DOI: 10.5862/JCSTCS.217-222.10
УДК 004.923

А.А. Хуршудов

ПОСТРОЕНИЕ ТРЕХМЕРНЫХ КАРТ ПРИЗНАКОВ НА ОСНОВЕ ВИДЕОФРАГМЕНТОВ МЕТОДОМ ОПТИЧЕСКОГО ПОТОКА

A.A. Khurshudov

CONSTRUCTING 3D FEATURE MAPS FROM VIDEO SEQUENCES BY OPTIC FLOW ESTIMATION

Рассмотрена обобщенная задача воссоздания трехмерной структуры объектов из набора видеофрагментов, на которых изображена заданная сцена. В отличие от распространенных методов фотограмметрии, широко используемых для решения подобных задач, рассматриваемый метод не требует знания параметров камеры, способен работать с любыми категориями согласованных видеофрагментов и справляться с высокими показателями шума. В процессе реконструкции объект представляется как совокупность устойчивых разреженных признаков в трехмерном пространстве, которые первоначально обнаруживаются в отдельных ключевых кадрах (например, с помощью детектора углов Ши-Томаси), затем с помощью разреженного оптического потока их перемещения отслеживаются в последующих кадрах. При наличии информации о движении камеры при помощи простейших геометрических расчетов становится возможным определить положение интересующей точки в пространстве, а многократное определение позиции для одной и той же точки в различных кадрах видео позволяет эффективно устранять погрешности оценки. Помимо разреженного оптического потока (метод Лукаса–Канаде), метод также использует некоторые свойства плотного потока (метод Фарнебака), позволяющие осуществлять сегментацию сцен и выделение объектов на разных уровнях глубины сцены. Полученные трехмерные карты признаков в дальнейшем могут использоваться в качестве макродетектора объектов на отдельных цифровых изображениях, устойчивого к инвариантным трехмерным трансформациям.

ОБНАРУЖЕНИЕ ОБЪЕКТОВ; ОПТИЧЕСКИЙ ПОТОК; АЛГОРИТМ ЛУКАСА–КАНАДЕ; РАЗРЕЖЕННЫЕ ПРИЗНАКИ; СЕГМЕНТАЦИЯ МЕТОДОМ ВОДОРАЗДЕЛА.

The study presents a general case of structure-from-motion problem where the given data consists of a bunch of video sequences filmed in the same scene. Unlike the popular methods of photogrammetry and bundle adjustment, the proposed solution does not required specific knowledge of intrinsic camera parameters, could be applied to any type of consistent motion pictures and can handle large amounts of noise. During the process of reconstruction an object is viewed as a 3D map of robust sparse features, which at first hand are discovered in certain key frames (using existent computer vision techniques like Shi-Tomasi corner detector) and afterwards tracked across the following frames using sparse optic flow method. When camera motion (egomotion) data is available, it is became possible to estimate each feature's depth by using simple geometric properties of two-image disparity, and having each feature estimated from multiple video frames allows to effectively filter out the noise. Apart from sparse Lucas-Kanade optic flow the study also makes use of some properties of dense optic flow (Gunnar Farneback's algorithm), which is used for scene segmentation during the camera motion. The resulting 3D feature maps are designed to be used as a macro object detector that could be applied to any previously unknown single digital images, representing structures

that are believed to store 3D visual memory of an object, and therefore being able to detect objects in spite of general invariant scene transformations.

OBJECT DETECTION; OPTICAL FLOW; LUCAS–KANADE ALGORITHM; SPARSE FEATURES; WATERSHED SEGMENTATION.

Задача определения объекта в сцене, представленной одиночным цифровым изображением, является одной из основных проблем, рассматриваемых в домене компьютерного зрения. Обобщенный подход к решению этой проблемы представляет собой поиск отдельных локальных участков изображения, однозначно соответствующих искомому объекту, так называемых признаков объекта. Существует значительное количество методов обнаружения признаков, устойчивых к различным искажениям, которым может подвергаться изображение, включающих в себя случайный шум, размытие, помехи и сдвиг камеры. Наиболее известные методы в этой области представлены детектором углов Харриса, детектором Ши-Томаси [1], методами SIFT, SURF и другими [2, 3].

