

УДК 004.932.2

РЕАЛИЗАЦИЯ АЛГОРИТМА ВИЗУАЛЬНОЙ ЛОКАЛИЗАЦИИ БЕСПИЛОТНОГО ЛЕТАТЕЛЬНОГО АППАРАТА НА ОСНОВЕ ИЗВЕСТНОЙ 3D МОДЕЛИ ОКРУЖЕНИЯ С ИСПОЛЬЗОВАНИЕМ CNN- СЕГМЕНТАЦИИ.

А.К. Буйвал (*alexbuyval@gmail.com*)

М.А. Гавриленков (*gavrilentkov@umlab.ru*)

Брянский Государственный Технический Университет,
Брянск

Аннотация. В статье рассматривается новый подход к визуальной локализации в помещении, основанный на использовании CNN-сегментации изображений для улучшения ранее разработанного алгоритма локализации. Авторы предлагают ряд нововведений для устранения таких проблем, как шум, создаваемый нестатическими объектами или группами непредвиденных объектов в помещении. Информацию о таких объектах в данной работе считается избыточной и исключается из алгоритма локализации. В данной статье рассматривается подход исключения избыточной информации с видеокамеры для локализации в пространстве посредством использования сверточных нейронных сетей для преобразования изображение в сегментированную карту поверхности, на которой хорошо прослеживаются отличительные признаки (плоскости и грани) изображения. В основе алгоритма лежит определение границ объектов на изображении с камеры и сопоставлении этих границ с моделью.

Ключевые слова: визуальная локализация, нейронные сети, ROS, фильтр частиц, SegNet.

Введение

Проблема локализации мобильного робота остается актуальной и по сей день. Локализация мобильного робота, а в частности БПЛА является критически важной функцией для функционирования робота внутри помещения. В отличие от роботов, которые предназначены для выполнения задач на открытой местности, роботы функционирующие внутри помещений, не могут воспользоваться датчиками глобальной навигации

(GPS/ГЛОНАСС). Для выполнения локализации, построения маршрутов, а также принятия решений очень часто используются инерциальные датчики, лазерные дальномеры или данные с видеокамеры.

Одним из прогрессивных подходов к решению задачи локализации является использование информации, полученной с бортовой камеры БПЛА и применении фильтра частиц. Данный подход основывается на сравнении изображения, полученного с камеры, со смоделированным изображением, полученным на основе 3D модели помещения. Этот подход во многом превосходит альтернативные решения, но, к сожалению, не лишен недостатков. Весомым недостатком становится избыточность получаемой с камеры информации.

В данной статье предлагается рассмотреть идею предобработки кадров с видеокамеры для дальнейшего использования в алгоритме локализации.

1. Базовый алгоритм локализации

В ранних работах [Буйвал, 2015, 2017] описывается алгоритм визуальной локализации с использованием граней изображений как основного характерного визуального признака. В данных работах приводится способ оценки схожести двух изображений (изображения с реальной камеры и изображения с камеры модели), как основного критерия оценки гипотез о местоположении робота в пространстве. Для обработки данных о схожести изображений, а также данных с других доступных датчиков используется алгоритм локализации, основанный на множестве частиц (гипотез о местоположении робота) – фильтр частиц (particle filter).

Для того, чтобы оценить вероятность каждой гипотезы (частицы) необходимо оценить близость границ, полученных из изображения с камеры, к границам, полученным из смоделированного изображения, соответствующего гипотезе.

В разработанной системе используется метод ближайших граней. Данный метод не имеет такого недостатка, как учет только совпадающих на гранях пикселей. Т.о. если грани одного и того же конструктивного элемента на изображении с камеры и на смоделированном изображении не совпадают, но проходят близко друг к другу, то это все равно будет учитываться в конечной вероятности, т.к. в данном методе предлагается оценивать близость смоделированной грани к грани на изображении с камеры с помощью набора нормалей, построенных от смоделированной грани.

В рассматриваемом методе выполняются последовательно следующие этапы:

1. Выделение границ на изображении, полученном с камеры.

