Санкт-Петербургский государственный университет

Синицын Даниил Дмитриевич

Оценка параметров систем камер без использования калибровочных паттернов

Магистерская диссертация

Научный руководитель:
д. ф.-м. н., профессор Терехов А. Н.

Рецензент:
к. т. н. Федоренко С. И.

Санкт-Петербург
2020

SAINT-PETERSBURG STATE UNIVERSITY

Daniil Sinitsyn

Pattern-free camera calibration

Master's Thesis

Scientific supervisor:
professor Andrey Terekhov

Reviewer:
Sergey Fedorenko

Saint-Petersburg
2020

Оглавление

1. Введение **5**

2. Задача **7**

2.1. Постановка задачи 7

3. Обзор литературы **9**

4. Модель камеры **14**

5. Парные оценки **17**

5.1. Фундаментальная матрица 17

5.2. Оценка 17

5.3. RANSAC и нелинейная оптимизация 18

5.4. Увеличение порядка полинома дисторсии 19

5.5. Ошибка Сэмпсона 19

6. Оценка фокусных расстояний и парных поз **21**

6.1. Фокусные расстояния и агрегация оценок 21

6.2. Существенная матрица 21

6.3. Разложение и оптимизация 23

7. Усреднение вращений **24**

7.1. Усреднение вращений в компоненте связности 24

7.1.1. Нелинейная IRLS оптимизация 25

7.2. Оценка ориентаций камер внутри системы 26

8. Результаты **29**

8.1. Парные оценки дисторсии и фундаментальных матриц . 30

8.1.1. Тестирование на реальных изображениях 31

8.2. Фокусные расстояния и агрегирование парных оценок . . 32

8.2.1. Фокусные расстояния 32

8.2.2. Агрегирование оценок 32

8.2.3. Нелнейная оптимизация с ошибкой Сэмпсона . . . 33

8.3. Усреднение вращений 34
8.3.1. Датасет KITTI 34
8.3.2. Датасет ladybug 36

Список литературы **41**

A. Оценка оставшихся ориентаций **45**
A.1. Оценка ориентаций системы из камер 45

1. Введение

В настоящее время распространено использование систем, состоящих из нескольких жёстко закрепленных камер. Такие системы могут быть найдены на автомобилях, квадрокоптерах, в устройствах для картографирования или устройствах для съемки панорам.

Для использования таких систем необходимо произвести так называемую калибровку системы. В калибровку системы входит оценка относительных позиций камер внутри системы, оценка преобразования, моделирующего проецирование точек реального мира в пиксели на изображении.

Большая часть подходов для калибровки систем из камер является калибровкой с использованием специальных паттернов [27][13]. Зная геометрию паттерна и координаты определенных точек паттерна на изображении, параметры камер оцениваются так, чтобы наилучшим способом описывать соответствие найденных точек калибровочному паттерну.

Как правило, калибровка производится единожды. Устройства, оснащённые системой камер, могут быть подвержены деформациям, замене элементов конструкции и иным возмущениям, приводящим к изменению расположения камер друг относительно друга, а следовательно и к приведению в негодность имеющейся калибровки; также, различные факторы, такие как температура окружающей среды – могут влиять на параметры оптических систем камер.

В случае автомобилей данная проблема могла бы быть решена наличием специальной “калибровочной” комнаты на каждом СТО, которая пустовала бы большую часть времени. Это было бы слишком дорого, поэтому активно развиваются альтернативные подходы к калибровке, не требующие калибровочных паттернов.

Помимо подходов, связанных с калибровкой по паттернам, имеется класс “автоматических” подходов к калибровке, как камер [23][20][25], так и систем из камер [17][16]. Данные подходы не требуют специальных паттернов.

Данные подходы, как правило, имеют некоторые ограничения: узкий угол обзора камеры, необходимость наличия пересечения области видимости камер, наличие калибровки внутренних параметров отдельных камер, предположения об особенностях расположения камер относительно окружающей среды — например, предположение о наличии достаточно плоской поверхности земли в области видимости. Ограничения сужают круг устройств, для которых может быть применён конкретный подход.

Камеры с широким углом обзора, именуемые fisheye-камерами, деформируют объекты на изображениях, искривляя прямые линии, но захватывают большую область пространства, потенциально уменьшая необходимое количество камер в системе. Данная деформация называется дисторсией.

В данной работе поднимается проблема калибровки жёсткой системы камер в общем случае: без ограничения на угол обзора камер и без необходиомсти пересечения областей видимости камер в системе; кроме того, в данной работе рассматривается в том числе и оценка не только внешних параметров камер, но и внутренних параметров, включая дисторсию.

2. Задача

2.1. Постановка задачи

С жёстко закреплённой системы, состоящей из N камер, в некоторые моменты времени одновременно с каждой из камер были получены изображения. Изображения содержат информацию о времени съёмки и о камере, с которой они были получены.

По данной коллекции изображений необходимо оценить внутренние параметры камер (параметры модели, проецирующей точки), внешние параметры камер в системе (отонсительные позы камер), а также позы системы камер в различные моменты времени.

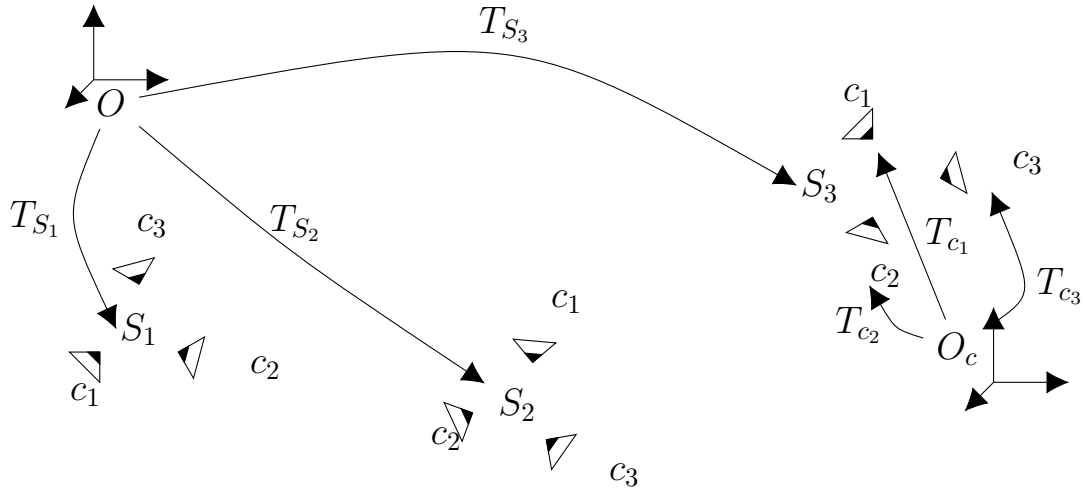


Рис. 1: Схематическое представление задачи

На рисунке 1 представлена схема поставленной задачи. O – мировая система координат, O_c – система координат системы из камер, S_i – система из камер в моменты времени i, c_j – камеры, T_{S_i} – переход из мировой системы координат в систему координат системы (поза) в момент времени i, T_{c_j} – позы камер в системе. Неизвестными являются позы T_{S_i}, T_{c_j} и внутренние параметры камер c_j.

Целью данной работы является разработка и реализация алгоритма калибровки жёстко закрепленной системы из камер по набору изображений (без использования калибровочных паттернов). Для достижения данной цели были сформулированы следующие задачи.

1. Реализовать одновременную оценку однопараметрической модели дисторсии и фундаментальных матриц для пар камер.

2. Реализовать оценку фокусных расстояний и существенных матриц, с последующим извлечением относительных поз.

3. Реализовать решение задачи усреднения вращений.

4. Реализовать решение задачи усреднения масштабов и переоценки смещений для системы.

5. Реализовать одновременную оптимизацию всех оцениваемых параметров.

В рамках данной работы рассматривается решение первых трёх задач.

3. Обзор литературы

Автоматической калибровке камер и калибровке систем камер посвящено множество работ. Многие подходы к автоматической калибровке, как говорилось ранее, опираются на соответствия ключевых точек, из которых в дальнейшем получаются трёхмерные точки. Как правило термин “ключевая точка” определяется для каждого алгоритма их поиска по своему. В общем случае это точка на изображении, содержащая достаточное количество информации в своей локальной окрестности.

Ключевые точки нужны для того, чтобы находить связь между изображениями. Качество ключевых точек характеризуется тем, сколько различных искажений можно применить к объекту, чтобы точка продолжала находиться в том же месте. Ключевые точки используются для сопоставления изображений, так называемый “матчинг” – процедура сопоставления ключевых точек одного изображения другому.

