**МИНОБРНАУКИ РОССИИ**

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

**«Ярославский государственный университет им. П. Г. Демидова» (ЯрГУ)**

Кафедра дискретного анализа

Реферат

на тему «Построение 3D модели на основе стереопары»

(специальность Прикладная математика и информатика)

|  |
| --- |
| ***Место***  ***для краткой рецензии научного руководителя с оценкой, датой проверки реферата и подписью научного руководителя*** |

Научный руководитель:

Шовгенов Д.А.

Ярославль

2019 г.

Оглавление

[Введение 3](#_Toc6157497)

[Обзор существующих решений 5](#_Toc6157498)

[Основы геометрии 6](#_Toc6157499)

[Эпиполярная геометрия 6](#_Toc6157500)

[Проективная геометрия 6](#_Toc6157501)

[Устройство камеры и стереокамеры 9](#_Toc6157502)

[Предварительная обработка изображений 12](#_Toc6157503)

[Преобразование в полутоновое изображение. 12](#_Toc6157504)

[Сегментация изображения. 12](#_Toc6157505)

[Размытие изображения. 13](#_Toc6157506)

[Выделение границ. 13](#_Toc6157507)

[Ректификация изображений. 14](#_Toc6157508)

[Поиск сопряженных точек 16](#_Toc6157509)

[Локальные алгоритмы. 16](#_Toc6157510)

[Глобальные алгоритмы. 19](#_Toc6157511)

[Алгоритм полуглобального сопоставления. 20](#_Toc6157512)

[Вычисление попиксельной стоимости 21](#_Toc6157513)

[Суммирование попиксельных стоимостей 22](#_Toc6157514)

[Вычисление карты диспаратности 22](#_Toc6157515)

[Заключение 24](#_Toc6157516)

[Используемая литература 25](#_Toc6157517)

# 

# Введение

Для человека визуальный канал получения информации один из самых главных, потому что зрение даёт приблизительно 80% всей информации. Так выходит и с компьютерами: системы, использующие визуальную информацию, владеют большей информативностью. Системы технического зрения относятся к группе бесконтактных пассивных информационных средств и занимают особое место.

Системы машинного зрения в современном мире находят повсеместное применение, поскольку связано с развитием технологий, повышением доступности высокопроизводительной вычислительной техники и усовершенствованием устройств получения цифровой информации. Сегодня не существует совершенно работающих алгоритмов и самих систем машинного зрения, которые используются для решения универсальных задач. Обычно в каждом отдельном случае разрабатывается алгоритм, подходящий для решения поставленной задачи с использованием популярных способов и алгоритмов.

В общих чертах системы технического зрения представляют собой взаимосвязанную последовательность технологий, включающую нижеследующие шаги:

- получение изображения предмета либо объекта видеокамерой;

- обработка (оцифровка) изображения;

- анализ цифрового изображения;

- перемещение камеры или изображения в пространстве.

Визуальная информация в системах технического зрения воспринимается и преобразуется в электрические сигналы через оптоэлектронные преобразователи, являющиеся главными преобразователями из лучей в электрические сигналы.

Системы технического зрения используются в различных сферах деятельности человека и выполняют много функций, что не может не привлечь внимания людей, заинтересованных в прогрессе, улучшении качества и упрощении своей работы. Например: трекинг объекта, помощь водителю, медицина.

Стоит отметить, что совместный прогресс в области компьютерного зрения связан не только с развитием аппаратного обеспечения. Оно открывает возможности для применения вычислительно затратных способов обработки изображений, однако сами эти методы нуждаются в разработке. За последние 10–15 лет были доведены до эффективного практического использования методы сравнения изображений трехмерных сцен, способы восстановления карт дальности на основе стереозрения, способы обнаружения и распознавания лиц и т. д. Общие принципы решения таких задач пополнились кругом специфических технических деталей и математических приемов, сделавших эти методы достаточно успешными в применении.

Благодаря развитию технологий и алгоритмов стало достижимым создание трехмерных моделей с помощью камеры. Построение моделей вручную сопряжено с крупными затратами, а применение специальной аппаратуры для сканирования трехмерных объектов не всегда возможно из-за того, что глубина сканирования такими устройствами ограничена. Совмещение изображений, позволяет человеку получить информацию о расстоянии до объектов по их расхождениям (disparity). Эта идея может использоваться в алгоритмах обработки изображений. Но чувство человека активное (то есть параметры оптической системы настраиваются под изображение) глаза вращаются в глазницах, меняется фокусное расстояние. Система пассивного стереозрения, обычно, включает в себя 2 камеры. В статье приводятся методы получения трехмерной модели сцены по стереопаре, т. е. парой плоских изображений объекта, полученных с немного отличающихся позиций.

В современном мире число актуальных приложений технического зрения непременно растет. Решение задач, связанных с анализом видеоданных, становится реальным. Такой востребованной задачей является конвертация видео 2D в 3D.

