

Санкт-Петербургский Государственный Университет
Математическое обеспечение и администрирование
информационных систем

Системное программирование

Миронович Елизавета Викторовна

Глобальная реконструкция и некалиброванные камеры

Курсовая работа

Научный руководитель:
Пименов А. А.

Консультант:
Корчёмкин Д. А.

Санкт-Петербург
2019

Оглавление

Введение	3
1. Обзор существующих решений	5
2. Постановка задачи	9
3. Реализация	10
Заключение	13
Список литературы	14

Введение

С развитием компьютерной графики и общедоступности фотокамер стала популярна задача 3D реконструкции. Она состоит в том, чтобы создать 3D модель объекта из набора его фотографий.

Процесс 3D реконструкции можно условно разделить на три этапа. На первом этапе задача состоит в том, чтобы из набора фотографий выделить условно похожие и определить, какие точки на фотографиях соответствуют одним и тем же точкам в реальном мире. На втором этапе, с помощью найденных соответствий возникает задача *structure-from-motion*: если каждому изображению условно сопоставить камеру, с помощью которой оно было сделано, то задачу можно сформулировать как оценку расположения камер, относительно друг друга. На третьем этапе, с помощью полученных оценок, восстанавливаются точки для модели и производится оценка поверхностей и текстур.

Данная работа будет посвящена задаче *structure-from-motion*.

Structure-from-motion

Существуют два принципиально разных подхода к *structure-from-motion*: итеративный и глобальный.

При использовании итеративных методов сначала оценивается расположение двух камер относительно друг друга и соответствующих им трёхмерных точек. Затем итеративно оценивается расположение каждой следующей камеры относительно существующего на текущий момент облака точек, облако пополняется и производится *Local/Global bundle adjustment* [9]. Недостаток такого решения в том, что *Bundle adjustment* занимает значительную долю времени, а также в том, что итеративные методы склонны к накоплению ошибки даже при наличии замыкания циклов.

В противовес итеративной реконструкции существует глобальная. Глобальные методы SfM одновременно оценивают расположение всех камер, учитывая все возможные попарные ограничения на их расположения. Данный метод основывается на использовании групповых свойств

$SO(3)$ и $SE(3)$, и не порождает трёхмерных точек на промежуточных шагах, что позволяет снизить размерности оптимизационных задач, а Bundle Adjustment выполняется лишь один раз, в качестве финального шага.

В данной работе метод глобальной реконструкции будет обобщён на случай некалиброванных камер.

Проективная геометрия

В данной работе, и в компьютерном зрении в принципе, широко используются концепции проективной геометрии, подробно описанные в [6]. Проективное пространство P^n можно воспринимать как расширение евклидова, полученное добавлением к точкам дополнительной координаты и отождествлением точек вида $(x_1, x_2, \dots, x_{n+1})$ и $(kx_1, kx_2, \dots, kx_{n+1})$, где $k \neq 0$. Так как можно построить соответствие между точками в евклидовом пространстве и в проективном, а работать в последнем удобнее, то используют точки в проективном пространстве.

1. Обзор существующих решений

Если модель камеры представлять pinhole моделью, т.е. точкой и плоскостью, на которой изображение получается с помощью центральной проекции, то эпиллярная геометрия для двух камер с центрами C и C' и соответствующими плоскостями α и α' описывает эиполи – проекции C на α' и C' на α .

Алгебраическим представлением эиполярной геометрии является фундаментальная матрица. Если есть соответствие $x \leftrightarrow x'$ координат в P^2 точек на двух изображениях, являющихся проекциями одной и той же точки X , то эиполярной прямой l' для x будет называться прямая проходящая через x' и эиполь e' в плоскости второго изображения (см. Рис. 1). Фундаментальная матрица связывает эти понятия соотношением $l' = Fx$ [6]. То что прямая l' проходит через x' можно записать как $x'^T l = 0$ в координатах P^2 , что приводит к уравнению $x'^T Fx = 0$.