Однако представление объекта в виде композиции признаков изображения имеет некоторые существенные недостатки, наиболее известным из которых оказывается изменение формы объекта в ходе инвариантных преобразований в трехмерном пространстве – перемещения, трансляции камеры, поворота угла обзора или масштабирования. Трехмерная природа окружающего мира вынужденно делает невозможным полноценное однозначное описание интересующего объекта с помощью двумерных признаков в отдельной проекции изображения. Существует ряд гибридных попыток обойти эту проблему. Так, дескриптор SIFT представляет собой один из первых высокоэффективных детекторов, инвариантных к масштабированию, а дескрипторы SURF и ORB обеспечивают устойчивость к повороту изображения. Для решения представленной задачи также с успехом используются методы машинного обучения, в которых признаки, как правило, подобраны вручную так, чтобы выражать наиболее существенные детали изображенного объекта. Но ни один из существующих методов не предназначен для обобщенного обнаружения объекта в случае его трансформации внутри сцены.

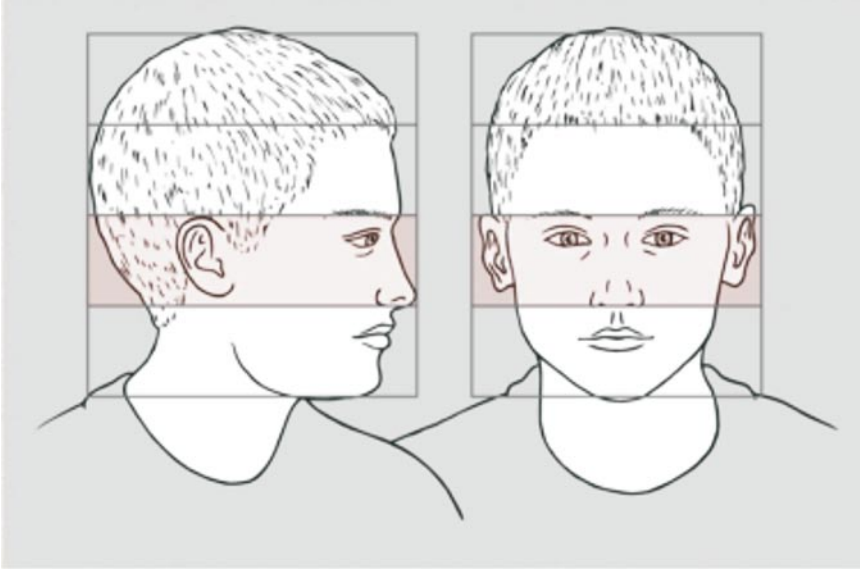

Рис. 1. Пример инвариантного преобразования, меняющего составной набор характерных признаков объекта

Так, если частным случаем рассматриваемой задачи является обнаружение человеческих лиц на фотографиях, адекватным способом выбора признаков представляется сочетание характерных элементов лица, таких как глаза, нос, рот. Простой поворот объекта, однако, способен изменить внешний вид перечисленных признаков, часть из которых пропадает из поля зрения целиком (рис. 1).

Пространство возможных инвариантных трансформаций отдельного объекта ограничено, и алгоритм обнаружения можно расширить экстенсивно, пополнив его необходимыми комбинациями признаков для каждой позиции камеры. Признавая общую жизнеспособность такого подхода, данная статья предлагает альтернативный способ, заключающийся в предварительном составлении модели объекта, представляющей собой карту признаков в трехмерном пространстве. Предложенный метод, таким образом, состоит из двух стадий.

1. Стадия накопления опыта. Объект изучается алгоритмом с различных углов зрения, составляется трехмерная модель с нанесенными на ее поверхность признаками.

2. Стадия обнаружения. Полученную модель можно использовать в качестве детектора для отдельных изображений, сопоставляя ключевые точки модели с признаками изображения. Для этой подзадачи существует ряд широко известных и эффективных методов, таких как «перспектива по *n* точкам» (Efficient PnP [4]). Успешное сопоставление модели позволяет не только обнаружить объект, но и оценить его удаленность от камеры и ориентацию в пространстве.

Для получения трехмерной модели в общем случае можно пользоваться множеством методов, включающих в себя специализированные устройства 3D-сканирования, бинокулярные камеры и дальномеры, но большая часть подобных решений ограничена как по стоимости, так и по области применения. Предлагаем использовать для этой цели извлечение информации о трехмерном положении точек с помощью оптического потока [5] – метод, имеющий потенциально наиболее широкий домен использования.

