

# Исследование форматов тестовых наборов данных для решения задачи одометрии

Е.В. Гошин<sup>1,2</sup>, П.Н. Вольхин<sup>1</sup>, А.П. Котов<sup>1,2</sup>

<sup>1</sup>Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

<sup>2</sup>Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

**Аннотация.** В настоящей статье проведён обзор различных наборов тестовых данных для проведения исследований по оценке качества алгоритмов определения движения и положения камеры по последовательности оптических изображений. На основании проведённого обзора сформулирована задача представления данных в заданном формате. Для решения данной задачи разработано программное средство преобразования форматов представления параметров исходных тестовых наборов на языке программирования Python.

## 1. Введение

Определение положения камеры в пространстве применяют в таких областях как навигация беспилотных аппаратов [1], локализация мобильных устройств [2], перемещение человека [3] и реконструкция трехмерных сцен [4][5].

В методах компьютерного зрения для получения количественных оценок эффективности используемых методов применяются заранее подготовленные тестовые наборы данных. Для сравнения методов в рамках одной задачи используются различные наборы данных, так как надежность методов может меняться от условий. В рамках задачи определения движения и ориентации камеры для получения оценок необходима информация о сдвиги и углах поворота. Как правило, параметры поворота предоставляются в виде матрицы поворота или кватерниона поворота, а также как композиция поворотов вокруг трех ортогональных осей. В ряде случаев параметры поворота не выделены отдельно, а объединены с параметрами движения и внутренними параметрами камеры в общей матрице проекции.

Настоящая работа посвящена сравнительному исследованию различных форматов представления параметров наборов тестовых данных и разработке программного средства автоматической подготовки этих параметров для проведения экспериментальных исследований, связанных с решением задачи одометрии.

## 2. Постановка задачи

Источником исходных данных для решения задачи, как правило, является монокулярная камера и неизменная сцена. Последовательность оптических изображений формируется посредством съемки объекта с различных ракурсов. Для описания процесса получения изображений используется модель камеры-обскуры.

В рамках модели камеры-обскуры вид сцены формируется проецированием точек пространства на плоскость и имеет вид [6]:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K[R|t] \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix},$$

где  $(X, Y, Z)$  – координаты точки в трехмерной системе координат,  $(u, v)$  – координаты проекции на изображении,  $K$  – верхне-треугольная матрица внутренних параметров камеры,  $R$  – матрица поворота камеры,  $t$  – вектор смещения камеры.

Преобразование кватерниона  $q = (w, x, y, z)$  к матрице поворота:

$$R = \begin{pmatrix} 1 - 2y^2 - 2z^2 & 2xy - 2zw & 2xz + 2yw \\ 2xy + 2zw & 1 - x^2 - 2z^2 & 2yz - 2xw \\ 2xz - 2yw & 2yz + 2xw & 1 - 2x^2 - 2y^2 \end{pmatrix}.$$

Преобразование вектора поворота к матрице поворота представляется формулой Родрига:

$$R = I \cos \theta + (1 - \cos \theta)rr^T + \sin \theta \begin{pmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{pmatrix}.$$

Преобразование композиции поворотов вокруг ортогональных осей в виде последовательного умножения матриц поворота вокруг осей  $Z, X, Y$  на углы  $\alpha, \beta, \gamma$  соответственно:

$$R = \begin{pmatrix} \cos \alpha \cos \gamma - \sin \alpha \cos \beta \sin \gamma & -\cos \alpha \sin \gamma - \sin \alpha \cos \beta \cos \gamma & \sin \alpha \sin \beta \\ \sin \alpha \cos \gamma + \cos \alpha \cos \beta \sin \gamma & -\sin \alpha \sin \gamma + \cos \alpha \cos \beta \cos \gamma & -\cos \alpha \sin \beta \\ \sin \beta \sin \gamma & \sin \beta \cos \gamma & \cos \beta \end{pmatrix}.$$

В настоящей работе авторы ставят задачей разработку программного решения для представления данных в едином заданном формате. Данная задача является актуально при апробации методов и её решение способствует повышению эффективности проведения научных исследований.