Особой популярностью пользуется детектор ключевых точек SIFT [15]. Данный алгоритм находит ключевые точки на разных разрешениях изображения, а так же формирует вектор, описывающий локальную окрестность – дескриптор ключевой точки. Дескриптор состоит из 128 чисел и инвариантен относительно поворота и изменения размера изображения, за что широко используется. Основываясь на работе [3], можно сделать вывод о том, что оригинальный дескриптор SIFT может быть изменен так, что качество матчинга между изображениями улучшится. RootSIFT – простая модификация дескриптора, приводящая к значительному улучшению качества. Данный дескриптор получается вычислением покомпонентного квадратного корня из l_1 нормализованных SIFT дескрипторов. Эта операция имеет смысл, так как SIFT является гистограммой ориентированных градиентов, а значит элементы дескриптора положительны. По своей сути это позволяет не меняя существующий $l2$ матчинг получать χ^2 метрику.

Получив сопоставления между ключевыми точками изображений, можно оценить геометрию сцены в виде набора фундаментальных или

существенных матриц. Помимо относительной позы изображений, из фундаментальной матрицы можно получить оценку внутренних параметров камер, на чём основаны многие работы.

Одним из ранних подходов является [23]. В данной статье авторы не учитывают дисторсию и оценивают внутренние параметры камеры в рамках pinhole модели. Помимо внутренних параметров камеры оценивается смещение камеры между снимками. Основная идея подхода состоит в модифицировании Absolute Dual Quadric из [20], путем использования изменяемых весов. Польза данного подхода в том, что он позволяет оценить фокусное расстояние, что может быть использовано в моей работе после оценки дисторсии.

Альтернативный подход к оценке фокусных расстояний – использование kruppa equations – соотношения между фундаментельной матрицей и матрицей внутренних параметров камеры. В [25] приводится вывод и решение kruppa equations для трёх кадров, снятых с одной камеры и с известными фундаментальными матрицами между ними. Авторы [11] предлагают решение задачи поиска фокусных расстояний для случая двух снимков с разных камер. В работе [11] предлагается неитеративный алгоритм оценки фокусных расстояний для двух изображений (в предположении, что центральная точка проекции известна).

Рассмотренные выше подходы работают в рамках pinhole модели без дисторсии. В случае нашей задачи предположение об отсутствии дисторсии неприемлемо. В [6] авторы занимаются проблемой оценки фундаментальной матрицы и параметров дисторсии между двумя снимками с разных камер. Авторы предлагают несколько решений, требующих 10 найденных пар соответствующих точек. Данные решения встраиваются в RANSAC [1] и по результатам, приведенным в статье, дают неплохой результат. Дисторсия оценивается в рамках division model [8].

Существует множество подходов, позволяющих калибровать системы из камер; как было указано ранее – они имеют некоторые ограничения применимости.

CamOdoCal[13] – фреймворк для калибровки систем камер. В данном решении приводится подход на основе использования калибровоч-

ных досок. Камеры калибруются по отдельности в проективную модель (Mei's model), а затем для каждой из камер запускается визуальная одометрия. Используя оценённую одометрию, происходит сопоставление ключевых точек, вычисляются как взаимные позиции камер, так и глобальные позиции всей системы из камер, затем выполняется нелинейная оптимизация всех оцененных параметров. Сопоставление ключевых точек происходит путем поиска в истории ключевых точек для каждой камеры в каждый из моментов времени.

Ключевой особенностью данного подхода является то, что он не требует пересечения полей видимости камер.

В работе [17] рассматривается проблема автокалибровки системы из нескольких шикроугольных камер. Инициализация внутренних параметров камер просиходит в предположении о том, что они все эквивалентны и имеют один и тот же угол обзора. Параметры дисторсии оцениваются исходя из угла обзора, остальные параметры через проективную реконструкцию. После инициализации производится построение 3d реконструкции, взаимное положение камер оценивается для каждых двух построенных реконструкций.

Автоматическую калибровку внутренних параметров камер возможно произвести в colmap [21] [19]. Это библиотека для построения реконструкций по коллекции изображений, которая поддерживает несколько моделей камер. Все парные оценки происходят в ректифицированной плоскости, что делает невозможным моделирование камер с углом обзора более 180°. В данном фреймворке нельзя задать условие постоянства относительных поз камер внутри системы. Ограничения, присутствующие в этом решении и то, что данный проект до сих пор поддержиается и разрабыватается, указывают на актуальность моей работы.

Отдельной частью моей работы является оценка относительных поз в системе камер. Она разбита на оценку вращений и смещений.

Основываясь на известных позах строится граф системы, где вершинами являются изображения, а рёбер – относительные позы. На данном графе решается задача поиска смещений и вращений камер внутри системы, а так же системы камер в разные моменты времени.

Задача поиска вращений более изучена, чем оценивание неизвестных смещений, поэтому поиск неизвестных поз предлагается произвести в два этапа: оценка неизвестных вращений, оценка неизвестных смещений. Поиск вращений является важной задачей, так как данным образом получаются начальные приближения для поз, впоследствии участвующих в глобальной нелинейной оптимизации. Поиск неизвестных вращений в вершинах графа называется задачей усреднения вращений.

В [5] использовалась IRLS процедура для дооптимизации решения, авторами были рассмотрены несколько функций, характеризующих расстояния между вращениями. Задача, решаемая авторами типична для усреднения вращений и выглядит следующим образом:

$$\arg\min \sum_{(i,j)\in\varepsilon} \rho\left(d\left(R_{ij}R_jR_i^{-1}\right)\right), \tag{1}$$

где R_i – неизвестные вращения, R_{ij} – известные парные позы. В качестве начального приближения использовалось l_1 решение на случайном остовном дереве графа вращений.

Зачастую для представления вращений используют группы Ли. В данном случае определены две взаимнообратные операции $e^v_{\mathfrak{so}(3)}, v \in R^3$ и $\log_{\mathfrak{so}(3)}(R), R \in SO(3)$. Тогда в качестве расстояния между вращениями можно использовать норму $\log_{\mathfrak{so}(3)}$, которая характеризует угол поворота вращения.

В [4] описывается двухэтапный подход к задаче усреднения вращений: l_1 решение, а затем IRLS[14] оптимизация. К недостаткам усреднения с помощью l_2 авторы относят сильную чувствительность к аутлаерам. Для улучшения предложенного решения авторы используют IRLS оптимизацию со следующей функцией потерь: $\rho(x) = \frac{x^2}{x^2+\sigma^2}$.

Авторы [24] анализируют устоичивость и разрешимость задачи усреднения вращений. Авторы показывают, как можно уменьшить количество степеней свободы задачи, зафиксировав вращение в одной из вершин. Данное действие также увеличивает область, где задача является выпуклой. Вывод из данной работы состоит в том, чтобы разбивать

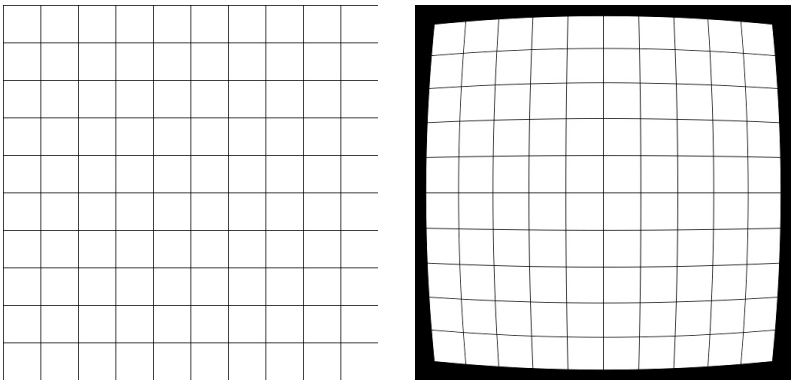
неравномерно распределённые компоненты на кластеры и решать их отдельно. Таким образом решается проблема зашумленности оценки на эпиполярную геометрию между кадрами.

4. Модель камеры

Калибровка камеры – задача оценки параметров модели камеры, описывающей проецирование точек.

В данной работе используются 2 параметрические модели камеры, одна из которых является расширением другой. В связи с выбраным решением задачи, в самом начале производится парная оценка дисторсии в рамках division model[8].

Данная модель параметризована 4-мя параметрами: фокусным расстоянием f; координатами центра проекции (c_x, c_y); одним коэфициентом для моделирования радиального искажения λ.



(a) $\lambda = 0$ (b) $\lambda = -0.625$

(c) $\lambda = -3.125$ (d) $\lambda = 3.125$

Рис. 2: Пример моделирования искажения с помощью division model.

Существенным преимуществом division model является возможность описания с помощью неё камер, с углом обзора превышающим 180 градусов, а также – полиномиальное выражение луча, соответствующего данной точке изображения, которое позволяет построить для данной

модели эффективные closed-form оценки [6] [7]. Основным недостатком модели ялвяется недостаточное количество параметров для качественного моделирования реальной дисторсии.

