Метод оценки движения камеры на основе компланарности, инвариантный к глубине сцены

Е.В. Гошин
Самарский национальный исследовательский университет им. академика С.П. Королева
Самара, Россия
goshine@ssau.ru

***Abstract*—В настоящей работе предлагается метод оценки параметров движения камеры по изображениям, полученным с этой камеры. Предложенный подход может применяться к трёхмерным сценам независимо от из глубины и позволяет избежать ограничений, накладываемых на набор исходных данных при использовании методов на основе существенной матрицы и матрицы проективного преобразования.**

Ключевые слова— оценка движения камеры, эпиполярная геометрия, проективная геометрия, компланарность.

1. Введение

В настоящей работе рассматривается задача монокулярной визуальной одометрии [1]. Исследуемая задача состоит в оценке параметров движения камеры [2], характеризующих ориентацию и положение этой камеры в пространстве в определенные моменты времени в процессе движения, по изображениям, регистрируемым этой камерой.

Исходными данными (при отсутствии информации о перемещении устройства) являются изображения, полученные с разных ракурсов. Традиционно задача оценки параметров поворота и сдвига сводится к определению так называемой существенной матрицы и последующему вычислению искомых параметров [3]. Однако этот подход имеет ряд недостатков. При нахождении исходных точек в одной плоскости вычисление существенной матрицы может осуществляться с ошибками, что, в свою очередь, приводит к ошибкам в определении параметров движения.

Настоящая работа исследует решение принципиальной проблемы поиска универсального способа оценки параметров движения, инвариантного к глубине сцены.

2. Математическая основа алгоритма

В основе предлагаемого подхода лежит тот факт, что даже при отсутствии информации о расстоянии до точки в трехмерном пространстве можно установить однозначную связь между положением точки после поворота, положением соответствующей точка на втором изображении и вектором сдвига.

Рассмотрим следующую постановку задачи. Пусть задано множество точек на предыдущем и текущем кадрах $\{\mathrm{p_i}\} = \{(\mathrm{x_i}, \mathrm{y_i}, 1)\}$ и $\{\mathbf{p}'_i\} = \{(x'_i, y'_i, 1)\}$, полученных в результате проецирования точек $\{\mathbf{M}_i\} = \{(X_i, Y_i, Z_i)\}$ и преобразованных точек $\{\mathbf{M}'_i\} = \{(X'_i, Y'_i, Z'_i)\} = \{\mathbf{R}(X_i, Y_i, Z_i) + \mathbf{t}\}$, соответственно. Необходимо вычислить оценку параметров движения $\mathbf{R}$ и $\mathbf{t}$.

Введем обозначения

$$x_i^R = R_{11}x_i + R_{12}y_i + R_{13},$$
$$y_i^R = R_{21}x_i + R_{22}y_i + R_{23},$$
$$z_i^R = R_{31}x_i + R_{32}y_i + R_{33}.$$

В предыдущей работе [4] было показано, что прямая на плоскости, проходящая через точки $\left(\frac{x_i^R}{z_i^R}, \frac{y_i^R}{z_i^R}\right)$ и (x'_i, y'_i),также проходит через точку$\left(\frac{t_x}{t_z}, \frac{t_y}{t_z}\right)$. Это требование можно записать в форме смешанного произведения:

$$\left(\left(\frac{x_i^R}{z_i^R}, \frac{y_i^R}{z_i^R}, 1\right) \times (x'_i, y'_i, 1)\right) \cdot \left(\frac{t_x}{t_z}, \frac{t_y}{t_z}, 1\right) = 0.$$

Так как умножение на константу не меняет уравнения, то и следующее выражение будет верным:

$$\left((x_i^R, y_i^R, z_i^R) \times (x'_i, y'_i, 1)\right) \cdot (t_x, t_y, t_z) = 0. \qquad (1)$$

Отметим, что вектор $\mathbf{t}$ в выражении (1) не зависит ни от исходных точек, ни от параметров поворота, поэтому это выражение может быть декомпозировано.