### 3. Обзор существующих форматов

Для апробации исследуемого алгоритма были подготовлены тестовые данные. Критериями отбора в тестовый набор являются наличие одного из следующих свойств: сферическая съемка, последовательная съемка, съемка со структурированным светом, съемка в инфракрасном диапазоне. Все оптические изображения имеют дополнительную информацию в виде наличия внутренних параметров, параметров ориентации и смещения камеры. Набор для апробации алгоритма был собран из наборов: Middlebury, ETH3D, DTU Robot Image, EPFL Multi-View Stereo, RGB-D SLAM, MVSIR. Соответствие наборов критериям приведено в таблице 1.

**Таблица 1.** Соответствие наборов критериям.

Вид съемки	Middlebury	ETH3D	DTU Robot Image	EPFL Multi-View Stereo	RGB-D SLAM	MVSIR
Сферическая	Да	Да	Да	Нет	Нет	Да
Последовательная	Нет	Да	Нет	Да	Да	Нет
Со структурированным светом	Нет	Нет	Да	Нет	Нет	Нет
В инфракрасном диапазоне	Нет	Нет	Нет	Нет	Нет	Да

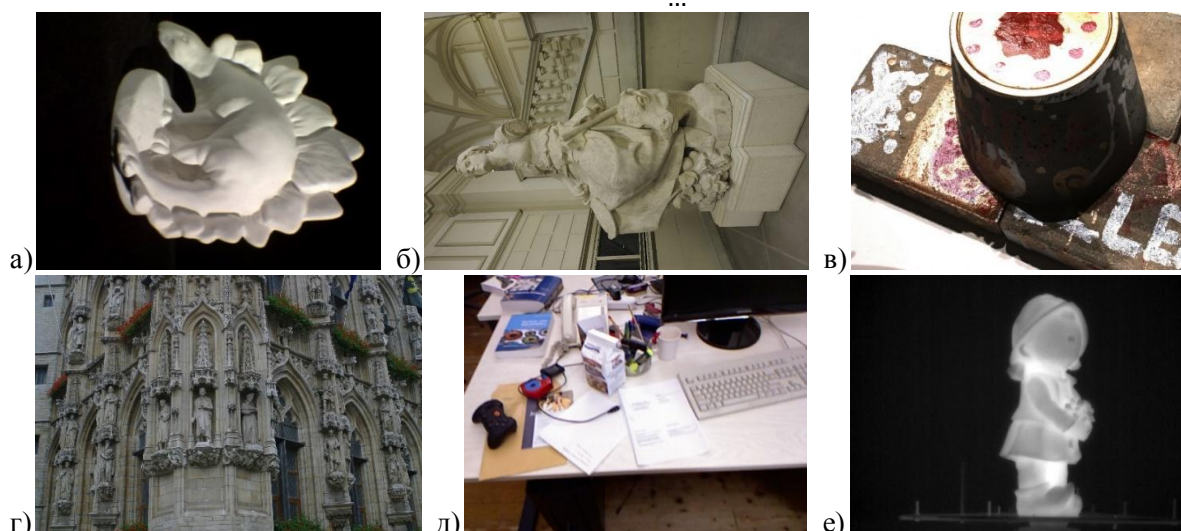
На рисунке 1 представлены примеры изображений для выбранных наборов данных.

#### 3.1 Набор данных Middlebury

Набор состоит из изображений объекта при различных положениях камеры на сфере, в центре которой находится этот объект. Они были получены в Стэнфордском университете, с помощью

камеры, которая может перемещаться по сфере на заданные углы широты и долготы [7]. Ниже представлен псевдокод структуры набора:

```
dataset_middlebury /  
  *_ang.txt  
  *_par.txt  
  *_XXX.png  
  ...
```



**Рисунок 1.** Примеры изображений из наборов: а) Middlebury; б) ETH3D; в) DTU Robot Image; г) EPFL Multi-View Stereo; д) RGB-D SLAM; е) MVSIR.

