

РАЗРАБОТКА КОМБИНИРОВАННОЙ АРХИТЕКТУРЫ ПОСТРОЕНИЯ КАРТЫ НА ОСНОВЕ SLAM-АЛГОРИТМОВ И ГЛУБОКОГО ОБУЧЕНИЯ

DEVELOPMENT OF A HYBRID MAPPING ARCHITECTURE FOR GROUND VEHICLE AUTOPILOT SYSTEMS

P. Yaroschuk

Summary. This study focuses on the development and evaluation of a hybrid architecture for map construction in ground vehicle autopilot systems. By combining deep learning, mathematical optimization, and advanced localization algorithms, we synthesized an algorithm designed to estimate the 3D position of dynamic objects, map static environments, and simultaneously localize the ego-object (SLAM). The proposed method demonstrates improved computational speed compared to existing approaches while maintaining competitive accuracy. Unlike methods such as SMOKE, it generates richer scene-understanding data, enabling broader application without additional modifications. The algorithm is highly configurable and human-interpretable, allowing adjustable trade-offs between processing speed and mapping precision. It can be integrated as a component of autopilot systems ranging from simple robotic platforms (e.g., warehouse forklifts) to complex consumer vehicle autonomy (e.g., adaptive cruise control). The solution is self-contained, requiring only fine-tuning of a 2D detector to adapt to specific operational conditions.

Keywords: mathematical optimization, deep learning, slam method, dynamic programming, pnp problem.

Ярошук Павел Олегович

Аспирант, ФГБОУ ВО «Воронежский
Государственный Университет»
pay.pray.spray@gmail.com

Аннотация. Данное исследование направлено на разработку и оценку эффективности комбинированной архитектуры для построения карты в автопилоте наземного транспорта. Благодаря комбинации глубокого обучения, математической оптимизации и передовых алгоритмов локализации был синтезирован алгоритм, задача которого состоит в определении 3D положения подвижных объектов, картировании неподвижных и одновременной локализации положения объекта. Метод показал прирост быстродействия по сравнению с другими рассмотренными и сравнительно высокую точность, кроме того, в отличие от таких методов как SMOKE, он генерирует больше полезной информации о наполнении сцены, что позволяет использовать его для более широкого круга задач без дополнительной модификации. Метод обладает высокой степенью конфигурируемости и человеко-интерпретируемости, так что можно настраивать баланс между скоростью обработки/точностью построения карты. Данный алгоритм может быть применен как компонент автопилота как простых роботов, таких как складской погрузчик, так и более сложных систем автопилота и круиз-контроля потребительского транспорта. Он полностью самодостаточен и требует только дообучения 2D детектора для адаптации к требуемым условиям эксплуатации.

Ключевые слова: математическое моделирование, математическая оптимизация, динамическое программирование, алгоритм реального времени, глубокое обучение.

Введение

Развитие автономных систем вождения является ключевым направлением современной робототехники и транспортной индустрии. Для безопасного и эффективного функционирования автопилотов наземного транспорта критически важна способность системы точно определять собственное положение, строить детальную карту окружающей среды и взаимодействовать с динамически изменяющимися объектами. Существующие методы картирования и локализации часто сталкиваются с ограничениями, такими как недостаточная точность в сложных условиях, высокая вычислительная стоимость или ограниченный объем генерируемой информации о сцене.

Современные системы автономного вождения требуют высокоточных и надёжных карт, способных учитывать как статические, так и динамические элементы окружаю-

щей среды. Традиционные методы, основанные на лидарах, обеспечивают высокую точность, но связаны с высокими затратами и ограничениями в условиях плохой видимости. Визуальные методы, использующие камеры и алгоритмы компьютерного зрения, предлагают более экономичное решение, но сталкиваются с проблемами при изменении освещения и погодных условий.

Научная значимость работы заключается в синтезе комбинированной архитектуры, объединяющей глубокое обучение, настраиваемую математическую оптимизацию и передовые алгоритмы локализации. Предлагаемый метод демонстрирует прирост быстродействия и сравнительно высокую точность по сравнению с аналогами, а также генерирует больше полезной информации о сцене, что расширяет его применимость без дополнительной модификации. Высокая степень конфигурируемости и человеко-интерпретируемости метода позволяет гибко настраивать баланс между скоростью

обработки и точностью построения карты. Данный алгоритм является самодостаточным и может быть применен как компонент автопилота для различных типов наземного транспорта, от складских погрузчиков до потребительских автомобилей, требуя лишь дообучения 2D-детектора для адаптации к условиям эксплуатации.

