

МИНОБРНАУКИ РОССИИ
Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Кафедра математического обеспечения и применения ЭВМ

ОТЧЕТ

«Использование методов ML в задачах визуальной одометрии»
Реферат-обзора предметной области

Студент гр. 3303

Гэ Као

Руководители

Кринкин К.В.

Санкт-Петербург

2018

1. Реферат-обзора предметной области

Визуальная одометрия (VO), как один из самых важных методов оценки позы и локализации роботов, в последние несколько десятилетий привлекла значительный интерес как в сообществах компьютерного зрения, так и в робототехнике [1]. Он широко применяется к различным роботам в качестве дополнения к GPS, Инерциальной навигационной системе (INS), одометрам колес и т. Д.

За последние тридцать лет была проделана огромная работа по созданию точной и надежной монокулярной системы VO. Как показано на рисунке 1, классический трубопровод [1], [2], который обычно состоит из калибровки камеры, обнаружения признаков, соответствия функций (или отслеживания), отклонения отбрасывания (например, RANSAC), оценки движения, оценки масштаба и локальной оптимизации (Bundle Adjustment), была разработана и широко признана золотым правилом. Хотя некоторые современные алгоритмы, основанные на этом конвейере, показали отличную производительность с точки зрения точности и надежности, они обычно жестко закодированы со значительными инженерными усилиями, и каждый модуль в трубопроводе должен быть тщательно спроектирован и доработан до обеспечения производительности. Кроме того, монокулярное VO должно оценивать абсолютную шкалу, используя некоторую дополнительную информацию (например, высоту камеры) или предварительные знания, делая ее склонной к большому дрейфу и более сложной, чем стерео VO.

Deep Learning (DL) в последнее время доминирует над многими задачами компьютерного зрения с многообещающими результатами. К сожалению, для проблемы VO это еще не пришло. Фактически, существует очень ограниченная работа над VO, даже связанная с проблемами 3D-геометрии. Мы исходим из того, что это связано с тем, что большинство существующих DL-архитектур и предварительно подготовленных моделей в основном предназначены для решения проблем распознавания и классификации, которые приводят в движение глубокие сверточные нейронные сети (CNN) для извлечения информации об изображении высокого уровня из изображений. Изучение представления внешнего вида ограничивает VO функционировать только в обученных средах и серьезно затрудняет популяризацию VO к новым сценариям. Вот почему алгоритмы VO сильно зависят от геометрических особенностей, а не от внешних. Между тем, алгоритм VO идеально должен моделировать

динамику движения, исследуя изменения и соединения на последовательности изображений, а не обрабатывая одно изображение. Это означает, что нам нужно последовательное обучение, которому CNN не соответствуют.

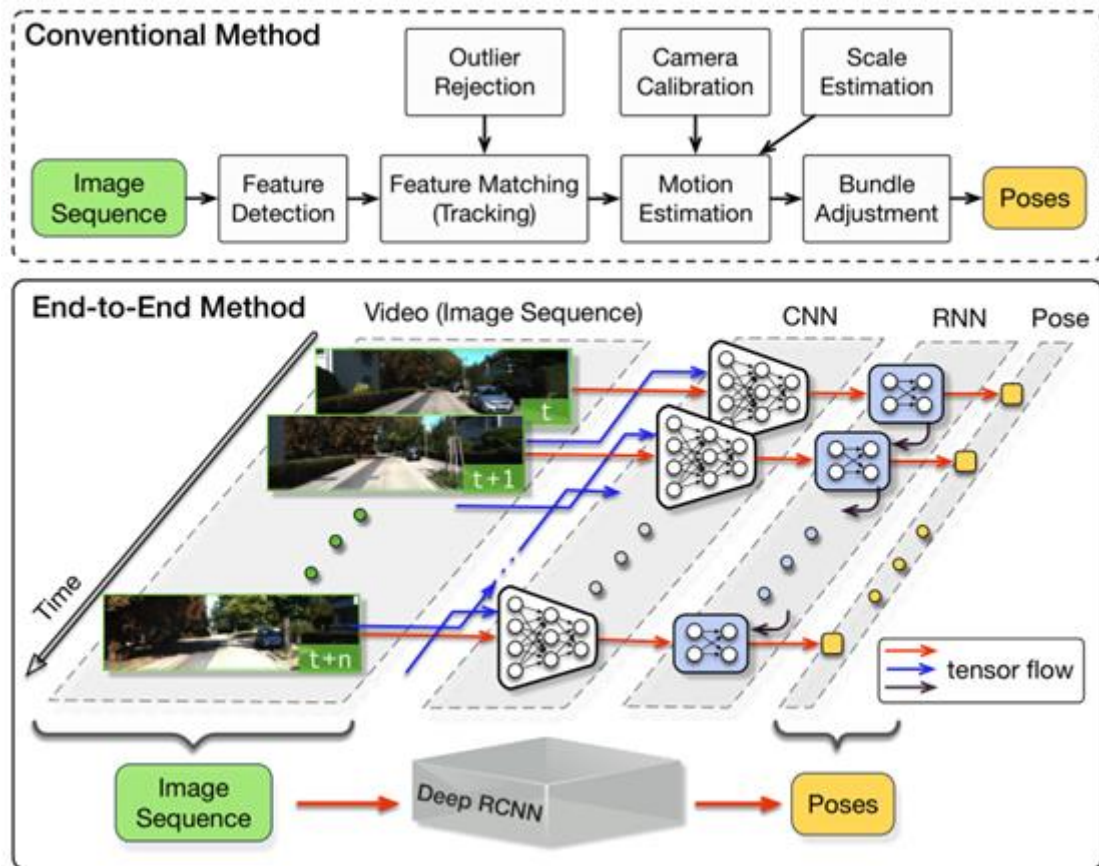


Рис.1. Архитектуры стандартного монокулярного VO и предлагаемого сквозного метода. В предлагаемом способе RCNN берет последовательность изображений RGB (видео) в качестве входных данных и изучает возможности CNN для последовательного моделирования на основе RNN для оценки позы. Кредит изображения камеры: набор данных KITTI [3].

REFERENCES

- [1] D. Scaramuzza and F. Fraundorfer, "Visual odometry: Tutorial," IEEE Robotics & Automation Magazine, vol. 18, no. 4, pp. 80–92, 2011.
- [2] F. Fraundorfer and D. Scaramuzza, "Visual odometry: Part II: Matching, robustness, optimization, and applications," IEEE Robotics & Automation Magazine, vol. 19, no. 2, pp. 78–90, 2012.
- [3] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