2. Нахождение прямых в выделенных границах с помощью преобразования Хафа. Полученные прямые используются как исходные данные для вычисления вероятности каждой частицы.
3. Рендеринг изображения согласно вектору состояния частицы.
4. Выделение границ на полученном изображении.
5. Нахождение прямых в выделенных границах с помощью преобразования Хафа.
6. На каждой прямой формируется набор точек с постоянным шагом. Из каждой точки строится нормаль некой предельной длины. Если эта нормаль пересекается с какой-либо прямой из полученных на основе изображения с камеры, то длина такой нормали учитывается в общем весе рассматриваемой прямой по следующей формуле:

$$g(d) = \exp\left(-\frac{d^2}{2\sigma^2}\right), \quad (1)$$

где d – длина нормали, σ – параметр определяющий вес нормали в зависимости от длины нормали (позволяет либо усиливать влияние длинных нормалей, либо уменьшать).

7. Для каждой прямой вычисляется общий вес путем суммирования всех весов нормалей по следующей формуле:

$$l = \frac{\sum_{i=0}^S g(d_i)}{S}, \quad (2)$$

где S – общее количество прямых в смоделированном изображении.

8. Вычисляется итоговая вероятность гипотезы на основе объединения весов каждой линии по следующей формуле:

$$W = \alpha \cdot \exp\left(k \frac{\sum_{n=0}^m l_n}{m}\right), \quad (3)$$

где m – количество прямых, α, k – параметры учета веса прямых в итоговом результате.

2. Сегментация изображений

Подавляющее большинство методов локализации основанных на обработке видеопотока используют так называемые характерные признаки изображения (ребра, грани, ключевые точки и т.д.). Соответственно, для локализации необходимо выделить эти признаки на изображении. Из этого вытекает проблематика избыточной/недостаточной информации по результатам выделения признаков. Зачастую эта разница становится ощутима при перемещении из меньшего помещения в большее, и наоборот, или же, в случаях сильной загруженности помещения. В частности, в работах [Буйвал, 2015, 2017] в качестве характерных признаков выступают ребра. Для нахождения ребер на изображении применяется связка «оператор Кэнни» + «преобразование Хафа». Результатом становится набор предполагаемых ребер изображения, соответствующих заданным

начальным условиям выборки (погрешность, минимальная длина и т.д.). В данной работе используются реализации оператора Кэнни и преобразований Хафа в пакете OpenCV. На рис. 1 представлен случай выявления граней изображения данным способом, но с неудачными начальными значениями. Как видно из рисунка, в тех случаях, когда изображение имеет четкие переходы цветов, и можно выделить непрерывные длинные ребра (левая часть изображения), алгоритм обрабатывает без проблем (красные круги оказались в точности на синих линиях). Тем не менее, в центральной и правой части ребра не были найдены, более того, в правой части изображения были сформированы избыточные ребра из-за изображенного там объекта. Подобные эффекты возникают из-за наличия множеств оттенков одного цвета на изображении.

Долгое время мы работали над различными подходами к улучшению нашего алгоритма локализации. Основной проблемой для нас было то, что мелкие предметы в помещении дают непропорционально большие помехи. Другой проблемой для нас стал выбор параметров инициализации для функции выделения граней. Здесь явно прослеживалась зависимость выбора критериев (порог толерантности, минимальная длина линии) с качеством локализации. К несчастью, золотой середины в выборе параметров найти не удалось. В зависимости от освещенности помещения, а также его загруженности посторонними предметами, выбор параметров определял качество локализации. К примеру, в загруженных предметами помещениях, следует использовать большее значение для параметра минимальной длины линии. В помещениях, где присутствуют отражающие поверхности, или комната наполнена солнечными бликами, следует повысить значение коэффициента толерантности. Отдельным случаем становятся комнаты, где поверхность пола или стен содержат крупные линии (например, крупная плитка на полу, обои с прямым рисунком).

Решением данной проблемы становится использование сверточных нейронных сетей. Одним из вариантов использования CNN является сегментация изображений. Сегментация изображений применяется для выделения отдельных объектов (человек1, человек2, стул1, стол1, стул2), или же для выделения групп объектов (люди, пол, потолок, мебель и т.д.). Оба эти подхода позволяют снизить количество используемых цветов в изображении.

В качестве базовой нейронной сети для модернизированного алгоритма была выбрана нейросеть SegNet [Kendall, 2015; Badrinarayanan, 2015, 2017]. Выбор нейросети обуславливается компромиссным оптимумом между качеством сегментации (82-85%) и скоростью сегментации (0.01-0.02с). На рисунке 2 представлено тоже самое помещение, но обработанное с помощью нейронной сети. Как видно, довольно четко остаются выражены

относительные границы между полом и потолком, а также дверь в левой части изображения.