Системы технического зрения будут развиваться и находить применение в других областях деятельности человека все с большей силой. Совершенствование технологий в этой сфере приведет к более полному контролю жизненно главных аспектов общества. Надеюсь, что новшества не заставят себя ждать, и мы в совершенной мере насладимся их качеством и удобством.

# Обзор существующих решений

Техническая часть гаджетов и алгоритмы развились уже настолько, что построение 3D моделей на основе стереопары используется везде. С помощью них выполняется довольно много задач. Рассмотрим некоторые из них.

Во-первых, распознавание жестов. Управлять устройством не касаясь к нему весьма удобный способ, поэтому мы последнее время стали чаще это замечать. Скажем, играть в игры значительно удобнее и полезнее двигаясь при этом. Ради этого компания XBOX создала сенсор Kinect, благодаря которому это стало возможным. Для большей точности и небольших требований к «железу» в этом устройстве используется датчик глубины, то есть в данном решении стереозрение не основной подход. Также, жестами можно управлять аудиосистемой автомобилей, что уже давно используется в топовых моделях Volkswagen, BMW, Mercedes.

Во-вторых, более быстрое построение 3D моделей. Все большей популярностью начинается пользоваться объемное представление данных. Для отображения объемных объектов необходимо построение их трехмерных моделей. И эту работу естественно хотелось бы упростить либо вообще автоматизировать. В этом нам может помочь стереозрение. Уже существует множество приложений способных это выполнять. Например, Photoscan или VSFM.

# 

# Основы геометрии

## Проективная геометрия

В геометрии стерео зрения вескую роль играет проективная геометрия. К проективной геометрии имеется несколько подходов: геометрический, аналитический, алгебраический. Далее в основном пригодится понимание аналитического подхода к проективной геометрии, и именно он и рассмотрен ниже.

**Точки проективной плоскости**. Рассмотрим двухмерное проективное пространство[2] (которое также называется проективной плоскостью). В то время как на обыкновенной Евклидовой плоскости точки описываются парой координат , на проективной плоскости точки описываются трехкомпонентным вектором . При этом для любого ненулевого числа a, векторы и соответствуют одной точке. А нулевой вектор не подходит никакой точке и выкидывается из рассмотрения. Такое описание точек плоскости называется однородными координатами (homogeneous coordinates).

Точкам проективной плоскости разрешено сопоставить точки Евклидовой плоскости. Векторупри w≠0 сопоставим точку Евклидовой плоскости с координатами . Если w = 0, т.е. вектор имеет вид , то станем говорить, что данная точка в бесконечности. Следовательно, проективную плоскость можно рассматривать как Евклидовую плоскость, дополненную точками из бесконечности.

Перейти от координат к Евклидовым можно путем деления вектора на последнюю компоненту и последующего её отбрасывания → . А от Евклидовых координатперейти к однородным можно путем добавления вектора единичкой: →

**Прямые на проективной плоскости**. Любая прямая на проективной плоскости описывается вектором . Вектор, описывающий прямую, определен с точностью до ненулевого множителя. При этом уравнение прямой будет иметь вид: .

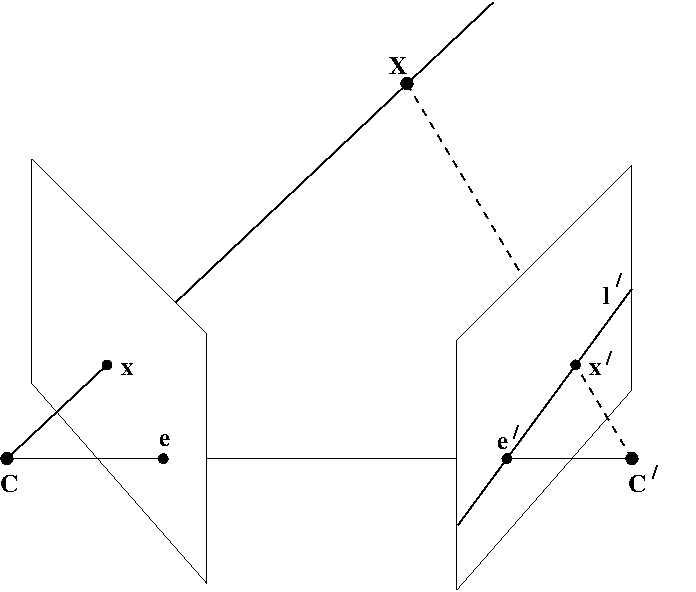
В случае, когда *a*2 + *b*2 ≠ 0 мы имеем аналог обыкновенной прямой ax + by + c = 0. А вектор (0,0,*w*) соответствует прямой лежащей в бесконечности.

**Трехмерное проективное пространство**. Аналогично с проективной плоскостью, точки трехмерного проективного пространства определяются вектором однородных координат . Также для любого ненулевого числа a, вектора и одной и той точке.

