

Московский государственный университет имени М.В. Ломоносова

Факультет вычислительной математики и кибернетики

Кафедра математических методов прогнозирования

Белобородов Дмитрий Сергеевич

Сегментация дальностных изображений и видеопоследовательностей

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Научный руководитель:

д.т.н., профессор

Л.М. Местецкий

Москва, 2017

Содержание

1	Введение	3
1.1	Определения и обозначения	5
1.2	Необходимые понятия	10
2	Метод выделения переднего плана	12
2.1	Схема метода	12
2.2	Выделение границ и бинаризация	13
2.3	Построение медиального представления	14
2.4	Построение триангуляции Делоне и определение переднего плана	16
2.5	Расширение переднего плана	19
2.6	Связные компоненты скелета	21
2.7	Восстановление областей переднего плана	22
2.8	Сложность алгоритма	23
3	Вычислительные эксперименты	25
3.1	Обсуждение и выводы	26
4	Заключение	27
	Список литературы	27

Аннотация

В данной работе рассматривается задача сегментации изображений, представленных картами глубины, путем выделения переднего плана. Передний план можно определить как множество областей сцены, расположенных ближе к камере, чем окружающие их объекты. Во многих случаях передний план соответствует искомым объектам в сцене и может быть использован для первичной обработки изображений — нахождения областей интереса. Знание переднего плана может быть также использовано при разрешении окклюзий — ситуаций, когда несколько объектов или частей объектов перекрывают друг друга. Обычно задача выделения объектов интереса решается путем накопления информации о последовательных кадрах видеопоследовательности или обучении классификаторов и детекторов на размеченных данных. Использование карт глубины, содержащих информацию о расстояниях до камеры для каждой точки изображения, позволяет учитывать важную информацию о взаимном расположении объектов. В настоящей работе предлагается альтернативный метод поиска переднего плана, основанный на использовании медиального представления силуэтов объектов сцены, заданной картой глубины. Данный метод способен оценить передний план сцены по единственной карте глубины, без использования процедур машинного обучения. Он устойчив к движениям и положению камеры и способен обрабатывать сложные сцены с множественными окклюзиями и пропусками в данных глубины в реальном времени. Работа данного метода демонстрируется на примере карт глубины, полученных с помощью устройства Microsoft Kinect v2.

1 Введение

Задача сегментации изображений и выделения объектов интереса встречается во многих приложениях, в том числе: дистанционное управление, распознавание жестов и поз, дополненная реальность, робототехника и многих других. В таких задачах обычно требуется обрабатывать последовательности кадров видео в реальном времени, иногда в высоком разрешении. При использовании вычислительно сложных алгоритмов, например, распознавания объектов, для уменьшения объема вычислений могут использоваться методы для первичного определения областей интереса, в которых и производится дальнейшая обработка. К таким методам можно отнести различные методы выделения переднего плана сцены.

Многие алгоритмы предобработки видеопоследовательностей основываются на анализе цветных изображений, полученных с камеры. Использование только цветовой информации делает методы чувствительными к текстуре и цвету объектов. В частности, сегментация изображений становится особенно сложной при наличии пестрого фона.

В последнее время продолжают активно развиваться так называемые RGB-D камеры (или камеры глубины), которые позволяют в реальном времени получать не только цветные изображения сцены, но и соответствующие карты глубины (дальностные изображения). В качестве примера таких устройств можно привести камеры глубины Microsoft Kinect, Asus Xtion и Intel Realsense. Карты глубины представляют собой массивы, содержащие информацию о расстоянии до камеры для каждой точки сцены. Они предоставляют важную дополнительную информацию о дистанциях до каждого объекта и между соседними объектами, что позволяет использовать принципиально новые подходы к выделению объектов интереса в сцене, основанные на их глубине.

Пример карты глубины можно видеть на изображении 1.

Часто используемые подходы определения переднего плана при использовании карт глубины включают:



Рис. 1: Пример карты глубины

Алгоритмы, основанные на накоплении информации о фоне. Модель фона оценивается по нескольким подряд идущим кадрам. Пиксели, плохо описываемые моделью фона, считаются передним планом. Например, подход, описанный в [1] использует как информацию о цвете, так и информацию о глубине. Более простой метод применяется в [2] для обнаружения пешеходов. У этого подхода есть следующий недостаток: такие модели способны находить только движущиеся объекты и накладывают ограничения на движение камеры.

Алгоритмы, использующие обученные классификаторы для выделения объектов интереса. Этот подход может использоваться как для сегментации, так и для непосредственного выделения объектов интереса. К примеру, метод, основанный на глубинном обучении и использующий данные карт глубины, описан в статье [3]. Такие подходы обычно требуют большого объема обучающих данных и являются вычислительно затратными.

Алгоритм RANSAC [4] может быть использован для выделения плоскости пола в сцене, либо для полной сегментации сцены на планарные области. Такой подход используется, например, в статье [5] для реконструкции трехмерной сцены, состоящей из планарных участков. Такой подход предполагает, что в сцене существуют планарные области, что не всегда выполняется для некоторых приложений.

Алгоритмы, считающие поток трехмерной сцены, например, описанные в [6, 7] используются для поиска динамических объектов. Их недостатками также являются невозможность выделения статичных объектов интереса и чувствительность к движению камеры.

В данной работе предлагается альтернативный подход к выделению переднего плана, основанный на медиальном представлении силуэтов сцены, заданной картой глубины. При этом искомый передний план представляет собой множество областей изображения, находящихся ближе к камере относительно локального фона. Примерами объектов переднего плана могут быть: тело человека на фоне интерьера помещения, руки на фоне тела, объекты на столе и так далее. Передний план, определенный таким образом, во многих случаях соответствует объектам интереса в прикладных задачах.

В предложенном алгоритме не вводятся ограничения на положение и движение камеры и структуру объектов сцены. Алгоритм использует в качестве входных данных единственную карту глубины и работает в реальном времени (до 20 кадров в секунду, в зависимости от сложности сцены).

Предложенный алгоритм выделения переднего плана также может быть использован для сегментации окклюзий — перекрывающихся объектов или частей объектов.

Данный алгоритм был реализован и протестирован на картах глубины, полученных с помощью устройства Microsoft Kinect v2.

