

СОДЕРЖАНИЕ

Введение.....	4
1 Анализ предметной области.....	6
1.1 Базовые понятия.....	6
1.1.1 Стереопара.....	6
1.1.2 Проективная геометрия и однородные координаты.....	6
1.1.3 Модель проективной камеры.....	8
1.1.4 Эпиполярная геометрия.....	12
1.1.5 Триангуляция точек (triangulation).....	14
1.1.6 Карта глубины.....	15
2 Существующие решения.....	17
2.1 Алгоритм ransac.....	17
2.2 Алгоритм sift.....	17
Заключение.....	18
Список использованных источников.....	19

ВВЕДЕНИЕ

Видео и фото тесно вплелись в нашу жизнь. Почти каждый мобильный телефон оснащен камерой. Почти каждая камера умеет записывать видео. Повсеместно распространилась 3D-графика. С развитием возможностей усиливается потребность в “дешевом” построении 3D-сцен. Самый очевидный из таких методов – стереозрение – получение трехмерной картины мира по видеоряду или нескольким изображениям.

Есть и другие возможные применения. В киноиндустрии – для создания спецэффектов и 3D-фильмов. В военном деле – для измерения расстояний. Как правило, камера расположена сбоку от монитора, где отображается лицо собеседника. Таким образом, другой участник смотрит куда-то в сторону. Построение трехмерного изображения лица позволяет синтезировать новое изображение, на котором собеседник смотрит прямо в глаза, что усиливает эффект присутствия.

Совмещение изображений, позволяет человеку получить информацию о расстоянии до объектов по их расхождениям (disparity). Эта идея может быть использована в алгоритмах обработки изображений. Хотя, конечно же, зрение человека активное (то есть параметры оптической системы настраиваются под изображение) – глаза вращаются в глазницах, меняется фокусное расстояние, как правило, построение системы активного зрения сложнее и дороже, чем зрения пассивного.

Система пассивного стереозрения, как правило, включает в себя 2 камеры и большинство алгоритмов решают общую задачу сопоставления 2-х изображений.

Существуют различные классификации алгоритмов сопоставления 2-х. Один из вариантов такой классификации представлен в [1]. Все алгоритмы делятся на локальные (в которых расхождение вычисляется в каждой точке на основе “похожести” окна вокруг этой точки и окна вокруг точек на другом изображении) и глобальные (основанные на ми-

нимизации функционала энергии – мы находим расхождение сразу для всех точек). Глобальные в свою очередь делятся по способу минимизации энергии. Как правило, это динамическое программирование [2, 3, 4, 5] или нахождение минимального разреза графа [6, 7]. Алгоритмы разреза графа называют также двумерными. Они дают довольно точные результаты, но имеют меньшую производительность. Алгоритмы, обрабатывающие строки изображений независимо друг от друга [2, 3], называют одномерными. Они работают быстрее, но подвержены эффекту гребенки, с которым борются с помощью разных ухищрений. Нечто среднее по производительности и качеству из себя представляют алгоритмы оптимизации на поддереве графа, построенного на пикселях изображения.

Многие алгоритмы построены на модели случайных полей Маркова (MRF). Основное предположение в такой модели: расхождение в любой точке зависит только от расхождений соседних точек (как правило, считают, что их 4, хотя можно соседними считать и 8 точек, оставаясь в рамках модели). Минимизация энергии в них производится на основе разрезов графа или распространения доверия [8].

...

Цель работы – анализ методов применяемых при построении объёмного изображения по стереопаре.

Задачи работы:

- описать термины предметной области и обозначить проблему;
- провести обзор существующих программных решений в области стереограмметрии;
- выбрать критерии для их оценки и сравнить;
- выбрать наиболее предпочтительный.

1 Анализ предметной области

...

1.1 Базовые понятия

...

1.1.1 Стереопара

Стереопара - это два изображения одного и того же предмета, рассматривая которые одновременно с помощью стереоскопа или же при помощи компьютера в виде последовательности изображений, к примеру, анимации.