Между точками трехмерного Евклидова пространства и трехмерного проективного пространства также можно установить соответствие. Вектору однородных координат при w ≠ 0 соответствует точка Евклидова пространства с координатами . А про точку с вектором однородных координат вида говорят, что она лежит в бесконечности.

**Проективное преобразование**. Еще одна вещь, которая будет нужно для дальнейшего изложения — это проективные преобразования (homography, projective transformation — в англ. литературе). С геометрической точки зрения, проективное преобразование — это обратимое преобразование проективной, которое переводит прямые в прямые. В координатах, проективное преобразование выражается в виде невырожденной квадратной матрицы H, при этом координатный вектор x переходит в координатный вектор x' по формуле: .

## Эпиполярная геометрия

Применение проективной геометрии в случае стерео задачи дает эпиполярную геометрию. Так давайте разберемся в ней.

Пускай есть 2 камеры, как изображено на рисунке 2. C — центр 1-ой камеры, C' — центр 2-ой камеры. Точка пространства X проецируется в x на плоскость изображения левой камеры и в x' на плоскость изображения правой камеры. Прототипом точки x на изображении левой камеры является луч xX. Данный луч проецируется на плоскость 2-ой камеры в прямую *l*', называемую эпиполярной линией. Образ точки X на плоскости изображения 2-ой камеры обязательно лежит на эпиполярной линии *l*'.

Поэтому, каждой точке x на левом изображении соответствует эпиполярная линия *l*' на правом изображении. При этом пара для x на правом изображении имеет возможность лежать лишь на соответствующей эпиполярной линии. Всякой точке x' на правом изображении подходит эпиполярная линия l на левом.

Эпиполярную геометрию[1] используют для поиска стереопар, и для проверки того, что две точки могут являться стереопарой (т.е. проекцией некоторой точки пространства). Эпиполярная геометрия имеет очень простую запись в координатах. Пускай есть две откалиброванных камеры, и пусть x — однородные координаты точки на изображении одной камеры, а x' — на изображении 2-ой. Есть такая матрица F размера 3 × 3, что две точек x, x' считаются стереопарой только, когда:

Матрица F называется фундаментальной матрицей (fundamental matrix). Ее ранг равен 2, она определена с точностью до ненулевого множителя и находится в зависимости от матриц исходных камер P и P'. В случае, когда матрицы камер имеют вид фундаментальная матрица вычисляется по формуле:

Через фундаментальную матрицу вычисляются уравнения эпиполярных линий. Для точки x, вектор, задающий эпиполярную линию, станет иметь вид , а уравнение эпиполярной линии: *l*'*T* *x*' = 0. Аналогично для точки x', вектор, задающий эпиполярную линию, станет иметь вид *l* = *FT*x'.

Кроме фундаментальной матрицы, есть еще такое понятие, как существенная матрица (essential matrix): *E* = *K*'*T* *F* *K*. Когда матрицы внутренних характеристик будут единичными, существенная матрица станет соответствовать фундаментальной. По существенной матрице можно восстановить положение и поворот 2-ой камеры относительно 1-ой.

Рассмотрим погрешности восстановления эпиполярных прямых[5] (ЭП). Уравнения ЭП будем задавать в общем виде:

, который просто привести к нормальному виду сделав нормирование. В данном случае погрешность нормированного коэффициента A3 станет погрешностью определения расстояния от начала координат до ЭП, а погрешности других коэффициентов – погрешностями угла меж осями координат и ЭП. Из *x*'*T* *F* *x* = 0 коэффициенты уравнения ЭП

Векторами , i = 1..3 станем обозначать столбцы фундаментальной матрицы. Из (2) построим приращение коэффициентов ЭП, подразумевая, что x никак не варьируется.

Из (3) получаем, что погрешность находится в зависимости от положения отметки на кадре, однако нас будет интересовать среднеквадратичная ошибка для изображения в целом:

Пускай отметки на изображении распределены случайным образом и известна нормированная функция распределения υ(u,v). Тогда можно перейти от суммы к интегралу:

(4)

Допустим для простоты, что отметки имеют нормальное распределение. Тогда после вычисления интеграла (4) получим:

# Устройство камеры и стереокамеры

CCD-камеры[3] имеют следующую модель, называемую проективной камерой (projective camera). Проективная камера определяется центром камеры, главной осью — лучом, исходящим из центра камера в сторону, куда смотрит камера, плоскостью изображения — плоскостью на которую производится проецирование точек.

Формула проецирования: , где X — координаты точки пространства, x — координаты точки плоскости, P— матрица камеры размера 3 × 4.

Матрица , где K — верхняя треугольная матрица внутренних характеристик камеры размера 3 × 3, R — ортогональная матрица размера 3 × 3, характеризующая поворот камеры относительно глобальной системы координат, I — единичная матрица размера 3 × 3, вектор c — координаты центра камеры, а t = −Rc.