Далее в главе «Метод выделения переднего плана» подробно и поэтапно описывается предложенный в работе алгоритм, а в главе «Вычислительные эксперименты» оценивается его работоспособность и эффективность реализации.

1.1 Определения и обозначения

Будем считать, что к переднему плану относятся объекты, расположенные ближе к камере, чем объекты локального фона. Это значит, что объекты переднего плана не обязательно являются самыми близкими объектами сцены; они могут располагаться

на разном удалении от камеры. Другими словами, объект переднего плана будет самым близким объектом в некотором выделенном фрагменте сцены.

Карты глубин сами по себе предоставляют полезную информацию о взаимном расположении объектов, но простого локального анализа карты глубины (например, определения глубин на стыках объектов) не будет достаточно для определения переднего плана, поскольку для этой задачи требуется сопоставлять между собой границы объекта, которые могут находиться на существенном удалении друг от друга. Также полезным свойством будет связность полученных областей переднего плана.

Вторая сложность заключается в том, что в картах глубины, получаемых с большинства RGB-D камер, практически всегда есть пропущенные значения. Это связано с общим принципом построения карт глубины современными RGB-D камерами: в них установлен излучатель сигнала и его приемник, разнесенные на некоторое расстояние. Именно эта дистанция между излучателем и приемником позволяет определить глубину точки по временной задержке сигнала или его пространственному сдвигу. Из-за такой конструкции, независимо от конкретной технологии вычисления глубины (например, Time of Flight [8], лежащей в основе функционирования Microsoft Kinect v2), некоторые области сцены могут попадать в тень относительно излучателя сигнала, отчего они становятся невидимыми для приемника. Обычно эти области расположены по границам объектов. Точки с неизвестной глубиной также появляются при выходе объектов за рабочий диапазон камеры или наличии хорошо отражающих поверхностей, расположенных под углом (сигнал не возвращается в приемник). Метод обработки карт глубин должен быть устойчив к наличию областей точек с неизвестной глубиной.

Использование медиального представления силуэтов сцены позволяет производить анализ объектов на глобальном уровне, с учетом их связности. Проблема обработки областей с неизвестной глубиной будет решаться с помощью триангуляции полученного медиального представления.

Для формального определения переднего плана необходимо воспользоваться понятием медиального представления.

Определение 1. *Область евклидовой плоскости — это непустое связное открытое множество точек.*

Определение 2. *Плоской фигурой называется ограниченная замкнутая область, граница которой представима конечным объединением замкнутых контуров, каждый из которых состоит из конечного числа аналитических кривых.*

Определение 3. *Скелетом плоской фигуры называется множество точек, состоящее из центров всех вписанных в эту фигуру кругов.*

Определение 4. *Радиальная функция определена в точках скелета. Она ставит в соответствие каждой точке скелета радиус вписанной окружности с центром в этой точке.*

Определение 5. *Совокупность скелета и радиальной функции фигуры называется ее медиальным представлением.*

Для плоских многоугольных фигур скелет представим в виде планарного геометрического графа. Его вершины являются центрами некоторых вписанных кругов многоугольной фигуры, а ребра являются отрезками прямых или парабол. При этом каждой точке скелета, включая точки ребер, соответствует значение радиальной функции — радиус вписанного круга с центром в этой точке.

Медиальное представление может быть также построено для бинарного изображения. Для этого сначала границы силуэтов бинарного изображения приближаются многоугольниками, после чего для данных многоугольников строится медиальное представление.

На изображении 2 приведен пример скелета бинарного изображения. Синим отмечены узлы скелета, зеленым — его ребра.

Для построения медиального представления многоугольных фигур существуют эффективные алгоритмы. Подробно медиальное представление и алгоритмы его построения описаны в книге [9].

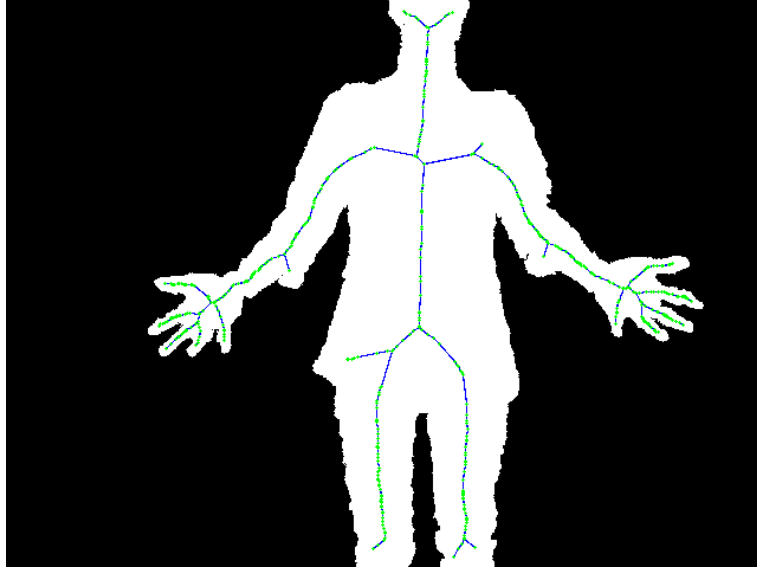


Рис. 2: Пример скелета бинарного изображения

Медиальное представление позволяет аппроксимировать фигуру объединением кругов, поэтому теперь можно определить понятие переднего плана для узлов скелета фигуры. Такое представление, в частности, позволяет учитывать связность объектов.

Рассмотрим узел скелета фигуры с координатами (x_c, y_c) и значением радиальной функции R . Он соответствует некоторому кругу, вписанному в данную фигуру, с центром (x_c, y_c) и радиусом R . Будем называть этот круг C . Теперь построим второй круг, концентричный с кругом C и обладающий радиусом $R + \varepsilon$, $\varepsilon > 0$. Обозначим его C_ε . При этом граница фигуры разобьет круг C_ε на несколько фигур A_0, \dots, A_n :

$$\bigcup_{i=0}^n A_i = C_\varepsilon \quad (1)$$

Причем только одна из них будет полностью содержать круг C , поскольку он вписан в фигуру и пересекает ее границы только в точках касания. Без ограничения общности будем считать, что $C \subset A_0$. Схема произведенных построений показана на изображении 3.

