

УДК 004.932.7, 004.4, 004.021

## ИССЛЕДОВАНИЕ МЕТОДОВ ОДНОВРЕМЕННОГО КАРТИРОВАНИЯ И ЛОКАЛИЗАЦИИ БЕСПИЛОТНЫХ ЛЕТАТЕЛЬНЫХ АППАРАТОВ ПО ВИДЕОПОТОКУ, ПОЛУЧЕННОМУ С ЕДИНСТВЕННОЙ КАМЕРЫ

А.В. Боковой (*bokovoy@isa.ru*)

Федеральный исследовательский центр «Информатика и  
управление» Российской академии наук, Москва

**Аннотация.** В работе исследуются современные методы одновременного картирования и локализации по видеопотоку (Visual-based Simultaneous Localization and Mapping - vSLAM), полученному с единственной камеры, применимые в задачах навигации беспилотных летательных аппаратов в неизвестной среде. Представлены данные экспериментального сравнения актуальных программных реализаций методов vSLAM с использованием среды Robot Operating System (ROS).<sup>1</sup>

**Ключевые слова:** компьютерное зрение, методы одновременного картирования и локализации, единственная камера, видеопоток.

### Введение

В настоящее время наблюдается существенное повышение интереса к беспилотным летательным аппаратам (БПЛА). Современные БПЛА используются для решения широкого круга задач: мониторинг объектов инфраструктуры, дистанционное зондирование земли, доставка грузов и т.д. Как правило, для навигации, а также повышения автономности, БПЛА оснащаются различными видами датчиков, которые позволяют отслеживать состояние объекта управления в каждый момент времени. К таким датчикам относятся датчики инерциальной навигационной системы (ИНС) – акселерометры, гироскопы, с помощью которых определяется отклонение корпуса БПЛА в земных координатах, приемники системы глобального позиционирования GPS/ГЛОНАСС, оптические датчики (видеокамеры, стереопары), ИК датчики, лазерные дальнометры и т.д.

Существует большое количество методов, позволяющих определять текущее местоположение беспилотного летательного аппарата, но, как

---

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ (проект № 15-07-07483).

правило, они используют информацию от спутниковых навигационных систем, которая доступна только на открытой местности. Для навигации в закрытых помещениях необходимо полагаться на ИНС, оптические и прочие датчики.

Одним из активно развивающихся направлений исследований в настоящее время является навигация БПЛА по данным, полученным с оптических датчиков. Одной из задач этого направления является разработка методов и алгоритмов одновременного картирования и локализации по видеопотоку [Lemaire, 2007], полученному с единственной камеры. Использование единственной камеры приводит к ряду проблем, которые необходимо рассматривать как отдельные подзадачи, а именно:

- выделение областей или особенностей (особых точек, углов и т.д.) изображений, которые будут использоваться как элемент построенной модели местности (карты);
- построение модели неизвестной местности по двум или более кадрам видеопотока;
- локализация системы в построенной модели;
- построение и корректировка пройденной траектории.

В связи с большими объемами данных, поступающих с камеры БПЛА, дополнительным условием (при решении указанных выше задач) является возможность обработки поступающих изображений, локализации и построения карты в режиме реального времени на современных бортовых вычислителях. Невыполнение этого условия может стать причиной потери актуальности данных и, следовательно, невозможности навигации БПЛА в неизвестной среде.

В настоящей работе представлен аналитический обзор методов одновременного картирования и локализации по видеопотоку, полученному с единственной камеры, а именно методов MonoSLAM [Davidson, 2007], PTAM [Klein, 2007] и LSD-SLAM [Engel, 2014]. Также приведены результаты экспериментального исследования, в котором сравнивались различные параметры (скорость обработки изображений, детализация полученной карты, скорость составления карты и т.д.) актуальных программных реализаций указанных алгоритмов.

## **1. Выделение областей и особенностей изображений**

Для построения модели неизвестной местности и локализации БПЛА в ней необходимо выделить устойчивые признаки (особенности) изображений, которые будут являться элементами будущей карты. Определяющим условием при выделении особенностей является возможность сравнивать полученные признаки для разных изображений и

находить однозначное соответствие между ними. В основном под особенностями понимаются точки локального градиента яркости и дескрипторы, которые представляют из себя вектор, элементами которого являются градиенты в некоторой области вокруг полученной особенности.