Определяем векторы $\mathbf{v}_i$ как:

$$\mathbf{v}_i = (x_i^R, y_i^R, z_i^R) \times (x'_i, y'_i, 1)$$

и матрицу, состоящую из этих векторов-строк как:

$$\boldsymbol{V}(R) = \begin{pmatrix} - \\ \mathbf{v}_i \\ - \end{pmatrix}.$$

Полученная матрица $\boldsymbol{V}(R)$ имеет размерность $n \times 3$, где n — количество пар исходных соответствующих точек.

Очевидно, что произведение каждой строки матрицы $\boldsymbol{V}(R)$ на вектор $\mathbf{t}$ будет равно нулю и, следовательно, $\mathbf{t}$ будет нуль-вектором полученной матрицы

3. Алгоритм

В общем случае вектор $\mathbf{t}$ предполагается ненулевым. Поскольку матрица имеет ненулевой нуль-вектор, ее ранг должен быть равен 2 или меньше. Другими словами, векторы матрицы $\mathbf{v}_i$ должны быть компланарными.

Очевидно, что в реальных условиях при наличии шума в исходных данных компланарность невозможна. Поэтому в качестве критерия соответствия параметра вращения исходным данным вместо этого оценим мультиколлинеарность строк матрицы $\boldsymbol{V}(R)$.

В качестве критерия может быть использовал любой из известных критериев мультиколлинеарности: минимальное собственное значение матрицы $\boldsymbol{V}(R)$, число обусловленности матрицы $\boldsymbol{V}(R)$ или определитель матрицы Грама $\boldsymbol{V}^T(R)\boldsymbol{V}(R)$.

После определения параметров поворота путем оптимизации выбранного критерия $J(\mathbf{R})$

$$\hat{R} = \underset{R}{\operatorname{argmax}} J(\mathbf{R})$$

значение вектора сдвига $\mathbf{t}$ с точностью до параметра масштаба можно определить путем решения переопределенной однородной системы нелинейных уравнений, исключив тривиальное решение:

$$\boldsymbol{V}(\widehat{\mathbf{R}}) \cdot \hat{\mathbf{t}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Можно показать, что для точек, не лежащих на плоскости, единственным значением, удовлетворяющим критерию, является $\widehat{\mathbf{R}} = \mathbf{R}$ и $\hat{\mathbf{t}} = \lambda\mathbf{t}$.

4. Результаты эксперимента

Предлагаемый подход был исследован на синтетических данных. Эксперимент заключался в оценке параметров камеры с использованием декомпозиции существенной матрицы и предложенного подхода. Мы использовали угол между точным вектором сдвига $\mathbf{t}_{true}$ и оценкой $\hat{\mathbf{t}}$ как меру ошибки оценки сдвига

$$\epsilon_{\mathbf{t}} = \frac{\mathbf{t}_{true} \cdot \hat{\mathbf{t}}}{\|\mathbf{t}_{true}\| \cdot \|\hat{\mathbf{t}}\|},$$

а также угол между истинным значением поворота $\mathbf{R}_{true}$ и оценкой $\widehat{\mathbf{R}}$ как меру ошибки

$$\epsilon_{\mathbf{R}} = \arccos\left(\frac{tr\left(\widehat{\mathbf{R}}^{-1}\mathbf{R}_{true}\right) - 1}{2}\right).$$

Результаты показаны на рисунке 1

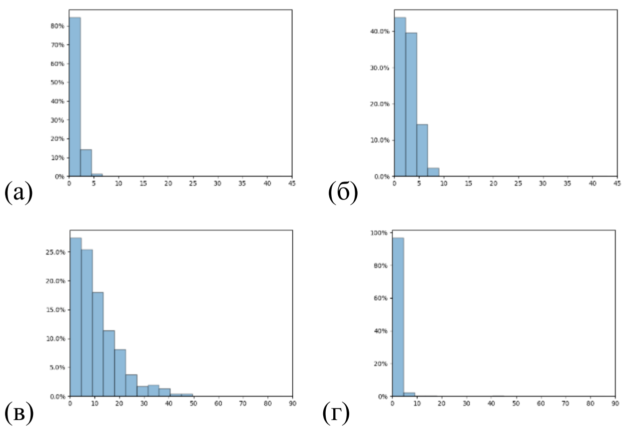

Рис. 1. Распределение ошибок полученных оценок на синтетических данных: (а) поворот, декомпозиция существенной матрицы, (б) поворот, предложенный подход, (в) сдвиг, декомпозиция существенной матрицы, (г) сдвиг, предложенный подход