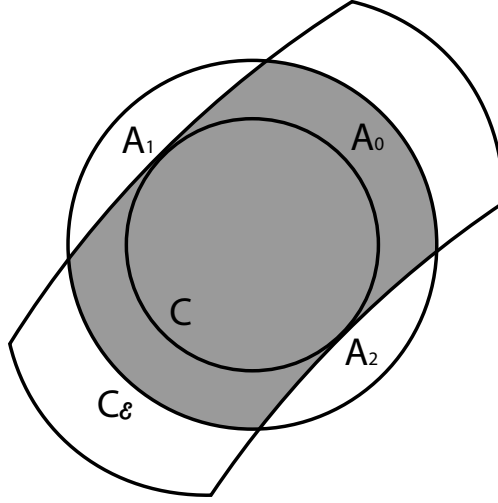


Рис. 3: Иллюстрация к определению переднего плана

Введем обозначение для объединения всех областей, кроме A_0 :

$$A = \bigcup_{i=1}^n A_i \quad (2)$$

Также обозначим глубину в точке $p = (x, y)$ как $Z(p)$ или $Z(x, y)$. Эти значения задаются картой глубины сцены.

Тогда передний план можно определить в терминах перепада глубин:

Определение 6. Будем говорить, что рассмотренный узел скелета относится к переднему плану с перепадом глубины d , если

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

Фактически это означает, что между точками вписанного круга, соответствующего данному узлу, и точками областей, отсеченных границами фигуры, существует перепад глубины величиной не менее d , причем точки круга C находятся ближе к камере.

Как уже было указано выше, в реальных картах глубины существуют точки с неизвестными значениями, что усложняет поиск переднего плана.

Далее будет описан алгоритм, позволяющий приближенно найти передний план сцены на основе скелетного представления объектов сцены, с учетом наличия зон с неизвестной глубиной.

1.2 Необходимые понятия

Предложенный алгоритм использует понятие триангуляции Делоне для поиска перепадов между узлами. Также одним из этапов алгоритма является поиск минимального разреза в потоковой сети с целью выделения связанных компонент переднего плана. Эти понятия будут рассмотрены в данном подразделе.

Определение 7. Пусть на евклидовой плоскости задано множество S из $n \geq 3$ точек, и не все они лежат на одной прямой. Тогда триангуляцией множества S называется плоский прямолинейный граф с вершинами из S , все конечные грани которого являются треугольниками, а все ребра выпуклой оболочки S входят в множество его ребер.

Определение 8. Триангуляция является триангуляцией Делоне, если внутри окружности, описанной вокруг любого построенного треугольника, нет ни одной из точек множества S .

Триангуляция Делоне обладает следующими свойствами:

- Если никакие четыре точки S не лежат на одной окружности, то триангуляция Делоне единственна;
- Для каждой точки из S ее ближайший сосед из S в метрике Евклида будет соединен с этой точкой некоторым ребром в триангуляции Делоне.

Впервые триангуляция Делоне была исследована в работе [10].

Далее рассмотрим задачу о поиске минимального разреза потоковой сети.

Определение 9. *Потоковой сетью называется набор $G = (V, E, c, s, t)$, где (V, E) – ориентированный граф с множеством вершин V и множеством ориентированных ребер (дуг) E , $c = \{c(e) | e \in E\}$ – вектор пропускных способностей для всех дуг, выделенные вершины s, t – исток и сток соответственно. Степень захода вершины s равна 0, степень исхода вершины t равна 0.*

Определение 10. *Разбиение множества V на два подмножества (V_s, V_t) : $V = V_s \cup V_t$, $V_s \cap V_t = \emptyset$, $s \in V_s$, $t \in V_t$ называется (s, t) -разрезом сети G .*

Для (s, t) -разреза обозначим множество дуг, идущих из V_s в V_t как

$$E_{st} = \{e | e = (u, v) \in E, u \in V_s, v \in V_t\} \quad (3)$$

Определение 11. *Пропускной способностью (s, t) -разреза называется величина*

$$c(V_s, V_t) = \sum_{e \in E_{st}} c(e),$$

где $c(e)$ – пропускная способность ребра e .

Определение 12. *Задача поиска минимального разреза потоковой сети ставится как задача поиска (s, t) -разреза сети, обладающего наименьшей пропускной способностью:*

$$c(V_s, V_t) \rightarrow \min_{V_s, V_t}$$

Для поиска минимального разреза в графе существуют различные эффективные алгоритмы. В данной работе для поиска минимального разреза применяется алгоритм Бойкова-Колмогорова, описанный в статье [11].

Также будет использована следующая известная теорема о планарных графах:

Теорема 1 (Эйлер). *Для планарного графа $G = (V, E)$ с множеством вершин V , множеством ребер E и множеством граней F (включая бесконечную внешнюю грань) справедливо равенство:*

$$|V| - |E| + |F| = 2$$

Следствие 1. Для планарного графа, в котором каждая грань ограничена не менее чем тремя ребрами, а каждое ребро разделяет две грани, справедливо неравенство:

$$|E| \leq 3|V| - 6$$

2 Метод выделения переднего плана

2.1 Схема метода

В данном разделе подробно описывается предложенный метод выделения переднего плана. Данный метод использует в качестве входных данных единственную карту глубины, описывающую текущее состояние сцены, и с помощью медиального представления силуэтов сцены оценивает передний план.

Поиск переднего плана включает три подзадачи:

1. Вычисление медиального представления сцены и локальное определение принадлежности к переднему или заднему плану для каждого узла медиального представления;
 - Выделение границ на карте глубины и ее бинаризация. Это необходимо для разделения различных объектов сцены;
 - Построение медиального представления для связных областей бинаризованной карты глубины;
 - Построение триангуляции Делоне на множестве вершин скелета сцены;
 - Определение факта принадлежности к переднему или заднему плану для каждого узла скелета;
2. Определение факта принадлежности к переднему плану для оставшихся узлов скелета и объединение всех узлов переднего плана в связные компоненты. Это делается путем поиска минимального разреза в потоковой сети, построенной по скелету;

3. Вычисление факта принадлежности к переднему плану для каждого пикселя исходной карты глубины. Области переднего плана исходной карты глубины оцениваются по медиальному представлению сцены.