Для получения стереопары необходимо снять одни и те же объекты с двух разных точек, отстоящих друг от друга по горизонтали на расстоянии стереобазиса.

1.1.2 Проективная геометрия и однородные координаты

В геометрии стерео зрения значительную роль играет проективная геометрия. К проективной геометрии есть несколько подходов: геометрический (подобно Евклидовой геометрии ввести понятие геометрических объектов, аксиом и из этого выводить все свойства проективного пространства), аналитический (рассматривать все в координатах, как в аналитическом подходе к Евклидовой геометрии), алгебраический.

Ниже изложен аналитический подход к проективной геометрии.

Точки проективной плоскости. Рассмотрим двухмерное проективное пространство (которое еще называется проективной плоскостью). В то время как на обычной Евклидовой плоскости точки описываются

парой координат $(x, y)^T$, на проективной плоскости точки описываются трехкомпонентным вектором $(x, y, w)^T$. При этом для любого ненулевого числа a , векторы $(x, y, w)^T$ и $(ax, ay, aw)^T$ соответствуют одной и той же точке. А нулевой вектор $(0, 0, 0)^T$ не соответствует никакой точке и выкидывается из рассмотрения. Такое описание точек плоскости называется однородными координатами (homogeneous coordinates).

Точкам проективной плоскости можно сопоставить точки обычной Евклидовой плоскости. Координатному вектору $(x, y, w)^T$ при $w \neq 0$ сопоставим точку Евклидовой плоскости с координатами $(x/w, y/w)^T$. Если же $w = 0$, т.е. координатный вектор имеет вид $(x, y, 0)^T$, то говорить, что эта точка в бесконечности. Таким образом, проективную плоскость можно рассматривать как Евклидовую плоскость, дополненную точками из бесконечности.

Перейти от однородных координат $(x, y, w)^T$ к обычным Евклидовым можно путем деления координатного вектора на последнюю компоненту и последующего ее отбрасывания $(x, y, w)^T \rightarrow (x/w, y/w)^T$. А от Евклидовых координат $(x, y)^T$ перейти к однородным можно за счет дополнения координатного вектора единицей: $(x, y)^T \rightarrow (x, y, 1)^T$.

Прямые на проективной плоскости. Любая прямая на проективной плоскости описывается, подобно точке, трехкомпонентным вектором $l = (a, b, c)^T$. Опять же вектор, описывающий прямую, определен с точностью до ненулевого множителя. При этом уравнение прямой будет иметь вид: $l^T x = 0$.

В случае, когда $a^2 + b^2 \neq 0$ мы имеем аналог обычной прямой $ax + by + c = 0$. А вектор $(0, 0, w)$ соответствует прямой лежащей в бесконечности.

Трехмерное проективное пространство. По аналогии с проективной плоскостью, точки трехмерного проективного пространства определяются четырехкомпонентным вектором однородных координат $(x, y, z, w)^T$. Опять же для любого ненулевого числа a , координатные вектора $(x, y, z, w)^T$ и $(ax, ay, az, aw)^T$ соответствуют одной и той же точке.

Как в случае проективной плоскости, между точками трехмерного Евклидова пространства и трехмерного проективного пространства можно установить соответствие. Вектору однородных координат $(x, y, z, w)^T$ при $w \neq 0$ соответствует точка Евклидова пространства с координатами $(x/w, y/w, z/w)^T$. А про точку с вектором однородных координат вида $(x, y, z, 0)^T$ говорят, что она лежит в бесконечности.

Проективное преобразование. (homography, projective transformation — в англ. литературе). С геометрической точки зрения, проективное преобразование — это обратимое преобразование проективной плоскости (или пространства), 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 следующей формуле: $x' = Hx$.

1.1.3 Модель проективной камеры

Современные CCD-камеры хорошо описываются с помощью следующей модели, называемой проективной камерой (projective camera, pinhole camera). Проективная камера определяется *центром камеры*, *главной осью* — лучом начинающимся в центре камеры и направленным туда, куда камера смотрит, *плоскостью изображения* — плоскостью на которую выполняется проецирование точек, и системой координат на этой плоскости. В такой модели, произвольная точка пространства X проецируется на плоскость изображения в точку x лежащую на отрезке CX , который соединяет центр камеры C с исходной точкой X (см. рис. 1.1).