Стоит отметить, что матрица камеры определена с точностью до постоянного ненулевого множителя, который никак не изменит итогов проецирования точек по формуле *x* = *PX*. Однако этот постоянный множитель обычно выбирается чтобы сетка камеры имела вышеописанный вид.

В самом простом случае, когда центр камеры лежит в начале координат, основная ось камеры сонаправлена оси Cz, оси координат на плоскости камеры имеют одинаковый масштаб, а центр изображения имеет нулевые координаты, матрица камеры будет P = K[I|0], где

У реальных CCD камер пикселы обычно незначительно отличаются от квадратных, а центр изображения имеет ненулевые координаты. В этом случае матрица внутренних характеристик станет:

Коэффициенты *f*, α*x*, α*y*  — называются фокусными расстояниями камеры.

Кроме этого, из-за неидеальности оптики, на изображениях, присутствуют искажения-дисторсии (distortion)[4]. Они имеют нелинейную математическую запись:

где k1, k2, p1, p2, k3 — коэффициенты дисторсии, являющиеся параметрами оптической системы; r2 = x'2 + y'2; (x', y') — координаты проекции точки относительно центра изображения при квадратных пикселях и отсутствии искажений; (x″, y″) — искаженные координаты точки относительно центра изображения при квадратных пикселях.

Дисторсии никак не зависят от расстояния до объекта, а находятся в зависимости лишь от координат точек, в которые проецируются пиксели объекта. Поэтому для компенсации дисторсий обычно производится преобразование исходного изображения. Преобразование одно для всех изображений при условии постоянства фокусного расстояния (математически — одинаковые матрицы внутренних характеристик). Если известны внутренние характеристики камеры и коэффициенты дисторсии, камера откалибрована.

Об определении трехмерных координат наблюдаемых точек можно говорить, когда имеется не менее 2 камер. Пусть есть 2 камеры, заданные своими матрицами P и P' в некоторой системе координат. Центры камер не должны совпадать.

Часто, система координат выбирается так, что матрицы камер равны *P* = *K*[*I*|0], *P*' = *K*'[*R*'|*t*']. Это всегда можно сделать, если выбрать начало координат совпадающее с центром 1-ой камеры, и направить ось Z вдоль её оптической оси.

Калибровка камер обычно производится, с помощью многократной съемки калибровочного шаблона. Дальше составляются и решаются системы уравнений, связывающие координаты проекций, матрицы камер и положения точек шаблона в пространстве.

# Предварительная обработка изображений

Предварительная обработка изображений стереопары включает в себя следующие шаги:

1. Преобразование в полутоновое изображение.

2. Сегментация изображения.

3. Размытие изображения.

4. Выделение границ.

5. Ректификация изображения.

Рассмотрим более подробно последовательность действий предварительной обработки изображения.

Преобразование в полутоновое изображение. Многие видеокамеры выдают цветное изображение в формате RGB. Преобразование цветного изображения в полутоновое выполняется по формуле

где r, g, b – цветовые компоненты исходного изображения, а y – выходное значение для пикселя изображения в оттенках серого.

Сегментация изображения. Сегментация изображения подразумевает разбиение изображения на множество его областей, для того чтобы можно было работать с объектами отдельно. Сегментация изображения выполнялась с помощью алгоритма MeanShift[6]. MeanShift группирует объекты с близкими признаками. Пиксели со схожими признаками объединяются в один сегмент, на выходе получаем изображение с однородными областями.

Например, в качестве координат в пространстве признаков можно выбрать координаты пикселя (x, y) и компоненты RGB пикселя. Изобразив пиксели в пространстве признаков, можно заметить сгущения в определенных местах.

Чтобы легче было описывать сгущения точек, вводится функция плотности:

- вектор признаков i-ого пикселя, d - количество признаков, N - число пикселей, h - параметр, отвечающий за гладкость, - ядро. Максимумы функции расположены в точках сгущения пикселей изображения в пространстве признаков. Пиксели, принадлежащие одному локальному максимуму, объединяются в один сегмент. Получается, чтобы найти к какому из центров сгущения относится пиксель, надо шагать по градиенту для нахождения ближайшего локального максимума.

При выборе в качестве признаков координат пикселей и интенсивностей по цветам в один сегмент будут объединяться пиксели с близкими цветами и расположенные недалеко друг от друга. Соответственно, если выбрать другой вектор признаков, то объединение пикселей в сегменты уже будет идти по нему. Например, если убрать из признаков координаты, то небо и озеро будут считаться одним сегментом, так как пиксели этих объектов в пространстве признаков попали бы в один локальный максимум.

