

П.А. Пантелюк

**ДЕТЕКТИРОВАНИЕ ВЫБРОСОВ В МЕТОДЕ ПРЯМОЙ
СТЕРЕО-ВИЗУАЛЬНОЙ ОДОМЕТРИИ НА БАЗЕ ИЕРАРХИЧЕСКОЙ
КЛАСТЕРИЗАЦИИ**

Представляется подход к стерео-визуальной одометрии без явного вычисления оптического потока. Визуальная одометрия – метод получения навигационной информации путем обработки последовательности кадров с бортовых камер. Существует два подхода к обработке видеoinформации – используя хорошо локализуемые участки изображения – признаковые точки и используя все высококонтрастные пиксели – прямой метод. Прямой метод работает, используя интенсивности всех высококонтрастных пикселей изображения, что позволяет снизить вычислительную сложность, затраченную на поиск, описание, сопоставление признаковых точек и повысить точность оценки движения. Однако методы подобного класса обладают недостатком – наличие движущихся объектов в кадре существенно снижают точность оценки параметров движения. Для избегания этого применяются методы детектирования выбросов. Классические методы детектирования выбросов во входных данных, такие как RANSAC плохо применимы, и имеют высокие вычислительные затраты из-за вычислительно сложной функции рейтингования гипотез. Целью данной работы является описание и демонстрация подхода детектирования выбросов на базе алгоритма иерархической кластеризации, который выделяет статистически наиболее вероятное решение, минуя этап рейтингования каждой гипотезы, что значительно снижает вычислительную сложность. Для иерархической кластеризации предлагается мера расстояния между гипотезами с низкой чувствительностью к ошибкам оценки параметров движения. Также предлагается расширение алгоритма стерео-визуальной одометрии для работы в более сложных условиях видимости благодаря переходу от интенсивностного представления изображения к многоканальному бинарному. Перевод изображения к многоканальному бинарному представлению дает инвариантность к изменениям яркости изображения, однако, требует модификации алгоритмов нелинейной оптимизации для работы с бинарными дескрипторами. В результате работы показано, что предложенный алгоритм детектирования выбросов способен работать в реальном масштабе времени на мобильных устройствах, и может служить менее ресурсоёмкой заменой алгоритма RANSAC в задачах визуальной одометрии и вычисления оптического потока. Качественные метрики предложенного решения демонстрируются на датасете KITTI. Приведены зависимости качества работы алгоритма от параметров алгоритма.

Навигация; визуальная одометрия; оптический поток; детектирование выбросов.

P.A. Panteliuk

**OUTLIERS DETECTION IN THE STEREO-VISUAL ODOMETRY BASED
ON HIERARCHICAL CLUSTERIZATION**

This paper presents an approach to stereo-visual odometry without explicitly calculating optical flow. Visual odometry is a method of obtaining navigation information by processing a sequence of frames from onboard cameras. There are two approaches to processing video information - using well-localized areas of the image - feature points and using all high-contrast pixels - a direct method. The direct method works by using the intensities of all high-contrast pixels in the image, making it possible to reduce the computational complexity spent searching, describing, and matching feature points and increasing the accuracy of motion estimation. However, methods of this class have a drawback - the presence of moving objects in the frame significantly reduces the accuracy of the estimation of motion parameters. Outlier detection techniques are used to avoid this problem. Classical methods for detecting outliers in input data, such as RANSAC, are poorly applicable and have high computational costs due to the computationally complex function of rating hypotheses. This work aims to describe and demonstrate an outlier detection approach

based on the hierarchical clustering algorithm, which selects the statistically most probable solution, bypassing the stage of rating each hypothesis, which significantly reduces the computational complexity. For hierarchical clustering, a measure of the distance between hypotheses with low sensitivity to errors in estimating motion parameters is proposed. An extension of the stereo-visual odometry algorithm is also proposed to work in more complex visibility conditions due to the transition from the intensity representation of the image to a multichannel binary one. Transforming an image to a multichannel binary representation gives invariance to changes in image brightness. However, it requires modification of nonlinear optimization algorithms to work with binary descriptors. As a result of the work, it has been shown that the proposed outlier detection algorithm can operate in real-time on mobile devices and can serve as a less resource-intensive replacement for the RANSAC algorithm in problems of visual odometry and optical flow eviction. Qualitative metrics of the proposed solution are demonstrated on the KITTI dataset. The dependences of the performance of the algorithm on the parameters of the algorithm are given.

Navigation; visual odometry; optical flow; outliers detection.