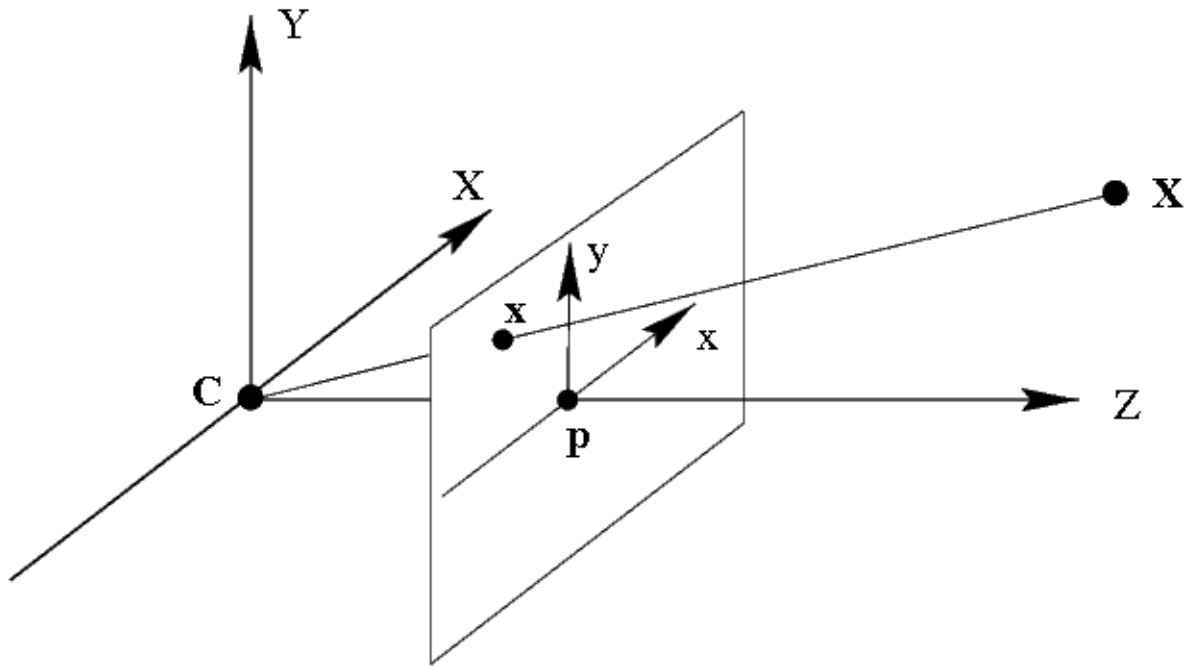


Рисунок 1.1 — **Модель камеры.** C — центр камеры, Cp — главная ось камеры. Точка X трехмерного пространства проецируется в точку x — на плоскости изображения.

Формула проецирования имеет простую математическую запись в однородных координатах:

$$x = PX$$

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

Матрица P выражается следующим образом $P = KR[I | -\mathbf{c}] = K[R|\mathbf{t}]$, где K — верхняя треугольная матрица внутренних параметров камеры размера 3×3 (конкретный вид приведен ниже), R — ортогональная матрица размера 3×3 , определяющая поворот камеры относительно глобальной системы координат, I — единичная матрица размера 3×3 , вектор \mathbf{c} — координаты центра камеры, а $\mathbf{t} = R\mathbf{c}$.

Стоит отметить, что матрица камеры определена с точностью до постоянного ненулевого множителя, который не изменит результатов проецирования точек по формуле $x = PX$. Однако этот постоянный

множитель обычно выбирается так, что бы матрица камеры имела выше-описанный вид.

