# Выделение связного переднего плана на картах глубины с помощью медиального представления сцены

Д.С. Белобородов, Л.М. Местецкий dmitribeloborodov@yandex.ru|mestlm@mail.ru МГУ им. М.В. Ломоносова, Москва, Россия

В данной статье рассматривается задача выделения переднего плана на картах глубины. Предлагается подход к выделению переднего плана, основанный на медиальном представлении силуэтов объектов сцены. Поиск минимального разреза в графе медиального представления позволяет выделять связные области, отвечающие отдельным объектам переднего плана. Предложенный метод работает в режиме реального времени, не требует обучения и является устойчивым относительно движений камеры и объектов. Работа метода демонстрируется на картах глубины, полученных с помощью устройства Microsoft Kinect v2.

Ключевые слова: карты глубины, передний план, медиальное представление, сегментация, Kinect.

# Connected foreground detection on depth maps using medial axes of the scene

D.S. Beloborodov, L.M. Mestetskiy dmitribeloborodov@yandex.ru|mestlm@mail.ru Lomonosov MSU, Moscow, Russia

This article considers the problem of foreground detection on depth maps. Proposed approach to foreground detection is based on medial representation of scene objects silhouettes. Minim cut calculation in medial axes graph allows to detect connected components responsible for separate foreground objects. Proposed method works in real time, it does not require learning procedures and is invariant to camera and objects movement. The method is demonstrated on depth maps obtained from Microsoft Kinect v2 device.

Keywords: depth maps, foreground, medial axes, segmentation, Kinect.

Выделение переднего плана и объектов интереса на изображениях и видео является одной из важных задач в областях компьютерного зрения, отслеживания и распознавания объектов, распознавания жестов и многих других. Эта задача заметно усложняется в условиях пестрого или сложно структурированного фона, при движении камеры, а также при наличии окклюзий среди объектов.

Карты глубины представляют собой массивы, хранящие информацию о расстоянии до камеры для точек сцены. Такая информация позволяет искать передний план в терминах взаимного расположения объектов сцены. На данный момент существуют устройства, позволяющие получать в реальном времени не только цветные изображения, но и соответствующие карты глубины сцен, например, Microsoft Kinect.

Известные подходы к выделению объектов интереса с использованием карт глубины включают:

Построение модели фона по нескольким подряд идущим картам глубины. Пиксели, которые плохо объясняются моделью фона, считаются передним планом. Такие подходы применяются, например, в статьях [4, 6]. Недостаток этого подхода заключается в том, что передним планом считаются только подвижные объекты, а метод чувствителен к движению камеры.

Подходы, основанные на вычислении потока сцены, например [9, 10], обладают теми же недостатками.

Другой подход — выделение в сцене плоскости пола либо полная сегментация сцены на планарные области с помощью алгоритма RANSAC [7]. Этот подход предполагает наличие в сцене планарных областей, то есть, накладывает сильные ограничения на структуру сцены. Например, такой подход применяется в статье [3].

Также для сегментации сцены и выделения объектов интереса применяются различные классификаторы, например, глубинные нейронные сети используются в статье [8]. Такой подход является вычислительно сложным и требует большого объема размеченных данных.

В предыдущей работе [1] был предложен метод выделения переднего плана на картах глубины, основанный на анализе медиального представления силуэтов сцены. Он использует единственную карту глубины в качестве входных данных и не требует обучения, а также инвариантен к движению камеры и не накладывает ограничений на структуру объектов сцены. Данная статья является развитием предыдущей работы. Предложен новый способ определения принадлежности узлов к переднему плану, устойчивый к пропускам в картах глубины, которые появляются из-за особенностей работы сенсоров. Также используется поиск минимального разреза в графе медиального представления для выделения связных компонент переднего плана.

Работа метода демонстрируется на примере карт глубины, полученных с помощью сенсора Microsoft Kinect v2.

# 1. Определение переднего плана

В данной статье используется то же определение переднего плана, что и в предыдущей работе [1], основанное на понятии медиального представления.

#### 1.1. Медиальное представление

Скелетом плоской фигуры называется множество всех центров максимальных вписанных в эту фигуру кругов. Радиальная функция ставит в соответствие каждой точке скелета радиус вписанного в фигуру круга с центром в этой точке. Совокупность скелета фигуры и ее радиальной функции называется медиальным представлением. Для многоуголных фигур скелет представлением. Для многоуголных фигур скелет представлением. Для многоуголных фигур скелет представим в виде плоского графа, узлы которого являются центрами некоторых вписанных кругов, а ребра являются отрезками прямых или парабол. Медиальное представление также может быть построено для бинарного изображения, в этом случае границы областей на изображении приближаются многоугольниками. Алгоритмы построения медиально представления для бинарных изображений описаны в книге [2].