На изображениях 4–8 схематично изображен метод поиска узлов переднего плана медиального представления.



Рис. 4: Карта глубины



Рис. 5: Силуэт

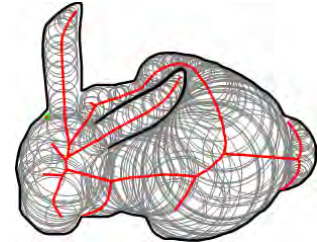


Рис. 6: Медиальное представление

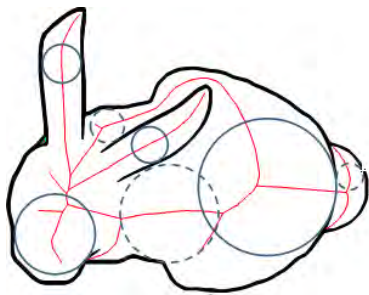


Рис. 7: Передний план



Рис. 8: Передний план

2.2 Выделение границ и бинаризация

Напомним, что карта глубины (дальностное изображение) представляет собой массив, хранящий информацию о расстоянии до камеры (или глубину) для каждой точки сцены. Карта глубины задает глубину $Z(p)$ для точек сцены p . При этом для некоторых точек это значение неизвестно из-за попадания объекта в тень относительно излучателя или выхода за рабочий диапазон камеры.

На первом этапе на карте глубины выделяются границы с помощью оператора Собеля (глубина точек с неизвестными значениями полагается равной нулю). Пиксель считается принадлежащим границе, если его значение после применения оператора Собеля превосходит некоторый заранее определенный порог (параметр алгоритма).

Затем изображение бинаризуется по следующим правилам: черным (значение 0) помечаются все пиксели, принадлежащие границе, а также все пиксели с неизвестной глубиной; белым (значение 1) помечаются все остальные пиксели.

На данном этапе получается бинарное изображение, на котором белым отмечены силуэты объектов, разделенные между собой границами.

На изображении 9 приведен пример бинаризованной карты глубины.

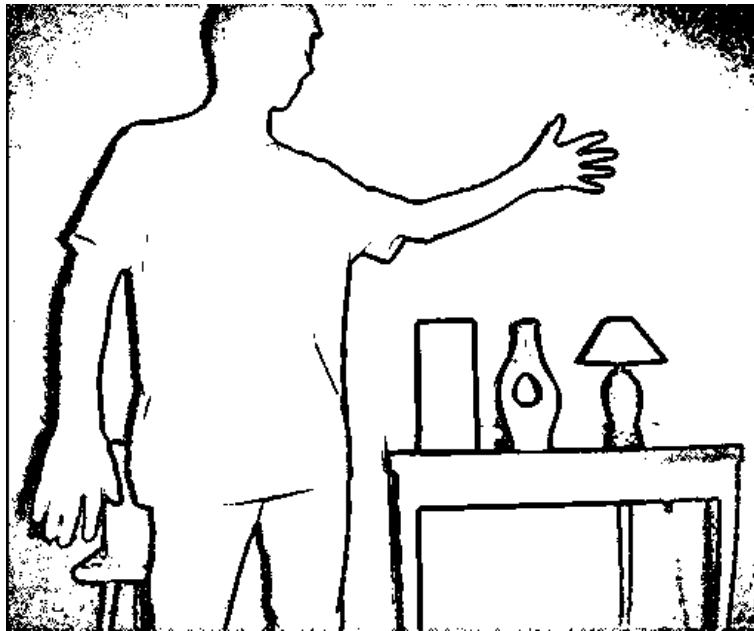


Рис. 9: Бинаризованная карта глубины

2.3 Построение медиального представления

Далее для полученного бинарного изображения строится медиальное представление.

Полученный скелет представляет собой планарный граф (возможно, имеющий несколько компонент связности). При этом каждый узел скелета характеризуется своими координатами внутри изображения и значением радиальной функции. Ребра являются отрезками прямых или парабол, соединяющих узлы.

В процессе построения отбрасываются связные компоненты, имеющие слишком маленькую площадь (меньше некоторого порога), а после построения скелета в нем отбрасываются узлы, имеющие слишком маленькое значение радиальной функции. Это необходимо для повышения устойчивости медиального представления.

Пример скелета сцены можно видеть на изображении 10. Синим отмечены узлы скелета, зеленым отмечены его ребра. На данном изображении скелет сцены для наглядности был наложен на RGB кадр камеры Microsoft Kinect, соответствующий обрабатываемой карте глубины. Все последующие построения, демонстрирующие работу метода на реальных картах глубины, также изображены поверх соответствующих RGB кадров для удобства демонстрации.

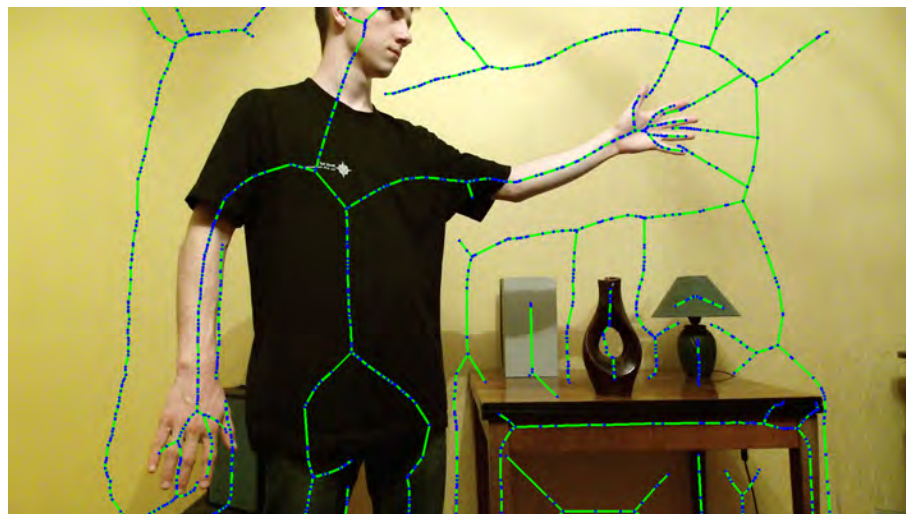


Рис. 10: Скелет сцены

2.4 Построение триангуляции Делоне и определение переднего плана

После построения скелета на множестве его узлов строится триангуляция Делоне. Таким образом, для каждого узла скелета становятся известны соседние по триангуляции узлы.