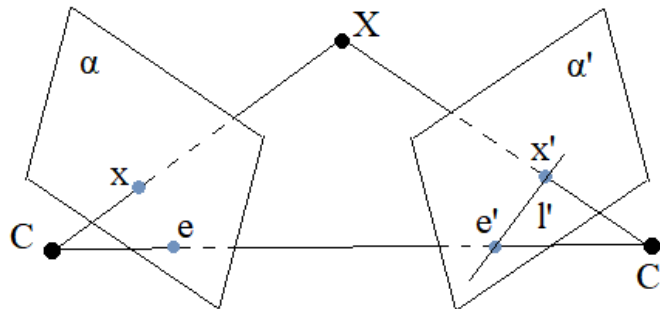


Рис. 1: e и e' – эиполи, l' – эиполярная прямая

Однако pinhole модель не подходит для описания изображений с радиальной дисторсией так как, например, предполагает, что прямые проецируются на изображение как прямые, а не «закругляются» (см. Рис. 2). Поэтому будет выбрана более общая модель камеры, описанная в [7], которая, в сущности, предполагает, что точки сначала проецируются на параметрически заданную поверхность и только потом на плоскость изображения. Данная модель может быть описана, используя N параметров, однако в данной работе мы ограничимся частным случаем – однопараметрической моделью, описанной в [4].



Рис. 2: Пример изображения с радиальной дисторсией, взятого из Интернета

Эффективное решение для оценки эпиполярных геометрий и параметров дисторсии камер предложен в [10]. В данной работе предложено эффективное решение для нахождения фундаментальных матриц для пары камер и оценки параметров дисторсии, с помощью 10 пар соответствий точек. Предложенный метод основан на преобразовании уравнения для фундаментальной матрицы, обсуждавшегося выше:

$$x_{u_i}^T(\lambda_1)Fx'_{u_i}(\lambda_2), i \in 1, \dots, 10 \quad (1)$$

где F – это фундаментальная матрица, λ_1 и λ_2 – это параметры дисторсии для первого и второго изображений соответственно, $x_{u_i}(\lambda_1)$ и $x'_{u_i}(\lambda_2)$ – это соответствующие координаты в P^2 изображений, преобразованные из данных координат с помощью модели, описанной в [4]. В данной статье равносильными алгебраическими преобразованиями уравнения (3) сводятся к нахождению корней многочлена 10 степени.

Ввиду того, что в наборах соответствий часто присутствуют неверно найденные соответствия, выбор фундаментальной матрицы для пары камер производится с помощью RANSAC (RANdom SAmple Consensus) метода. Данный метод является итеративным, на каждой итерации метода выбираются случайные десять пар из набора соответствий, для которых вычисляются оценки параметров дисторсии и фундаментальная матрица. Найденные матрица и параметры проверяются для остальных точек, и подсчитывается количество пар соответствий, удовлетво-

щих гипотезе о том, что данные найдены верно, т.е. удовлетворяющих уравнению (3) с некоторым порогом. В конце выбираются матрица и параметры, удовлетворяющие наибольшему количеству соответствий. Количество итераций в методе зависит от желаемой вероятности ошибки: чем больше итераций, тем, соответственно, меньше вероятность нахождения неудовлетворительной оценки эпиполярной геометрии.

Следует заметить, что для описанной оценки эпиполярной геометрии минимальное требуемое количество известных соответствий равняется девяти, а не десяти. Использование дополнительного соответствия уменьшает вероятность построения правильной модели, ввиду возможности наличия неверно найденных соответствий, однако значительно ускоряет оценочный алгоритм. Несмотря на то, что для достижения той же вероятности ошибки, которая была бы при использовании меньшего количества соответствий, требуется большее количество RANSAC-итераций, выигрыш по времени от ускорения алгоритма оценки параметров оказывается больше.

В работе [3] предлагается метод глобальной реконструкции, сводящий задачу к задаче усреднения вращений, предложенным в статье [2].

Рассмотрим постановку задачи усреднения вращений. Дано множество камер и для некоторых пар камер i и j известна R_{ij} – матрица поворота камер j относительно i . Для каждой камеры i требуется найти R_i – абсолютный поворот. Тогда ясно, что

$$R_{ij} = R_j R_i^{-1}, \quad (2)$$

для пары камер i и j .