Использование оптического потока для определения смещений точек

Рассмотрим фрагмент видео: последовательность кадров, ориентированную по оси времени, и некоторые точки изображения, представленные значениями координат и времени $I(x, y, t)$. Тогда при условии, что видео изображает движение в рамках одной и той же сцены, можно ожидать, что в кадре $t + \Delta t$ точка I сместится на некоторые значения Δx и Δy:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t).$$

Применение к этому выражению аппроксимации с помощью ряда Тейлора дает уравнение оптического потока [6]:

$$f_x u + f_y v + f_t = 0,$$

где $f_x = \frac{\partial f}{\partial x}, f_y = \frac{\partial f}{\partial y}, u = \frac{\partial x}{\partial t}, v = \frac{\partial y}{\partial t}$.

Если описанный видеофрагмент получен при помощи плавного движения камеры в статичной сцене, то знание смещений камеры в сторону дает возможность выразить расстояние до точки (рис. 2):

$$d = \frac{Tf}{\Delta x}, \qquad (1)$$

где f – фокусное расстояние камеры; T – шаг камеры; Δx – смещение точки на изображении, значение оптического потока по оси x.

Для получения абсолютного значения дистанции до объекта необходимы, таким образом, следующие данные:

- направление движения камеры. Можно оценить по направлению общего смещения объектов в кадре, с помощью оптического потока (задача усложняется, если в кадре присутствуют автономно движущиеся объекты). Для отдельных классов задач информацию о движении камеры можно получать напрямую от моторных элементов, управляющих движением;
- шаг камеры. Приняв допущение о том, что покадровая съемка ведется с высокой частотой, в упрощенном случае можно считать шаг камеры равным единице.

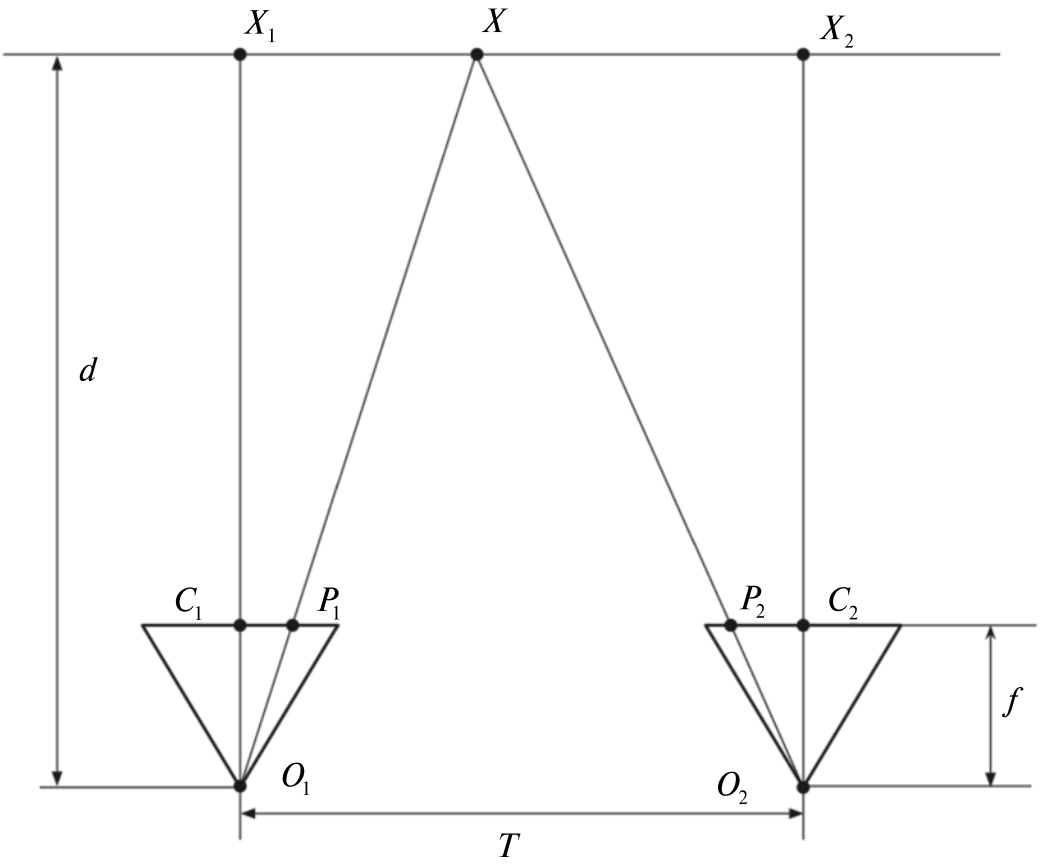

Рис. 2. Горизонтальный сдвиг камеры $\Delta O_1C_1P_1 \Delta O_1X_1X, \Delta O_2C_2P_2 \Delta O_2X_2X$

Отклонения в движении и рывки камеры должны обрабатываться отдельно либо отслеживаться и отбрасываться;

• фокусное расстояние камеры. Для случаев, когда информация о внутренних технических параметрах камеры недоступна, оценить фокусное расстояние можно при повороте камеры (рис. 3).