Целью данного исследования является разработка и оценка эффективности комбинированной архитектуры для построения карты в автопилоте наземного транспорта, способной одновременно определять 3D-положение подвижных объектов, картировать неподвижные поверхности и выполнять одновременную локализацию положения транспортного средства.

Для достижения поставленной цели были сформулированы следующие задачи исследования:

- Синтезировать алгоритм, интегрирующий методы 2D-детекции и сегментации, ORB-SLAM для локализации и картирования неподвижных поверхностей, а также метод 3D-эстимации движущихся объектов.
- Оценить быстродействие и точность разработанного алгоритма в сравнении с существующими подходами.
- Продемонстрировать способность алгоритма генерировать обширную и полезную информацию о сцене для широкого спектра задач автопилотирования.
- Исследовать степень конфигурируемости метода и его человеко-интерпретируемость для настройки баланса между скоростью обработки и точностью построения карты.
- Проверку провести экспериментальную эффективность разработанной архитектуры на датасете KITTI.

Литературный обзор

Монокулярная 3D-детекция объектов в контексте автономного транспорта продолжает оставаться предметом активных исследований из-за её низкой стоимости и простоты интеграции, особенно по сравнению со стерео- и LiDAR-системами. Современные подходы к решению этой задачи, как правило, основываются на глубоком обучении и используют либо прямую регрессию параметров 3D-боксов, либо двухэтапные схемы с разделением задач детекции и оценки позы.

Одним из наиболее известных подходов является MonoDLE, предлагающий прямую регрессию параметров 3D-бокса и положения в пространстве на основе изображений [1]. Другой интересный метод, MonoFlex, объединяет многоуровневые признаки и использует гибкую систему потерь для улучшения точности предсказаний [2]. Тем не менее, большинство методов подоб-

ного типа страдают от высокой вычислительной нагрузки и требуют больших объемов обучающих данных.

Работы, направленные на повышение эффективности, включают FCOS3D [3], который базируется на безъякорном подходе и показывает высокую точность при разумной скорости обработки. Тем не менее, такие методы редко интегрируются в реальные SLAM-системы или учитывают геометрию сцены. В то же время, более геометрически ориентированные подходы, такие как MonoPair [4], используют отношения между объектами и их симметрией, но демонстрируют снижение производительности на плотных и неоднородных сценах.

Некоторые недавние исследования показывают, что использование PnP-подобных методов для оценки позы объекта на основе выделенных 2D-точек может повысить как интерпретируемость, так и устойчивость алгоритма к обобщению на новых сценах [5]. Это хорошо сочетается с идеей биологически вдохновлённой обработки, где отдельные признаки объектов (например, передняя часть автомобиля) играют роль якорей для локализации.

Интеграция методов SLAM и 3D-детекции также активно развивается. Например, в работе [6] описано объединение ORB-SLAM с сетевыми методами, что позволяет одновременно проводить построение карты и отслеживание объектов. Однако их подход ориентирован на статичные сцены и не решает проблему ориентации движущихся объектов.

Таким образом, текущая работа отличается от существующих решений тем, что она:

- использует эффективную монокулярную 2D-детекцию (например, YOLO) с выделением характерных признаков;
- сочетает геометрический вывод с PnP-механикой для оценки глубины;
- включает SLAM-подсистему для оценки плоскости дороги и положения наблюдателя;
- направлена на реальное применение в условиях ограниченных ресурсов (вплоть до работы на мобильных GPU).

Такое комбинирование обеспечивает баланс между вычислительной эффективностью, точностью и универсальностью в разнообразных дорожных условиях.

Материалы и методы

Для построения метода за основу брались биологические принципы визуального восприятия информации, подробнее в работе [7]. Полученная архитектура системы включает две параллельные подсистемы: локализацию и картирование дороги с использованием ORB-

SLAM, а также оценку положения объектов. Последняя состоит из следующих этапов:

- 2D-детекция объектов и их характерных признаков,
- сопоставление признаков с априорными 3D-точками модели,
- построение 3D-луча через изображённую 2D-точку и определение положения объекта на плоскости дороги (из SLAM),
- оптимизация ориентации объекта по проекции 3D-модели (рис. 1).

Входные данные:

- Монокулярное RGB-изображение
- Калибровочная матрица камеры
- Набор априорных 3D-моделей объектов

Выходные данные:

- Спллайн дороги и доступная для движения зона
- 3D-координаты центра объектов в системе камеры
- Угол ориентации объектов (вдоль оси Y)

Локализация и построение карты статических объектов осуществляется алгоритмом SLAM, после чего происходит интерполяция полученных 3Д-точек. На основе полученных данных в дальнейшем осуществляется поиск глубины двумерного якоря.