Отображения в данной модели выполняются следующим образом:

- отображение кооридanты пикселя (u, v) в луч : $\pi^{-1} : \mathbb{R}^2 \to \mathbb{R}^3$

$$\pi^{-1}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} u - c_x \\ v - x_y \\ f\left(1 + \lambda r^2\right) \end{pmatrix}, \quad r = \sqrt{(u - c_x)^2 + (v - c_y)^2} \tag{2}$$

- отображение трёхмерной точки в координаты на изображнии : $\pi : R^3 \to R^2$

$$\pi\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \alpha f \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix}, \tag{3}$$

$$\alpha = \frac{z - \sqrt{z^2 - 4\lambda r^2}}{2\lambda r^2}, \quad r = \sqrt{x^2 + y^2} \tag{4}$$

Данная модель камеры может быть расширена путём увеличения степени полинома $1 + \lambda r^2$. Для обеспечения гладкости в нуле, коэффициент перед первой степенью принимается равным нулю. Модель с полиномом степени больше 2-ой в данной работе будет называться polynomial model.

Отображение точки изображения в луч практически не изменится, третья координата будет формироваться полиномом степени n от радиуса.

$$z = f\left(1 + a_2 r^2 + ... + a_n r^n\right) \tag{5}$$

Обратное отображение выглядит несколько сложнее (в предположе-

нии, что $\|(x, y)\|_2 = 1$):

$$\pi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = f\alpha \begin{pmatrix} x \\ y \end{pmatrix} \tag{6}$$

$$\alpha = \text{root}\left(1 - zr + a_2 r^2 + ... + a_n r^n\right) \tag{7}$$

Для его реализации необходимо найти корень полинома. В данной работе это производится с помощью поиска собственных чисел сопровождающей матрицы.

5. Парные оценки

5.1. Фундаментальная матрица

Фундаментальная матрица F– матрица, связывающая нормализованные координаты точек двух изображений (с устранённой дисторсией), принадлежащих одной и той же трёхмерной точке.

$$x^\intercal F x' = 0 \tag{8}$$

Ранг фундаментальной матрицы равен 2, количество степеней свободы – 7. Фундаментальная матрица определена с точностью до умножения на ненулевую константу.

Так как фундаментальная матрица F имеет неполный ранг, то с точностью до умножения на некоторый скаляр она может быть представлена в виде

$$F = R_1 \operatorname{diag}(\cos(\theta), \sin(\theta), 0) R_2 \tag{9}$$

где R_1, R_2 – матрицы вращения; ввиду этого разложения, разумно параметризовывать её во время оптимизации как $SO(3) \times SO(3) \times S^1$.

5.2. Оценка

Для пары кадров, имеющих точки соответствия, производится одновременная оценка фундаментальной матрицы и параметров дисторсии в рамках division model. Метод, используемый для оценки, описан в [6].

Авторы в статье приводят метод решения для 10 соответствий. Данный метод не является минимальным так как неизвестных параметров в системе 9 – фундаментальная матрица (7), параметры дисторсии (2). В статье приводится сравнение скорости роботы методов, 10 точечный метод (F10) работает в $\sim$ 1000 раз быстрее чем 9 точечный, поэтому имеет смысл взять его, так как он будет более эффективен в паре с RANSAC.

В своей статье авторы решают следующую систему:

$$\left(u_{p_i},\ \ v_{p_i},\ \ 1+\lambda_1\left(u_{p_i}^2+v_{p_i}^2\right)\right) F \begin{pmatrix} u_{q_i} \\ v_{q_i} \\ 1+\lambda_2\left(u_{q_i}^2+v_{q_i}^2\right) \end{pmatrix} = 0, \quad i = 1,\ldots,10 \tag{10}$$

где координаты точек p_i и q_i отцентрированы и нормированы на диагональ изображения, так как в данном случае решения получаются более устойчивыми, неизвестными параметрами выступают λ_1, λ_2 и элементы матрицы F.

5.3. RANSAC и нелинейная оптимизация

Соответствий точек между кадрами ожидается сильно больше чем 10. Алгоритмы поиска соответствий ключевых точек могут неверно найти соответствующую парную точку, поэтому 10-ти точечный решатель встраивается в процедуру RANSAC[1].

Идея RANSAC состоит в том, чтобы N раз случайно сэмплировать точки из найденных соответствий, оценивать параметры модели и проверять, насколько хорошо они описывают все найденные точки. Выбираются те параметры модели, для которых находится большее количество инлаеров.

Инлаерами являются те пары точек, для которых некоторый функционал имеет значение менее заранее заданного порога. Предлагается использовать оценку ошибки репроекции, схожую с Sampson Error[12], но учитывающую наличие дисторсии. Вывод и вид данной ошибки будет представлен в разделе 5.5.

Имея начальные приближения фундаментальной матрицы и параметров дисторсии, а так же множество инлаеров – производится нелинейная оптимизация оценок, с целью лучшего описания точек.

Ошибкой в нелинейной оптимизации выбрана так же sampson error.

Оптимизационная задача выглядит следующим образом:

$$\arg\min_{F,\lambda_1,\lambda_2} \sum_i \rho\left(se\left(p_i, p_i', F, \lambda_1, \lambda_2\right)\right), \tag{11}$$

где se – sampson error, ρ – робастная функция потерь.

5.4. Увеличение порядка полинома дисторсии

После начальной инициализации параметров дисторсии камер предлагается увеличить порядок полинома дисторсии для более качественного описания радиального искажения.

Для вычисления значения sampson error, необходимо лишь вычисление луча по проекциям точки, данное действие – вычисление значения полинома в точке, поэтому можно использовать методы автоматического дифференцирования для решения задачи оптимизации.

$$\arg\min_{F,P_1,P_2} \sum_i \rho\left(se\left(p_i, p_i', F, P_1, P_2\right)\right), \tag{12}$$

где P_1 и P_2 – полиномы, со свободным коэфициентом 1 и с нулевым коэфициентом при первой степени.

5.5. Ошибка Сэмпсона

В данном разделе приводится распространение идеи ошибки Сэмпсона[12] на случай камер с радиальной дисторсией. Воспользуемся эпиполярным ограничением от фундаментальной матрицы в качестве функции ошибки:

$$C_F(X) = x_1^\intercal F x_2 \tag{13}$$

где x_1, x_2 – лучи, полученные из пикселей на изображениях, учитывая дисторсию, а X – (X_1, X_2) – агрегированные координаты пикселей на изображении.

В связи с наличием искажения, координаты это некая функция от

x, y:

$$C_F(X) = \begin{pmatrix} x_1, & y_1, & p_1(x_1, y_1) \end{pmatrix} F \begin{pmatrix} x_2 \\ y_2 \\ p_2(x_2, y_2) \end{pmatrix} \tag{14}$$

Пусть $X+\delta$ – правильные координаты точек с устранённой ошибкой. Разложив функцию ошибки в координатах в ряд Тейлора, до первого члена получим:

$$C_F(X + \delta) = C_F(X) + J\delta \tag{15}$$

Из условия оптимальности $X + \delta$ следует $J\delta = -C_F(X)$. С помощью множителей Лагранжа находим минимальный по норме δ, удовлетворяющий полученному выше ограничению.

$$\delta_{opt} = -J^\intercal \cdot \frac{C_F(X)}{JJ^\intercal} \tag{16}$$

Sampson error будем называть квадрат нормы δ_{opt}:

$$se = \frac{C_F(X)^2}{JJ^\intercal} \tag{17}$$

В случае division model якобиан по x_1 получается как:

$$\begin{gathered} \frac{\partial C_F(X)}{\partial x_1} = \begin{pmatrix} Tr\left[\left(\frac{\partial (p_1(x_1)^\intercal F p_2(x_2)}{\partial p_1(x_1)}\right)^\intercal \frac{\partial p_1(x_1)}{\partial (x_1)_1}\right] \\ Tr\left[\left(\frac{\partial (p_1(x_1)^\intercal F p_2(x_2)}{\partial p_1(x_1)}\right)^\intercal \frac{\partial p_1(x_1)}{\partial (x_1)_2}\right] \end{pmatrix} = \\ Fp_2(x_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{\partial p_1(r)}{\partial r}\frac{\partial r}{\partial (x_1)_1} & \frac{\partial p_1(r)}{\partial r}\frac{\partial r}{\partial (x_1)_2} \end{pmatrix} \end{gathered} \tag{18}$$

для производных по x_2 – аналогично.