Пример триангуляции Делоне, построенной на множестве узлов скелета сцены, можно видеть на изображении 11.

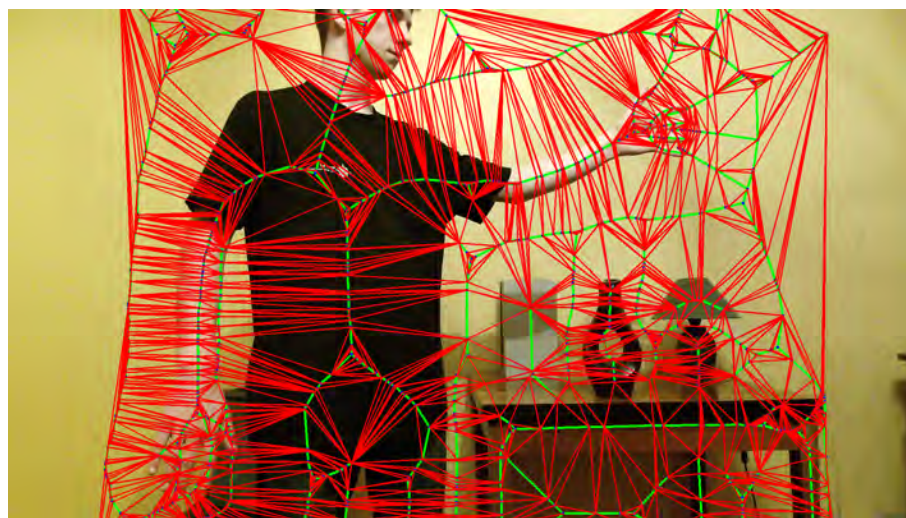


Рис. 11: Триангуляция Делоне скелета сцены

Далее рассмотрим два произвольных узла скелета C_1, C_2 с центрами $p_1 = (x_1, y_1)$, $p_2 = (x_2, y_2)$ и радиусами R_1, R_2 соответственно. Будем рассматривать особенности их взаимного расположения в сцене.

Введем обозначения:

$\Delta Z = |Z(p_1) - Z(p_2)|$ — абсолютный перепад глубин между узлами;

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

Произведенные построения схематично изображены на изображении 12.

Также введем два параметра: $a > 0$ и $b > 0$, смысл которых будет объяснен далее.

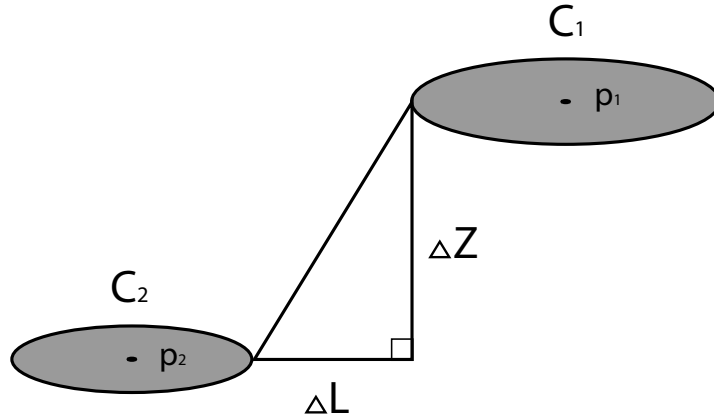


Рис. 12: Иллюстрация к алгоритму определения переднего плана

Будем говорить, что узел C_1 является передним планом относительно узла C_2 с параметрами a, b , если:

$$\begin{cases} 0 < \Delta L \leq b \\ \frac{\Delta Z}{\Delta L} \geq a \\ Z(p_1) < Z(p_2) \end{cases} \quad (4)$$

При этом будем говорить, что узел C_2 является задним планом относительно узла C_1 .

Если хотя бы одно из условий не выполнено, то ни один из этих узлов не является передним или задним планом относительно другого.

Первое неравенство означает, что зазор между кругами, соответствующими узлам, положителен (они не пересекаются) и не превосходит порога b . Таким образом, мы позволяем двум кругам находиться на некотором ограниченном удалении друг от друга. Необходимость анализировать круги, находящиеся на удалении друг от друга, связана с тем, что на карте глубин могут присутствовать области с неизвестной глубиной, соответствующие точкам сцены, попавшим в тень относительно излучателя камеры. Такие области, как правило, появляются на границах выступающих

объектов — то есть, объектов переднего плана. Поэтому два круга, соответствующие соседним объектам, но находящиеся на разной глубине, могут иметь между собой зазор из неопределенных значений. Параметр b регулирует максимальный допустимый зазор между кругами — то есть, максимальную ширину области из неопределенных значений на границе двух объектов. Наблюдения показали, что у датчиков глубины (в частности, Microsoft Kinect) ширина таких областей действительно ограничена некоторым небольшим значением.

Второе неравенство требует, чтобы переход с одного круга на другой был достаточно крутым. Если отношение $\frac{\Delta Z}{\Delta L}$ будет слишком маленьким, это будет означать либо то, что круги имеют маленький перепад по глубине, либо то, что они имеют большой зазор. Мы требуем, чтобы значение этого отношения было достаточно большим.

Третье неравенство требует, чтобы узел C_1 находился ближе к камере, чем C_2 (имел меньшее значение глубины).

Будем говорить, что узлы C_1 и C_2 являются соседними, если они соединены ребром в построенной триангуляции Делоне, но при этом не соединены ребром в скелете сцены (можно считать, что если два узла соединены ребром в скелете, то между ними нет резкого перепада глубины, поэтому такие пары узлов на данном этапе нам не интересны).

Далее приведем правило для отнесения узлов скелета к переднему плану сцены:

Будем относить узел скелета к переднему плану сцены, если он является передним планом относительно некоторых своих соседей, и при этом у данного узла нет соседей, являющихся передним планом относительно него.

Будем относить узел к заднему плану сцены, если он является задним планом относительно некоторых своих соседей.