Размытие изображения. В связи с тем, что алгоритмы выделения границ чувствительны к наличию «шума» на изображении, на этом шаге алгоритма выполняется размытие изображения с помощью фильтра Гаусса[7]. Размытие по Гауссу — это характерный фильтр размытия изображения, который использует нормальное распределение для вычисления преобразования, применяемого к каждому пикселю изображения. Уравнение распределения Гаусса

где r — это радиус размытия, , σ — стандартное отклонение распределения Гаусса. Формула задает поверхность, имеющей вид концентрических окружностей с распределением Гаусса от центральной точки. Пиксели, где распределение отлично от нуля используются для построения матрицы свертки, которая применяется к исходному изображению. Значение каждого пикселя становится средне взвешенным для окрестности. Исходное значение пикселя принимает наибольший вес (имеет наивысшее Гауссово значение), и соседние пиксели принимают меньшие веса, в зависимости от расстояния до них.

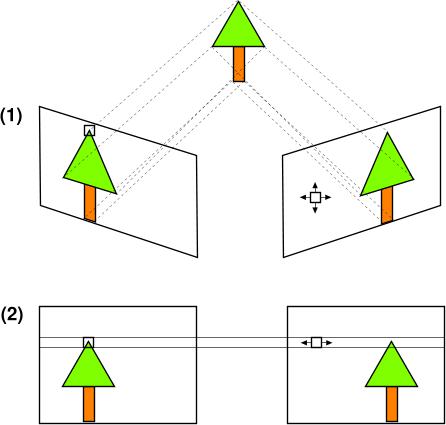
Выделение границ. Выделения границ выполняется методом Канни[8]. Этот метод зарекомендовал себя как эффективный алгоритм выделения границ. В этом методе выполняется многошаговая обработка изображения, которая включает сглаживание обрабатываемого изображения, поиск границ на основе градиентного метода, удаление ложных максимумов и формирование контуров. Поиск градиентов выполняется с помощью оператора Собеля[9].

Оператор Собеля - это дискретный дифференциальный оператор, вычисляющий приближенное значение градиента яркости изображения. Результатом применения оператора Собеля в каждой точке изображения является либо вектор градиента яркости в этой точке, либо его норма. Оператор Собеля основан на свёртке изображения небольшими целочисленными фильтрами в вертикальном и горизонтальном направлениях, поэтому его относительно легко вычислять. С другой стороны, используемая им аппроксимация градиента достаточно грубая, особенно это сказывается на высокочастотных колебаниях изображения.

Для подавления ложных максимумов пикселями границ объявляются пиксели, в которых достигается локальный максимум градиента в направлении вектора градиента. Значение направления должно быть кратно 45°. Удаляются пиксели, которые не являются частью края, следовательно, останутся только тонкие линии.

Следующий шаг — применение порога, чтобы определить находится или нет граница в данной точке изображения. Чем меньше порог, тем больше границ будет находиться, но тем более восприимчивым к шуму станет результат, выделяя лишние данные изображения. Наоборот, высокий порог может проигнорировать слабые края или получить границу фрагментами. Выделение границ Канни использует два порога фильтрации: если значение пикселя выше верхней границы – он принимает максимальное значение (граница считается достоверной), если ниже – пиксель подавляется, точки со значением, попадающим в диапазон между порогов, принимают фиксированное среднее значение (они будут уточнены на следующем этапе).

Ректификация изображений. Ректификация (rectification) стереопары - преобразование, при котором и правое, и левое изображение проецируются на плоскость, параллельную базовой линии (линии, соединяющей оптические центры камер). Тогда если (x, y) точка левой проекции, а (, ) соответствующая ей точка правой проекции, то y = .



Изображения с камер (1) проектируем на плоскость, параллельную базовой линии (2). Каждая строчка левого изображения соответствует той же строчке правого.

Получив матрицу P для каждой из камер, а также матрицу F, мы можем перейти к ректификации изображений (матрицы смотреть в 1.3). Под ректификацией в данном случае подразумевается установление соответствий между эпиполярными линиями на стереопаре, вдоль которых будет выполняться поиск в ходе работы алгоритма стереосопоставления. Для нахождения этих сечений мы использовали следующий метод. Исходя из общей особенности матрицы P, все точки , лежащие на некоторой прямой L, проходящей через центр линзы, будут иметь в проекции на изображение одни и те же координаты I. Этим, в частности, объясняется эффект наложения (occlusion effect). Используя выражение и считая координаты I какого-либо пикселя изображения известными, мы можем восстановить уравнение прямой L, такой, что

Выберем линию поиска (scanline) нa левом изображении. Обычно, для удобства и ускорения работы алгоритма, выбирают прямую, совпадающую с горизонтальной строчкой пикселей изображения. Далее для двух точек и , принадлежащих этой прямой N( , ) рассчитаем уравнения соответствующих им прямых и . Используя матрицу F, повернем эти прямые в координатную систему правой камеры. Описанная полученными прямыми плоскость сечет плоскость правого изображения по прямой N( , ). Эта прямая и будет искомой линией поиска для правой камеры. Получив уравнения соответствующих прямых для стереопары, мы можем построить дискретную функцию ,определяющую для каждой пары (, ) значение . Получившуюся прямоугольную матрицу можно рассчитать однократно и использовать для быстрой ректификации изображений.