Преимущество sampson error заключается в том, что она даёт хорошее приближение ошибки репроекции трёхмерных точек, не вводя их в рассмотрение как параметры оптимизационной задачи, существенно уменьшая размерности оптимизационных задач

6. Оценка фокусных расстояний и парных поз

6.1. Фокусные расстояния и агрегация оценок

После нелинейной оптимизации фундаментальных матриц вместе с полиномами дисторсии более высокого порядка производится получение фокусных расстояний f_1, f_2 из оцененной матрицы $\tilde{F}$. Используя алгоритм, описанный в [11]:

$$
\begin{aligned}
&\tilde{e_1} : \tilde{F}\tilde{e_1} = 0, \quad \tilde{e_2} : \tilde{F}^\intercal\tilde{e_2} = 0 \\
&T_1 : T_1\tilde{e_1} = (e_1, 0, e_2), \quad T_2 : T_2\tilde{e_2} = (e_1', 0, e_2'), \\
&F = [f_{ij}] = T_2\tilde{F}T_1^\intercal, \\
&f_1 = -\frac{f_{11}f_{21}e_1^2}{f_{11}f_{21}e_3^2 + f_{12}f_{22}}, \\
&f_2 = -\frac{f_{11}f_{12}e_1'^2}{f_{11}f_{12}e_3'^2 + f_{21}f_{22}}
\end{aligned}
\tag{19}
$$

После получения оценок фокусных расстояний, различные оценки каждой из камер объединяются в одну:

1. Фокусное расстояние выбирается, как медиана всех оценок фокусных расстояний для каждой модели.

2. Агрегирование оценок полиномов дисторсии:

$$
p_{opt} = \arg\min_p \sum_i \int_0^r r\,(p - p_i)^2\,, \tag{20}
$$

где p_i – оценка полинома с i изображения для данной камеры.

6.2. Существенная матрица

После агрегации оценок из оцененных ранее фундментальных матриц можно получить существенные матрицы, которые раскладываются

на вращения и направления смещений. Если преобразование системы координат второго кадра в первый записано как

$$x_1 = Rx_2 + T, \tag{21}$$

то существенная матрица может быть выражена, как:

$$E = R\,[T]_\times\,, \tag{22}$$

где $[\cdot]_\times$ – матрица векторного произведения. Вращение и направление смещения может быть получено из существенной матрицы с точностью до 4 решений[12].

Существенная матрица получается из фундаментальной с помощью полученных агрегированных оценок на фокусные расстояния($f_1,\ f_2$):

$$E = \operatorname{diag}(f_1, f_1, 1)\, F \operatorname{diag}(f_2, f_2, 1) \tag{23}$$

Полученные из (23) оценки существенных матриц могут не удовлетворять всем свойствам существенных матриц (в частности, построением гарантируется лишь ранг матрицы E, но не равенство её ненулевых сингулярных чисел), поэтому далее используется следующая оценка существенной матрицы (получаемая из SVD разложения[12]):

$$\widetilde{E} = U \operatorname{diag}(1, 1, 0) V^\intercal, \tag{24}$$

После получения начальных приближений существенных матриц, выполняется нелинейная оптимизация существенных матриц и фокусных расстояний, с использованием ценовой функции:

$$\underset{E_i, f_{i1}, f_{i2}}{\arg\min} \sum_i \min \left(\left\| E_i - \sqrt{2} \frac{\widetilde{E}_i}{\|\widetilde{E}_i\|_F} \right\|_F^2, \left\| E_i + \sqrt{2} \frac{\widetilde{E}_i}{\|\widetilde{E}_i\|_F} \right\|_F^2 \right) \tag{25}$$

где $\widetilde{E}_i$ получается из F_i с помощью (23). Так как существенная матрица получается из (22), то $\|E\|_F = \sqrt{2}\|T\|$, а у T поддерживается единичная норма, т.к. нам важно лишь направление смещения.

6.3. Разложение и оптимизация

Каждой существенной матрице соответствуют 4 различных движения камер[12]. Из возможных движений выбирается максимизирующее количество соответствий, для которых направление на результат триангуляции (с использованием коррекции из ошибки Сэмпсона (16)) совпадает с лучами, соответствующими точкам на изображении.

После получения позы из существенных матриц, производится нелинейная оптимизация поз и оценок на параметры камер. В качестве функции ошибки в данной задаче используется ошибка сэмпсона (5.5). Используя (23) и (22), фундаментальная матрица получается из относительной позы и фокусных расстояний как:

$$F = \mathrm{diag}(1, 1, f_1) R \left[T\right]_\times \mathrm{diag}(1, 1, f_2),$$

7. Усреднение вращений

В результате предыдущей части работы для некоторых пар изображений были получены оценки относительных поз. Используя их, множество камер в различные моменты времени может быть представлено в виде графа, где вершины – изображения. Каждая вершина характеризуется вращением и смещением относительно фиксированной системы координат.

Данный граф разбивается на компоненты связности. Задача усреднения вращений решается последовательно, сначала внутри компонент связности, затем (путём наложения ограничений движения твёрдого тела) производится оценка ориентаций камер внутри системы и системы камер в мире для максимально разрешимого количества моментов времени.

7.1. Усреднение вращений в компоненте связности

Рассмотрим компоненту связности, т.е. связный граф $\{E, V\}$. Необходимо оценить вращения каждой из вершин (изображения) относительно одной системы отсчёта. Для разрешимости системы вращение одной из вершин выбирается единичным. Для всех рёбер составим систему уравнений.

$$R_{ab}R_a - R_b = 0, \tag{26}$$

где R_{ab} – вращение из вершины V_a в вершину V_b, R_a, R_b – неизвестные вращения в вершинах. Матрица системы блочная, блок, соответствующий (26) представлен в (27).

Данная система линейна по неизвестным вращениям, рассматриваемым как 3x3 матрицы (отбросив, временно, ограничения матриц вра-

щения).

$$\begin{bmatrix} R_{ab} & 0 & 0 & & -I & 0 & 0 \\ 0 & R_{ab} & 0 & \cdots & 0 & -I & 0 \\ 0 & 0 & R_{ab} & & 0 & 0 & -I \end{bmatrix} \tag{27}$$

Легко видеть, что в каждой строке матрицы данной системы линейных уравнений – не более 4 ненулевых элементов; что делает выгодным использование методов, предназначенных для разреженых систем линейных уравнений. В данной работе используется метод сопряженных градиентов для наименьших квадратов с диагональным предобуславливателем.

Решения данной системы могут не удовлетворять условиям, налагаемым на матрицы вращения. В связи с этим каждая из оценок заменятся на ближайшую по норме Фробениуса матрицу вращения (с помощью метода, описанного в [26]).

После получения линейных оценок и приближения их ближайшей матрицей вращения при помощи SVD, можно удалить некоторые рёбра и вершины графа, ошибка в которых более какого-то порога. Данное действие помогает избавиться от некоторого количества аутлаеров среди парных поз. В качестве ошибки используется норма $\log_{\mathfrak{so}3}$, которая характеризует угол вращения, то есть её цель – отсеивать вращения сильно отличные от единичного поворота.

$$\theta = \|\log_{\mathfrak{so}(3)}(R_b^{\mathsf{T}} R_{ab} R_a)\| \tag{28}$$

7.1.1. Нелинейная IRLS оптимизация

Решения каждой из систем целесообразно нелинейно дооптимизировать, так как системы были решены без каких либо ограничений на матрицы вращений. Каждая система представима в виде:

$$R_b^{\mathsf{T}} R_{ab} R_a = E, \tag{29}$$

где R_b, R_a – необходимо дооценить, а R_{ab} – считается точно известной.

В качестве ошибки используется $\log_{\mathfrak{so}(3)}$.

$$f\left(R_a, R_b\right) = \log_{\mathfrak{so}(3)}\left(e^{-\Delta b} R_b^\intercal R_{ab} R_a e^{\Delta a}\right), \tag{30}$$

где под e подразумевается $e_{\mathfrak{so}(3)}$.

Для решения данной системы используется IRLS[14], с функцией потерь Хьюбера:

$$r : r_i = f\left(X_{n-1}\right) \tag{31}$$

$$w_i = \frac{1}{\|r_i\|^2 + \sigma^2} \tag{32}$$

$$W_n = \operatorname{diag}\begin{pmatrix} w_1 & \cdots & w_n \end{pmatrix} \otimes I_3 \tag{33}$$

$$\Delta X : J^\intercal W_n J \Delta X = J^\intercal W_n r \tag{34}$$

$$X_n = X_{n-1} - \Delta X, \tag{35}$$

где X – R_a и R_b записаны в столбец в виде элементов $\mathfrak{so}(3)$, f – функция, вычисляющая остатки каждого уравнения системы, для удобства записана таким образом. Для решения разреженной линейной системы на каждом шаге используется метод сопряженных градиентов.

7.2. Оценка ориентаций камер внутри системы

Оценив ориентации всех вершин для каждой из компонент связности, можно перейти к оценке ориентаций камер внутри системы. Для всех возможны пар камер оцениваются относительные вращения. Возможны 2 вида ограничений.

Внтури компоненты связности в один момент времени

$$R_{ab} = V_{s,b}^\intercal V_{s,a}, \tag{36}$$

где $V_{s,c}$ – ориентация камеры c в момент времени s (задана из вершины в мир), R_{ab} – искомое относительное вращение между двумя камерами.