Рис. 1. Пример медиального представления фигуры.

Обозначим как Z(p) глубину в точке p, заданную картой глубины.

Рассмотрим некоторый вписанный круг C узла скелета с центром в точке  $(x_c, y_c)$  и радиусом R. Построим второй круг  $C_{\epsilon}$ , концентричный с кругом C и имеющий радиус  $R + \epsilon$ ,  $\epsilon > 0$ . При таком построении границы фигуры разобьют  $C_{\epsilon}$  на несколько фигур  $A_0, \ldots, A_n$ :  $\bigcup_{i=0}^{n} A_i = C_{\epsilon}$ . При этом только одна из областей  $A_i$  будет полностью содержать в себе C, поскольку этот круг вписанный. Пусть это область  $A_0 : C \subset A_0$ .



**Рис. 2.** Схема построений для определения переднего плана.

**Определение 1.** Будем говорить, что рассмотренный узел лежит на переднем плане с перепадом глубины d, если

$$\exists d > 0, \; \exists \epsilon > 0: \; \max_{p \in C} Z(p) + d \le \min_{p \in C_{\epsilon} \setminus A}$$

Это значит, что между вписанным кругом узла переднего плана и внешними областями, отсеченными границами фигуры, есть некоторый перепад глубины, причем круг находится ближе к камере.

# 2. Структура метода

Предложенный метод выделения переднего плана содержит три подзадачи и имеет следующую структуру:

- Построение медиального представления сцены и разметка некоторых узлов:
  - выделение границ и бинаризация карты глубины;
  - построение медиального представления силуэтов сцены;
  - разметка части узлов медиального представления как передний или задний план.
- Разметка оставшихся узлов медиального представления на передний и задний план с помощью минимального разреза.
- Восстановление переднего плана в форме пиксельной маски.

Первые две подзадачи позволяют найти передний план в форме размеченного графа медиального представления, с помощью которого при необходимости можно вычислить пиксельную маску переднего плана для карты глубины.

# 3. Медиальное представление и первичная разметка

#### 3.1. Бинаризация

На первом шаге на карте глубины вычисляются границы оператором Собеля, затем производится ее бинаризация по следующему правилу: черным (значение 0) помечаются все пиксели, лежащие на границе, а также пиксели с неизвестной глубиной, остальные пиксели помечаются белым (значение 1). Это необходимо для разделения силуэтов объектов сцены.

Затем строится медиальное представление всех силуэтов сцены.

Порог, используемый при бинаризации границ, влияет на метод следующим образом: более низкие значения порога приводят к повышению детализации границы и медиального представления. При этом излишняя детализация может приводить к появлению шума и неустойчивости. Этот параметр можно подобрать эмпирически: таким образом, чтобы можно было различать близко находящиеся объекты при низком уровне шума.



Рис. 3. Бинаризованная карта глубины.

## 3.2. Триангуляцией Делоне

называется триангуляция, для которой выполнено условие Делоне: круг, описанный вокруг любой ее грани, не содержит внутри себя точек триангуляции. Если никакие четыре точки множества не лежат на одной окружности, то триангуляция Делоне единственна. Для каждой точки ее ближайший сосед в евклидовой метрике будет соединен с ней некоторым ребром в триангуляции Делоне. Построение триангуляции Делоне на множестве всех узлов скелета позволяет задать систему соседства на множестве узлов скелета.



**Рис. 4.** Триангуляция Делоне на мнножестве узлов скелета.

#### 3.3. Первичная разметка узлов

Рассмотрим два произвольных узла скелета  $C_1, C_2$  с центрами в точках  $p_1, p_2$  и радиусами  $R_1, R_2$  соот-

ветственно. Для этой пары узлов рассмотрим величины:

 $\Delta Z = |Z(p_1) - Z(p_2)|$  — абсолютный перепад глубин между узлами; здесь Z(p) — глубина точки p, задаваемая картой глубины;

 $\Delta L = ||p_1 - p_2|| - R_1 - R_2$  — зазор между кругами (отрицательный, если круги пересекаются).



**Рис. 5.** Схема построений для первичной разметки узлов.

Определение 2. Будем говорить, что узел  $C_1$  является передним планом относительно  $C_2$  с параметрами a > 0, b > 0, если выполнены следующие условия:

$$\begin{cases} 0 < \Delta L \le b \\ \frac{\Delta Z}{\Delta L} \ge a \\ Z(p_1) < Z(p_2) \end{cases}$$

При этом будем говорить, что узел  $C_2$  является задним планом относительно  $C_1$ . Если хотя бы одно из неравенств не выполнено, ни один из узлов не является передним/задним планом относительно другого.