В файле «\*\_par.txt» хранится информация о внутренних и внешних параметрах камеры для всех изображений набора, а общее количество изображений записано в первой строке. Каждая строка данного файла, исключая первую (с общим количеством изображений), имеет следующий формат: название изображения, 9 элементов матрицы камеры, 9 элементов матрицы поворота камеры, 3 элемента вектора смещения камеры.

### 3.2 Набор данных ETH3D

Данный набор охватывает различные сцены в помещении и на улице. Изображения получены с помощью зеркальной камеры, а также синхронизированной многокамерной установки с переменным зрением [8, 9]. Псевдокод структуры наборов продемонстрирован ниже:

```
dataset_ETH3D /  
  dslr_calibration * /  
  camera.txt  
  images.txt  
  point3D.txt  
  dslt_images * /  
  * XXX.jpg  
  ...
```

В файле «cameras.txt» приводится информация о количестве камер, используемых для получения набора, а также их внутренние параметры. Каждая строка имеет следующий формат: индекс камеры, название камеры, ширина, высота, фокусное расстояние по ширине, фокусное расстояние по высоте, координата принципиальной точки по X, координата принципиальной точки по Y. Файл «images.txt» содержит информацию о количестве изображений в наборе, а также сопоставления внешних параметров камеры с внутренними параметрами и изображениями. Каждая строка файла имеет формат, где все данные записаны через пробел: идентификатор изображения, 4 компоненты кватерниона, 3 элемента вектора смещения, дополнительные параметры.

### 3.3 Набор данных DTU Robot Image

Набор был получен при помощи автоматизированной руки робота со структурированным световым сканером. Это позволило сканировать структурированное освещение, соответствующее каждому изображению из набора данных. Изображения были сделаны одной из камер в сканере структурированного света [10, 11]. Ниже представлен псевдокод, описывающий структуру набора.

```
dataset_MVS /
  calibration /
    cal * /
  Calib_Result_left.mat
  Calib_Result_right.mat
  Calib_Result_stereo.mat
  cleaned /
    scan * /
    * XXX.png
    ...
```

В «\*.mat» файлах описаны внешние и внутренние параметры камеры для каждого из изображений. Поворот камеры в данном наборе описан с помощью вектора поворота.

### 3.4 Набор данных EPFL Multi-View Stereo

Набор состоит из изображений внешнего фасада зданий с внешними и внутренними параметрами камеры. Ниже приведена структура набора в виде псевдокода:

```
dataset_EPFL /
  * XXX.ppm
  * XXX.ppm.camera
  * XXX.ppm.3Dpoints
  ...
```

Каждому изображению в формате ppm соответствует файл «\*.camera», в котором описаны параметры камеры в следующем формате: с 1 по 3 строку –  $3 \times 3$  матрица камеры, в 4 строке – вектор радиального искажения, с 5 по 7 строку –  $3 \times 3$  матрица поворота камеры, в 8 строке – вектор смещения камеры.

### 3.5 Набор данных RGB-D SLAM Dataset

В процессе создания набора использовался Kinect, параметры камеры которого приводились отдельно и представлены матрицей камеры:

$$K = \begin{pmatrix} 525 & 0 & 319,5 \\ 0 & 525 & 239,5 \\ 0 & 0 & 1 \end{pmatrix}.$$

Псевдокод структуры набора показан ниже:

```
dataset_RGB_D_SLAM /
  depth /
    * XXX.png
    ...
  rgb /
    * XXX.png
    ...
  groundtruth.txt
```

В текстовом файле «groundtruth.txt» указаны параметры, описывающие ориентацию камеры в пространстве для каждого изображения набора. Все параметры конкретного изображения записаны в строку через пробелы и представлены в виде следующей последовательности: временная метка, 3 элемента вектора смещения, 4 компоненты кватерниона поворота.