Для оценки воспользуемся двумя значениями дистанции d и d_{rot}, рассчитанными по формуле (1). Таким образом, координата искомой точки может быть выражена как

$$X = \frac{d\left(x - \frac{w}{2}\right)}{f};$$

$$X_{rot} = \frac{d_{rot}\left(x_{rot} - \frac{w}{2}\right)}{f}.$$

где X, X_{rot} – координата искомой точки в трехмерном пространстве по оси x относительно положения камеры; w – ширина изображения в пикселях.

Если абсолютные координаты центра камеры равны (X_0, Y_0, Z_0), то абсолютную координату X_a искомой точки можно выразить двумя способами:

$$X_a = X + X_0,$$

$$X_a = X_{rot}\cos\alpha - d_{rot}\sin\alpha + X_0,$$

$$\frac{d\left(x - \frac{w}{2}\right)}{f} = \frac{d_{rot}\left(x_{rot} - \frac{w}{2}\right)}{f}\cos\alpha - d_{rot}\sin\alpha.$$

Таким образом, фокусное расстояние можно определить следующим образом:

$$f = \frac{d_{rot}\left(x_{rot} - \frac{w}{2}\right)\cos\alpha - d\left(x - \frac{w}{2}\right)}{d_{rot}\sin\alpha}. \quad (2)$$

Для видеофрагментов, где отсутствуют

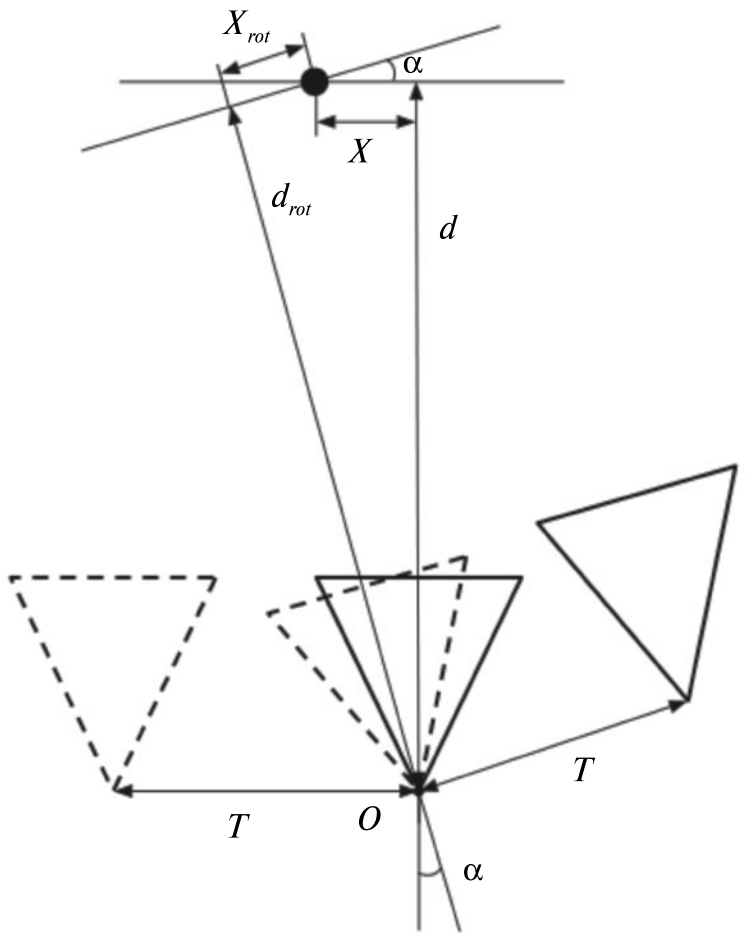

Рис. 3. Поворот камеры в движении на угол α

повороты камеры, значение фокусного расстояния влияет только на пропорцию общей глубины сцены и не приводит к рассогласованиям оценок положения точек. Для таких случаев фокусное расстояние может быть выбрано произвольно.