Между компонентами связности в два общих момента времени:

$$R_{ab}V_{t_1,a}^{\intercal}V_{t_2,a} = V_{t_1,b}^{\intercal}V_{t_2,b}R_{ab}, \tag{37}$$

где камеры a и b лежат в разных компонентах связности, а t_1, t_2 – моменты времени в которые производилась съемка.

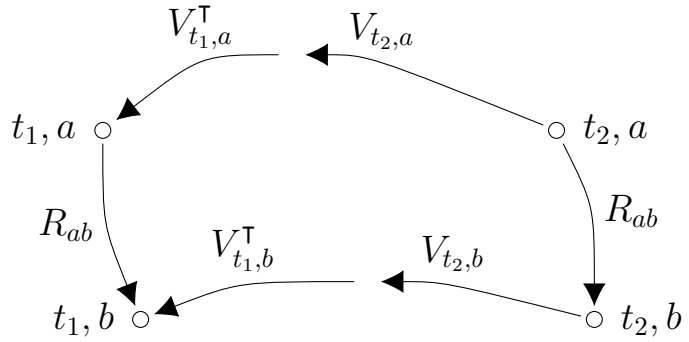


Рис. 3: Схематическое представление ограничения (37)

Используя в данном случае параметризацию вращений в виде кватернионов, каждое уравнение представляется матричной системой размера 4x4. Полученная после агрегации всех наблюдений для каждой пары из камер система может получиться большого размера, поэтому разумно сразу собирать нормальные уравнения.

При решении задачи данным подходом, возникает проблема с тем, что единичные кватернионы дважды покрывают $SO(3)$. Вращение получаемое действием кватерниона q соответствует вращению, получаемому с помощью $-q$. То есть (36) и (37) на самом деле задают следующие ограничения соответственно:

$$\begin{aligned} q_{R_{ab}} &= \pm q_{V_{s,b}^{\intercal}} q_{V_{s,a}} \\ q_{R_{ab}} q_{V_{t_1,a}^{\intercal}} q_{V_{t_2,a}} &= \pm q_{V_{t_1,b}^{\intercal}} q_{V_{t_2,b}} q_{R_{ab}} \end{aligned} \tag{38}$$

В случае однородной системы мы имеем уравнение вида $qq_a = q_b q$, где q – неивзвестный кватернион. Заметив, что в qq_aq^* действительная часть не отличается от таковой у q_a, можно решить проблему со знаками, выбрав q_a и q_b такими, что у них совпадает знак действительной части.

Для решения данной системы используется SVD-разложение. Систе-

ма приводится к виду $\begin{bmatrix} A & -b \end{bmatrix} x = 0$. Решением будет один из правых сингулярных векторов. При ненулевой части b следует взять последний вектор. Во втором случае, то есть при однородной системе, решением будет правый сингулярный вектор, соответствующий второму минимальному сингулярному числу.

В качестве кватерниона берутся первые 4 числа из решения и нормируются.

Сформулировав задачу в смысле нелинейных наименьших квадратов : $\arg\min\|f(x)\|^2$, можно дооптимизировать полученный ответ, а также оценить ковариацию полученной оценки. Ковариация считается для $\mathfrak{so}(3)$ параметризации, в предположении о том, что все остатки гомоскедастичны. Тогда оценка ковариация выражается следующим образом:

$$C = 4\sigma^2 \frac{(J(\hat{x})^\intercal J(\hat{x}))^{-1}}{N}, \tag{39}$$

где $\hat{x}$ – текущая оценка вращения, $J(x)$ – якобиан $f()$ в точке x, σ^2 – остатки, N – количество остатков.

В связи с несмещенностью оценок, математическое ожидание ошибки находится в нуле, поэтому математическое ожиданеие квадрата нормы вектора в $\mathfrak{so}(3)$ представлении : $\mathrm{Tr}(C)$, что по своей сути является величиной квадрата угла поворота.

Используя данную оценку, можно выбрать те относительный оценки, математическое ожидание квадрата угла ошибки которых менее определенного порога. Далее ориентации части камер вычисляются используя (26).

Предположение о том, как можно оценить оставшиеся параметры, описано в приложении А.

8. Результаты

Данная работа покрывает оценку внутренних параметов камер; ориентаций камер внутри системы и системы камер во времени. В рамках данной работы достигнуты следующие результаты:

- Реализована и протестирована парная оценка фундаментальных матриц и полиномов дисторсии.
- Реализовано и протестировано получение фокусных расстояний и существенных матриц из оценённой фундаментальной матрицы.
- Построена и апробирована ошибка Сэмпсона, учитывающая дисторсию.
- Реализована оценка ориентаций изображений, начальная оценка ориентаций некоторых камер внутри системы

Реализация всех описанных алгоритмов выполнена на языке C++ с использованием следующих дополнительных библиотек:

- Eigen [10] – для линейной алгебры
- Ceres Solver [2] – для нелинейной оптимизации
- Sophus [22] – для параметризации поз

8.1. Парные оценки дисторсии и фундаментальных матриц

Реализация алгоритма [6], выполненная в данной работе, позволяет получать 30000 решений для 10-точечной задачи в секунду.

В качестве синтетического теста случайным образом генерируются трехмерные точки (p_{3d}), матрица поворота (R), вектор смещения(T), параметры калибрвоки (π_1, π_2); на основании этих данных точки проецируются на изображения.

$$p_1 = \pi_1(p_{3d})$$
$$p_2 = \pi_2(R^\intercal\,(p_{3d} - T))$$

По сгенерированным и спроецированным точкам оценивается фундаментальная матрица и 2 параметра дисторсии.

На синтетических тестах с нормально распределенным шумом оценки выглядят следующим образом:

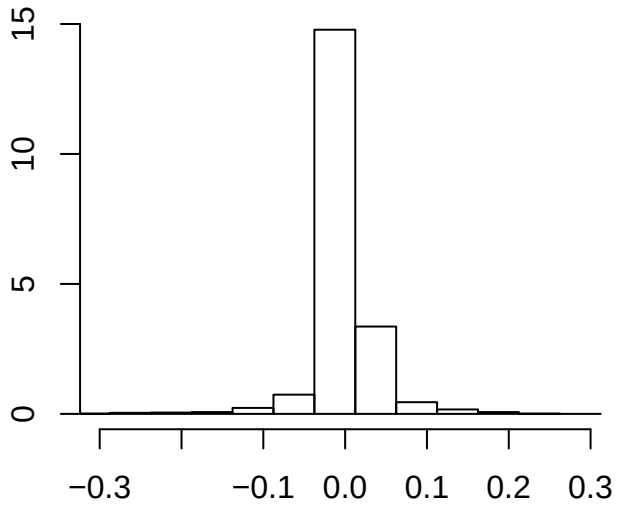
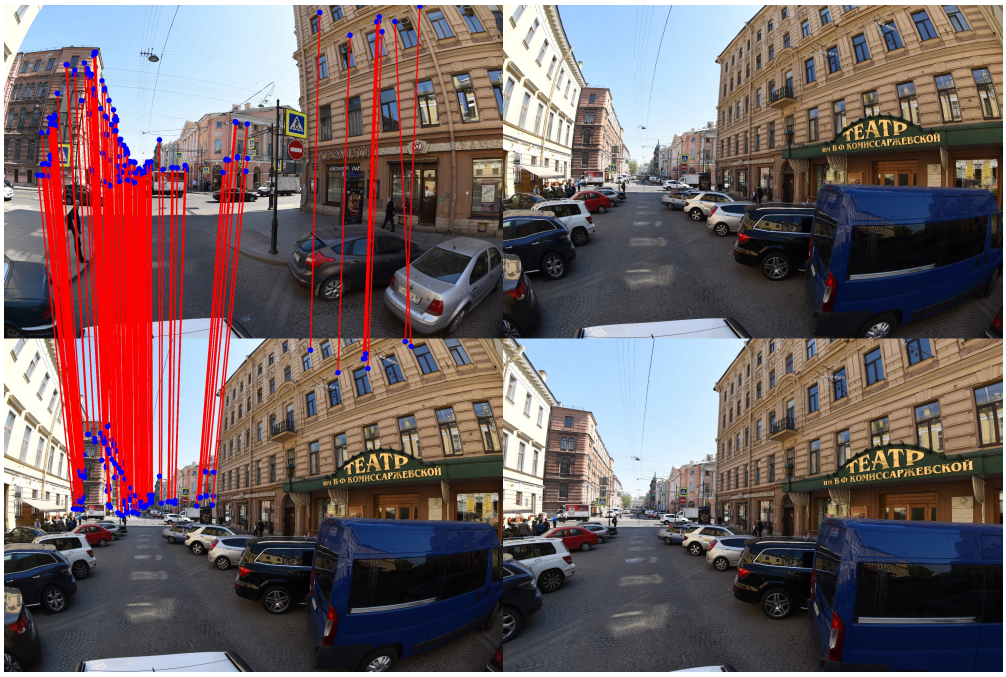


Рис. 4: Гистограмма относительной разницы λ оцененной и заданной

8.1.1. Тестирование на реальных изображениях



(a) (b)

Рис. 5

Для поиска и сопоставления ключевых точек была использована библиотека OpenCV. Был использован детектор ключевых точек SIFT [15], в качестве дескрипторов был реализован RSIFT дескрипторы [3]. Для матчинга был использован brute-force mathcing, также реализованный в OpenCV.