В самом простейшем случае, когда центр камеры лежит в начале координат, главная ось камеры сонаправлена оси Cz , оси координат на плоскости камеры имеют одинаковый масштаб (что эквивалентно квадратным пикселям), а центр изображения имеет нулевые координаты, матрица камеры будет равна $P = K[I|\mathbf{0}]$, где

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

У реальных CCD камер пикселы обычно незначительно отличаются от квадратных, а центр изображения имеет ненулевые координаты. В таком случае матрица внутренних параметров примет вид:

$$K = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Коэффициенты $f_{,x,y}$ — называются фокусными расстояниями камеры (соответственно общим и вдоль осей x и y).

Помимо этого, в силу неидеальности оптики, на изображениях, полученных с камер, присутствуют искажения-дисторсии (distortion). Данные искажения имеют нелинейную математическую запись:

$$x'' = x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x' y' + p_2 (r^2 + 2x'^2)$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1 (r^2 + 2y'^2) + 2p_2 x' y'$$

где k_1, k_2, p_1, p_2, k_3 — коэффициенты дисторсии, являющиеся параметрами оптической системы; $r^2 = x'^2 + y'^2$; (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 не зависят от расстояния до объекта, а зависят только от координат точек, в которые проецируются пиксели объекта. Соответственно для компенсации дисторсий обычно выполняется преобразование

исходного изображения полученного с камеры. Это преобразование будет одним и тем же для всех изображений, полученных с камеры, при условии постоянства фокусного расстояния (математически — одной и той же матрицы внутренних параметров).

В ситуации, когда известны внутренние параметры камеры и коэффициенты дисторсии говорят, что камера откалибрована.

Об определении трехмерных координат наблюдаемых точек можно говорить, когда есть как минимум две камеры.

Матрицы пары камер, калибровка. Пусть имеются две камеры, заданные своими матрицами P и P' в некоторой системе координат. В таком случае говорят, что имеется пара откалиброванных камер. Если центры камер не совпадают, то эту пару камер можно использовать для определения трехмерных координат наблюдаемых точек.

Зачастую, система координат выбирается так, что матрицы камер имеют вид $P = K[I|0]$, $P' = K'[R'|t']$. Это всегда можно сделать, если выбрать начало координат совпадающее с центром первой камеры, и направить ось Z вдоль ее оптической оси.

Калибровка камер обычно выполняется, за счет многократной съемки некоторого калибровочного шаблона, на изображении можно легко выделить ключевые точки, для которых известны их относительные положения в пространстве. Далее составляются и решаются (приближенно) системы уравнений, связывающие координаты проекций, матрицы камер и положения точек шаблона в пространстве.

Существуют общедоступные реализации алгоритмов калибровки, например, Matlab Calibration toolbox. Так же библиотека OpenCV включает в себя алгоритмы калибровки камер и поиска калибровочного шаблона на изображении.

1.1.4 Эпиполярная геометрия

Перед тем как перейти к описанию собственно метода вычисления трехмерных координат точек, я опишу некоторые важные геометрические свойства, связывающие положения проекций точки трехмерного пространства на изображениях с обеих камер.