Введение. Оценка собственного движения объекта управления на основе последовательности кадров с бортовых камер называется визуальная одометрия [1–4]. Это одна из основных тем исследований в области робототехники и компьютерного зрения. Визуальная одометрия широко применяется в большом количестве задач, в числе которых детектирование и сопровождение объектов и навигация.

В частности, визуальная одометрия на базе стереокамеры активно исследуется поскольку позволяет точно и быстро оценить свое собственное движение в метрическом масштабе, что имеет большое значение в автономном вождении [5].

Существующие и активно развивающиеся методы визуальной одометрии можно разделить на две группы: базирующиеся на признаковых точках и прямые. Методы, базирующиеся на признаковых точках, используют в расчетах детектированные и сопровождаемые на последовательности кадров признаковые точки. Функция невязки для такого метода определяется как ошибки проецирования признаковых точек, а оценка параметров движения осуществляется путем минимизации функции невязки [6]. Однако, при таком подходе локальное изменение внешности признаковых точек на последовательности кадров приводит к дрейфу позиций признаковых точек от начальных позиций, что приводит к накоплению ошибки оценки собственного движения [7]. Напротив, прямые методы оценивают параметры движения путем минимизации разностей интенсивностей пикселей между двумя последовательными кадрами [8]. Использование большего количества пикселей в оценке параметров движения. Недостатком прямого метода является сложность детектирования и обработки выбросов, а также чувствительность к изменению яркости кадров [9]. По этой причине большинство новых методов визуальной одометрии использует поход на основе признаковых точек и детектированию выбросов алгоритмом RANSAC [10–14], последовательно выбирая минимальных набор признаковых точек, вычислением оптимальных параметров движения на их основе, и рейтингованию вычисленных параметров движения по оставшимся признаковым точкам.

В данной работе предлагается подход, который последовательно выбирает минимальный набор признаковых точек на изображении, на подобии того, как это происходит в подходах на базе признаковых точек, и вычисляет параметры движения используя интенсивности пикселей в окрестностях этих точек как в прямых методах. Рейтингование гипотез осуществляется не на основе фотометрической разности оставшиеся окрестностей признаковых точек, а на основе принадлежности гипотез некому множеству, большинству гипотез со схожими признаками.

Предложенный метод. Функцию невязки прямого метода визуальной одометрии можно записать в следующем виде:

$$error(R, \mathbf{t}) = \sum_i^N \sum_j^W \|T(K \times \mathbf{p}_i + win(j)) - I(K \times (R \times \mathbf{p}_i + \mathbf{t}) + win(j))\|^2, \quad (1)$$

где R – матрица поворота камеры, \mathbf{t} – вектор трансляции камеры, K – матрица внутренних параметров камеры, описывающая камеру Обскура, \mathbf{p}_i – позиция обзереваемой точки в пространстве, $win(j)$ – функция, задающая окрестность проекции точки \mathbf{p}_i с аппаратурой W , $I(\mathbf{x}\mathbf{y})$ – функция задающая текущий кадр изображения, $T(\mathbf{x}\mathbf{y})$ – функция задающая предыдущий кадр изображения.

Уравнение (1) описывает процесс движения камеры в пространстве. В момент времени t_0 камера снимает изображение T , в момент времени t_1 камера снимает изображение I . За время $t_1 - t_0$ камера поворачивается (это поворот описывается матрицей поворота R) и смещается на вектор \mathbf{t} . При этом камера наблюдает статическую точку окружающего пространства \mathbf{p}_i . Координаты точки \mathbf{p}_i вычисляются посредством стерео-триангуляции алгоритмом стерео-зрения [15]. В идеальном случае, если отсутствуют какие-либо шумы и искажения камеры, точке \mathbf{p}_i будет соответствовать пиксели на изображениях I и T которые будут иметь одинаковую яркость. Поскольку камера обзеревает большое количество точек пространства (по сути, все что снимает камера) и эти точки имеют разное свечение (наблюдаемая сцена обладает контрастностью и неоднородностью), и работа камеры вносит шумы и искажения в наблюдаемые яркости, можно выразить R, \mathbf{t} как параметры движения, которые минимизируют функцию (1).

Выразив матрицу поворота как функцию от углов Родрига [16]:

$$R = Rodrig(\alpha, \beta, \gamma), \quad (2)$$

задача оценки движения сводится к задаче минимизации нелинейной функции:

$$\alpha, \beta, \gamma, \mathbf{t} = \arg \min[error(Rodrig(\alpha, \beta, \gamma), \mathbf{t})]. \quad (3)$$

Сравнение результата работы вычисления оптического потока классическим методом [17] и базирующемся на предложенном методе приведены на рис. 1.