На рисунке 5 показаны 2 изображения и получившиеся ключевые точки с их сопоставлениями; справа продемонстрировано устранение дисторсии с использованием полученой оценки (изображение с устранённой дисторсией – снизу). Данные изображения харкатеризуют модель без увеличения степени полинома дисторсии. При корректной отработки алгоритма, прямые линии должны выпрямиться на изображении.

8.2. Фокусные расстояния и агрегирование парных оценок

8.2.1. Фокусные расстояния

На тестах без шума, когда фундаментальная матрица генерируется из фокусных расстояний, матрицы поворота и смещеня:

$$F = \operatorname{diag}(1, 1, f_1) R \left[T\right]_\times \operatorname{diag}(1, 1, f_2) \tag{40}$$

Относительная ошибка вычисления каждого фокусного расстояния составляла порядка 1^{-10}.

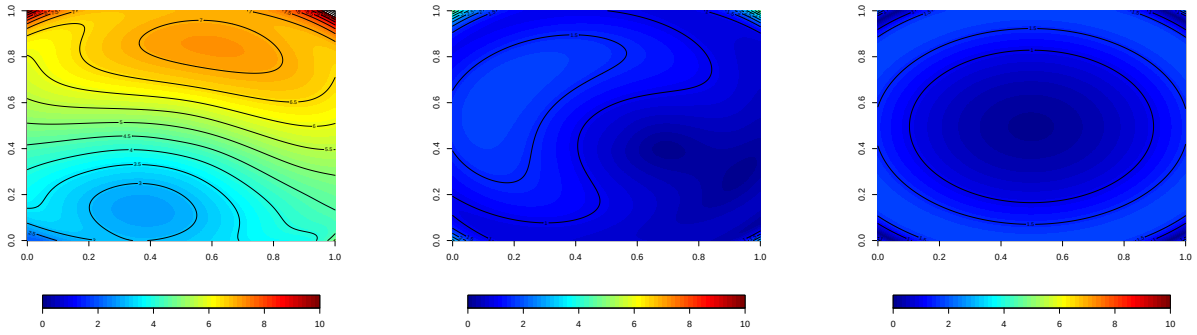
Данная работа не содержит тестирование данного алгоритма на устойчивость, результаты, полученные с различным шумом, приведены в статье [11]. В [11] описаны ситуации, когда фокусные расстояния невозможно вычислить. Пары изображений, для которых невозможно вычислить фокусные расстояния, не участвуют далее в работе алгоритма.

8.2.2. Агрегирование оценок

На датасете, используемом для тестирования, количество инлаеров среди матчей для пар изображений $\sim 40\%$; для достижения вероятности отсутствия корректной гипотезы примерно равной 0.05% (в рамках RANSAC), производилось 50000 итераций.

Для оценки точности калибровки, оценённая калибровка сравнивалась с калибровкой камер по паттернам из датасета. В качестве функции ошибки выступает $\|\pi_{gt} \circ \pi_0^{-1}(u) - u\|$, π_{gt} – калибровка по паттернам, π_0^{-1} – полученная алгоритмом калибровка, u – координата пикселя на изображении.

Центр проекции в алгоритме инициализируется в центре изображения, изменяясь в дальнейшем только в нелинейной оптимизации. Есть возможность заранее задать известное начальное предположение и зафиксировать его, поэтому сравнение калибровки происходило в трёх вариантах:



(a) Центр проекции фиксирован в центре изображения

(b) Центр проекции инициализирован в центре изображения

(c) Центр проекции фиксирован в центре из ground-truth

Рис. 6: Ошибка калибровки в пикселях по плоскости изображения.



(a) (b)

Рис. 7: Пример выпрямленных изображений после агрегирования оценок

8.2.3. Нелнейная оптимизация с ошибкой Сэмпсона

В вычислении остатков при решении оптимизационной задачи, описанной в разделе 6.3, присутствует множество общих вычислений: существенная матрица и производные её элементов по параметризации вращения и направления смещения, производные ненормированного луча по параметрам дисторсии. Вычисления существенных матриц и их производных были вынесены в отдельную подготовительную фазу, выполняющуюся перед каждым вычислением остатков и якобианов параллельно по рёбрам.

8.3. Усреднение вращений

Для вычисления ошибки вращений находится оптимальное вращение между двумя множествами: полученных оценок и ground-truth. Имея пару вращений (в кватернионах) $\tilde{q_i}$ – полученная оценка, q_i – вращение из ground-truth, оптимальное вращение q может быть получено с помощью решения следующей оптимизационной задачи, составленной из всех пар i:

$$\begin{aligned} q &= \arg\min \sum_i \|q - \tilde{q_i} q_i^{-1}\|^2 \\ q_x^2 + x_y^2 + q_z^2 + q_w^2 &= 1 \end{aligned} \tag{41}$$

Решая данную систему с помощью метода множителей Лагранжа, получается аналитическое решение:

$$q = \frac{\sum\limits_i \tilde{q_i} q_i^\intercal}{\|\sum\limits_i \tilde{q_i} q_i^\intercal\|} \tag{42}$$

Затем в качестве параметризации рассматриваются матрицы вращения и нелинейно оптимизирется $R\ (q)$ в смысле $\log_{\mathfrak{so}3}$. Ошибка e для каждой пары:

$$e = \|\log_{\mathfrak{so}3}\left(R R_i \tilde{R_i}^\intercal\right)\| \tag{43}$$

Ниже представлены результаты с использованием ceres-solver [2], для нелинейной оптимизации вращений.

8.3.1. Датасет KITTI

Тестирование точности получаемых предлагаемым алгоритмом относительных и абсолютных ориентаций изображений производилось на последовательности 00 из набора данных KITTI[9], предназначенного для тестирования алгоритмов визуальной одометрии. Поиск соответствий между изображениями производился реализованным в colmap[21] матчингом, сопоставляя каждую пару изображений из последовательности.

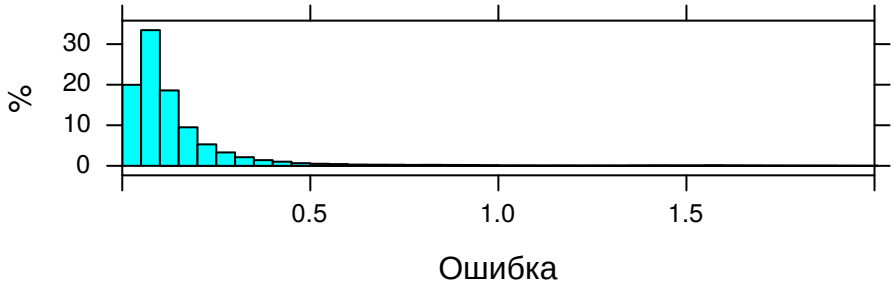


Рис. 8: Пример кадра из последовательности 00 для одометрии

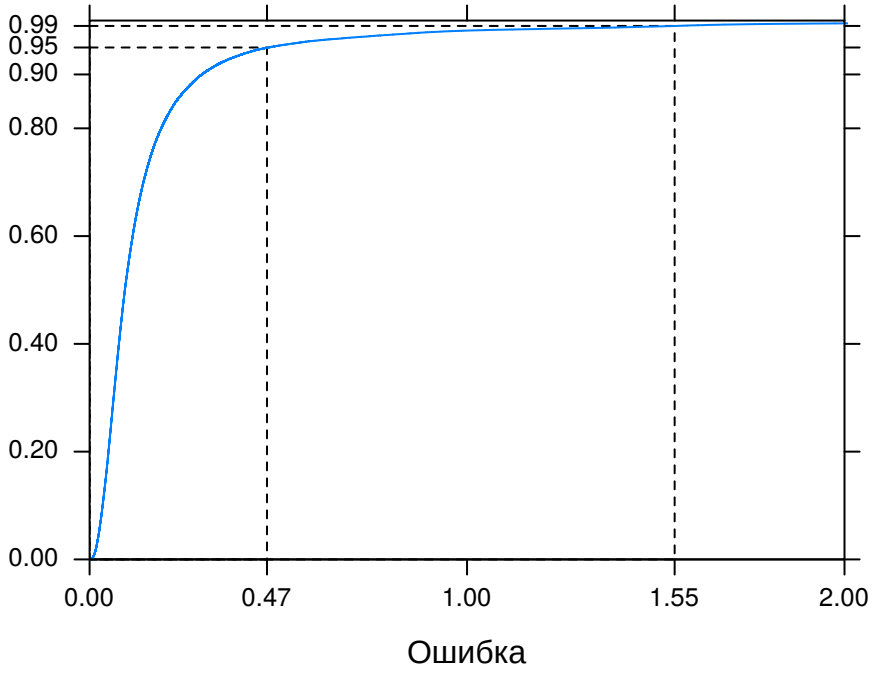
На рисунке 9 показано распределение ошибки в оценке относительных ориентаций изображений.