### 3.6 Набор данных MVSIR Dataset

Данный набор содержит в себе инфракрасные изображения трехмерных объектов. Ниже приведен псевдокод структуры набора.

```
dataset_MVSIR /  
  * XXX.txt  
  * XXX.bmp  
  ...
```

В текстовых файлах описаны следующие параметры: фокусное расстояние, номинальная фокусная длина, углы поворота Эйлера вокруг осей  $X$ ,  $Y$ ,  $Z$  (строго в данном порядке), масштабирование пикселей по  $X$  и  $Y$ , положение главной точки в пространстве изображения. В данном файле каждый параметр записан в две строки: название и значение.

## 4. Разработка программного решения

В качестве единого формата, описывающего поворот и смещение камеры, был выбран формат, состоящий из матрицы поворота и вектора сдвига. Выбор данного формата обусловлен тем, что он может быть получен из любого другого без потери информации. Обратное преобразование (из матрицы поворота к кватерниону) является неоднозначным, поэтому представление в виде кватернионов не может являться таким форматом. Преобразование из матрицы поворота к композиции произведения поворотов вокруг трех ортогональных осей имеет проблему определения порядка произведения матриц, а использование вектора поворота требует сведений об оси поворота поэтому такие варианты также были исключены.

Работа программы состоит из трех этапов:

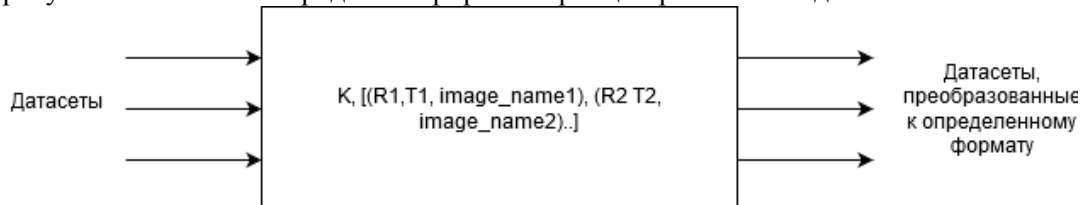
- 1) определение текущего формата данных;
- 2) преобразование текущего формата во внутреннее представление;
- 3) конвертация внутреннего представления в выбранный формат.

На основе вышеописанных наборов данных были сформированы шаблоны для определения текущего формата. Во время анализа текущего формата происходит сравнение с шаблонами, и если он совпадает с одним из них, то начинается процесс преобразования данных во внутреннее представление в соответствии с его спецификой.

Внутреннее представление данных описывается матрицей параметров камеры и списком моделей, которые представляют собой композицию информации по конкретному изображению: название изображения, матрица поворота, вектор сдвига.

Данный формат внутреннего представления был выбран по следующим причинам. Первой причиной являлась необходимость выбора промежуточной формы для различных форматов. Вторая причина заключалась в том, чтобы это представление могло выступать в качестве самостоятельного формата. Данный формат планируется использовать для проведения исследований по разработке алгоритма определения движения и ориентации камеры.

На рисунке 2 схематично продемонстрирован принцип работы метода.



**Рисунок 2.** Принцип работы скрипта по преобразованию форматов наборов данных:  $K$  –  $3 \times 3$  матрица внутренних параметров камеры,  $R_i$  –  $3 \times 3$  матрица поворота камеры для  $i$  изображения,  $T_i$  –  $1 \times 3$  вектор смещения камеры для  $i$  изображения,  $image\_name$  – имя файла изображения.

