы восстановить геометрию динамических элементов сцены. В некоторых работах, например в [8], для реконструкции движения человека используется специализированное оборудование для измерения исходных данных – множественность камер, специальные системы освещения, высокая частота съемки. Здесь был разработан

метод, основанный на построении карт нормалей со слиянием сеток (meshes) на этапе реконструкции методом «volumetric method».

В большинстве работ предлагаемая методология базируется на контролируемом окружении, где известен фон и используются калиброванные и фиксированные камеры. Более общий подход к реконструкции и сегментации сложных динамических сцен с помощью множества перемещающихся камер без априорных знаний о сцене предлагается в работе [11]. Согласно методу оцениваются плотные карты глубин для каждой камеры с последующим их объединением для реконструкции каждого ДО. Особенностью подхода является метод грубой начальной сегментации и алгоритм плотной реконструкции. Эффективным в плане алгоритмической простоты, сравнительно небольших вычислительных затрат и возможности использовать стандартное графическое ускорение представляется подход, основанный на точечном представлении объектов. Например, в работе [12] используется глобальная точечная модель для представления статических и ДО. Каждая точка характеризуется постоянно обновляемым во времени показателем степени доверия (confidence counter), по которому можно вероятностным образом судить о принадлежности точки «статике» или «динамике».

<sup>1</sup> 690041, Владивосток, ул. Радио, 5. Тел.: +7 (423) 231-04-39.  
E-mail.: bobkov@iacp.dvo.ru

На первом этапе те точки, которые определяются во входных снимках с помощью алгоритма ICP (Iterative Closest Point) (как outliers, рассматриваются с большой вероятностью как динамические и служат seed-точками для последующего формирования ДО). А на втором этапе с помощью point-based region growing procedure непосредственно формируются ДО путем объединения точек на основе сходства по текстурным и пространственным геометрическим характеристикам. Эффективный подход был предложен в [13] к реконструкции сложных динамических сцен. Особенностью подхода является использование темпоральной когерентности и multi-view сегментации с применением geodesic star convexity constraint, позволяющей лучше выделять геометрию динамических объектов. В предлагаемом в работе [14] подходе к реконструкции и визуализации в режиме реального времени акцент делается на использовании вычисляемой кривизны поверхности. По мнению авторов, этот подход уменьшил ошибку траектории камеры и повысил качество реконструкции. В работах, использующих видео с одной камеры применительно к динамическим сценам, решение задачи восстановления структуры по движению основывается на обработке оптического потока точечных особенностей с кластеризацией треков и последующей оценкой отдельных ДО [15]. Следует заметить, что в такой постановке из-за ограниченного использования 3D информации вычислительная схема усложняется, что неоправданным образом повышает вычислительную трудоемкость и снижает качество 3D реконструкции.