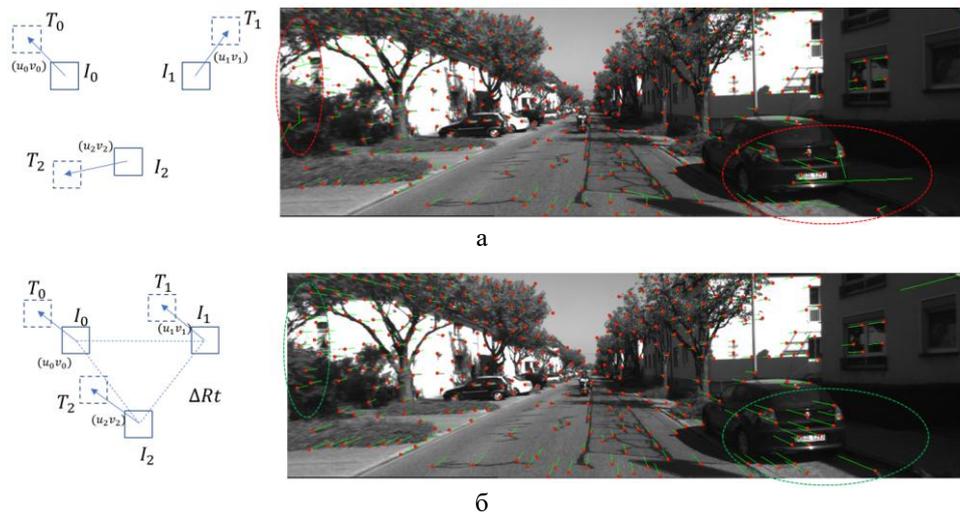


Рис. 1. Вычисление оптического потока классическим методом (а), вычисление оптического потока на базе предложенного метода (б)

Как показано на рис. 1, вычисление оптического потока путем минимизации фотометрической ошибки и определение оптимальных параметров движения – $\Delta R \mathbf{t}$ даст меньше выбросов по сравнению с методом вычисления оптического потока, который минимизирует фотометрическую ошибку и определяет оптимальные смещения – (u_i, v_i) индивидуально для каждой признаковой точки на изображении.

Цифровые видеокамеры искажают измеряемое свечение нелинейной передаточной функцией сенсора камеры, поскольку не способны охватить весь динамический диапазон, это выражается в том, что, если камера обзрывает окружение с неизменяемой «яркостью», полученные последовательные кадры могут значительно отличаться по суммарной яркости. Это явление нарушает предположение о постоянстве яркости обзриваемых точек, на основе которого задана функция (1). В случаях, когда два последовательных кадра будут сильно отличаться по яркости из-за резкого изменения времени экспонирования сенсора камеры, оценка параметров $\alpha, \beta, \gamma, \mathbf{t}$ будет содержать значительную ошибку.

Для нивелирования влияния описанных выше эффектов предлагается использовать непараметрическое преобразования изображения известное как Цензус-преобразование (*англ. Census transform*) [18], которое обладает инвариантностью к локальным изменениям яркости изображения, и зарекомендовало себя в алгоритмах компьютерного зрения. Цензус-преобразование строит бинарный дескриптор для каждого пикселя изображения путем сравнения яркости пикселя с яркостями его локального соседства (восемь соседних пикселей). В результате каждого сравнения формируется один бит дескриптора. Цензус-преобразованные изображения сравниваются, используя расстояние Хемминга, в отличие от оригинальных изображений, которые сравниваются, используя евклидово расстояние.

Расстояние по Хэммингу не применимо в нелинейной оптимизации в исходном виде, однако расстояние Хемминга можно линеаризовать. Представив каждый бит Цензус-дескриптора как отдельный канал в изображении, Уравнение (1) можно записать в следующем виде:

$$error(R, \mathbf{t}) = \sum_c^8 \sum_i^N \sum_j^W \left\| T(K \times \mathbf{p}_i + win(j))_c - I(K \times (R \times \mathbf{p}_i + \mathbf{t}) + win(j))_c \right\|^2, \quad (4)$$

где c – номер канала изображения.

Функция (4) обладает инвариантностью к локальным изменениям яркости изображения, а также более устойчива в условиях плохой видимости камерами.

Поскольку обзриваемые точки \mathbf{p}_i могут принадлежать динамическим объектам, или они не обзриваются в оба момента времени t_0 и t_1 , их использование в выборке для минимизации функции (4) приводит к значительному увеличению ошибки оценки параметров движения.