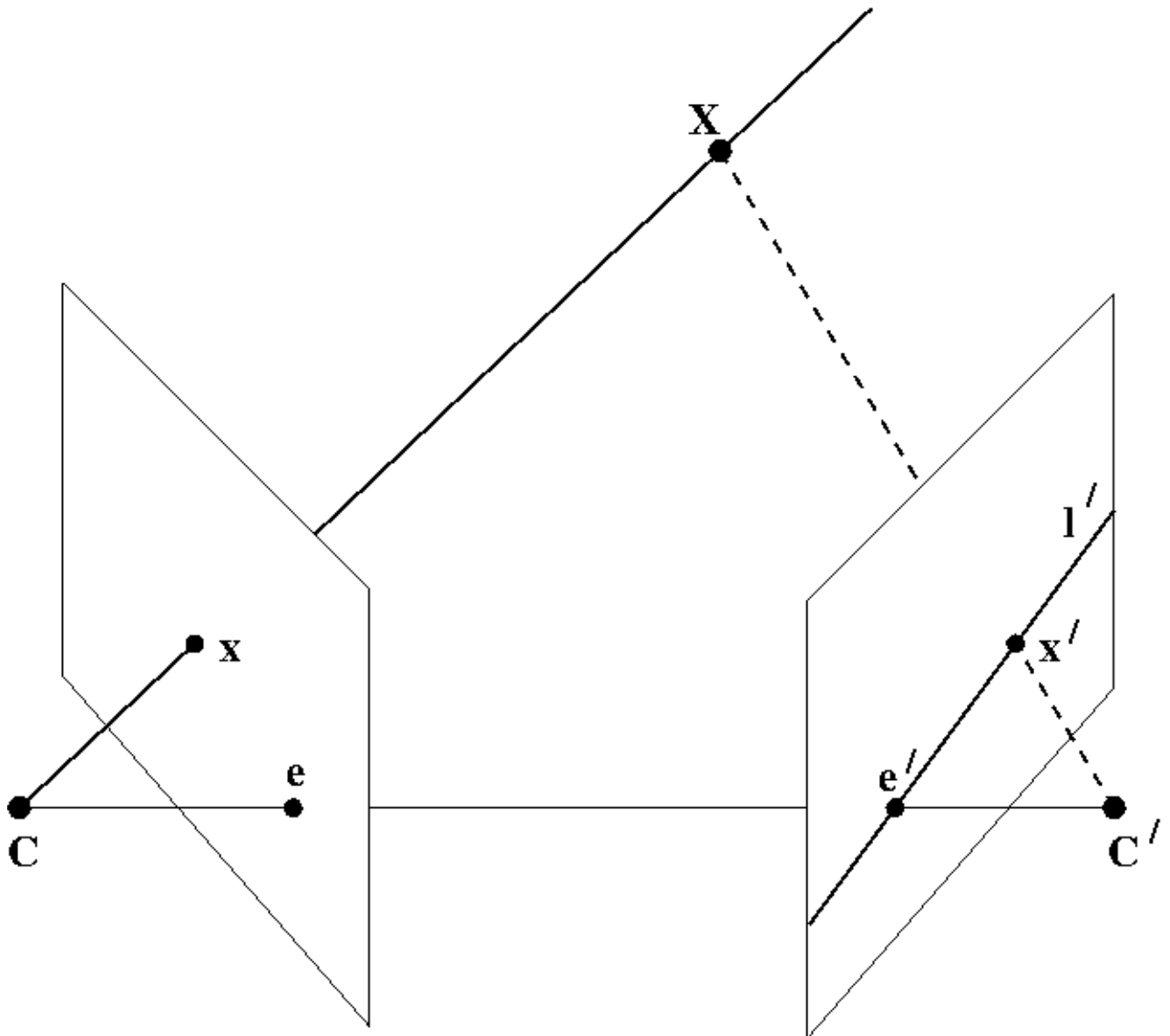


Рисунок 1.2 — Эпиполярная геометрия

Пусть имеются две камеры, как изображено на рисунке 1.2. C — центр первой камеры, C' — центр второй камеры. Точка пространства X проецируется в x на плоскость изображения левой камеры и в x' на плоскость изображения правой камеры. Прообразом точки x на изображении левой камеры является луч xX . Этот луч проецируется на плоскость

второй камеры в прямую l' , называемую эпиполярной линией. Образ точки X на плоскости изображения второй камеры обязательно лежит на эпиполярной линии l' .

Таким образом, каждой точке x на изображении левой камеры соответствует эпиполярная линия l' на изображении правой камеры. При этом пара для x на изображении правой камеры может лежать только на соответствующей эпиполярной линии. Аналогично, каждой точке x' на правом изображении соответствует эпиполярная линия l на левом.

Эпиполярную геометрию используют для поиска стереопар, и для проверки того, что пара точек может быть стереопарой (т.е. проекцией некоторой точки пространства).

Эпиполярная геометрия имеет очень простую запись в координатах. Пусть имеется пара откалиброванных камер, и пусть x — однородные координаты точки на изображении одной камеры, а x' — на изображении второй. Существует такая матрица F размера 3×3 , что пара точек x, x' является стереопарой тогда и только тогда, когда:

$$x'^T F x = 0$$

Матрица F называется фундаментальной матрицей (fundamental matrix). Ее ранг равен 2, она определена с точностью до ненулевого множителя и зависит только от матриц исходных камер P и P' .