Идеальный случай

В идеальном случае видеофрагмент должен удовлетворять следующим условиям:

искомый объект изображен на нем достаточно полно и со всех сторон;

камера перемещается одинаковыми шагами;

направление движения камеры всегда известно, при поворотах камеры известен точный угол поворота.

Рассмотрим следующий алгоритм:

1. Получить для первого кадра набор точек f с помощью детектора углов (детектор Харриса или детектор Ши-Томаси). Определить пустой массив R.

2. До тех пор, пока количество кадров не исчерпано:

2.1. Для текущей ориентации (угла) камеры задать стек устойчивых признаков S, инициализировать его полученным набором точек p. Установить $c = 0$.

2.2. До тех пор, пока S не пусто или камера не совершила поворот:

2.2.1. В каждом последующем кадре найти смещения для последнего элемента R с помощью оптического потока Лукаса–Канаде [7].

2.2.2. Те точки, смещение которых найти не удалось, исключаются из S вместе с предшествующими им.

2.2.3. Добавить обнаруженные смещения в S.

2.2.4. Инкрементировать c.

2.3. Полученный массив R содержит некоторое количество оценок смещений для одних и тех же точек, наблюдавшихся в пределах с последовательных кадров. С помощью формулы (1), принимая $T = 1$ и произвольное f, перевести значения смещений в значения глубины d.

2.4. Полученный массив значений dc отфильтровать по Гауссу, оставив значения в пределах трех стандартных отклонений. Получить усредненную оценку координат искомых точек, усреднив значения dc. Добавить полученные точки в общий массив R.

2.5. При повороте камеры с помощью формулы (2) определить фокусное расстояние. При необходимости скорректировать обнаруженные на предыдущем шаге координаты на новое значение f.

3. Для каждого элемента R_i массива R:

3.1. Для каждой точки в R_i найти ближайшего соседа среди других элементов R_k, где $k \neq i$.

3.2. Получить окончательные значения пространственных координат обнаруженных точек, усреднив координаты ближайших соседей.

Результаты работы алгоритма (воссозданные карты признаков) приведены на рис. 4. Для оценки результатов использовались модели-образцы, полученные при помощи компьютерной графики.

Варьирование параметров детектора углов способно дать на выходе как максимально разреженную карту (так, в примере куб представляется исключительно своими углами), так и относительно плотную, в за-