Для уменьшения влияния выбросов (точек \mathbf{p}_i принадлежащих динамическим объектам или не обзриваемых в оба момента времени t_0 и t_1) выполняется следующая процедура:

- 1) из множества признаковых точек выбирается подмножество случайных $M = 3$ точек;
- 2) выполняется минимизация функции (4);
- 3) оцененные параметры движения $\alpha, \beta, \gamma, \mathbf{t}$ заносятся в список.

Выполнив описанную процедуру несколько раз (в реальных экспериментах число повторений равно 1000), формируется набор гипотез параметров движения, часть из которых будет построена с использованием верных точек \mathbf{p}_i , а часть на выбросах.

Для того что бы разделить условно верные гипотезы от условно ложных, выполняется иерархическая кластеризация гипотез [19]. Гипотезы, которые были получены и использованием «истинных» точек p_i будут образовывать кластер «истинных» гипотез, которые отражают параметры движения камеры. Используя предположение что «истинных» точек p_i больше, чем ложных (иначе вычислить параметры движения практически невозможно), «истинных» гипотез будет тоже больше, и кластер с наибольшим количеством гипотез является кластером «истинных» гипотез.

Выделив кластер «истинных» гипотез, гипотезы входящие в этот кластер усредняются, что дает оценку параметров движения камеры. Точки p_i , которые использовались с построением «истинных» гипотез, взвешиваются согласно частоте их использования, и используются в дальнейшем уточнении параметров движения камеры.

Принцип работы метода детектирования ложных гипотез приведен на рис. 2.

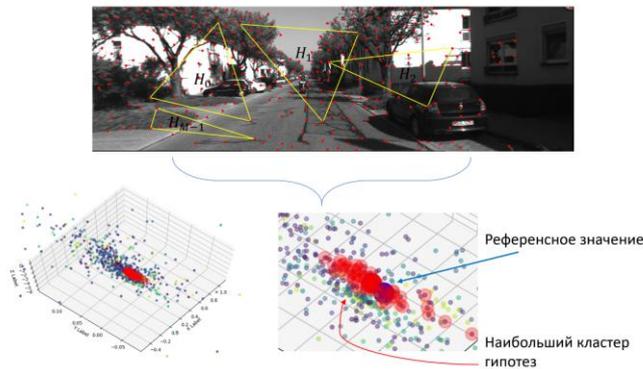


Рис. 2. Принцип работы метода детектирования ложных гипотез

Гипотеза H_i – набор из трех точек используемых для оценки параметров движения. Каждая точка на нижнем графике – параметры движения, построенные на основе данных гипотез H_i . Гипотезы, принадлежащие наибольшему кластеру выделены красным цветом

Испытание работы алгоритма. Испытание точности работы алгоритма осуществлялось датасете KITTI [20]. Датасет предоставляет истинную траекторию движения автомобиля и изображения с бортовой стереокамеры. Изображения используются для оценки параметров движения предложенным алгоритмом, а истинная траектория используется для оценки точности работы метода.

Точность оценки параметров движения зависит от количества используемых гипотез. График зависимости точности оценки позиции от количества гипотез приведен на рис. 3.

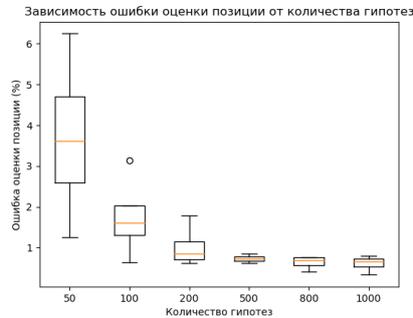


Рис. 3. Зависимость точности оценки позиции от количества гипотез

Из графика видно, что для данной конфигурации достаточно 500 гипотез
Траектория движения, полученная предложенным методом для датасета
KITTI с номером 00 приведена на рис. 4.

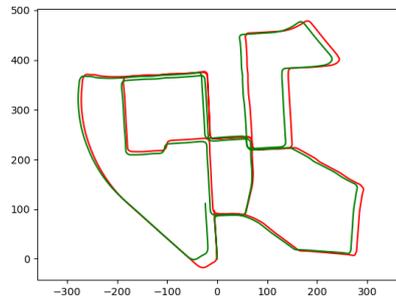


Рис. 4. траектория движения для датасета KITTI с номером 00. Красная линия – референсная траектория, зеленая линия – полученная траектория.
Ошибка в конечной точке – 0.6 % от пройденного пути