В статье [2], используя отображения между группой Ли $SO(3)$ и соответствующей ей алгебре Ли, находится набор R_1, \dots, R_n абсолютных поворотов, минимизирующий сумму квадратов расстояний¹ между R_{ij} и $R_j R_i^{-1}$. Задача формулируется в терминах нахождения минимизирующего набора, а не обнуляющего, так как на практике набор относительных поворотов извлекается из множества соответствий точек на

¹Расстояние в $SO(3)$ между R_1 и R_2 выбрано как $d(R_1, R_2) = \frac{1}{\sqrt{2}} \|\log(R_1 R_2^{-1})\|_F$, где F – норма Фробениуса.

изображениях, в которых часто существуют неверно найденные пары, что приводит к найденным с ошибками относительным поворотам.

В работах [8] и [1] рассмотрены теоретическая часть вопроса метода глобального усреднения вращений. В частности показано, что если граф для камер является достаточно плотным, то решение будет глобально оптимальным, что делает данную группу методов привлекательной для использования.

2. Постановка задачи

Традиционная формулировка задачи глобальной реконструкции подразумевает знание внутренних параметров камеры для оценки попарных поз. Цель данной работы – обобщить задачу глобальной реконструкции до использования некалиброванных камер с опциональной оценкой дисторсии, с использованием усреднения подобий. Для достижения цели были сформулированы следующие задачи:

1. Реализовать метод closed-form оценки эпполярных геометрий и параметров дисторсии, описанный в [10] и объединить его с извлечением фокусных расстояний из фундаментальной матрицы [5]
2. Используя подход описанный в [3] свести задачу глобальной реконструкции к задаче усреднения подобий.
3. Реализовать усреднение вращений опираясь на [2] и теоретические выкладки представленные в [8] и [1].
4. Провести тестирование полученного решения.

3. Реализация

В данной работе была реализована алгоритм, описанный в [10]. В качестве языка для реализации был выбран MATLAB, так как он предоставляет широкий набор инструментов для решения часто возникающих математических задач.

Начальными данными являются множество соответствий точек на паре изображений с координатами (x_{d_i}, y_{d_i}) и (x'_{d_i}, y'_{d_i}) . Согласно модели описанной в [4], преобразование координат $X_{u_i} = (x_{d_i}, y_{d_i}, 1 + \lambda(x_{d_i}^2 + y_{d_i}^2))$ позволяет получить координаты изображения без дисторсии, таким образом переходя к модели pin-hole камеры. В данной задаче λ – неизвестный параметр дисторсии изображения.

Проведя данное преобразования можно говорить о фундаментальной матрице $F = \{F_{ij}\}_{i,j=1}^4$ для получившихся изображений. Данная матрица является неизвестной и находится из 10 уравнений:

$$X_{u_i}^T(\lambda_1) F X'_{u_i}(\lambda_2), i \in 1, \dots, 10, \quad (3)$$

где $X_{u_i} = (x_{d_i}, y_{d_i}, 1 + \lambda_1(x_{d_i}^2 + y_{d_i}^2))$, $X'_{u_i} = (x'_{d_i}, y'_{d_i}, 1 + \lambda_2(x_{d_i}^2 + y_{d_i}^2))$ – координаты с неизвестными параметрами дисторсии λ_1 и λ_2 первого и второго изображений соответственно.

Для решения данной системы уравнений, для начала, предлагается выписать его в матричной форме

$$MX = 0, \quad (4)$$

где $X = (F_{11}, F_{12}, F_{21}, F_{22}, \lambda_1 F_{31}, F_{31}, \lambda_1 F_{32}, F_{32}, \lambda_2 F_{13}, F_{13}, \lambda_2 F_{23}, F_{23}, F_{33}, \lambda_1 F_{33}, \lambda_2 F_{33}, \lambda_1 \lambda_2 F_{33})$

– вектор со всеми неизвестными параметрами.

Во избежание увеличения времени работы программы, данное преобразование выполнено вручную и результат записан в код функции.