В результате выполнения предварительного этапа каждое изображение стереопары содержит бинарное ректифицированное изображение с выделенными границами.

# Поиск сопряженных точек

Проблема установления соответствия состоит в идентификации характерных фрагментов на двух изображениях, являющихся проекциями одного и того же фрагмента трехмерного мира. Если это сделано, то можно вычислить расстояние до этого фрагмента. При некоторых положениях камеры точка поверхности объекта может оказаться за пределами изображения, однако если она видна на обоих изображениях, то обе эти точки должны лежать на соответствующих эпиполярных линиях. Как можно отождествить точки? В одном из подходов предлагают анализировать изображения порознь, выделяя на них характерные особенности. Ими могут быть либо объекты, подлежащие идентификации, либо особые области полутонового изображения, которые мы можем с уверенностью узнать. Удобной для этой цели особенностью являются края. Можно выделить также «углы», где график функции яркости обладает ненулевой гауссовой кривизной. Можно ожидать, что яркостная картина не меняется или меняется мало в небольшой окрестности каждой точки. В принципе возможен случай, когда в двух точках уровни яркости совпадают. Однако обычно таких точек несколько на соответствующих эпиполярных линиях.

Методы поиска сопряжённых точек делятся на 2 вида:

* Локальные
* Глобальные

Локальные алгоритмы.Этот класс алгоритмов подсчитывает диспаратность каждого пикселя в отдельности, используя окна фиксированного или адаптируемого размера для корреляции. Выбор формы окна и его размеров является непростой задачей. Это связано с тем, что, с одной стороны, корреляция предполагает, что глубина всех пикселей внутри окна не терпит разрывов. Это значит, что увеличение размеров окна ведет к нарушению этого условия и, как следствие, к некорректной работе алгоритма. С другой стороны, уменьшение размеров окна ведет к увеличению влияния шума, что приводит к уменьшению корректных совпадений. Для локальных алгоритмов стоимость соответствия определяется как схожесть между двумя областями, одна из которых находится в базовом изображении, а другая в парном. Форма этих областей зависит от конкретного алгоритма. Стандартные подходы, появившиеся на заре исследований проблемы соответствия, используют прямоугольное окно фиксированного размера. Размеры этого окна определяются опытным путем. Объем вычислений, в этом случае, сильно сокращается по сравнению с современными методами, но существует ряд проблем, избежать которых, используя фиксированный размер окна, практически невозможно. Для того чтобы понять: почему фиксированный размер окна порождает проблемы, мы рассмотрим, как, в общем, работают локальные алгоритмы. Рассмотрим произвольный пиксель базового изображения, вокруг него строится окно. Для данного пикселя базового изображения существует ряд подозрительных на соответствие пикселей парного изображения. Вокруг каждого такого пикселя строится такое же окно. Подобная ситуация показана на Рис. 2.1.

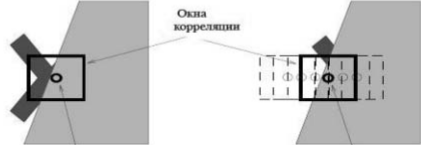


Рисунок 2.1. Рассматриваемый пиксель левого и правого изображения

Возможные позиции пикселей в парном изображении определены минимально допустимым расстоянием между камерами и объектом, что дает нам максимальную диспаратность. Для каждой возможной пары окон подсчитывается стоимость соответствия. Наиболее распространенными стоимостями являются квадрат разности интенсивностей (SD)[10] и абсолютная разность интенсивностей (AD). Пара окон с наибольшей или наименьшей, в зависимости от алгоритма вычисления, стоимостью и определяют пиксель, соответствующий данному. Тем не менее, если в окне нарушается условие на непрерывность глубины, то его часть будет влиять на результирующую стоимость неопределенным образом.

Левая часть корреляционного окна, как в базовом изображении, так и в парном, содержит фон. Но этот фон разный для обоих окон. Более того, стоит отметить, что левая часть окна в левом изображении, по крайней мере частично, заслонена на правом битмапе. Размеры заслоненной части зависят от разности диспаратностей и размеров окна. Все это приводит нас к тому, что та часть окна, которая содержит фон, вносит ошибки в подсчет стоимостей. Возможным решением проблемы будет уменьшение размеров окна, но, как я уже говорил, это приведет к возрастанию влияния шума при подсчете стоимости.