Заключение. В ходе проведенной работы был предложен метод оценки параметров движения камеры, который способен работать в условиях ухудшенной видимости, а также меньше подвержен влиянию выбросов во входных данных что ведет к меньшей ошибке оценке параметров движения. Время генерации 1000 гипотез и выделение «истинных» гипотез, с использованием функции невязки (1) составляет 50 мс. на процессоре core i7 2.6 ГГц, с использованием функции невязки (4) – 270 мс. в однопоточном режиме, что позволяет оценивать параметры движения в реальном масштабе времени. На датасетах KITTI предложенный метод достиг в среднем 2% ошибки от пройденного пути. В дальнейших работах по развитию предложенного метода предлагается улучшить способ выбора подмножества входных точек, выбирать такие подмножества точек, которые дадут «истинную» гипотезу с большей вероятностью. Это позволит формировать кластера гипотез с более четкими границами, что приведет к более точному выбору «истинных» гипотез и повысит точность оценки параметров движения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Howard A. Real-time stereo visual odometry for autonomous ground vehicles, 2008 *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2008 з. 3946-3952.
2. Scaramuzza D., Fraundorfer F. Visual odometry [tutorial], *IEEE robotics & automation magazine*, 2011, Vol. 18, No. 4, pp. 80-92.
3. Fraundorfer F., Scaramuzza D. Visual odometry: Part ii: Matching, robustness, optimization, and applications, *IEEE Robotics & Automation Magazine*, 2012, Vol. 19, No. 2, pp. 78-90.
4. Maimone M., Cheng Y., Matthies L. Two years of visual odometry on the mars exploration rovers, *Journal of Field Robotics*, 2007, Vol. 24, No. 3, pp. 169-186.
5. Aqel M.O.A. et al. Review of visual odometry: types, approaches, challenges, and applications, *SpringerPlus*, 2016, Vol. 5, No. 1, pp. 1-26.
6. Badino H., Yamamoto A., Kanade T. Visual odometry by multi-frame feature integration, *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 222-229.
7. Liu H. et al. Navigational drift analysis for visual odometry, *Computing and Informatics*, 2014, Vol. 33, No. 3, pp. 685-706.
8. Usenko V. et al. Direct visual-inertial odometry with stereo cameras, 2016 *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1885-1892.
9. Usenko V. et al. Direct visual-inertial odometry with stereo cameras, 2016 *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1885-1892.

10. *Derpanis K.G.* Overview of the RANSAC Algorithm, *Image Rochester NY*, 2010, Vol. 4, No. 1, pp. 2-3.
11. *Zuliani M.* RANSAC for Dummies, *Vision Research Lab, University of California, Santa Barbara*, 2009.
12. *Nistér D.* Preemptive RANSAC for live structure and motion estimation, *Machine Vision and Applications*, 2005, Vol. 16, No. 5, pp. 321-329.
13. *Raguram R., Frahm J. M., Pollefeys M.* A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus, *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2008, pp. 500-513.
14. *Li H. et al.* An efficient image matching algorithm based on adaptive threshold and RANSAC, *IEEE Access*, 2018, Vol. 6, pp. 66963-66971.
15. *Haller I., Nedeveschi S.* GPU optimization of the SGM stereo algorithm, *Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing*. IEEE, 2010, pp. 197-202.
16. *Piña E.* Rotations with Rodrigues' vector, *European journal of physics*, 2011, Vol. 32, No. 5, pp. 1171.
17. *Suhr J.K.* Kanade-lucas-tomasi (klt) feature tracker, *Computer Vision (EEE6503)*, 2009, pp. 9-18.
18. *Hafner D., Demetz O., Weickert J.* Why is the census transform good for robust optic flow computation?, *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, Berlin, Heidelberg, 2013, pp. 210-221.
19. *Johnson S.C.* Hierarchical clustering schemes, *Psychometrika*, 1967, Vol. 32, No. 3, pp. 241-254.
20. *Geiger A. et al.* Vision meets robotics: The kitti dataset, *The International Journal of Robotics Research*, 2013, Vol. 32, No. 11, pp. 1231-1237.

Статью рекомендовал к опубликованию д.т.н., профессор В.В. Курейчик.

Пантелюк Павел Александрович – Южный федеральный университет; e-mail: pavel-panteluk@mail.ru; г. Таганрог, Россия; тел.: +79185693138; кафедра синергетики и процессов управления; аспирант.

Pantelyuk Pavel Alexandrovich – Southern Federal University; e-mail: pavel-panteluk@mail.ru; Taganrog, Russia; phone: +79185693138; the department of synergetics and management processes; postgraduate student.