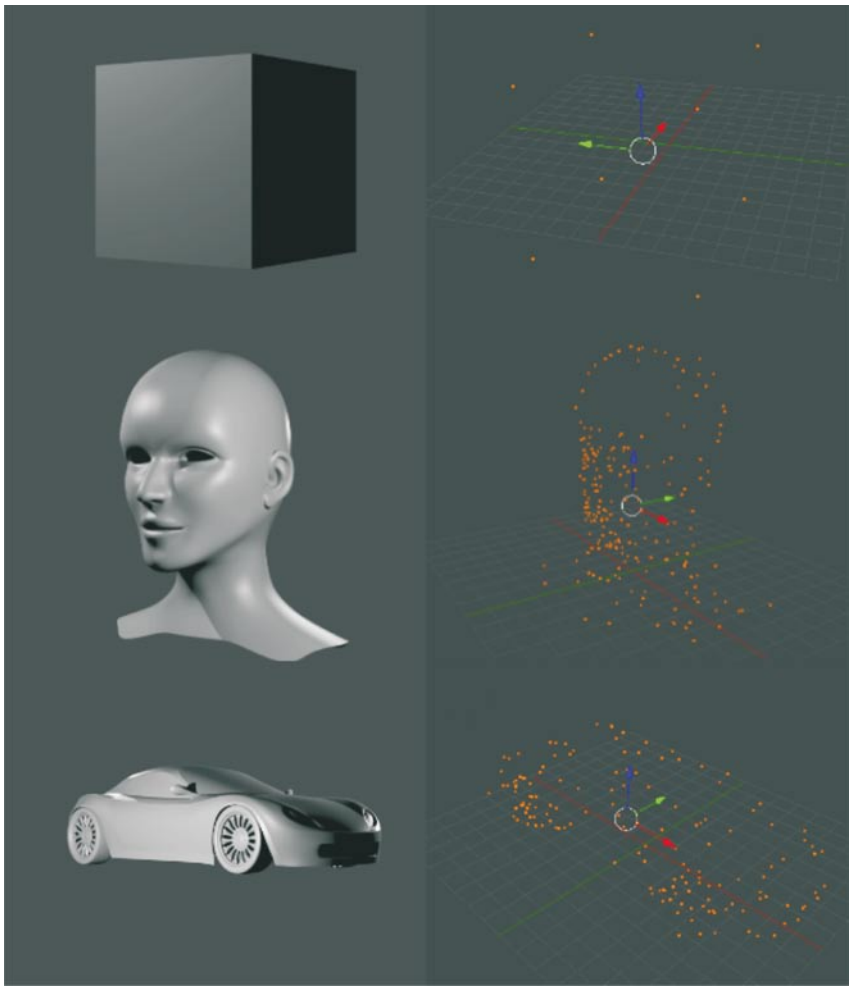

Рис. 4. Некоторые примеры карт признаков, восстановленных по изображениям. Слева – искомый трехмерный объект, справа – карта признаков, представляющая собой разреженное облако точек

висимости от количества точек, обнаруженных детектором углом. Следует отметить, что для задачи классификации и распознавания получение максимально плотного облака точек не только не является желаемым результатом, но и представляет потенциальные проблемы при использовании полученных данных в классификаторе. Идеальная карта признаков должна содержать только то минимальное количество точек, которое с гарантией позволяет идентифицировать объект в кадре.

Прикладное использование и сегментация объектов

При использовании алгоритма в реальных условиях, когда входной видеофрагмент представляет собой данные физического устройства камеры, указанные выше ограничения идеального случая с высокой вероятностью будут нарушаться. Кроме того, съемка естественных сцен, как правило, содержит несколько объектов в одной сцене, что делает необходимым проведение предварительной сегментации сцены и выделение отдельных предметов. Для решения этих подзадач воспользуемся методом расчета плотного оптического потока Фарнебака [8].

В отличие от метода Лукаса–Канаде, метод Фарнебака рассчитывает значение оптического потока для каждой точки изображения. В силу недостаточно точных приближений, он непригоден для воссоздания трехмерной формы объекта, но может использоваться для оценки движения камеры.

На рис. 5 изображены примеры рассчитанного потока Фарнебака для различных случаев движения, где цвет пикселя определяет направление вектора потока в точке. Для оценки движения камеры необходимо определить локальные максимумы потока, после чего, в зависимости от выбранного пространства возможных смещений камеры, использовать подходящий метод принятия решений. В данной работе для шести типов движения (четыре указанных выше плюс повороты камеры влево и вправо) высокую эффективность показал метод логистической регрессии [9], обученной на предварительно составленной выборке, позволивший добиться точности определения направления в 87 %. Результат можно улучшить, принимая во внимание инерцию движения камеры: таким образом, несколько соседних кадров с высокой вероятностью должны принадлежать одному и тому же вектору смещения. Усреднение показателей регрессии с окном в три кадра повышает точность до 94 %.

Для определения длины шага между двумя отдельным камерами, в случае существования соответствующей неопределенности, необходимо отслеживать точки, для которых разреженный поток рассчитывается с высокой надежностью (т. е. направление вектора потока Лукаса–Канаде согласуется с эвристическим предсказанием движения камеры по методу Фарнебака). В ситуации, когда шаги камеры одинаковы,

Рис. 5. Плотный оптический поток для различных случаев движения камеры, последовательно: влево, вправо, вперед и назад.
Цвета обозначают направление движения потока