Первое неравенство накладывает ограничения на зазор между вписанными кругами двух узлов: он должен быть положительным (круги не пересекаются) и не превосходящим параметра b. Необходимость анализировать круги, находящиеся на расстоянии друг от друга, возникает из-за возможности наличия пропусков (неопределнных значений) в карте глубины. Такие области неопределенных значений возникают из-за особенностей работы датчиков глубины и обычно располагаются по краям выступающих объектов — то есть, объектов переднего плана. Параметр b накладывает ограничения на максимальный зазор между кругами ограничивает максимальную ширину области неопределенных значений вокруг объектов. Наблюдения показывают, что в датчиках глубины ширина таких областей действительно ограничена некоторым небольшим значением. Второе неравенство ограничивает отношение перепада глубин и зазора между кругами — то есть, регулирует крутизну перепада между ними. Если перепад достаточно крутой, круги будут являться передним/задним планом относительно друг друга. Третье неравенство означает, что узел С1 находится ближе к камере, чем узел  $C_2$ .

Определение 3. Будем говорить, что узлы  $C_1 u C_2 яв-$ ляются соседями, если они соединены некоторым ребром в построенной триангуляции Делоне, но при этом не соединены ребром в графе скелета сцены.

Если два узла соединены ребром в скелете сцены, то можно считать, что между ними нет значительного перепада глубины.



**Рис. 6.** Белым отмечены пары соседних узлов, являющихся передним/задним планом относительно друг друга.

Определение 4. Будем говорить, что узел С относится к переднему плану сцены, если он является передним планом относительно некоторых своих соседей и не является задним планом относительно всех своих соседей.

Будем говорить, что узел С относится к заднему плану сцены, если он является задним планом относительно некоторых своих соседей.

Используя это правило, узлы скелета сцены размечаются на передний план, задний план и узлы без меток.

#### 4. Разметка оставшихся узлов

Каждый узел скелета соответствует некоторому вписанному в силуэт кругу — фрагменту объекта. На прошлом этапе некоторые фрагменты были помечены как передний план, а некоторые — как задний, но до этого момента не была использовано свойство связности объектов сцены. При разметке часть узлов могла быть пропущена, поскольку все их соседи имеют маленький перепад глубины. Следующий шаг направлен на исправление разметки с использованием свойства связности объектов.

Построим потоковую сеть на основе скелета сцены: пусть граф  $G_0 = (V_0, E_0)$  — скелет сцены ( $V_0$  — вершины,  $E_0$  — ребра). Построим на его основе орграф  $G_1 = (V_1, E_1)$ , где  $V_1 = V_0$ , а  $E_1$  получается из  $E_0$  заменой каждого ребра (u, v) на пару ориентированных ребер (дуг) (u, v), (v, u). Добавим вершины s, t — исток и сток соответственно. Далее все узлы переднего плана  $u_i$  соединим дугами  $(s, u_i)$  с истоком, а узлы заднего плана  $v_j$  соединим дугами  $(v_j, t)$  со стоком. Получим орграф (V, E),  $V = V_1 \cup \{s, t\}$ , E =

 $E \cup \{(s, u_i)|i\} \cup \{(v_j, t)|j\}$ . Зададим пропускные способности всех дуг. Для дуг, идущих от стока или к стоку, определим пропускающую способность как некоторое большое число  $c_0$ . Выбор пропускных способностей для остальных дуг будет влиять на выделяемый передний план. Если для дуг, инцидентных узлам переднего плана, задать пропускную способность  $c_1$ , а для остальных дуг  $c_2$ :  $c_1 < c_2 << c_0$ , то при поиске минимального разреза его граница будет проходить ближе к первичным узлам переднего плана, при этом объединяя некоторые узлы переднего плана.

Получим потоковую сеть G = (V, E, s, t, c), где  $c = \{c(e) | e \in E\}$  — заданные пропускные способности. Найдем минимальный разрез  $(V_s, V_t)$  в этой сети:

$$\begin{cases} \sum_{e=(u,v), u \in V_s, v \in V_t} c(e) \to \min_{V_s, V_t} \\ V_s \cup V_t = V, \ V_s \cap V_t = \emptyset, \ s \in V_s, \ t \in V_t \end{cases}$$

В данной работе для поиска минимального разреза используется алгоритм Бойкова-Колмогорова, описанный в [5].

Таким образом, множество узлов скелета будет разделено на два множества  $V_s, V_t$ . Все узлы, вошедшие в  $V_s$ , отнесем к переднему плану, остальные отнесем к заднему плану. Такой подход позволяет разметить все узлы скелета на передний/задний план с учетом связности объектов.