Детекция

Особенностью метода является использование характерных 2D-признаков объектов для точной привязки к их 3D-модели. Признаки могут выделяться либо основной нейросетью, либо отдельной специализированной моделью. Сопоставление признаков позволяет выбрать 2D-якорь, соответствующий известной точке на канонической 3D-модели (точка сопоставляется по априорному правилу, заданному на этапе настройки для конкретного объекта). Пример детекции и сегментации на рис. 2 — обнаруживаются только автомобили (только для них составлялся датасет).

Поиск якоря

Для определения глубины строится 3D-луч, исходящий из камеры и проходящий через 2D-якорь. Точка пересечения луча с плоскостью дороги (определённой ORB-SLAM) используется как оценка положения объекта.

Оптимизация ориентации

Зная положение объекта и его 3D-модель, оптимизируется ориентация вокруг вертикальной оси. Это достигается путём минимизации функции ошибки между проекцией повернутой 3D-модели и наблюдаемой 2D-рамкой. Используется итерационный метод (напри-

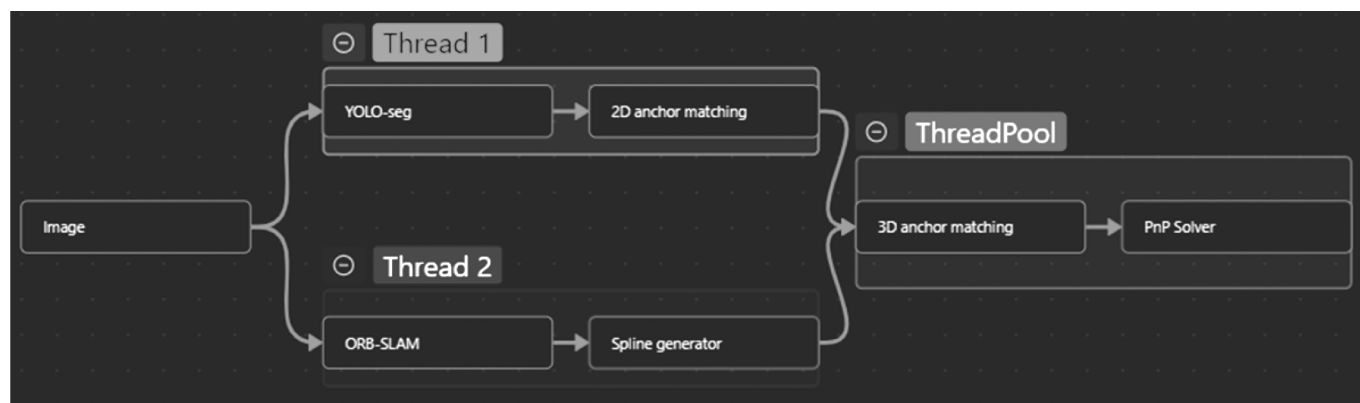


Рис. 1. Схема алгоритма

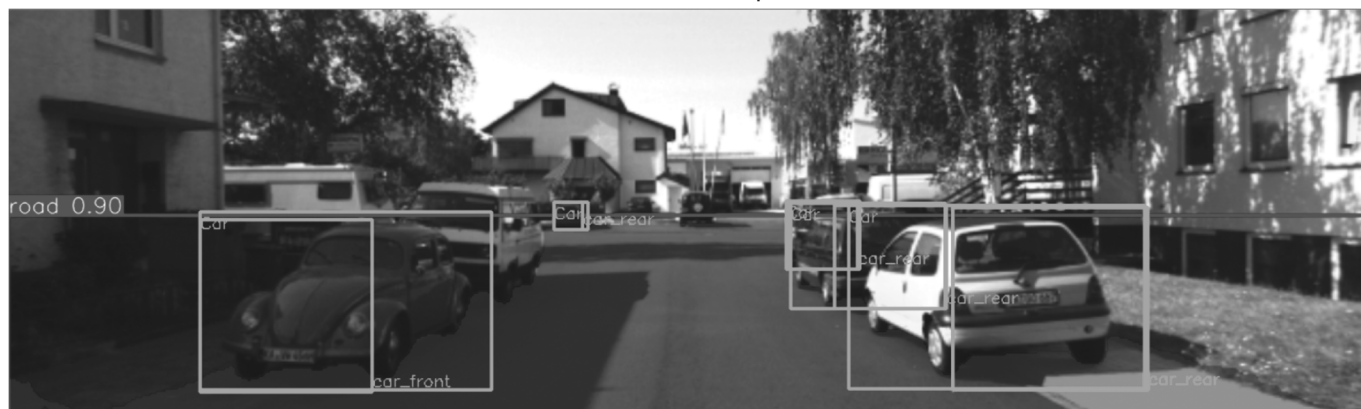


Рис. 2. Результат детекции и сегментации

мер, Гаусса-Ньютона), где на каждом шаге происходит численная оптимизация положения бокса на основе информации о боксе объекта.