## 5. Заключение

В рамках данной работы проведено исследование наборов данных для разработки алгоритма определения движения и ориентации в пространстве камеры по последовательности оптических изображений. Предложена информационная технология обработки различных

форматов данных с целью удобной апробации методов определения параметров движения. В результате исследования разработана программа на языке программирования Python. Данное программное средство может преобразовывать форматы наборов один в другой. Поддерживает следующий список форматов наборов: Middlebury; ETH3D; DTU Robot Image; EPFL Multi-View Stereo; RGB-D SLAM; MVSIR.

## 6. Благодарности

Работа выполнена в рамках государственного задания по теме FSSS-2020-0017 при частичной поддержке РФФИ (проект № 17-29-03112). Авторы выражают благодарность доктору технических наук, профессору Фурсову В.А. за ряд ценных замечаний при подготовке настоящей работы.

## 7. Литература

- [1] Fu, C. Efficient visual odometry and mapping for unmanned aerial vehicle using ARM-based stereo vision pre-processing system / C. Fu, A. Carrio, P. Campoy // International Conference on Unmanned Aircraft Systems (ICUAS), 2015. – P. 957962
- [2] Мясников, В.В. Исследование зависимости точности одновременной реконструкции сцены и позиционирования камеры от погрешностей, вносимых датчиками мобильного устройства / В.В. Мясников, Е.А. Дмитриев // Компьютерная оптика. – 2019. – Т. 43, № 3. – С. 492-503. DOI: 10.18287/2412-6179-2019-43-3-492-503.
- [3] Chen, X. An inertial-based human motion tracking system with twists and exponential maps / X. Chen, J. Zhang, W.R. Hamel, J. Tan // IEEE international conference on robotics and automation (ICRA), 2014. – P. 5665-5670.
- [4] Fursov, V.A. The hybrid CPU/GPU implementation of the computational procedure for digital terrain models generation from satellite images / V.A. Fursov, Ye.V. Goshin, A.P. Kotov // Computer Optics. – 2016. – Vol. 40(5). – P. 721-728. DOI: 10.18287/2412-6179-2016-40-5-721-728.
- [5] Кириш, Д.В. Алгоритм реконструкции трёхмерной структуры кристалла по двумерным проекциям / Д.В. Кириш, А.С. Широкаев, А.В. Куприянов // Компьютерная оптика. – 2019. – Т. 43, № 2. – С. 324-331. DOI: 10.18287/2412-6179-2019-43-2-324-331.
- [6] Библиотека обработки изображений OpenCV [Электронный ресурс]. – Режим доступа: <http://opencv.org> (20.11.2019).
- [7] The Middlebury Computer Vision Pages [Electronic resource]. – Access mode: <http://vision.middlebury.edu/mview/data> (20.11.2019).
- [8] ETH3D [Electronic resource]. – Access mode: <https://www.eth3d.net/datasets> (20.11.2019).
- [9] Aanæs, H. Large-Scale Data for Multiple-View Stereopsis / H. Aanæs, R.R. Jensen, G. Vogiatzis, E. Tola, A.B. Dahl // International Journal of Computer Vision. – 2016. – Vol. 120(2). – P. 153-168.
- [10] Schops, T. A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos / T. Schops, J.L. Schönberger, S. Galliani, T. Sattler, K. Schindler, M. Pollefeys, A. Geiger // IEEE Conference on Computer Vision and Pattern Recognition, 2017. – P. 209-217.
- [11] DTU Robot Image Data Sets [Electronic resource]. – Access mode: <http://roboimagedata.compute.dtu.dk> (20.11.2019).

# Research of formats of test data sets for solving the odometry problem

E.V. Goshin<sup>1,2</sup>, P.N. Volkhin<sup>1</sup>, A.P. Kotov<sup>1,2</sup>

<sup>1</sup>Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

<sup>2</sup>Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

**Abstract.** This article provides an overview of various sets of test data for conducting researches to assess the quality of algorithms for determining the movement and position of a camera from a sequence of optical images. Based on the review, the task of presenting data in a given format is formulated. To solve this problem, a software tool has been developed for converting the formats for representing the parameters of the initial test suites in the Python.