Легко заметить, что, хотя распределение ошибки вращения несколько хуже для предложенного подхода, оно остается в допустимом диапазоне значений. С другой стороны, при использовании предлагаемого подхода значительно снижается ошибка оценки сдвига. В целом, предлагаемый метод обеспечивает большую надежность.

Также предложенный метод был протестирован на наборе данных MIT Blackbird [5]. Эксперименты проводились на парах изображений из набора данных. Для получения соответствующей пары точек была использована стандартная процедура, включающая обнаружение и сопоставление точек. На рисунке 2 показано распределение ошибок перевода для оценки с использованием декомпозиции существенной матрицы и предлагаемого метода.

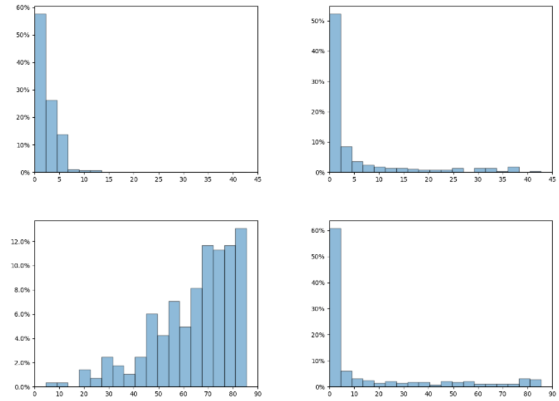

Рис. 2. Распределение ошибок полученных оценок на данных MIT Blackbird: (а) поворот, декомпозиция существенной матрицы, (б) поворот, предложенный подход, (в) сдвиг, декомпозиция существенной матрицы, (г) сдвиг, предложенный подход

Показано, что традиционный подход к оцениванию методом декомпозиции существенной матрицы демонстрирует около 95% неправильных оценок поворота (более 30 градусов). Предлагаемый подход дает приемлемые результаты более чем в 75% случаев.

5. Заключение

Предложенный подход к оценке движения камеры показывает гораздо более стабильные результаты для сложных сцен, когда количество соответствующих точек невелико, и большинство этих точек лежат в одной плоскости или в одной и той же части изображения.

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках государственного задания по теме 0777-2020-0017 и программы «Приоритет-2030».

Литература

[1] He, M. A review of monocular visual odometry / M. He, C. Zhu, Q. Huang, B. Ren, J. Liu // The Visual Computer. – 2019. – Vol. 36(5). – P. 1053-1065. DOI: 10.1007/s00371-019-01714-6.

[2] Scaramuzza, D. Visual Odometry / D. Scaramuzza, F. Fraundorfer // IEEE Robotics' Automation Magazine. – 2011. – Vol. 18(4). - P. 80-92. DOI: 10.1109/mra.2011.943233.

[3] Longuet-Higgins, H. A computer algorithm for reconstructing a scene from two projections / H. Longuet-Higgins // Nature. – 1981. – Vol. 293(5828). – P. 133-135. DOI: 10.1038/293133a0.

[4] Goshin, Y. Estimating Intrinsic Camera Parameters Using the Sum of Cosine Distances / Y. Goshin // Journal of Physics: Conference Series. – 2018. – Vol. 1096. – P. 012092. DOI: 10.1088/1742-6596/1096/1/012092.

[5] Antonini, A. The Blackbird UAV dataset / A. Antonini, W. Guerra, V. Murali, T. Sayre-McCord, S. Karaman // The International Journal of Robotics Research. – 2020. – Vol. 39(10-11). – P. 1346-1364. DOI: 10.1177/0278364920908331.