Результаты

Для всесторонней оценки эффективности разработанного монокулярного алгоритма были проведены экспериментальные исследования на широко используемом в задачах автономного вождения наборе данных KITTI. В частности, использовалась подвыборка для 3D-детекции объектов, включающая синхронизированные данные с монокулярных камер и лидара, а также точную разметку трехмерных ограничивающих рамок. Основное внимание уделялось классу «автомобиль» как наиболее показательному для данной задачи.

Для количественной оценки производительности предложенной системы использовались стандартные метрики для задачи 3D-детекции объектов на датасете KITTI:

- Точность 3D-локализации (BEV IoU): Оценивается путем сравнения проекций 3D-боксов на горизонтальную плоскость (вид сверху), измеряя Intersection over Union (IoU) в Bird's Eye View.
- Точность оценки ориентации: Измеряется с помощью метрики Average Orientation Similarity (AOS) из протокола KITTI, которая учитывает как точность 2D-детекции, так и ошибку ориентации.

Помимо качества, оценивалась скорость обработки данных. Проводилось сравнение полученных показателей с другими монокулярными методами.

Оценка ошибки для трехмерной ориентации бокса по оси рыскания показала, что алгоритм достаточно точно отражает реальный поворот автомобиля. Сравнение с другими передовыми методами (Таблица 1) демонстрирует конкурентоспособные показатели точности ориентации.

Таблица 1.

Сравнение методов по точности 3D-ориентации

Метод	Easy AOS (%)	Moderate AOS (%)	Hard AOS (%)
Deep3DBox [8]	92.90	88.75	76.76
SMOKE [9]	92.94	87.02	77.12
Mono3D [10]	91.01	86.62	76.84
Твой метод	82.49	71.27	64.38

Для достижения высокой точности 3D-позиционирования, разработанный метод эффективно использует значения глубины в точках установки якоря, полученные из ORB-SLAM. Точность алгоритма выше или сравнима с передовыми методами, при этом стабильность

детекции выше. Значения Hard, Moderate, Easy схожи, при этом на средних случаях точность даже выше, чем на простых (Таблица 2).

Таблица 2.

Сравнение показателей точности совпадения теней 3D-боксов

Метод	Easy BEV AP (%)	Moderate BEV AP (%)	Hard BEV AP (%)
SMOKE	21.08	15.13	12.91
OPA-3D	17.05	24.60	14.25
DID-M3D	16.29	24.40	13.75
Твой метод	17.02	19.27	19.09

При этом, как видно из Таблицы 3, разработанный метод значительно превосходит аналоги по скорости обработки данных, достигая 25–33 FPS на мобильной видеокарте NVIDIA GTX 1650. Это преимущество в производительности делает его пригодным для интеграции в менее мощные и более энергоэффективные системы реального времени.

Таблица 3.

Сравнение производительности методов

Метод	Среднее FPS на GTX 1650
Mono3D	1 FPS
Deep3DBox	2–3 FPS
RTM3D	5 FPS
SMOKE	13 FPS
Ваш метод	20–25 FPS

На рис. 3 можно видеть пример работы алгоритма

Обсуждение

Метод не только демонстрирует хорошие показатели быстродействия, но и позволяет получать богатый набор данных для построения карты. Совмещение сегментации дороги с построением трехмерных точек позволяет алгоритму обрабатывать сложные ситуации, когда детекция объектов напрямую не срабатывает. Например, если на дороге лежит объект, не обнаруживаемый детекцией, он будет одновременно сегментирован YOLO-seg и локализован SLAM, поскольку будет выделяться на общем фоне. Это позволит, ко всему прочему, избежать ситуации, недавно произошедшей с автономным автомобилем Tesla, когда он проигнорировал сетчатое ограждение аэропорта и проехал насквозь. В случае с предложенным методом сетка создала бы много трехмерных опорных точек, которые можно было бы применить для построения карты.