Как было показано выше, фиксированный размер окна не дает необходимой точности при решении проблемы соответствия. Hirschmuller создал метод коррекции ошибок на границе объектов. Данная коррекция применяется на этапе постпроцесса и позволяет уменьшить количество ошибок на границах на 50%. Идея алгоритма состоит в использовании нескольких (а именно 5) окон для уменьшения количества ошибок, а также в предоставлении функции, которая определяет неверные совпадения. Наряду с исследованиями, связанными с формой и размерами окна, существуют различными подходы к измерению стоимости соответствия. Уже были упомянуты два стандартных подхода SD (разность интенсивностей в квадрате) и AD (абсолютная разность интенсивностей)

В обоих случаях, подсчет стоимости можно сделать очень эффективным. Однако, в случае выбора этих формул для вычисления, неявно делается предположение, что для данного пикселя базового изображения соответствующий пиксель имеет примерно такую же интенсивность. Это значит, что если у двух камер была выставлена разная экспозиция или на фотографии, сделанной одной из камер, появился блик, то эти формулы будут давать неверный результат. Также, оба этих подхода довольно сильно подвержены влиянию шума и случайных выбросов в данных.

Zabih представил измерение, основанное не на значениях интенсивностей, а на их числовом порядке.

Здесь С(W) – это трансформация данных корреляции, а R(W) - упорядочивание данных корреляции. На их основе вычисляется итоговая стоимость соответствия. Такой подход позволяет уменьшить воздействие шума и случайных выбросов данных по сравнению со стандартным подходом. Тем не менее, различия в настройке камер по-прежнему могут сильно влиять на итоговый результат.

После того, как для каждого пикселя p в базовом изображении и для каждой возможной диспаратности d, от 0 до максимальной диспаратности, подсчитана стоимость соответствия, локальный алгоритм, обычно, переходит к шагу суммирования стоимостей. Стандартный подход состоит в суммировании или усреднении стоимостей соответствия в некоторой области. Усреднение может производиться с помощью свертки с каким-нибудь ядром, чаще всего Гауссовым.

Наконец, финальным шагом является вычисление диспаратности. На основе просуммированных или усредненных стоимостей нахождение соответствия становится тривиальной задачей. Диспаратность для данного пикселя p базового изображения – это такое d, на котором достигается минимальное значение стоимости. Такой подход получил название «победитель получает все» (WTA). Проблема, возникающая при таком подходе, каждой точке выбирается уникальное соответствие в парном изображении. На самом же деле, одной точке может соответствовать сразу несколько. К сожалению, эта проблема возникает не только у локальных алгоритмов, но и у подавляющего большинства методов, целью которых является построение карты диспаратности.

Просуммировав все вышесказанное про локальные алгоритмы, можно сказать, что существует огромное число локальных методов, решающих проблему соответствия. Тем не менее, если методу удается решить задачу быстро, то это означает большое количество ошибок в вычислениях, что, конечно, не всегда приемлемо.

Глобальные алгоритмы. В отличие от локальных алгоритмов, где нахождение диспаратности происходит для каждого пикселя отдельно, целью глобального подхода является поиск наилучшей карты диспаратности для всего изображения сразу. Глобальные методы практически всю работу выполняют на шаге вычисления диспаратностей, часто пропуская шаг суммирования стоимостей соответствия. Чаще всего глобальные алгоритмы решают проблему соответствия путем минимизации функционала энергии.

Допустим, мы хотим найти наилучшую конфигурацию f. Подходящая конфигурация будет минимизировать функционал Е – функционал глобальной энергии

Первое слагаемое показывает: насколько хорошо данная конфигурация согласуется с парой входных изображений. Для того чтобы подсчитать первое слагаемое, используются попиксельные стоимости

Второе слагаемое отвечает за штраф, налагаемый на данную конфигурацию, в случае, когда она нарушает непрерывность диспаратностей. Чаще всего рассматриваются соседние пиксели.

Здесь, в качестве функции ρ выступает некоторая монотонно-возрастающая функция штрафа.

После того, как функционал энергии определен, существует целый ряд алгоритмов, позволяющих найти его минимум. Классическими подходами являются: Марковские сети, максимальный поток, graph-cuts. К сожалению, поиск минимума — это очень сложная вычислительная задача. Одним из методов, призванных решить эту проблему, является динамическое программирование. Вместо поиска минимума для всего изображения сразу, поиск ведется построчно, причем строки обрабатываются независимо друг от друга. Нахождение минимума для одной строки возможно, в этом случае, за полиномиальное время. Проблемы данного подхода очевидны, при вычислении карты диспаратностей мы совершенно не учитываем связей между пикселями в разных строках. Кроме того, этот подход требует, чтобы относительный порядок пикселей в строке был одинаковым для левого и правого изображения, что невозможно выполнить в том случае, когда на сцене находится узкий объект на переднем плане.

Глобальные алгоритмы дают отличные результаты в нахождении карты диспаратностей. Количество ошибок, допускаемых ими, относительно невелико. Тем не менее, все глобальные алгоритмы требуют больших вычислительных затрат и времени расчёта и ни один не может быть реализован в реальном времени.

Алгоритм полуглобального сопоставления.