Такой алгоритм позволяет оценить передний план с перепадом глубины ab .

Очевидно, что узел не может быть одновременно отнесен к переднему и заднему плану сцены. Также в скелете могут присутствовать узлы, которые не являются ни передним, ни задним планом сцены.

Результатом данного этапа является размеченный скелет сцены: узлы скелета помечены как передний план или задний план, либо не имеют меток.

Пример размеченного скелета сцены можно видеть на изображении 13. Здесь синим цветом отмечены узлы переднего плана сцены, зеленым — ребра, инцидентные двум узлам переднего плана; красным отмечены узлы заднего плана сцены, желтым — ребра, инцидентные двум узлам заднего плана; серым отмечены остальные узлы и ребра скелета. Также красным отмечены ребра триангуляции, соединяющие пары узлов, один из которых является передним планом относительно другого.

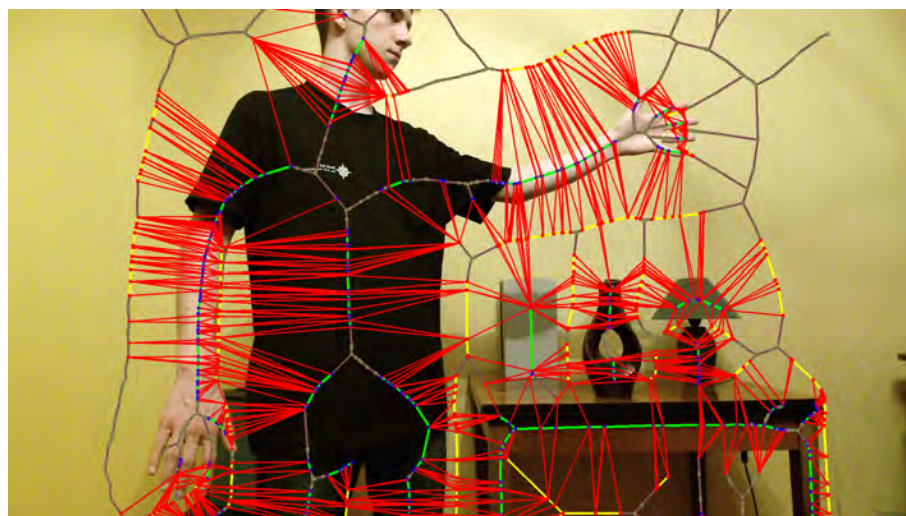


Рис. 13: Размеченный скелет сцены

2.5 Расширение переднего плана

Размеченный передний план, полученный на предыдущем этапе, может содержать узлы, которые лежат на переднем плане, но не были помечены. Это связано с тем, что у некоторых узлов все соседи лежат на том же уровне глубины, поэтому заметного перепада глубины нет. На данном этапе требуется разметить на передний/задний план все узлы, которые не были помечены на предыдущем этапе.

Для этого используется алгоритм поиска минимального разреза сети.

Построим сначала потоковую сеть $G = (V, E, c, s, t)$. Рассмотрим скелет сцены как неориентированный граф и построим на его основе ориентированный граф (V_0, E_0) , заменив каждое неориентированное ребро (u, v) на два ориентированных ребра (дуги) (u, v) и (v, u) . Добавим две дополнительные вершины: исток и сток s, t . Обозначим $V = V_0 \cup \{s, t\}$. Все узлы u_i , помеченные как передний план сцены, соединим дугами (s, u_i) с истоком; узлы заднего плана v_j соединим дугами (v_j, t) со стоком. Обозначим $E = E_0 \cup \{(s, u_i) \mid i\} \cup \{(v_j, t) \mid j\}$.

Осталось лишь задать пропускные способности дуг. Пропускные способности для дуг, ведущих от истока или к стоку, можно положить равными любому достаточно большому числу c_0 . Далее, по-разному определяя пропускные способности дуг скелета, можно выделять области переднего плана разным образом. Например, если для дуг, инцидентных узлам переднего плана, установить значения c_1 , а для остальных дуг $c_2 : c_1 < c_2 \ll c_0$, то минимальный разрез будет проходить ближе к первоначальным областям переднего плана, но при этом будет объединять некоторые отдельные области.

Найдем минимальный разрез (V_s, V_t) построенной сети:

$$\begin{cases} \sum_{e=(u,v):u \in V_s, v \in V_t} c(e) \rightarrow \min_{V_s, V_t} \\ V_s \cup V_t = V, V_s \cap V_t = \emptyset, s \in V_s, t \in V_t \end{cases} \quad (5)$$

В данной работе для этого используется алгоритм Бойкова-Колмогорова, подробно описанный в статье [11]. По сути, минимальный разрез, полученный таким образом, разделит все пары узлов переднего/заднего плана набором дуг с минимальной суммой пропускных способностей.

Теперь к переднему плану отнесем все узлы, попавшие в множество V_s , а все остальные отнесем к заднему плану. Таким образом, окончательно определено множество узлов скелета, соответствующих переднему плану сцены.

Будем относить ребро скелета к переднему плану, если обе вершины, инцидентные этому ребру, были отнесены к переднему плану.

Результатом этого шага является новая разметка скелета сцены. Каждый узел отмечен как передний или как задний план. Иллюстрация 14 показывает пример такой разметки. Здесь синим и зеленым отмечены узлы и ребра скелета переднего плана соответственно, а серым отмечены узлы и ребра заднего плана.