Распространенные методы выделения особенностей изображения, такие как SIFT [Lowe, 2004] и SURF [Bay, 2006], используемые в MonoSLAM, позволяют получать инвариантные повороту, масштабированию и (в некоторой степени) перспективным искажениям дескрипторы размером 64 или 128 элементов, которые соответствуют каждой выделенной особенности. Модель местности в таком случае модель местности (карта) представляет из себя облако точек, элементами которой и являются эти особенности.

Метод FAST-10 [Rosten, 2006], используемый в алгоритме PTAM, выделяет углы в качестве особенностей изображения и, соответственно, вычисляет дескриптор каждого угла.

Главными преимуществами вышеописанных методов выделения особенностей является скорость обработки поступающих изображений и составления дескрипторов полученных особенностей, позволяющая выполнять эти алгоритмы в реальном времени даже на слабых бортовых вычислителях и использовать их как составной элемент алгоритмов vSLAM.

Недостатком использования алгоритмов SIFT, SURF и FAST-10, является скорость сопоставления дескрипторов разных изображений видеопотока. По причине того, что на изображении могут быть найдены тысячи особенностей, каждой из которых соответствует дескриптор размером от 64-х элементов, то их полный перебор и сравнение всех дескрипторов может занимать длительное время. Для сокращения времени сопоставления наборов дескрипторов между собой используется представление дескрипторов в виде K-мерного дерева и, соответственно, применяются алгоритмы поиска на K-мерных деревьях.

В качестве входных данных для составления модели местности могут также использоваться наиболее контрастные области (а не отдельные точки) изображения. Такой подход впервые был представлен авторами метода LSD-SLAM (см. рис. 2). Он позволяет строить детализированные модели, а алгоритм, его реализующий, может выполняться в реальном времени без использования ускорения на GPU.



Рис 1. Карта, построенная с использованием контрастных областей изображений (LSD-SLAM Foodcourt Dataset)

## 2. Построение модели местности и локализация

Для построения модели местности (карты) с помощью единственной камеры используются алгоритмы, позволяющие находить положение определенной точки в пространстве, используя 2 кадра, на которых эта точка обозревается. В качестве точек карты используются вышеописанные особенности изображений.

Для начала, необходимо откалибровать камеру. Это позволит найти внутренние параметры камеры, такие как фокусные расстояния, углы наклона пикселей и принципиальную точку. Эти параметры используются как для устранения искажений на изображениях, так и для построения карты глубины изображения (т.е. матрицы, каждый элемент которой соответствует расстоянию от камеры до пикселя).

Зная параметры калибровки камеры, можно найти фундаментальную матрицу и существенную матрицу. Фундаментальная матрица, в данном случае, необходима для вычисления уравнений эпиполярных линий, которые необходимы для составления карты глубины. Для вычисления фундаментальной матрицы обычно используется Eight-point algorithm [Hartley, 1997] и его вариации, которые позволяют найти фундаментальную матрицу по 8-ми или более сопоставленных точек на двух изображениях. Существенная матрица позволяет восстановить положение и поворот камеры относительно ее предыдущего состояния по двум изображениям. Для нахождения такой матрицы достаточно знать внутренние параметры камеры, полученные на этапе калибровки камеры.

Процесс вычисления расстояний до особых точек и нахождения положения камеры происходит в каждый для каждого вновь поступившего изображения с камеры. Текущий снимок сравнивается с предыдущим.

Данные о положении камеры поступают на вход расширенному фильтру Калмана (Extended Kalman Filter) [Julier, 2004], где вычисляется текущее положение относительно построенной карты, а также производится корректировка предыдущих положений согласно полученным данным.

Вышеописанные вычисления производятся для каждого вновь поступившего изображения с камеры, что накладывает требования к скорости выполнения каждого алгоритма. В совокупности, все алгоритмы должны выполняться до поступления нового изображения с камеры. Иначе возможна потеря актуальности данных и, следовательно, затрудненность или невозможность навигации БПЛА или иного робототехнического устройства.