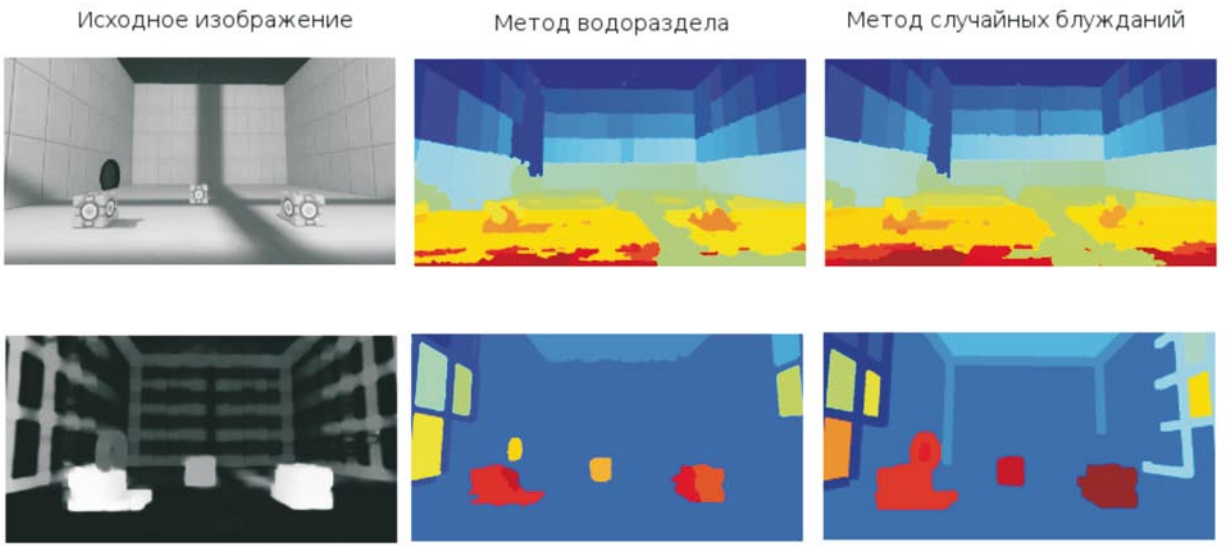

Рис. 6. Сравнение классической сегментации методом водораздела
и совмещенной с потоком Фарнебака

смещения надежных точек должны быть также одинаковы. В случае, когда это условие не выполняется в определенных парах кадров, соответствующий шаг между ними должен быть отмасштабирован на относительную разницу смещения. Для определения надежных точек и их смещений могут использоваться как подмножество устойчивых точек, определяемых детектором Ши-Томаси на первом шаге алгоритма.

Для сегментации сцены на отдельные объекты существует несколько распространенных методов. Учитывая специфику задачи (необходимость обеспечить автономную работу алгоритма и устойчивость к инвариантным условиям среды, таким как освещение), исключим из рассмотрения методы, требующие участия человека (GrabCut), и методы на основе спектральных кластеров (K-means, spectral decision boundary) и рассмотрим методы управляемого водораздела (watershed segmentation) [10] и случайных блужданий (random walker segmentation) [11]. Предлагаем совместить использование сегментации с данными потока Фарнебака, воспользовавшись эффектом параллакса движения, который дает возможность окрашивания объектов в зависимости от их удаления от камеры. Этот метод позволяет снизить погрешность сегментации и избавиться от ложных сегментов, полученных в ходе наличия на объектах поверхностных текстур. Таким образом, различимыми объектами в ходе совмещенной сегментации должны оказываться предметы, испытывающие различные значения смещений в кадре.

Рис. 6 позволяет оценить эффективность предложенного метода. Верхняя часть демонстрирует результат работы приведенных выше алгоритмов на исходном изображении, а нижняя – на изображении, полученном из абсолютных значений потока Фарнебака. Видно, что совмещенная сегментация позволяет с уверенностью выделять отдельные объекты в ситуациях, где классическая сегментация не способна их обнаружить: например, в случае, когда цвет объекта схож с цветом фона.

Предложено решение задачи воссоздания структуры объектов по вторичным данным. Рассмотренный метод может использоваться для задач компьютерного зрения, таких как обнаружение препятствий, 3D-реконструкция сцен, создание эффектов дополненной реальности. Однако основным прикладным применением метода, по нашему мнению, является возможность использования полученных трехмерных

карт признаков для задач распознавания и машинного обучения. Полученные карты признаков могут использоваться в качестве входных данных для интеллектуальных анализаторов, таких как нейронные сети и модели глубокого обучения. Учитывая то, что предложенный метод требует наличия только сенсора захвата движений, построение карт признаков может производиться в течение всего цикла деятельности интеллектуального агента (беспилотного аппарата, камеры слежения и т. д.), играя, таким образом, роль массива визуального опыта. Способность опознавать объект в кадре и сопоставлять его с изученной заранее разреженной трехмерной моделью дает потенциальную возможность полностью решить проблему инвариантных трехмерных преобразований.

Дальнейшим логическим шагом развития метода построения трехмерных карт признаков представляется разработка модели пространственной иерархии, делающая возможным применение отдельной карты для распознавания множества объектов одного класса и допускающая изменчивость отдельных признаков на локальном уровне.