Ошибка оценки ориентаций вершин приведена на рисунке 10.

Данный этап работы алгоритма нужен для того, чтобы получить необходимое количество “правильных” поз для вычисления ориентаций камер и системы камер во времени.



(a)



(b) Функция распределения ошибки

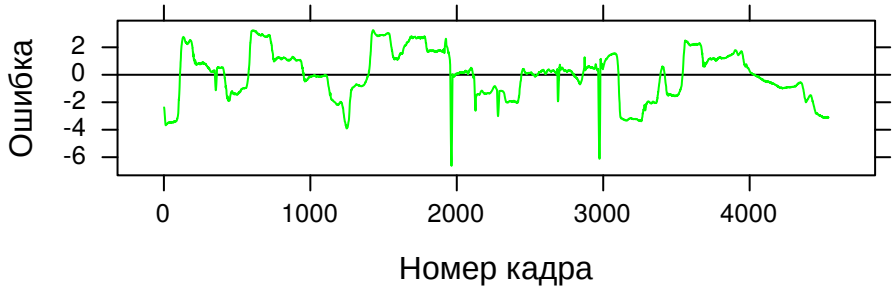
Рис. 9: Ошибка вычисления относительных ориентаций (градусы)

8.3.2. Датасет ladybug

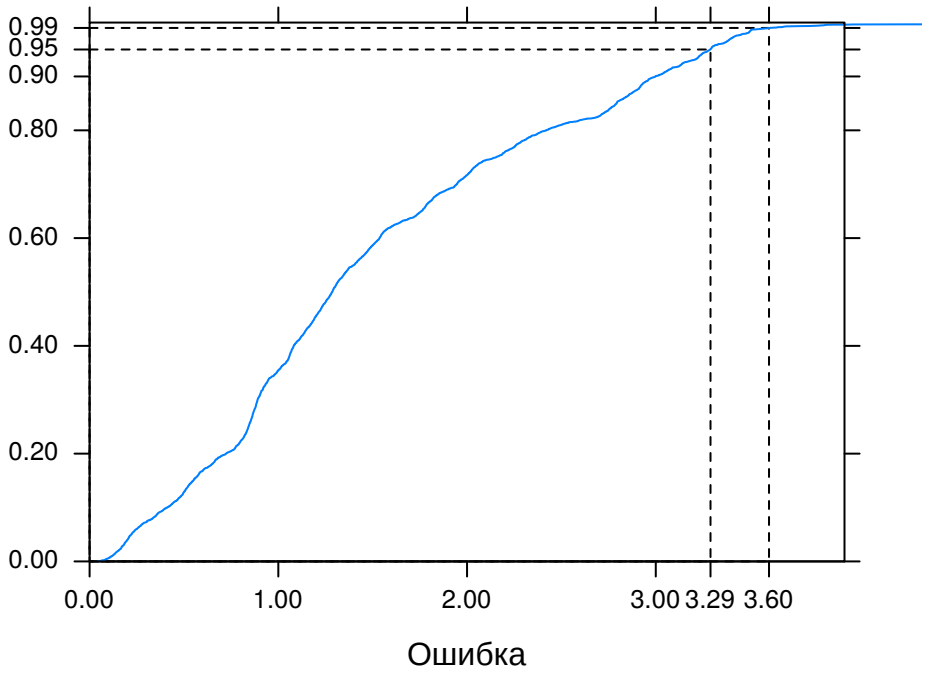
Тестирование ориентаций также проводилось на снятых с ladybug изображениях. Пример кадров представлен на рисунке 11.

В отличии от KITTI камеры в этом датасете подвержены значительной дисторсии. Данный датасет содержит 6 разнонаправленных камер, использовались все, за исключением ориентированной вверх камеры.

Распределение ошибки относительных вращений показано на рисунке 12. Как видно подавляющее большинство ориентаций было оценено



(a) Азимутальная ошибка от номера кадра



(b) Функция распределения ошибки

Рис. 10: Ошибка ориентаций изображений (градусы)

с ошибкой менее 3.5°(95-ый квантиль). Согласно результатам из [18], это хорошее начальное приближение для дальнейшей оценки поз.

На рисунке 13 показана гистограмма ошибки ориентаций изображений.

Данный проезд оказался цельным, то есть после оценки относительных поз и агрегации параметров не возникло разбиение на несколько компонент связности. Единственная компонента связности содержит 7715 рёбер и 1494 вершины.



Рис. 11: Пример кадров с ladybug

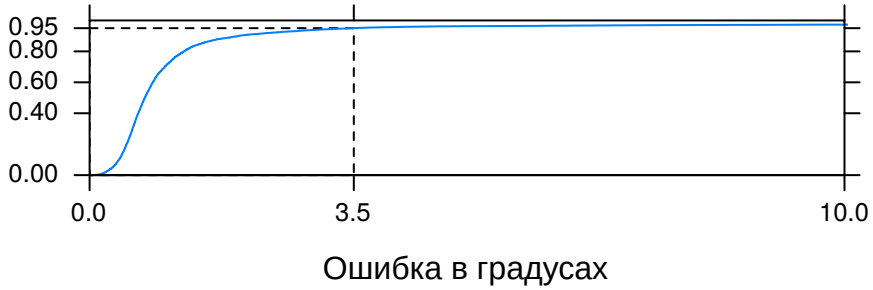


Рис. 12: Функция распределения ошибки относительных вращений

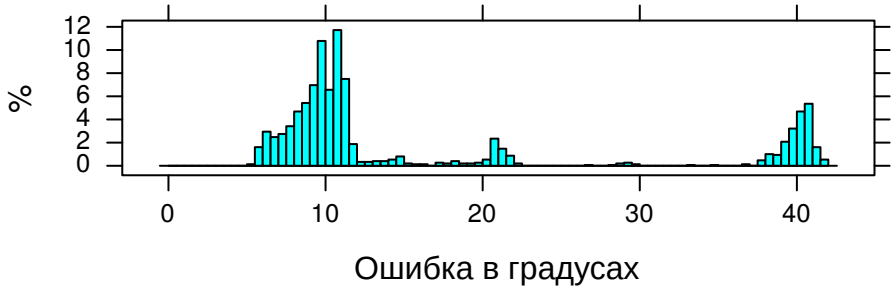


Рис. 13: Ошибка ориентаций изображений

Как видно из графика 14, медиана ошибка оценки ориентаций 10.6°, что сопоставимо с медианной ошибкой, полученной в [4] и ошибкой рассматриваемых в [5] подходов. Данная оценка не является финальным результатом, полученные оценки используются в качестве начального приближения для получения дальнейших оценок на ориентации камер и системы из камер.

Для полученных оценок ориентаций изображений можно оценить

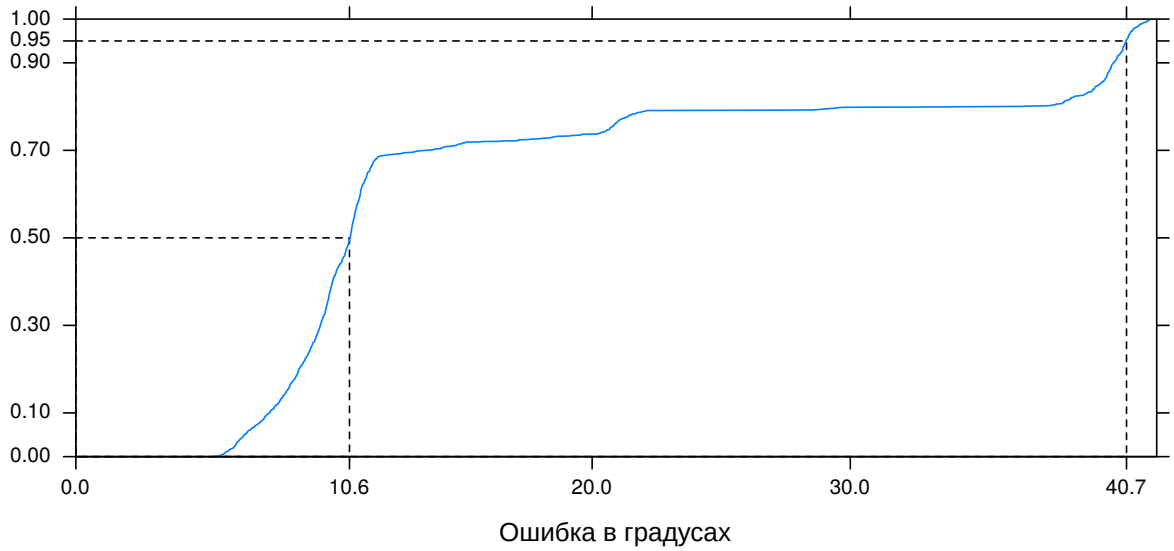


Рис. 14: Функция распределения ошибки

относительные ориентации камер внутри системы, описанным в 7.2 способом. Полученные на датасете ladybug относительные ориентации представлены в таблице 1.