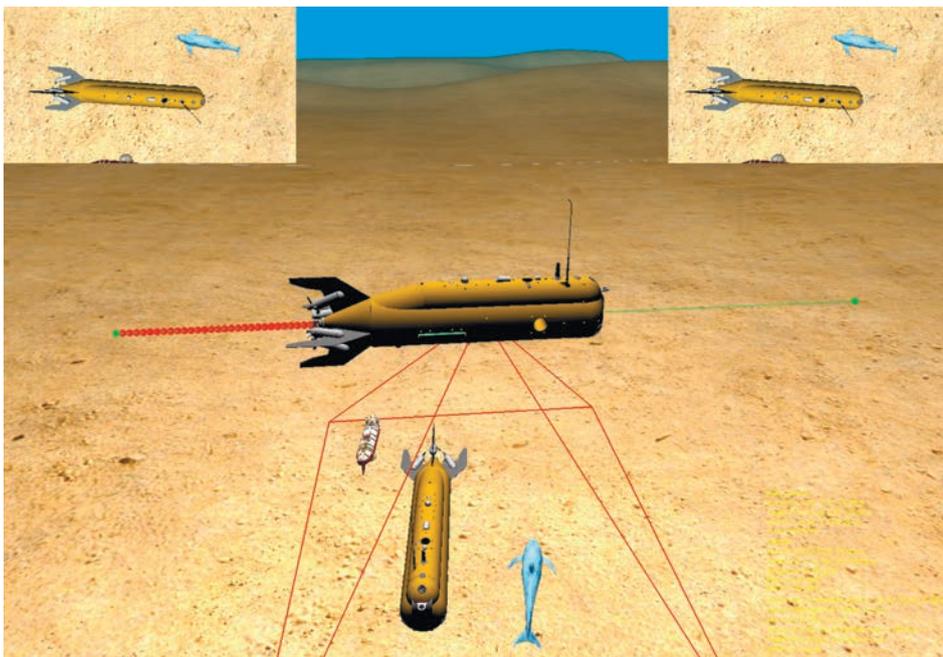


Рис. 1. Модельная подводная сцена: камера «видит» 3 ДО

Анализируя существующие результаты исследований в данной области в целом, можно констатировать пока еще недостаточную степень универсальности существующих решений, использование во многих работах ряда постановочных ограничений, высокую вычислительную трудоемкость алгоритмов, и не всегда приемлемую эффективность программно-алгоритмических средств для практических приложений.

В настоящей статье в рамках развития авторского подхода, описанного в [16, 17], предлагается оригинальный метод решения задачи пространственной реконструкции динамических сцен. В реализованной вычислительной схеме используется ранее разработанный авторами метод визуальной навигации по стереоизображениям [18, 19]. В компьютерных экспериментах при тестировании разработанной алгоритмической базы применялся имитационный моделирующий комплекс [20].

## 1. Описание общего подхода. Постановка задачи

В работах [16, 17] представлен общий подход к решению проблемы, известной в области компьютерного зрения как «проблема восстановления структуры и движения по изображениям». Движение ДО осуществляется в априори неизвестной обстановке и фиксируется установленной на борту автономного робота стереокамерой (рис. 1). Кадр из фиксируемой видеопоследовательности приведен на рис. 2.

Решение указанной проблемы подразумевает наряду с восстановлением траектории движения камеры и реконструкции статической части сцены, также и вычисление траекторий движения ДО с построением их 3D геометрических моделей.

Подход основывается на применении точечного представления объектов сцены, визуального метода [18, 19] расчета траекторий движения камеры и ДО, и на разработке оригинальной алгоритмической базы, направленной на решение задач идентификации объектов, отслеживания их движения и пространственной реконструкции.

Предполагается, что ДО являются недеформируемыми, и в сцене они появляются в произвольные моменты времени. Расчет траекторий движения камеры и ДО выполняется пошагово – для каждого участка между соседними позициями траектории камеры (робота) вычисляются матрицы относительного движения камеры и ДО. Основными этапами вычислительной схемы в целом являются:

- генерация точечного представления сцены по стереоизображениям в каждой позиции траектории камеры. На этом этапе применяется детектор SURF для сопоставления точечных особенностей на снимках и метод триангуляции лучей для построения 3D координат точек;

- выделение объектов сцены (3D точечные представления) на основе применения: а) критерия «связности» (близость соседних точек поверхности объекта) и б) критерия «жесткости» (характеристика недеформируемых объектов), анализируемого при оценке движения ДО. Применение этих критериев с использованием измерений лазерного дальномера (или его виртуального аналога [21]) позволяет строить плотные точечные представления видимых поверхностей объектов;

- выделение статической части сцены и вычисление траектории движения камеры;

- выделение ДО;

- вычисление траекторий собственного движения ДО на основе декомпозиции комплексного движения, фиксируемого камерой (движение камеры плюс собственное движение объекта);

- генерация видов ДО (3D точечное представление видимой поверхности ДО) в каждой позиции траектории камеры;

- 3D-реконструкция ДО, т.е. построение пространственных граничных моделей (поверхности) объектов по полученным видам.

В настоящей статье описывается предлагаемый метод решения задачи последнего этапа обозначенной схемы – задачи 3D реконструкции ДО, с учетом того, что предыдущие этапы описаны в [16, 17]. Метод базируется на двух основных алгоритмических решениях:

- использование вычисленных матриц собственного относительного движения каждого ДО по всей длине траектории для помещения всех видов в единую систему координат (СК);



Рис. 2. Снимок виртуальной подводной сцены в одной из позиций траектории камеры/АПР: в поле зрения камеры находятся 3 ДО на фоне рельефа дна

- объединение полученных видов в единой СК «воксельным методом», основанном на взвешенном объединении видов в воксельном пространстве сцены.