Рис. 1. Результат поиска границ на изображении и их сопоставления с границами, полученными из смоделированного изображения (синие линии – ребра из моделируемого изображения, зеленые линии и красные круги – найденные на изображении с камеры)

Применение нейронной сети позволило снизить количество используемых цветов от нескольких миллионов до 12 основных. Таким образом отпадает потребность в подборе параметров для применения преобразований Хафа. Теперь достаточно выбрать оптимальные параметры, которые будут применяться во всех операциях.

На рис. 2 также представлены результаты применения связки «оператор Кэнни» + «преобразование Хафа» к сегментируемому изображению. Видно, что количество найденных ребер, которые соответствуют смоделированному изображению, превышает количество ребер на изображении без предварительной сегментации. В процессе сегментации кадра на изображении могут возникать артефакты (правый нижний угол), это обуславливается недостаточным опытом нейронной сети и информационной энтропией. Тем не менее, при достаточном уровне обучения нейронной сети, такие артефакты слабо влияют на локализацию. В первую очередь следует обращать внимание на стык стен и пола – такие границы наиболее четко описывают помещение.

3. Реализация модернизированного алгоритма в ROS

Базой для разработки подсистемы локализации выбрана система ROS. Robotic Operation System позволяет разрабатывать масштабируемые приложения, развертываемые на платформе Ubuntu или Debian.

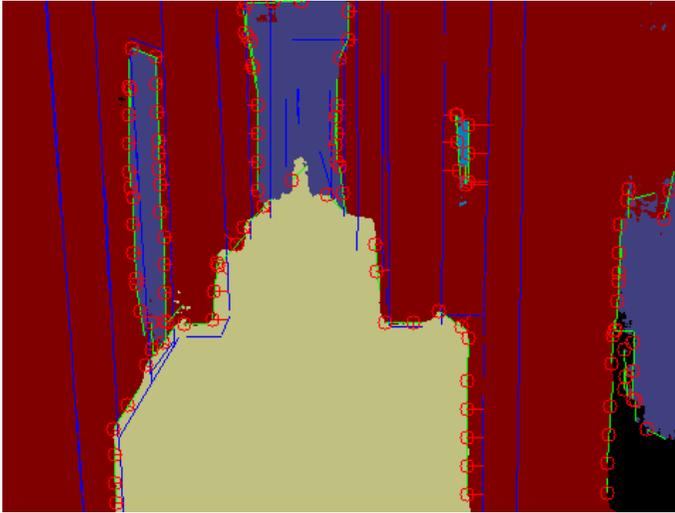


Рис. 2. Результат сегментации изображения с последующим поиском и сравнением найденных граней

На рис. 3 показана схема подсистемы локализации БПЛА без дополнительных модернизаций. Основным модулем (узлом, в терминологии ROS) является `edge_localization_node`. Данный узел получает изображение с камеры БПЛА, после выделения ключевых линий на изображении пересылает их в узел `rendering_node`.

В качестве программной реализации улучшения базового алгоритма выступает новый узел `neironet_adapter_node`. Данный узел служит связью между нейронной сетью и подсистемами ROS. Текущий узел получает от узла `edge_localization_node` исходное изображение, после чего к нему применяется сегментация. Сегментированное изображение возвращается в `edge_localization_node` и базовый алгоритм продолжается [Буйвал, 2017].

4. Экспериментальный анализ алгоритма

Для проверки работоспособности системы был проведен следующий эксперимент.

Была создана модель офисного коридора протяженностью 10 метров, протяженная сторона ориентирована по оси Y. В качестве объекта исследования используется БПЛА типа квадрокоптер. Задача состоит в том, чтобы БПЛА пролетел от начала коридора до его конца, используя только визуальные датчики. Во время эксперимента квадрокоптер управлялся в ручном режиме и преимущественно двигался вдоль коридора.

Целью эксперимента является определение минимального времени и точности локализации.

В данном эксперименте подсистема работает в режиме «offboard», это означает, что все вычисления происходят на внешнем управляющем компьютере со следующими характеристиками: *ОЗУ: 8Гб; Процессор: Intel Core i7 7700HQ 2.8ГГц; видеокарта: NVIDIA GeForce GTX 1050(m) 2ГБ*. Изображение с камеры квадрокоптера поступало в режиме реального времени на внешний компьютер.

В результате эксперимента были получены числовые значения распределения временной задержки по каждому из этапов локализации.