Далее в статье предлагается привести матрицу M к диагональному виду с помощью метода Гаусса, однако, так как данный метод не является эффективным и вычислительно устойчивым, была использована стандартная функция MATLAB `mldivide()`, которая находит решения

систем линейных уравнений.

Затем в статье описываются алгебраические преобразования, поочередно исключая неизвестные переменные, приводящие систему к решению уравнения $\det(R(\lambda_1)) = 0$, где $R(\lambda_1)$ – матрица 4x4, элементами которой являются многочлены от λ_1 , степени не выше третьей. Так как выполнять данные преобразования во время исполнения программы, с помощью символьных переменных MATLAB, очень долго, а для выполнения вручную данные вычисления слишком громоздки, была написана программа на MATLAB производящая данные вычисления символьно и выводящая их в файл. Затем данный файл был преобразован в программу для функции на MATLAB. Таким образом, матрица R находится напрямую, с помощью функции, высчитывающей её элементы, которыми являются вектора коэффициентов многочленов.

Далее исполняется программа, высчитывающая определитель матрицы R , также написанная вручную, с оптимизациями, учитывающими нулевые вхождения и подобные элементы. Результатом выполнения программы являются коэффициенты многочлена λ_1 10 степени. Решение данного уравнения находится с помощью стандартной функции MATLAB, которая находит решения как собственные числа сопровождающей матрицы.

После нахождения решений λ_1 уравнения из них откидываются комплексные, так как параметр дисторсии должен быть вещественным. Затем с помощью найденного λ_1 находится λ_2 и F . Таким образом, результатом работы программы является набор из не более чем 10 троек $(\lambda_1, \lambda_2, F)$.

Описанная программа используется в методе RANSAC. На каждой своей итерации метод выбирает десять случайных точек из данного набора и, с помощью описанного выше алгоритма, находит $(\lambda_1, \lambda_2, F)$. Данный метод проверяет, удовлетворяет ли остальные точки полученным моделям, построенным по выбранным 10 и затем, выбирает модель, которая удовлетворяет наибольшему количеству точек.

Корректность работы данной программы проверялась на сгенерированных данных. Генерация пар соответствий точек производилась с

помощью генерации точек в трёхмерном пространстве, которые затем проецировались на плоскость камеры, с использованием выбранной модели дисторсии со случайным параметром λ . Соответственно для пары камер также генерировались поворот и смещение относительно друг друга, а также фокусные расстояния.

Заключение

В данной работе предоставлен обзор задачи глобальной реконструкции для некалиброванных камер с неизвестным параметром дисторсии и предложен план по её решению. Для данной задачи на языке MATLAB реализованы алгебраические преобразования описанные в статье [10] для closed-form оценок эпполярных геометрий и параметров дисторсии.

Список литературы

- [1] Anders Eriksson Carl Olsson Fredrik Kahl, Chin Tat-Jun. Rotation Averaging and Strong Duality. — 2018.
- [2] Chatterjee Avishek, Govindu Venu Madhav. Efficient and robust large-scale rotation averaging. — 2013.
- [3] Cui Zhaopeng, Tan Ping. Global Structure-from-Motion by Similarity Averaging. — 2015.
- [4] Fitzgibbon Andrew W. Simultaneous Linear Estimation of Multiple View Geometry and Lens Distortion. — 2001.
- [5] Hartley Richard. Extraction of focal lengths from the fundamental matrix. — 1993.
- [6] Hartley Richard, Zisserman Andrew. Multiple View Geometry in Computer Vision. — Cambridge University Press, 2004.
- [7] Scaramuzza Davide Agostino Martinelli, Siegwart Roland. A flexible technique for accurate omnidirectional camera calibration and structure from motion. — 2006.
- [8] Wilson Kyle David Bindel, Snavely Noah. When is Rotations Averaging Hard? — 2016.
- [9] Wu Changchang. Towards linear-time incremental structure from motion. — 2013.
- [10] Zuzana Kukelova Jan Heller Martin Bujnak Andrew Fitzgibbon Tomas Pajdla. Efficient Solution to the Epipolar Geometry for Radially Distorted Cameras. — 2015.