Решение этой задачи замыкает общую вычислительную схему, обеспечивающую решение поставленной задачи в целом по восстановлению структуры и движения применительно к динамической сцене.

## 2. Реконструкция динамической сцены

3D реконструкция каждого ДО в сцене выполняется на основе объединения видов объекта, помещаемых в единую СК, связанную с данным объектом. Термин «единая СК» здесь употребляется, чтобы подчеркнуть, что все виды должны быть помещены в одну СК, связанную с объектом (например, в СК объекта в момент его появления в сцене). Виды генерируются во всех позициях траектории движения камеры/робота. Под термином «вид» здесь понимается 3D точечное представление участка поверхности объекта, видимого камерой из конкретной позиции ее траектории движения (т.е. с определенного ракурса). Здесь следует учитывать, что камера фиксирует на изображениях результат комплексного движения – собственного движения ДО и движения камеры (робота). А вычисляемые по стереопаре снимков координаты 3D точек вида определены в СК камеры. Поэтому для осуществления пространственной реконструкции ДО необходимо решение следующих основных локальных задач:

- восстановление движения камеры и собственного движения ДО;

- вычисление матриц геометрического преобразования координат точек наблюдаемых камерой видов ДО в единую СК, связанную с ДО;

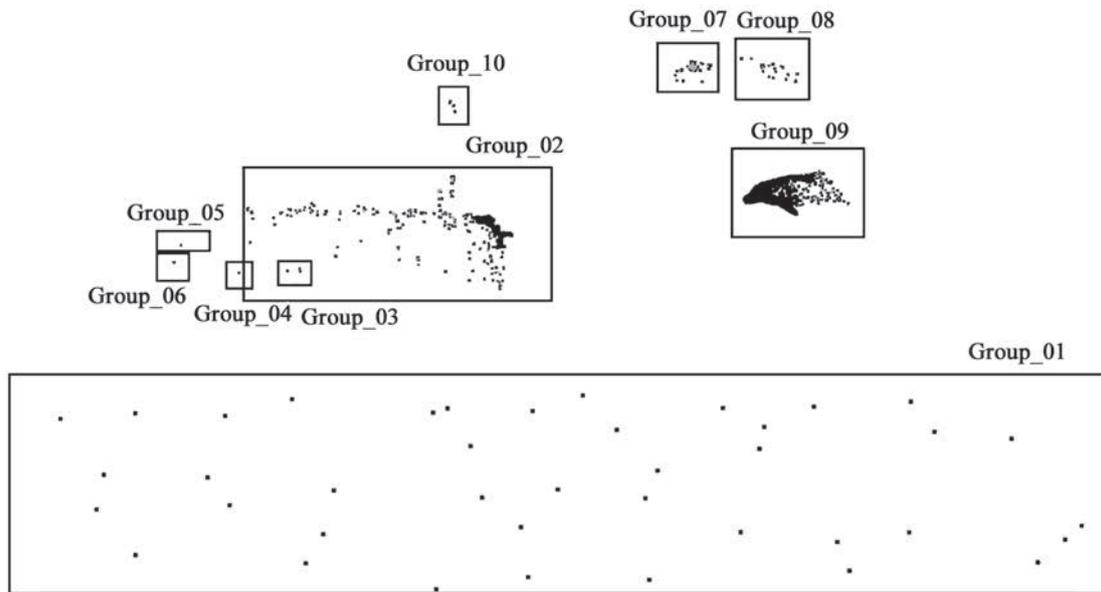


Рис. 3. Показана проекция «вид сбоку». Каждая из 9 сформированных групп по критерию «связности» точек выделена отдельным прямоугольником

- построение полной 3D точечной модели для каждого ДО на основе объединение видов;
- генерация граничного представления ДО (полигональной поверхности).