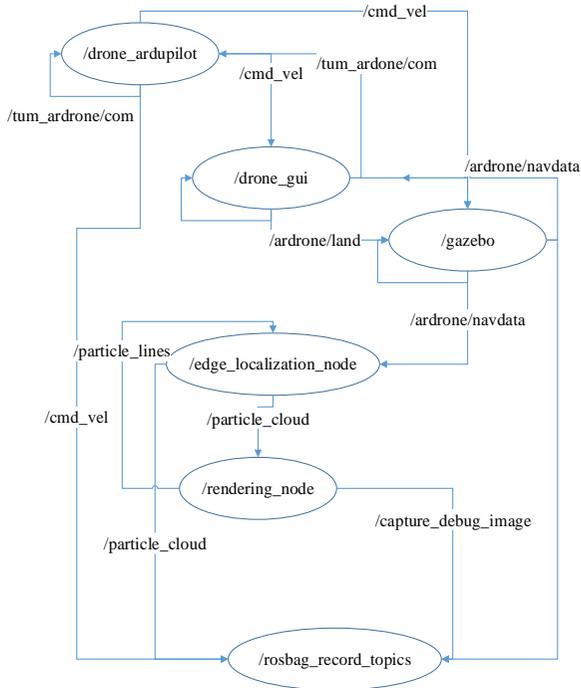


Рис. 3. Схема подсистемы локализации

Как видно из таблицы, наибольшая доля времени уходит именно на сегментацию изображения с помощью нейронной сети. Таким образом, время локализации для одного кадра не превышает 200 мс. В целом, данное время является допустимым для локализации при условии медленного движения БПЛА. Предполагается, что сократить это время возможно за счет увеличения вычислительной мощности графического процессора и использовании более прогрессивных архитектур нейронных сетей.

Таблица 1

Распределение времени работы в модуле сегментации

	Съемка кадра, мс	Изменение размера изображения, мс	Обработка в SegNet, мс	Представление результатов, мс	Суммарный результат, мс
min	4,43	1,35	165,74	4,51	176,05
avr	4,47	1,54	167,10	4,23	177,36
max	4,50	1,74	168,47	3,94	178,67

На рис. 4,5 представлено сравнение рассчитанных и реальных значений позиции БПЛА по оси X и Y соответственно. В качестве ground truth выступает информация от ультразвуковых датчиков, расположенных в помещении.

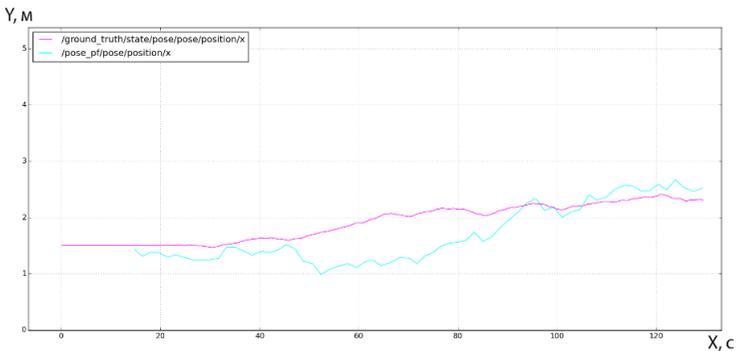


Рис. 4. Сравнение рассчитанных значений позиции квадрокоптера с реальными по оси X (красный - реальное значение, синий - рассчитанное значение)

На графике по оси Y (длинная сторона коридора) видно, что графики находятся в ощутимой близости друг к другу. Рассчитанное положение

робота практически совпадает с реальным на протяжении всего времени эксперимента.

Схожая ситуация наблюдается по оси X, но в середине эксперимента наблюдается значительное отклонение, порядка 0.4 метра. Тем не менее, на всех остальных участках отклонения не превышает ожидаемого.

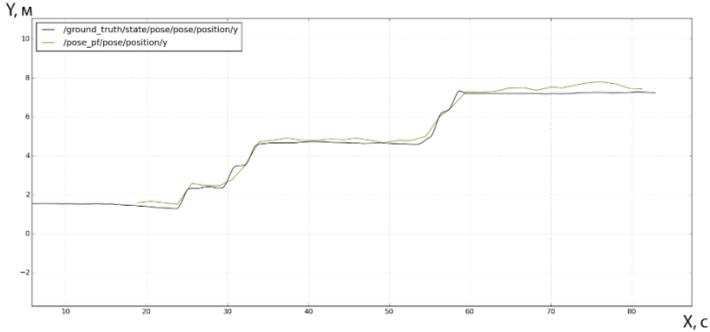


Рис. 5. Сравнение рассчитанных значений позиции квадрокоптера с реальными по оси Y (зеленый - реальное значение, черный - рассчитанное значение)

На рис. 6, 7 представлены аналогичные графики, но с использованием метода визуальной локализации, который не использует сегментацию изображения.