На данный момент, существуют программные реализации описанных алгоритмов, выполняющихся в режиме реального времени для видеоряда с частотой до 30-ти кадров в секунду.

### **3. Экспериментальные исследования методов vSLAM**

Для проведения экспериментов была сформирована коллекция, состоящая из 63 видеозаписей помещений коридорного типа. Были смоделированы движения БПЛА, оснащенного камерой, а именно движения вперед/назад, изменение высоты, повороты вокруг своей оси и т.д. Все изображения видеопотока имеют разрешение 640x480 pix. Частота кадров – 30 кадр./с. Камера была откалибрована алгоритмами открытой библиотеки компьютерного зрения OpenCV [Datta, 2009] с помощью паттерна типа “шахматная доска” размером 9x6.

Для эксперимента были выбраны разные помещения коридорного типа. 18 видеозаписи были произведены с учетом замыкания траектории (т.е., возвращения камеры в исходную позицию). Помещение на этих видеозаписях имеет квадратную форму с квадратной перегородкой внутри. Остальные видеозаписи были сняты в коридоре, имеющим один поворот. Длины траекторий – 12,8 м для первых 18 видео и от 5 м до 9,8 м для всех остальных.

Для сравнения реальной траектории камеры и траектории, построенной алгоритмами, была создана 3D модель описанных помещений и реальной траектории. Сравнение производилось путем наложения построенной модели и карты с траекторией, полученной с помощью каждого из методов.

Платформа, на которой производился эксперимент, представляет собой персональный компьютер со следующими основными характеристиками: CPU – Intel Core i7 3770K 3,4 GHz, 16 GB RAM, ОС – Linux

Для тестирования были выбраны свободные реализации алгоритмов MonoSLAM, PTAM и LSD-SLAM на языке C++ для фреймворка Robot

Operating System (ROS). Исходные коды методов скомпилированы с использованием набора компиляторов GCC 4.9.2 без использования возможности ускорения параллельных вычислений (таких, как OpenCL или CUDA).

Параметрами для сравнения являлись время обработки одного изображения (пре-процессинг + выделение особенностей), время обновления карты и текущего положения системы, количество устойчивых особенностей (данный параметр также характеризует детализацию карты) и среднее отклонение от действительной траектории. Примеры карт, построенных с помощью каждого из методов представлены на рисунке 2.

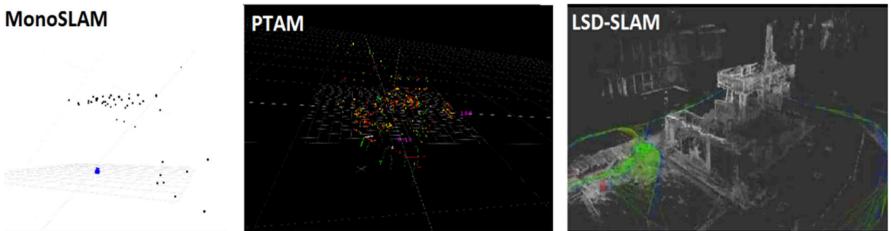


Рис 2. Примеры карт, построенных с помощью методов MonoSLAM, PTAM и LSD-SLAM

Результаты эксперимента приведены в таблице 1.

Метод	Время обработки кадра (мс)	Время обновления карты (мс)	Кол-во устойчивых признаков	Среднее отклонение от траектории (в %)
MonoSLAM	8.35	4.7	473	2.1
PTAM	14.7	32.95	5837	1.9
LSD SLAM	12.4	20	~19 000*	0.9

\*Количество устойчивых признаков различается в следствии вероятностного характера алгоритма выделения областей интереса.

По этим данным можно заключить, что метод MonoSLAM является самым быстрым из 3-х представленных методов, однако обладает наименьшей детализацией карты и наибольшим отклонением построенной траектории от действительной. Метод PTAM – самым медленным, но обладающим большей детализацией карты, нежели

MonoSLAM и более высокой точностью построенной траектории. Метод LSD-SLAM обладает средними среди этих методов характеристиками по скорости выполнения алгоритма, но зато позволяет строить высокодетализированные карты и его среднее отклонение от заданной траектории – наименьшее.