### 2.1. Построение 3D точечных представлений ДО

Идентификация и построение 3D точечных представлений объектов (в каждой позиции движения камеры) осуществляются последовательным выполнением нескольких этапов обработки данных. Вначале работает алгоритм с применением критерия

«связности» точек поверхности, который формирует несколько групп точек, представляющих отдельные фрагменты видимых поверхностей объектов (рис. 3). Затем отдельные группы точек со сходным движением (уже на основе критерия «жесткости») объединяются в полное точечное представление объекта (рис. 4). Согласно критерию «жесткости» расстояние между любой парой точек, принадлежащих недеформируемому объекту, сохраняется при его движении. Группы точек, которые нельзя соотнести ни с одним из объектов, исключаются из дальнейшей обработки. Затем выполняется идентификации ста-

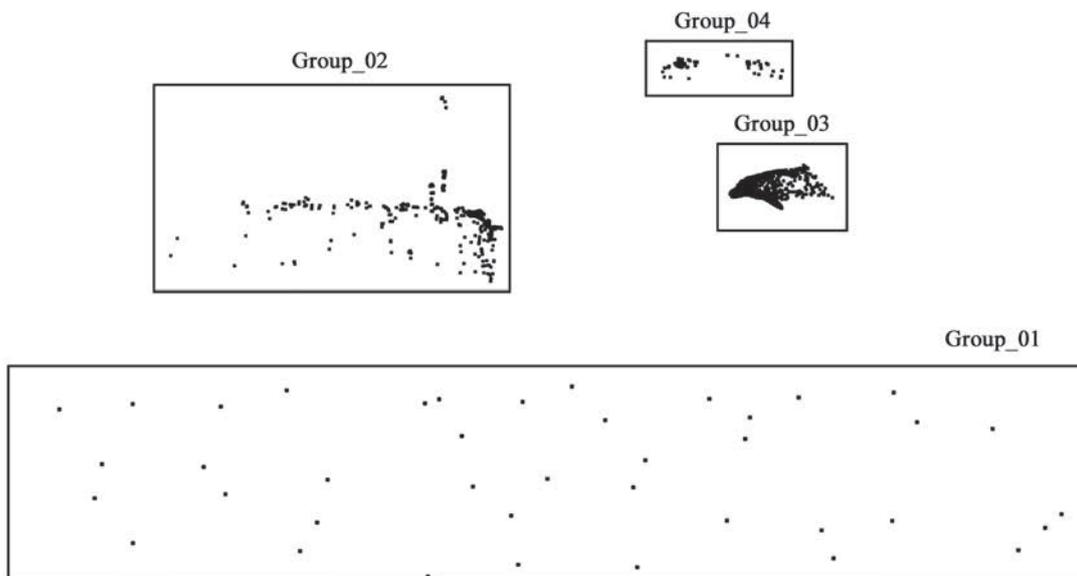


Рис. 4. Выделены 3 группы точек, относящихся к ДО, и группа точек, представляющих рельеф дна (после работы алгоритма объединения) по критерию «жесткости»

тической части сцены (морское дно) и ДО. Результатом выполнения такой последовательности этапов обработки на каждом шаге является множество видов всех ДО во всех позициях движения камеры/робота.

## 2.2. Вычисление матриц движения камеры и динамических объектов

Перемещение ДО, фиксируемое камерой в позиции  $i$ , является комплексным движением камеры и самого объекта. Поэтому матрицу  $H_{i-1, i}^{joint, dyn\_ob}$  совместного движения объекта и камеры на шаге  $[i-1, i]$  можно представить как:

$$H_{i-1, i}^{joint, dyn\_ob} = H_{i-1, i}^{camera} \cdot H_{i-1, i}^{dyn\_ob},$$

здесь  $H_{i-1, i}^{dyn\_ob}$  – матрица собственного движения ДО, а  $H_{i-1, i}^{camera}$  – матрица относительного движения камеры/АПР на шаге  $[i-1, i]$ .

Матрица совместного движения  $H_{i-1, i}^{joint, dyn\_ob}$  вычисляется указанным методом визуальной навигации по точечным особенностям, принадлежащим данному ДО.

Матрица  $H_{i-1, i}^{camera}$  относительного движения камеры/АПР вычисляется методом визуальной навигации [18, 19] по сопоставленным точечным особенностям, принадлежащим статической части сцены. Сопоставление выполняется на двух стереопарах снимков, относящихся к этим позициям траектории.