	0	1	2	3	4
0					
1	0.64				
2	0.28	0.73			
3	0.49	1.13	0.50		
4	0.49	1.12	0.40	0.22	

Таблица 1: Ошибка относительных ориентаций камер в градусах

Как видно из таблицы, оценки относительных ориентаций камер имеют ошибку менне 1.13°, что является хорошим результатом для данного этапа получения оценок. Используя оценённые относительные ориентации оценивается ориентация каждой из камер внутри системы, ошибки ориентации каждой из камер приведены в таблице 2.

0	1	2	3	4
0.1544935	0.7023890	0.1361514	0.3847801	0.3780175

Таблица 2: Ошибка ориентаций камер в градусах

Полученные оценки ориентаций камер внутри системы сравнимы с получаемыми в [13] с единственным отличием в том, что в случае данной работы они получены до финальной оптимизации с введением трёхмерных точек.

Список литературы

[1] A. Fischler Martin, C. Bolles Robert. Random Sample Consensus: A Paradigm for Model Fitting with Applications To Image Analysis and Automated Cartography // Communications of the ACM. — 1981. — 06. — Vol. 24. — P. 381–395.

[2] Agarwal Sameer, Mierle Keir, Others. Ceres Solver. — `http://ceres-solver.org`.

[3] Arandjelović Relja, Zisserman Andrew. Three things everyone should know to improve object retrieval // 2012 IEEE Conference on Computer Vision and Pattern Recognition / IEEE. — 2012. — P. 2911–2918.

[4] Chatterjee A., Govindu V. M. Efficient and Robust Large-Scale Rotation Averaging // 2013 IEEE International Conference on Computer Vision. — 2013. — P. 521–528.

[5] Chatterjee A., Govindu V. M. Robust Relative Rotation Averaging // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2018. — Vol. 40, no. 4. — P. 958–972.

[6] Efficient Solution to the Epipolar Geometry for Radially Distorted Cameras / Zuzana Kukelova, Jan Heller, Martin Bujnak et al. — 2015. — 12. — P. 2309–2317.

[7] Fast and robust numerical solutions to minimal problems for cameras with radial distortion / M. Byrod, Z. Kukelova, K. Josephson et al. // 2008 IEEE Conference on Computer Vision and Pattern Recognition. — 2008. — June. — P. 1–8.

[8] Fitzgibbon A. W. Simultaneous linear estimation of multiple view geometry and lens distortion // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. — Vol. 1. — 2001. — Dec. — P. I–I.

[9] Geiger Andreas, Lenz Philip, Urtasun Raquel. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite // Conference on Computer Vision and Pattern Recognition (CVPR). — 2012.

[10] Guennebaud Gaël, Jacob Benoît et al. Eigen v3. — http://eigen.tuxfamily.org. — 2010.

[11] Hartley Richard. Extraction of focal lengths from the fundamental matrix // Unpublished manuscript. — 1993.

[12] Hartley Richard, Zisserman Andrew. Multiple view geometry in computer vision. — Cambridge university press, 2003.

[13] Heng Lionel, Li Bo, Pollefeys Marc. CamOdoCal: Automatic intrinsic and extrinsic calibration of a rig with multiple generic cameras and odometry // IROS. — IEEE, 2013. — P. 1793–1800.

[14] Holland Paul W., Welsch Roy E. Robust regression using iteratively reweighted least-squares // Communications in Statistics - Theory and Methods. — 1977. — Vol. 6, no. 9. — P. 813–827. — https://doi.org/10.1080/03610927708827533.

[15] Lowe David G. Distinctive image features from scale-invariant keypoints // International journal of computer vision. — 2004. — Vol. 60, no. 2. — P. 91–110.

[16] MultiCol bundle adjustment: a generic method for pose estimation, simultaneous self-calibration and reconstruction for arbitrary multi-camera systems / Steffen Urban, Sven Wursthorn, Jens Leitloff, Stefan Hinz // International Journal of Computer Vision. — 2017. — Vol. 121, no. 2. — P. 234–252.

[17] Nguyen Thanh Tin, Lhuillier Maxime. Self-calibration of omnidirectional multi-cameras including synchronization and rolling shutter // Computer Vision and Image Understanding. — 2017. — 08. — Vol. 162.

[18] Optimizing the Viewing Graph for Structure-from-Motion / C. Sweeney, T. Sattler, T. Höllerer et al. // 2015 IEEE International Conference on Computer Vision (ICCV). — 2015. — P. 801–809.

[19] Pixelwise View Selection for Unstructured Multi-View Stereo / Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, Jan-Michael Frahm // European Conference on Computer Vision (ECCV). — 2016.

[20] Pollefeys M., Koch R., Van Gool L. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters // Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). — 1998. — Jan. — P. 90–95.

[21] Schönberger Johannes Lutz, Frahm Jan-Michael. Structure-from-Motion Revisited // Conference on Computer Vision and Pattern Recognition (CVPR). — 2016.

[22] Strasdat Hauke. Sophus. — `https://github.com/strasdat/Sophus`.

[23] Thormählen Thorsten, Broszio Hellward, Mikulastik Patrick. Robust Linear Auto-calibration of a Moving Camera from Image Sequences. — 2006. — 01. — P. 71–80.

[24] Wilson Kyle, Bindel David, Snavely Noah. When is rotations averaging hard? // European Conference on Computer Vision / Springer. — 2016. — P. 255–270.

[25] Xu Gang, Sugimoto Noriko. Algebraic derivation of the Kruppa equations and a new algorithm for self-calibration of cameras // JOSA A. — 1999. — Vol. 16, no. 10. — P. 2419–2424.

[26] Zhang Z. A flexible new technique for camera calibration // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2000. — Nov. — Vol. 22, no. 11. — P. 1330–1334.

[27] A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern / Bo Li, Lionel Heng, Kevin Koser, Marc Pollefeys // 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems / IEEE. — 2013. — P. 1301–1307.

A. Оценка оставшихся ориентаций

Оценив некоторые ориентации камер, можно переоценить ориентации вершин, путём введения новых ограничений.

Оценки на ориентации камер дают возможность ввести в рассмотрение ориентацию системы камер.

$$R_v = R_s R_c, \tag{44}$$

где R_v – ориентация вершины, R_s – ориентация системы в момент времени изображения с вершины, R_c – ориентация камеры.

Вводя данную факторизацию в уравнения задачи усреднения вращений, можно переоценить ориентации вершин. Дополнительно на данном этапе можно связать некоторые компоненты связности, в которых в один и тот же момент имеются камеры с имеющейся оценкой на ориентацию. В данном случае вместо фиксации ориентации одной из вершин следует зафиксировать ориентацию системы из камер в один из таких моментов времени.

После переоценки ориентаций вершин, можно составить систему, схожую с (37) и (36), чтобы оценить неизвестные ориентации камер.

Этот итеративный процесс предполагается остановить, если никакая из полученных оценок не удовлетворяет ограничению, наложенным следом ковариационной матрицы (39).

A.1. Оценка ориентаций системы из камер

Оценив ориентации камер и связав с помощью этого компоненты связности, можно переоценить и оценить оставшиеся ориентации системы камера в моменты ремени.

$$R_s = R_v^\intercal R_c, \tag{45}$$

где R_v – ориентация вершины, R_s – ориентация системы в момент времени изображения с вершины, R_c – ориентация камеры.

Составляя данную систему, параметризируя кватернионами, получаем следующую систему:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} q_w \\ q_x \\ q_y \\ q_z \end{bmatrix} = q_v^\intercal q_c, \tag{46}$$

Формируя для каждого момента времени из всех вершин и соответствующих камер, задача формулируется в смысле наименьших квадратов аналогично (41) и её решением будет (42).

Таким образом можно переформулировать (26), факторизуя R_a на композицию вращения камеры и системы из камер:

$$R_{ab} = R_{c_a}^\intercal R_{s_a}^\intercal R_{s_b} R_{c_b}, \tag{47}$$

где c_a – камера соответствующая вершине a, s_a – момент времени вершины a.

Оценив ориентации системы камер из (46), все оценки на вращения оптимизируются, аналогично оптимизации в усреднении вращений:

$$min \sum_{a,b} \rho \left(\| \log_{\mathfrak{so}(3)} \left(R_{ab}^\intercal R_{c_a}^\intercal R_{s_a}^\intercal R_{s_b} R_{c_b} \right) \|^2 \right) \tag{48}$$