В каждом из методов, на разных этапах (выделение особенностей, построение карты и т.д.) используются разные алгоритмы, реализации которых влияют на скорость выполнения и обработки данных. Некоторые данные об использовании конкретных алгоритмов в каждом из методов приведены в таблице 2.

Этап	MonoSLAM	PTAM	LSD-SLAM
Выделение особенностей	SIFT	FAST-10	Выделение контрастных областей
Построение фундаментальной матрицы	8-point algorithm	5-point algorithm	5-point algorithm
Построение карты и траектории	Extended Kalman Filter	JCBB*	Extended Kalman Filter
Алгоритм корректировки замыкания траектории	-	-	+

\*JCBB – Joint Compatibility Branch and Bound [Neira, 2001]

В ходе эксперимента было показано, что все 3 метода подходят для использования в задачах одновременного картирования и локализации БПЛА в помещениях. Использование каждого из методов обуславливается контекстом решаемой задачи. При малых вычислительных мощностях робототехнического устройства целесообразно использование метода MonoSLAM. Возможно улучшить точность алгоритма при использовании данных с инерциальных датчиков. Если данные с видеокамеры отсылаются на наземную станцию, то точной метода LSD-SLAM даст достаточно точные результаты для возможности навигации БПЛА. Использование метода PTAM является уместным при использовании полученной карты другими, более ресурсоемкими алгоритмами (т.к. количество выделенных точек в разы меньше, нежели в методе LSD-SLAM).

## Заклучение

В работе рассмотрены методы одновременного картирования и локализации по видеопотоку MonoSLAM, PTAM и LSD-SLAM. В результате эксперимента установлена возможность применения их в задачах навигации и повышения автономности беспилотных летательных аппаратов в условиях нехватки датчиков (наличия только информации с единственной камеры).

В дальнейшем планируется доработка метода LSD-SLAM, как наиболее современного и перспективного, с учетом полученных данных исследования, программная оптимизация отдельных компонентов алгоритма и интеграция метода в систему управления беспилотными летательными аппаратами малого размера.

## Список литературы

- [**Bay, 2006**] Bay H., Tuytelaars, T., Van Gool, L., SURF: Speeded Up Robust Features. // Proceedings of the ninth European Conference on Computer Vision, 2006.
- [**Datta, 2009**] A. Datta, J.S. Kim, T. Kanade. Accurate camera calibration using iterative refinement of control points. // Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on 2009, p. 1201-1208
- [**Davison, 2007**] Davison A, Reid I, Molton N. and Stasse O. MonoSLAM: Real-Time Single Camera SLAM // IEEE Transactions on Pattern Analysis and Machine Intelligence archive, Volume 29 Issue 6, J, pp 1052-1067, 2007
- [**Engel, 2014**] Engel J., Schops T. and Cremers D. LSD-LSAM: Large-Scale Monocular SLAM // In Proc. European Conference on Computer Vision (ECCV), 2007
- [**Hartley, 1997**] Hartley I, In Defence of the 8-point Algorithm. // Pattern Analysis and Machine Intelligence, Vol. 19, Issue 6, pp. 580-593, 1997
- [**Julier, 2004**] Julier, S.J.; Uhlmann, J.K.. Unscented filtering and nonlinear estimation. // Proceedings of the IEEE, pp. 401–422, 2004
- [**Klein, 2007**] Klein G. and Murray D. Parallel Tracking and Mapping for Small AR Workspaces // In Proc. International Symposium on Mixed and Augmented Reality, ISMAR'07, 2007
- [**Lemaire, 2007**] Lemaire T., Berger C., Jung I. and Lacroix S. Vision-Based SLAM: Stereo and Monocular Approaches // International Journal of Computer Vision, Volume 74, Issue 3, pp 343-364, 2007
- [**Lowe, 2004**] Lowe G., Distinctive Image Features from Scale-Invariant Keypoints. // International Journal of Computer Vision, pp. 91-110, 2004
- [**Rosten, 2006**] Neira J, Tardos J. Data Association in Stochastic Mapping using the Joint Compatibility Test. // Robot and Automation, IEEE Transactions pp. 890-897, 2001
- [**Rosten, 2006**] Rosten E. and Drummond T., Machine learning for high-speed corner detection. // In Proc. 9th European Conference on Computer Vision (ECCV'06), 2006.