Будем относить ребро скелета к переднему плану, если оба инцидентных ему узла отнесены к переднему плану.

#### 5. Пиксельная маска переднего плана

Скелетного представления переднего плана сцены уже достаточно для выделения отдельных объектов переднего плана, вычисления их координат и ограничивающих прямоугольников, но для некоторых задач, например, дополненной реальности, может понадобиться пиксельная маска переднего плана. При наличии размеченного скелета, маску переднего плана можно вычислить, разметив множества пикселей, покрываемых каждым ребром переднего плана. Подробно эта процедура описана в предыдущей работе [1].

# 6. Сложность алгоритма

Для карты глубины, содержащей N пикселей, при построении скелета и триангуляции Делоне порождается O(N) элементов, а вычислительная сложность составляет  $O(N \log N)$ . Построенная потоковая сеть содержит O(N) элементов, а сложность поиска минимального разреза составляет  $O(N^4)$ , согласно [5]. Таким образом, минимальный разрез является теоретическим узким местом. На практике основная вычислительная нагрузка приходится на построение скелета, а метод способен работать в режиме реального времени.

#### 7. Реализация алгоритма

Метод был протестирован на картах глубины, полученных с помощью сенсора Microsoft Kinect v2. По результатам экспериментов, метод способен адекватно оценивать передний план.

Метод работает в режиме реального времени: до 20 кадров в секунду (на процессоре Intel Core i3 2.3 ГГц, оперативная амять 8 Гб). Размер карт глубины составляет  $512 \times 424$  пикселей. По сравнению с методом, предложенным в предыдущей работе [1], данный метод более устойчив к наличию пропусков в картах глубины и дает более стабильный результат, хотя он также восстанавливает пиксельную маску переднего плана неточно. Метод адекватно работает с окклюзиями в сцене. Пример работы метода можно видеть на изображении 7.



**Рис. 7.** Передний план сложной сцены. Светлым обозначена пиксельная маска переднего плана.

# 8. Выводы

В данной статье развивается метод выделения переднего плана на картах глубины, предложенный в предыдущей работе [1]. Модифицированный метод сохраняет полезные свойства: работает в реальном времени, использует единственную карту глубины в качестве входных данных. В отличие от других известных подходов, метод не накладывает ограничений на структуру сцены и движение камеры, а также не требует обучения. Кроме того, модификация метода более устойчива к пропускам в данных и позволяет выделять объекты переднего плана в виде связных компонент графа.

Данный метод может применяться в задачах отслеживания и распознавания объектов, дистанционного управления и дополненной реальности в качестве начального этапа для определения областей интереса, а также для определения порядка расположения объектов сцены.

Дальнейшая работа будет включать применение метода в задачах распознавания жестов и отслеживания объектов.

# 9. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов 17-01-00917 и 16-57-45054.

# 10. Литература

- [1] Белобородов Д.С., Местецкий Л.М. Выделение переднего плана на картах глубины с помощью скелетного представления силуэтов объектов // Труды международной конференции ГрафиКон-2016, 2016. – с.378–383.
- [2] Местецкий Л. М. Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры // Litres, 2017.
- [3] Bokaris P., Muselet D., Trémeau A. 3D reconstruction of indoor scenes using a single RGB-D image // 12th International Conference on Computer Vision Theory and Applications (VISAPP 2017), 2017.
- [4] Bondi E., Seidenari L., Bagdanov A.D., Del Bimbo A. Real-time people counting from depth imagery of crowded environments // Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference, 2014. – pp.337–342.
- [5] Boykov Y., Kolmogorov V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision // IEEE transactions on pattern analysis and machine intelligence, 2004. – Vol.26, No.9. – pp.1124–1137.
- [6] Fernandez-Sanchez E.J., Diaz J., Ros E. Background subtraction based on color and depth using active sensors // Sensors, 2013. – Vol.13, No.7. – pp.8895–8915.
- [7] Fischler M.A., Bolles R.C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM, 1981.
  Vol.24, No.6. pp.381–395.
- [8] Gupta S., Girshick R., Arbeláez P., Malik J. Learning rich features from RGB-D images for object detection and segmentation // European Conference on Computer Vision, 2014. – pp.345– 360.
- [9] Hornacek M., Fitzgibbon A., Rother C. SphereFlow: 6 DoF scene flow from RGB-D pairs // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. – pp.3526–3533.
- [10] Jaimez M., Souiai M., Gonzalez-Jimenez J., Cremers D. A primal-dual framework for realtime dense RGB-D scene flow // Robotics and Automation (ICRA), 2015 IEEE International Conference, 2015. – pp.98–104.