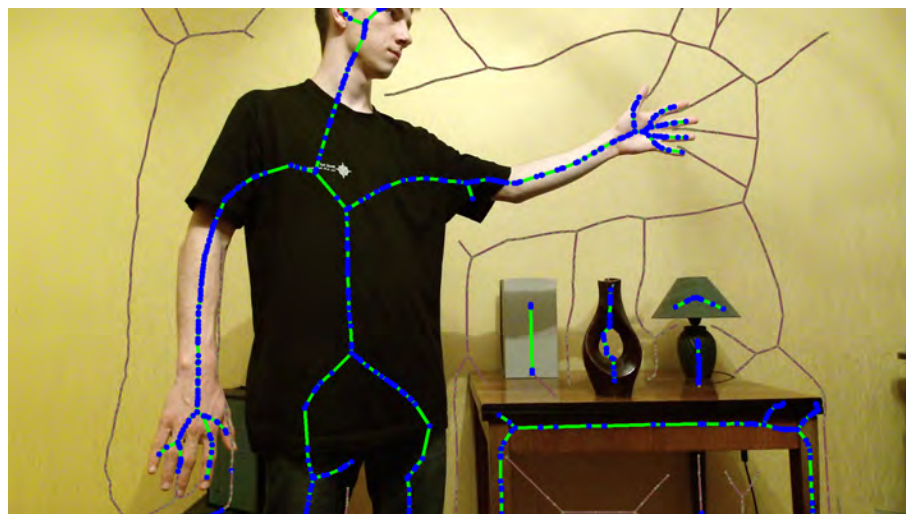


Рис. 14: Передний план

2.6 СВЯЗНЫЕ КОМПОНЕНТЫ СКЕЛЕТА

В результате выделения переднего плана скелет сцены был размечен на передний план и задний план. Теперь, используя простой рекурсивный алгоритм, можно разделить скелет на компоненты связности по типу меток.

При этом стоит заметить, что триангуляция хранит информацию о взаимном расположении соседних узлов скелета — то есть, информацию о том, какие узлы являются передним планом относительно своих соседей. Следовательно, теперь можно для каждой компоненты связности определить, относительно каких соседних (по триангуляции) компонент она является передним планом, а какие соседние компоненты являются передним планом относительно нее. Для этого достаточно рассмотреть все узлы данной компоненты.

Таким образом, для множеств узлов скелета будет определен не только факт принадлежности к переднему плану, но и все соответствующие компоненты заднего плана. Это можно использовать, например, в приложениях дополненной реальности для определения порядка расположения объектов в сцене.

2.7 Восстановление областей переднего плана

К этому моменту передний план был вычислен для скелетного представления сцены, он представлен в виде разметки скелета сцены. Этого уже достаточно, например, для вычисления ограничивающих прямоугольников областей переднего плана.

Для других приложений необходимо будет восстановить области переднего плана для пикселей на карте глубины.

Рассмотрим ребро скелета, инцидентное двум узлам с координатами (x_a, y_a) , (x_b, y_b) и значениями радиальной функции R_1, R_2 соответственно. Каждая точка на ребре, являясь точкой скелета, является также центром некоторого вписанного в фигуру круга. Будем считать, что радиусы таких кругов меняются линейно вдоль ребра от R_a в точке (x_a, y_a) до R_b в точке (x_b, y_b) . Ребра, являющиеся отрезками парабол, для простоты также будем считать прямыми. Это упрощение практически не сказывается на форме скелета в реальных приложениях.

Точку на карте глубины с координатами (x, y) отнесем к переднему плану, если для нее найдется окружность на ребре, отнесенном к переднему плану, такая, что данная точка попадает внутрь окружности.

Обозначим кратчайшее расстояние от точки (x, y) до данного ребра как ρ .

Введем обозначения:

$$t = \frac{(x_a - x)(x_a - x_b) + (y_a - y)(y_a - y_b)}{(x_a - x_b)^2 + (y_a - y_b)^2} \quad (6)$$

$$R = \begin{cases} (1-t)R_a + tR_b, & t \in [0, 1] \\ R_a, & t < 0 \\ R_b, & t > 1 \end{cases} \quad (7)$$

Тогда будем относить точку (x, y) к переднему плану, если для нее найдется ребро скелета, для которого выполнено неравенство:

$$\rho \leq R \quad (8)$$

Здесь величина t параметризует положение проекции точки (x, y) на ребро, а R — радиус вписанной окружности для этой проекции.

Для упрощения вычислений можно для каждого ребра переднего плана рассмотреть множество точек:

$$\begin{cases} \min(x_a - R_a, x_b - R_b) \leq x \leq \max(x_a + R_a, x_b + R_b) \\ \min(y_a - R_a, y_b - R_b) \leq y \leq \max(y_a + R_a, y_b + R_b) \end{cases} \quad (9)$$

и проверить выполнение неравенства $\rho \leq R$ относительно данного ребра только для этих точек. Все точки, для которых это неравенство оказалось верным, помечаются как передний план.

В результате на данном этапе выделяется множество пикселей исходной карты глубины, отмеченных как передний план. Это множество можно в том числе рассматривать как маску переднего плана.

Иллюстрация 15 демонстрирует скелет переднего плана сцены и результат восстановления областей переднего плана для пикселей изображения сцены.

2.8 Сложность алгоритма

Пусть N — число пикселей карты глубины. Оценим сложность алгоритма.

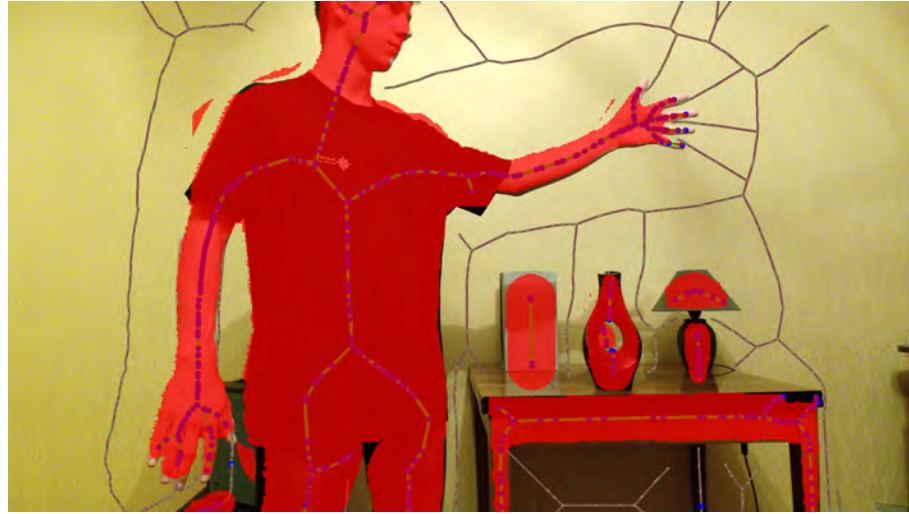


Рис. 15: Маска переднего плана

Бинаризация и выделение границ осуществляются за $O(N)$ операций.