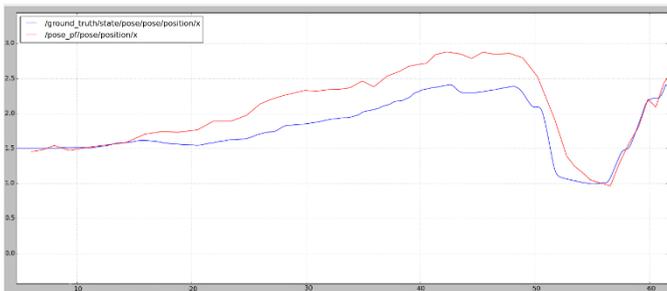


Рис. 6. Сравнение рассчитанных значений позиции квадрокоптера с реальными по оси X (без сегментации)

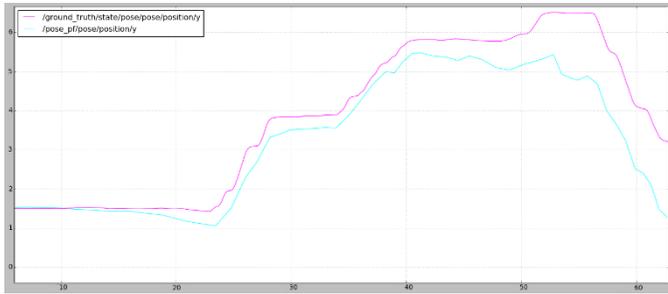


Рис. 7. Сравнение рассчитанных значений позиции квадрокоптера с реальными по оси Y (без сегментации)

Направление дальнейших исследований

В качестве направления дальнейших исследований предполагается анализ влияния архитектуры нейронной сети на качество локализации. Предполагается создание собственной нейронной сети на базе архитектуры применяемой в нейронной сети ResNet.

Заключение

В работе продемонстрирована состоятельность изложенного подхода для локализации. Пока рано говорить о завершенности подсистемы визуальной локализации, но на основании полученных результатов, можно сделать вывод, что использование нейросетевой сегментации позволяет повысить качество локализации по сравнению с классическими подходами. Данная задача требует дальнейшего изучения, что позволит в обозримом будущем повысить быстродействие системы и качество локализации.

Список литературы

- [Буйвал, 2015] Alexander Buyval, Mikhail Gavrilentov, Robust indoor visual localization of unmanned micro aerial vehicle using a particle filter and 3d edge model of environment, IX международная научно-практическая «современные проблемы машиностроения» конференция СПИМ - 2015, 1–4 декабря 2015 г. Томск, Россия
- [Буйвал, 2017] Buyval A., Gavrilentov M., Magid E. A multithreaded algorithm of UAV visual localization based on a 3D model of environment: implementation with CUDA technology and CNN filtering of minor importance objects. // In Proceedings of the 2017 International Conference on Artificial Life and Robotics (ICAROB 2017). January 19-22, 2017, Miyazaki, Japan. P.356-359.

- [**Kendall, 2015**] Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding." arXiv preprint arXiv:1511.02680, 2015.
- [**Badrinarayanan, 2017**] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." PAMI, 2017. arXiv:1511.00561v3 [cs.CV] 10 Oct 2016
- [**Badrinarayanan, 2015**] Vijay Badrinarayanan, Ankur Handa and Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling." arXiv preprint arXiv:1505.07293, 2015.
- [**Nuske, 2008**] Nuske, S., Roberts, J., & Wyeth, G. (2008, May). Outdoor visual localization in industrial building environments. // IEEE International Conference on Robotics and Automation, Pasadena, CA., 2008, 544-550
- [**Nuske, 2009**] Nuske, S., Roberts, J., & Wyeth, G. Robust outdoor visual localization using a three- dimensional- edge map // Journal of Field Robotics. 2009 №26(9), 728-756.
- [**Redmon, 2015**] Redmon, J., Allen, S. D. Unified, Real-Time Object Detection. arXiv preprint arXiv: 1506.02640v4, 2015.