В случае, когда матрицы камер имеют вид $P = K[I|0]$, $P' = K'[R|t]$ фундаментальная матрица может быть вычислена по формуле:

$$F = K'^{-1T} R K^T [K R^T t] x$$

где для вектора e обозначение $[e]_X$ вычисляется как

$$[e]_X = \begin{bmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 1 \end{bmatrix}$$

С помощью фундаментальной матрицы вычисляются уравнения эпиполярных линий. Для точки x , вектор, задающий эпиполярную линию, будет иметь вид $l' = Fx$, а уравнение самой эпиполярной линии: $l'^T x' = 0$.

Аналогично для точки x' , вектор, задающий эпполярную линию, будет иметь вид $l = F^T x'$.

Помимо фундаментальной матрицы, существует еще такое понятие, как существенная матрица (essential matrix): $E = K'^T F K$. В случае, когда матрицы внутренних параметров будут единичными существенная матрица будет совпадать с фундаментальной. По существенной матрице можно восстановить положение и поворот второй камеры относительно первой, поэтому она используется в задачах, в которых нужно определить движение камеры.

1.1.5 Триангуляция точек (triangulation)

Процесс определения трехмерных координат точки по координатам ее проекций в литературе называется триангуляцией.

Пусть имеются две откалиброванные камеры с матрицами P_1 и P_2 . x_1 и x_2 — однородные координаты проекций некоторой точки пространства X . Тогда можно составить следующую систему уравнений:

$$\begin{cases} x_1 = P_1 X \\ x_2 = P_2 X \end{cases}$$

На практике для решения этой системы применяется следующий подход. Векторно умножают первое уравнение на x_1 , второе на x_2 , избавляются от линейно зависимых уравнений и приводят систему к виду $AX = 0$, где A имеет размер 4×4 . Далее можно либо исходить из того, что вектор X является однородными координатами точки, положить его последнюю компоненту равной 1 и решать полученную систему из 3х уравнений с тремя неизвестными. Альтернативный способ — взять любое ненулевое решение системы $AX = 0$, например вычисленное, как сингулярный вектор, отвечающий наименьшему сингулярному числу матрицы A .

1.1.6 Карта глубины

Карта глубины (depth map) — это изображение, на котором для каждого пикселя, вместо цвета, хранится его расстояние до камеры. Карта глубины может быть получена с помощью специальной камеры глубины (например, сенсор Kinect является своего рода такой камерой), а так же может быть построена по стереопаре изображений.

Идея, лежащая в основе построения карты глубины по стереопаре очень проста. Для каждой точки на одном изображении выполняется поиск парной ей точки на другом изображении. А по паре соответствующих точек можно выполнить триангуляцию и определить координаты их прообраза в трехмерном пространстве. Зная трехмерные координаты прообраза, глубина вычисляется, как расстояние до плоскости камеры.

Парную точку нужно искать на эпиполярной линии. Соответственно, для упрощения поиска, изображения выравнивают так, что бы все эпиполярные линии были параллельны сторонам изображения (обычно горизонтальны). Более того, изображения выравнивают так, что бы для точки с координатами (x_0, y_0) соответствующая ей эпиполярная линия задавалась уравнением $x = x_0$, тогда для каждой точки соответствующую ей парную точку нужно искать в той-же строчке на изображении со второй камеры. Такой процесс выравнивания изображений называют ректификацией (rectification). Обычно ректификацию совершают путем ремеппинга изображения и ее совмещают с избавлением от дисторсий. Пример ректифицированных изображений приведен на рисунке 1.3.



Рисунок 1.3 — Пример ректифицированных картинок, и соответствующей им disparity map.