Тогда матрицу собственного движения ДО, с учетом вышеприведенного соотношения и предварительно вычисленной матрицы движения камеры, можно вычислить как:

$$H_{i-1, i}^{dyn\_ob} = H_{i-1, i}^{joint, dyn\_ob} \cdot (H_{i-1, i}^{camera})^{-1}$$

Имея цепочку геометрических преобразований собственного движения ДО (т.е. матрицы  $H_{i-1, i}^{dyn\_ob}$ ) на протяжении всего отрезка времени, когда объект наблюдается камерой, можно связать СК объекта в каждой позиции траектории его движения с СК объекта в начальный момент времени. Это позволяет объединить виды ДО в одной СК (СК начального положения объекта). Более детальное описание вычисления матриц движения ДО можно найти в [17].

## 2.3. Объединение видов

После построения видов каждого ДО (для последовательности позиций, пока объект находится в области видимости камеры) и вычисления матриц собственного движения ДО, можно построить полные



Рис. 5. Результирующие облака точек двух ДО

точечные модели ДО. Для этого виды каждого ДО помещаются в одну СК, что позволяет затем по множеству точек всех видов построить граничное представление объекта (в виде полигональной поверхности с нанесением текстуры).

На рис. 5 приведен пример получения 3D точечных моделей ДО после совмещения видов в единой СК. Объект «батискаф» содержит 1832 точки, полученные по 32 видам сцены. Можно заметить, что из-за равномерной текстуры объекта точки распределены неравномерно и в основном сосредоточены вокруг граней и других контрастных деталей. Объект «дельфин» содержит 3211 точек, которые получены по 44 видам и распределены более равномерно по всей поверхности объекта. У обоих объектов не наблюдается нижней части и меньше точек на боковых гранях из-за того, что эти объекты наблюдались только сверху.

## 2.4. Построение граничной модели ДО

По множеству точек всех видов ДО (для каждого объекта) необходимо построить поверхность, максимально точно описывающую этот объект. Предлагаемый метод построения полной трехмерной модели объекта является модификацией ранее предложенного метода, где используется обработка данных в воксельном пространстве сцены [22, 23]. Отличие данной реализации от вычислительной схемы для статической сцены заключается в том, что исходные виды ДО генерируются с учетом собственного движения (т.е. используются предварительно вычисленные

матрицы геометрических преобразований координат, определяющие последовательные относительные перемещения ДО).

В воксельном пространстве сцены задается непрерывная неявная функция  $D(V_i)$ , представленная значениями в узлах воксельной решетки. Функция конструируется как взвешенная сумма получаемых для  $n$  видов расстояний  $d_1(V_i), d_2(V_i), \dots, d_n(V_i)$  от точки  $V_i$  до ближайшей поверхности. Расстояние  $d_j$  берется на луче, направленном из центра проекций вида  $j$  в точку  $V_i$ , и является величиной со знаком (положительным для точек, находящихся перед поверхностью, и отрицательным для точек за поверхностью). В качестве весовой функции  $W(V_i)$  берется скалярное произведение вектора нормали к поверхности и вектора направления на точку наблюдения. Такой выбор исходит из предположения, что при корреляционном подходе к определению расстояния до поверхности степень неопределенности для наклонных поверхностей выше, чем для поверхностей, наблюдаемых под прямым углом. Для вычисления  $D$  и  $W$  используются следующие инкрементальные правила:

$$D_{j+1}(V_i) = [W_j(V_i)D_j(V_i) + w_{j+1}(V_i)d_{j+1}(V_i)] / [W_j(V_i) + w_{j+1}(V_i)],$$

$$W_{j+1}(V_i) = W_j(V_i) + w_{j+1}(V_i),$$

где  $i$  – номер вокселя, а  $j$  – номер вида,  $d_{j+1}$  – расстояние от вокселя до поверхности,  $w_{j+1}$  – скалярное произведение нормали к поверхности и вектора направления на точку наблюдения. Весовая функция  $W(V_i)$  – константа для всех вокселей до видимой поверхности (для данного вида), а для вокселей за поверхностью линейно убывает до нуля в пределах  $\epsilon$ -окрестности. Такой выбор области определения весовой функции направлен на предотвращение возникновения ложных поверхностей. В результате последовательной обработки всех видов каждому вокселю присваиваются значение интегральной функции расстояния  $D(V_i)$  (величина со знаком) и значение интегральной весовой функции  $W(V_i)$ . Тогда построение изоповерхности  $D(V_i) = 0$  и является решением нашей задачи. Преимуществом сведения исходной задачи к такой постановке является то, что построение изоповерхности скалярного поля, заданного на воксельной решетке, можно выполнить с помощью известного алгоритма маршрутизирующих кубиков [24].