Рассмотрим этап построения контуров связных областей бинаризованного изображения. Число вершин построенных границ не превосходит N . Построенные контуры представимы в виде планарного графа, и по следствию теоремы Эйлера число ребер в них $O(N)$. Всего в контурах $O(N)$ элементов (вершин и ребер). Число операций, затраченное на построение, также равно $O(N)$.

Сложность построения скелета равна $O(N \log N)$. При этом число полученных вершин будет $O(N)$. Число ребер скелета также будет $O(N)$, поскольку скелет является планарным графом.

Построение триангуляции Делоне требует $O(N \log N)$ операций. Число ребер триангуляции будет равно $O(N)$, поскольку триангуляция является планарным графом. Таким образом, определение взаимного расположения соседних узлов потребует $O(N)$ операций.

На этапе построения потоковой сети в скелет с $O(N)$ вершинами и ребрами было добавлено две вершины и не более $O(N)$ ребер, ведущих от истока или к стоку, а также $O(N)$ внутренних ребер. Итоговое число вершин и ребер $O(N)$. Согласно [11],

сложность поиска минимального разреза графа с n ребрами, m вершинами и ценой разреза C составляет $O(n^2mC)$, в случае скелета — $O(N^4)$.

Итоговая теоретическая сложность алгоритма $O(N^4)$. Теоретическим узким местом является поиск минимального разреза. На практике скелет имеет достаточно простую структуру, и поиск минимального разреза происходит существенно быстрее.

3 Вычислительные эксперименты

Предложенный алгоритм был реализован на языке C++ с использованием библиотеки скелетизации Л.М. Местецкого (http://www.machinelearning.ru/wiki/images/d/da/SkeletonMaker_C.rar); для поиска минимального разреза использовался алгоритм Бойкова-Колмогорова [11]; для построения триангуляции Делоне, а также для операций обработки изображений была использована библиотека OpenCV [12].

Алгоритм был применен к картам глубины, полученным с помощью сенсора Microsoft Kinect v2. Пример выделенного переднего плана для сцены с окклюзиями можно видеть на изображении 16.



Рис. 16: Передний план сцены с окклюзиями

Можно заметить, что на данном изображении присутствуют ошибки в восстановлении переднего плана для пикселей — неверно отмеченные области на удалении от объектов. В данном случае появление таких артефактов связано не с особенностями работы предложенного метода, а с необходимостью переводить координаты пикселей из пространства карты глубины в пространство RGB кадра в целях демонстрации. При таком преобразовании координат могут возникать ошибки сопоставления пикселей карты глубины и RGB кадра из-за погрешностей измерения глубины.

Метод корректно определяет передний план сложных сцен и работает в реальном времени (до 20 кадров в секунду, в зависимости от сложности сцены).

Наиболее вычислительно затратным является построение скелета сцены.

3.1 Обсуждение и выводы

Поскольку данный метод использует новый подход к определению переднего плана, сравнить его с другими подходами можно лишь качественно.

Предложенный метод использует единственную карту глубины для вычисления переднего плана, поэтому для его работоспособности не требуется фиксированное положение камеры, что делает возможным его применение в задачах робототехники. Также не накладывается никаких ограничений на структуру объектов сцены. Метод устойчив к наличию областей с неизвестной глубиной, появление которых неизбежно при обработке карт глубины. С помощью такого подхода возможно как представить объекты переднего плана в виде скелета, так и вычислить маску переднего плана.

Недостатком этого метода является приближенное восстановление переднего плана по скелету: при построении скелета часть информации о локальных неровностях объектов не учитывается, и в итоге восстановленная маска переднего плана вычисляется с некоторой погрешностью.

Данный метод может быть применен в задачах компьютерного зрения и распознавания жестов для обнаружения первоначальных областей интереса, а также в

задачах дополненной реальности для сопоставления порядка расположения объектов в сцене и последующей визуализации виртуальных объектов с учетом окклюзий.

4 Заключение

В данной работе был предложен новый подход к выделению переднего плана, основанный на построении медиального представления объектов сцены. Метод работает в реальном времени и позволяет определять передний план сцен с окклюзиями, как в форме скелета, так и в форме пиксельной маски. Метод является устойчивым к движениям камеры и наличию областей с неизвестной глубиной.

Данный метод может быть применен в задачах компьютерного зрения, распознавания жестов и поз, робототехники и дополненной реальности для выделения областей интереса и определения порядка расположения объектов в сцене.

На защиту выносятся:

- Новый подход к определению переднего плана;
- Алгоритм вычисления переднего плана для узлов скелета;
- Метод объединения узлов переднего плана;
- Метод восстановления областей переднего плана на карте глубины.

Дальнейшая работа будет направлена на повышение точности определения переднего плана, полноценную сегментацию сцены на логически разделенные по глубине области, а также применение этого метода в задачах дополненной реальности.

Список литературы

- [1] *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.

- [2] Real-time people counting from depth imagery of crowded environments / E. Bondi, L. Seidenari, A. D Bagdanov, A. Del Bimbo // Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on / IEEE. — 2014. — Pp. 337–342.
- [3] Learning rich features from RGB-D images for object detection and segmentation / S. Gupta, R. Girshick, P. Arbeláez, J. Malik // European Conference on Computer Vision / Springer. — 2014. — Pp. 345–360.
- [4] *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.
- [5] *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.
- [6] *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.
- [7] A primal-dual framework for real-time dense RGB-D scene flow / M. Jaimez, M. Souiai, J. Gonzalez-Jimenez, D. Cremers // Robotics and Automation (ICRA), 2015 IEEE International Conference on / IEEE. — 2015. — Pp. 98–104.
- [8] Time-of-flight cameras: principles, methods and applications / M. Hansard, S. Lee, O. Choi, R. P. Horaud. — Springer Science & Business Media, 2012.
- [9] *Местецкий Л. М.* Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры. — Litres, 2016.
- [10] *Делоне Б. Н.* О пустом шаре // *Изв. АН СССР, серия физ.-матем.* — 1934. — no. 6. — Pp. 793–800.

- [11] *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.
- [12] OpenCV library [Электронный ресурс]. — Электрон.дан. — URL: <http://opencv.org/>.