После того как изображения ректифицированы, выполняют поиск соответствующих пар точек. Самый простой способ проиллюстрирован на рисунке 1.4 и состоит в следующем. Для каждого пикселя левой картинки с координатами (x_0, y_0) выполняется поиск пикселя на правой картинке. При этом предполагается, что пиксель на правой картинке должен иметь координаты $(x_0 - d, y_0)$, где d — величина называемая несоответствие/смещение (disparity). Поиск соответствующего пикселя выполняется путем вычисления максимума функции отклика, в качестве которой может выступать, например, корреляция окрестностей пикселей. В результате получается карта смещений (disparity map), пример которой приведен на рис. 1.3.

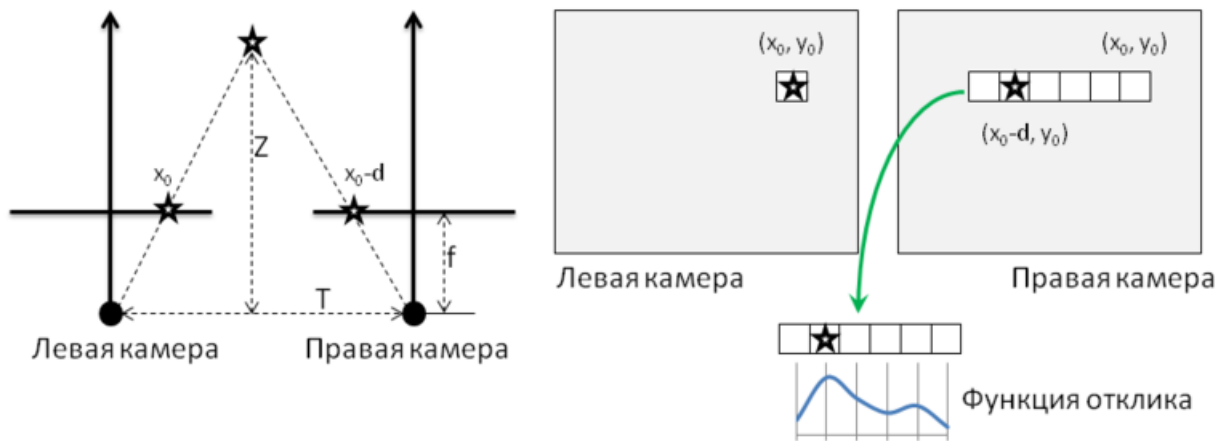


Рисунок 1.4 — Вычисление карты глубины

Собственно значения глубины обратно пропорциональны величине смещения пикселей. Если использовать обозначения с левой половины рисунка 1.4, то зависимость между disparity и глубиной можно выразить следующим способом:

$$\frac{T - d}{Z - f} = \frac{T}{Z} \rightarrow Z = \frac{fT}{d}$$

Из-за обратной зависимости глубины и смещения, разрешающая способность систем стерео зрения, которые работают на основе данного метода, лучше на близких расстояниях, и хуже на далеких.

2 Существующие решения

...

2.1 Алгоритм ransac

...

2.2 Алгоритм sift

...

ЗАКЛЮЧЕНИЕ

...

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Scharstein D., Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // Int. Journal of Computer Vision 47. April-June 2002. С. 7–42.
2. Bobick A. F., Intille S. S. Large occlusion stereo // Int. Journal of Computer Vision. 33(3). 1999. С. 181–200.
3. Ohta Y., Kanade T. Stereo by intra- and inter- scanline search using dynamic programming // IEEE TPAMI. 7(2). 1985. С. 139–154.
4. Veksler O. Stereo correspondence by dynamic programming on a tree // Proc. CVPR. Vol. 2. 2005. С. 384–390.
5. Naveed I. R., Huijun Di, GuangYou Xu Refine stereo correspondence using bayesian network and dynamic programming on a color based minimal span tree // ACIVS. 2006. С. 610–619.
6. Kolmogorov V., Zabih R. Computing visual correspondence with occlusions using graph cuts // ICCV. Vol. 2. 2001. С. 508–515.
7. Boykov Y., Veksler O., Zabih R. Fast approximate energy minimization via graph cuts // IEEE TPAMI 23(11). 2001. С. 1222–1239.
8. Sun J., Shum H., Zheng N Stereo matching using belief propagation // In ECCV. 2002. PP. 510–524.