Таким образом, строится поверхность ДО по всем точкам со всех видов, на которых этот объект наблюдался. Следует отметить, что если объект исчезнет из вида, а потом появится, то он будет считаться новым объектом. Пример реконструкции приведен на рис. 6.

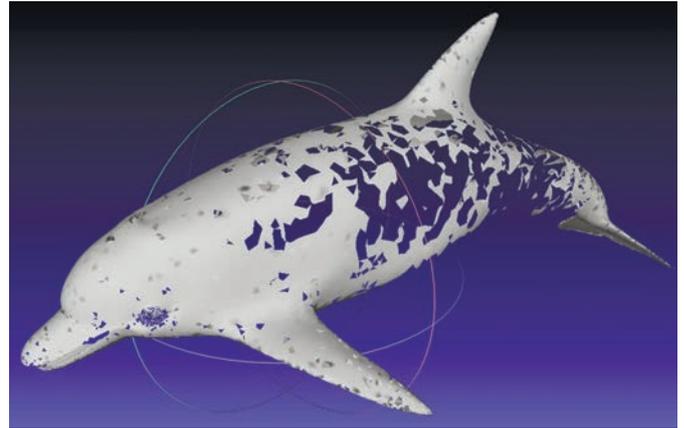


Рис. 6. Трехмерная реконструкция ДО

## ЗАКЛЮЧЕНИЕ

Представленный в настоящей работе метод построения 3D-моделей ДО является завершающим этапом в реализации общей вычислительной схемы, направленной на решение задачи восстановления структуры и движения по стереоизображениям применительно к динамической сцене. Отличительная особенность метода заключается в методике объединения сгенерированных видов ДО с учетом их собственного движения. Тестирование разработанных программно-алгоритмических средств на виртуальных сценах подтвердило правомерность подхода в целом и принципиальную работоспособность метода 3D-реконструкции динамических сцен. Дальнейшее развитие работы предполагает эксперименты с реальными данными и совершенствование алгоритмической базы.

Работа выполнена при частичной поддержке РФФИ (проект № 18-07-00165) и Программы «Дальний Восток» (проект 18-5-014).