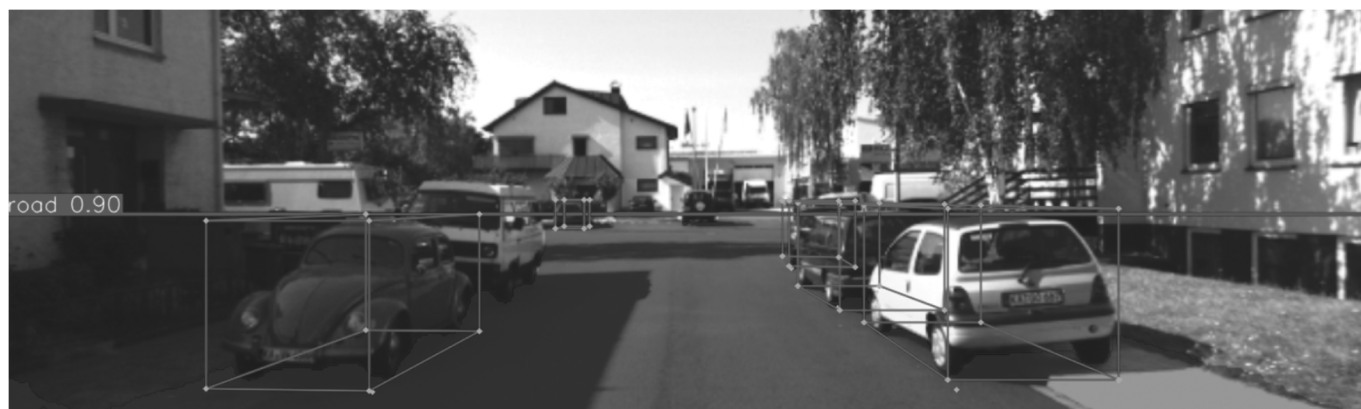


Рис. 3. Пример работы алгоритма

Метод позволяет получать «сырые» результаты до полного завершения работы, что может быть полезно в критических ситуациях, когда доли секунды определяют исход аварии и количество потенциальных жертв. Точность и скорость работы может быть дополнительно задана в настройках, а не только лишь зависит от весов и архитектуры обученной нейронной сети.

Кроме того, точность может быть улучшена путем дополнительной классификации автомобилей по габаритам, что позволит сопоставлять боксы более точно.

Заключение

В результате исследования был разработан высокопроизводительный алгоритм, пригодный для постро-

ения карты и локализации положения объекта в пространстве. Метод по производительности превосходит другие методы определения трехмерных объектов, кроме того, имеет преимущества в виде одновременной локализации и гибкого построения карты. Необходимо провести дальнейшие исследования по интеграции сегментации (например, для интеграции в карту ям и уточнения положения пешеходов по пробелам сегментации дороги) и улучшению производительности за счёт детекции большего количества видов габаритов автомобилей.

ЛИТЕРАТУРА

1. Ma X., Wang Z., Li H., Wang Y. MonoDLE: Monocular 3D Object Detection via Local Depth Estimation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. С. 3126–3135.
2. Liu Z., Wang D., Shen Y., Wang Y., Lu H. MonoFlex: Multi-task Regression for Monocular 3D Object Detection // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. С. 9130–9140.
3. Wang Z., Wang Y., Zhan W., Tomizuka M. FCOS3D: Fully Convolutional One-Stage Monocular 3D Object Detection // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021. С. 913–922.
4. Chen Z., Huang X. MonoPair: Monocular 3D Object Detection Using Pairwise Spatial Relationships // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. С. 12093–12102.
5. Simonelli A., Buló S.R., Porzi L., López-Antequera M., Kotschieder P. Disentangling Monocular 3D Object Detection // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. С. 1991–2000.
6. Sun J., Zhao Y., Zhang L., Yang M., Lin W. SLAM-aware Monocular 3D Object Detection for Dynamic Environments // Advances in Neural Information Processing Systems (NeurIPS). 2022. № 35. С. 13760–13772.
7. Ярошук П.О. Метод анализа трехмерного положения объектов на двумерном изображении // Актуальные проблемы прикладной математики, информатики и механики: сборник трудов Международной научной конференции. — Воронеж, 2024. — С. 812–817.
8. Mousavian A., Anguelov D., Flynn J., Kosecka J. Deep3DBox: 3D Bounding Box Estimation Using Deep Learning and Geometry // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. С. 7074–7082.
9. Liu Z., Wang Y., Wu S., Su H., Zhu S., Bao H. SMOKE: Single-Stage Monocular 3D Object Detection via Keypoint Estimation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2020.
10. Chen X., Kundu K., Zhu Y., Ma H., Fidler S., Urtasun R. Mono3D: Monocular 3D Object Detection for Autonomous Driving // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. С. 2147–2156.

© Ярошук Павел Олегович (pay.pray.spray@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»