Алгоритм полуглобального сопоставления или Semi-Global Matching (SGM)[11] был предложен Heiko Hirschmuller в 2005 году. Преимущества этого подхода, по сравнению с другими решениями, состоит в том, что, с одной стороны, он дает очень неплохие результаты с точки зрения качества, так как не является уже локальным алгоритмом. С другой стороны, скорость его работы существенно выше, чем у глобальных подходов. То есть, предложенный алгоритм SGM представляет собой гибридный метод, при котором сочетается качество глобальных алгоритмов со скоростью локальных подходов. В качестве примера работы разных алгоритмов приводятся карты диспаратности:

Этапы алгоритма:

1. Вычисление попиксельной стоимости
2. Суммирование попиксельных стоимостей
3. Вычисление карты диспаратности

### Вычисление попиксельной стоимости

Стоимость считается следующим образом: для пикселей и базового и парного изображений соответственно, определим следующие значения:

где и — какие-то функции интенсивности. Получается, что - линейно-интерполированная интенсивность между и пикселем слева. Аналогичным образом определяется и

Получаем функцию:

Итого,

### Суммирование попиксельных стоимостей

После того как вычислены попиксельные стоимости, необходимо совершить последний шаг работы алгоритма SGM, а именно, суммирование стоимостей. Стандартным подходом для глобальных алгоритмов является нахождение такой конфигурации, которая минимизирует некий функционал энергии вида:

Классическим оптимизационным подходом является динамическое программирование. В этом случае суммирование происходит вдоль одного направления, однако в этом случае связи между пикселями в разных рядах практически, а часто и вообще, не учитываются, поэтому для увеличение точности давайте суммировать по всем направлениям.

Пусть мы хотим просуммировать стоимости в направлении r для данного пикселя p и диспаратности d. Суммирование происходит рекурсивно:

Здесь первое слагаемое – это попиксельная стоимость, а второе – это минимум из четырех чисел, которые зависят от суммированной стоимости для предыдущего пикселя в данном направлении. P1 - это константный штраф, который налагается в том случае, если диспаратности у двух соседних пикселей отличаются на один. P2 - это константный штраф, который налагается в том случае, когда диспаратности у двух соседних пикселей отличаются больше, чем на один. Вычитаемое слагаемое константно для всех диспаритетов пикселя p.

### Вычисление карты диспаратности

После того, как подсчитаны суммарные стоимости для каждого из направлений, вычисляется общая сумма всех стоимостей для каждого из пикселей:

Для каждого пикселя базового изображения p, соответствующая ему диспаратность находится как минимум по всем диспаратностям в общем массиве стоимостей:

В итоге получаем на выходе карту диспаратности, которую можно преобразовать в 3D объект в программе для 3D-моделирования.

# Заключение

В представленной работе задача стереореконструкции решена с помощью алгоритма Semi – Global Matching – полуглобального сопоставления, который дал лучшие результаты поиска и сопоставления сопряжённых точек на стереопаре, по сравнению с другими видами алгоритмов. Предложенный алгоритм пространственной реконструкции выгодно отличается от алгоритмов построчной обработки, так как даёт более гладкую модель восстановленной поверхности. Отличительная особенность алгоритма заключается в том, что при восстановлении структуры объекта учитывается его непрерывность как в горизонтальном, так и в вертикальном направлении.

Несмотря на качество, его производительность оставляет желать лучшего (несколько секунд на строчку изображения 512x512). Впрочем, одномерные алгоритмы обладают возможностью быть распараллеленными (в идеале: по строке на вычислительное устройство), что очень хорошо в условиях наиболее популярного метода наращивания мощностей вычислительной техники.

Качество работы алгоритма можно улучшить, если известна информация о сегментации изображения. Так соседним пикселям из одного сегмента, скорее всего, должны быть присвоены похожие расхождения. Низкая производительность не позволяет использовать в лоб алгоритмы стереозрения для получения данных из видеоряда.

# Используемая литература

1. <https://courses.graphicon.ru/files/courses/vision/2010/cv_2010_12.pdf>
2. vuz.exponenta.ru/pdf/book/PROEKTIV.pdf
3. https://courses.graphicon.ru/files/courses/vision/2010/cv\_2010\_13.pdf
4. “Multiple View Geometry in Computer Vision” Richard Hartley, Andrew Zisserman, Cambridge University Press, March 2004.
5. <https://courses.graphicon.ru/files/courses/vision/2010/cv_2010_12.pdf>
6. <http://robot-develop.org/wp-content/uploads/2012/03/seg3.pdf>
7. <https://www.nist.gov/sites/default/files/documents/calibrations/10-srfce.pdf>
8. <https://perso.limsi.fr/vezien/PAPIERS_ACS/canny1986.pdf>
9. Дуда Р., Харт П. Распознавание образов и анализ сцен. 1976
10. H.Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. IEEE Conference on Computer Vision and Pattern Recognition, 2006
11. <http://se.math.spbu.ru/SE/YearlyProjects/2012/YearlyProjects/2012/445/445_Mokaev_report.pdf>