## ЛИТЕРАТУРА

1. Kim H., Sarim M., Takai T., Guillemaut J.Y., Hilton A. Dynamic 3d scene reconstruction in outdoor environments. *Three-Dimensional Data Processing // Proceedings of Int. Symp. «Visualization and Transmission» (3DPVT)*. Paris, France, 2010.
2. Guan L., Franco J.S., Pollefeys M. Multi-object shape estimation and tracking from silhouette cues // *Proceedings of IEEE Conf. «Computer Vision and Pattern Recognition» (CVPR)*. Anchorage, USA, 2008. P. 283–303.
3. Franco J.S., Boyer E. Fusion of multi-view silhouette cues using a space occupancy grid // *Proceedings of Int. Conf. «Computer Vision» (CCV)*. Beijing, China, 2005. P. 1747–1753.
4. Furukawa Y., Ponce J. Dense 3d motion capture for human faces // *Proceedings of IEEE Conf. «Computer Vision and Pattern Recognition» (CVPR)*. Miami, USA, 2009. P. 1674–1681.
5. Tung T., Nobuhara S., Matsuyama T. Complete multi-view reconstruction of dynamic scenes from probabilistic fusion of narrow and wide baseline stereo // *Proceedings of Int. Conf. on Computer Vision (ICCV)*. Kyoto, Japan, 2009.
6. Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R.: A comparison and evaluation of multi-view stereo reconstruction algorithms // *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. New York, USA, 2006. P. 519–528.
7. Seitz S.M., Dyer C.R. Photorealistic scene reconstruction by voxel coloring // *Proceedings of Conf. on Computer Vision and Pattern Recognition (CVPR)*. San Juan, Puerto Rico, 1997. P. 1067–1079.
8. Vlasic D., Peers P., Baran I., Debevec P., Popovi C.J., Rusinkiewicz S., Matusik W. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. Graphics // Proceedings of Conf. and Exhibition on Computer Graphics & Interactive Techniques Asia (SIGGRAPH)*. Yokohama, Japan, 2009. P. 1–11
9. Ahmed N., Theobalt C., Dobrev P., Seidel H.P., Thrun S. Robust fusion of dynamic shape and normal capture for high-quality reconstruction of time-varying geometry // *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Anchorage, USA, 2008.
10. Vedula S., Baker S., Seitz S., Kanade T. Shape and motion carving in 6d // *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Hilton Head, USA, 2000.
11. Armin Mustafa, Hansung Kim, Jean-Yves Guillemaut, Adrian Hilton. General Dynamic Scene Reconstruction from Multiple View Video // *Proceedings of Int. Conf. on Computer Vision (ICCV)*. Santiago, Chile, 2015. P. 900–908.
12. Keller M., Lefloch D., Lambers M., Izadi S., Weyrich T., Kolb A. Real-time 3d reconstruction in dynamic scenes using point-based fusion // *Proceedings of Joint 3DIM/3DPVT Conf. (3DV)*. Seattle, USA, 2013. P. 1–8.
13. Armin Mustafa, Hansung Kim, Jean-Yves Guillemaut Adrian Hilton. Temporally coherent 4D reconstruction of complex dynamic scenes // *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA, 2016.
14. Lefloch D., Kluge M., Sarbolandi H., Weyrich T., Kolb A. Comprehensive Use of Curvature For Robust And Accurate Online Surface Reconstruction // *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39, No. 1. P. 2349–2365.
15. Ranftl R., Vineet V., Chen Q., Koltun V. Dense Monocular Depth Estimation in Complex Dynamic Scenes // *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, USA, 2016.
16. Bobkov V., Kudryashov A., Mel'man S. On the Recovery of Motion of Dynamic Objects from Stereo Images // *Programming and Computer Software*. 2018. Vol. 44, Iss. 3. P. 148–158.
17. Бобков В.А., Кудряшов А.П. Идентификация и расчет траекторий динамических объектов по стереоизображениям // *Программирование*. 2020. № 1. С. 3–14.
18. Бобков В.А., Машенцев В.Ю. Навигация подводного робота по стереоизображениям // *Мехатроника, автоматизация, управление*. 2016. Т. 17, № 2. С. 101–109.
19. Bobkov V.A., Kudryashov A.P., Mel'man S.V., Shcherbatyuk A.F. Autonomous Underwater Navigation with 3D Environment Modeling Using Stereo Images // *Gyroscopy and Navigation*. 2018. Vol. 9, No. 1. P. 67–75.
20. Мельман С.В., Бобков В.А., Инзарцев А.В., Павин А.М., Черкашин А.С. Программный моделирующий комплекс для автономных подводных аппаратов на базе многопроцессорной архитектуры // *Подводные исследования и робототехника*. 2015. № 1 (19). С. 23–32.
21. Bobkov V.A., Ronshin Y.I. GPU Implementation of Depth Map Algorithm // *The First Russia and Pacific Conf. on Computer Technology and Applications*. Vladivostok, Russia, 2010. P. 382–387.
22. Бобков В.А., Кудряшов А.П. Построение трехмерной модели морского дна воксельным методом // *Подводные исследования и робототехника*. 2016. № 2. С. 13–18.
23. Curless B., Levoy M. A Volumetric Method for Building Complex Models from Range Images, *Computer Graphics // Proceedings of 23th Int. Conf. on Computer Graphics and Interactive Technique SIGGRAPH*. New Orleans, USA, 1996. P. 303–312.
24. Lorensen W.E., Cline H.E. Marching cubes: A high resolution 3D surface construction algorithm // *Proceedings of Computer Graphics SIGGRAPH*. 1987. Vol. 21. P. 163–169.