СПИСОК ЛИТЕРАТУРЫ

1. **Tommasini T. et al.** Making good features track better //Computer Vision and Pattern Recognition. Proc. IEEE Computer Society Conf., 1998. Pp. 178–183.

2. **Lowe D.G.** Object recognition from local scale-invariant features // Computer vision. Proc. of the 7th IEEE Internat. Conf. 1999. Vol. 2. Pp. 1150–1157.

3. **Bay H., Tuytelaars T., Van Gool L.** Surf: Speeded up robust features // Computer vision–ECCV. Springer Berlin–Heidelberg, 2006. Pp. 404–417.

4. **Nistér D.** An efficient solution to the five-point relative pose problem // Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2004. Vol. 26. No. 6. Pp. 756–770.

5. **Barron J.L., Fleet D.J., Beauchemin S.S.** Performance of optical flow techniques // Internat. J. of Computer Vision. 1994. Vol. 12. No. 1. Pp. 43–77.

6. **Warren W.H., Hannon D.J.** Direction of self-motion is perceived from optical flow // Nature. 1988. Vol. 336. No. 6195. Pp. 162–163.

7. **Lucas B.D. et al.** An iterative image registration technique with an application to stereo vision // IJCAI. 1981. Vol. 81. Pp. 674–679.

8. **Farnebäck G.** Two-frame motion estimation based on polynomial expansion // Image Analysis. Springer Berlin–Heidelberg, 2003. Pp. 363–370.

9. **Giunta G., Mascia U.** Estimation of global motion parameters by complex linear regression // Image Proc. IEEE Transactions on. 1999. Vol. 8. No. 11. Pp. 1652–1657.

10. **Beucher S. et al.** The watershed transformation applied to image segmentation //Scanning Microscopy-Supplement. 1992. Pp. 299–299.

11. **Grady L.** Random walks for image segmentation // Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2006. Vol. 28. No. 11. Pp. 1768–1783.

REFERENCES

1. **Tommasini T. et al.** Making good features track better, *Computer Vision and Pattern Recognition, Proceedings IEEE Computer Society Conference on*, 1998, Pp. 178–183.

2. **Lowe D.G.** Object recognition from local scale-invariant features, *Computer vision. The Proceedings of the 7th IEEE International Conference on*, 1999, Vol. 2, Pp. 1150–1157.

3. **Bay H., Tuytelaars T., Van Gool L.** Surf: Speeded up robust features, *Computer vision–ECCV*, Springer Berlin–Heidelberg, 2006, Pp. 404–417.

4. **Nistér D.** An efficient solution to the five-point relative pose problem, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2004, Vol. 26, No. 6. Pp. 756–770.

5. **Barron J.L., Fleet D.J., Beauchemin S.S.** Performance of optical flow techniques, *International Journal of Computer Vision*, 1994, Vol. 12, No. 1, Pp. 43–77.

6. **Warren W.H., Hannon D.J.** Direction of self-motion is perceived from optical flow, *Nature*, 1988, Vol. 336, No. 6195, Pp. 162–163.

7. **Lucas B.D. et al.** An iterative image registration technique with an application to stereo vision, *IJCAI*, 1981, Vol. 81, Pp. 674–679.

8. **Farnebäck G.** Two-frame motion estimation based on polynomial expansion, *Image Analysis*. Springer Berlin–Heidelberg, 2003, Pp. 363–370.

9. **Giunta G., Mascia U.** Estimation of global motion parameters by complex linear regression, *Image Processing, IEEE Transactions on*, 1999, Vol. 8, No. 11, Pp. 1652–1657.

10. **Beucher S. et al.** The watershed transformation

applied to image segmentation, *Scanning Microscopy-Supplement*, 1992, Pp. 299–299.

11. **Grady L.** Random walks for image segmentation, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2006, Vol. 28, No. 11, Pp. 1768–1783.

ХУРШУДОВ Артем Александрович – *аспирант кафедры информационных систем и программирования Кубанского государственного технологического университета.*
350072, Россия, Краснодарский край, г. Краснодар, ул. Московская, д. 2.
E-mail: art1783@gmail.com

KHURSHUDOV Artem A. *Kuban State Technological University.*
350072, Moskovskaya Str. 2, Krasnodar, Krasnodar krai, Russia.
E-mail: art1783@gmail.com