

Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Высшая школа программной инженерии

Работа допущена к защите

Директор ВШПИ

_____ П.Д.Дробинцев

« ____ » _____ 2019г

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

БАКАЛАВРА

**Применение методов машинного обучения для решения задачи
обнаружения препятствий с помощью стереозрения**

по направлению 09.03.04«Программная инженерия»

по образовательной программе

09.03.04_0 1«Технологии разработки и сопровождения
качественного программного продукта»

Выполнил

студент гр. 43504/1

Дейлид И.А.

Руководитель

Профессор

Молодяков С.А.

Санкт-Петербург

2019

Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Высшая школа программной инженерии

Утверждаю
Директор ВШПИ
_____ П.Д.Дробинцев
« ____ » _____ 2019г

З А Д А Н И Е

на выпускную квалификационную работу бакалавра

Студенту группы 43504/1 Дейлид Иван Анатольевич

1. Тема проекта (работы) Применение методов машинного
обучения для решения задачи обнаружения препятствий с
помощью стереозрения

(утверждена распоряжением по ИКНТ
от _____ № _____)

2. Срок сдачи студентом оконченого проекта (работы)

3. Исходные данные к проекту (работе)

Централизованная система обмена данных, видеопоток

сетевых камер

видеопоток камеры машинного зрения

симулятор датчиков транспортных средств

4. Содержание расчетно-пояснительной записки (перечень подлежащих разработке вопросов)

Изучение предметной области

Организация аппаратной базы

Анализ методов для решения задачи

Разработка архитектуры ПО

Программная реализация

Анализ полученных результатов

5. Перечень графического материала (с точным указанием обязательных чертежей)

6. Перечень графического материала (с точным указанием обязательных чертежей)

7. Дата выдачи задания _____

Руководитель Молодяков С.А. _____

Задание принял к исполнению _____

_____ (дата)

(подпись студента)

РЕФЕРАТ

Отчёт содержит 60 страниц, 19 рисунков.

Ключевые слова: стереозрение, машинное обучение, беспилотный транспорт, оптический поток.

В данной работе освещается вопрос обнаружения препятствий в беспилотном транспорте. Акцент сделан на использовании алгоритмов стереозрения. Будут описаны достоинства и недостатки данных подходов. Для компенсации недостатков предлагается использовать ряд подходов, изменяющий архитектуру блока обнаружения препятствий, и включающий в себя применение методов классического и глубокого машинного обучения, с целью повышения робастности системы и частоты получения данных

ABSTRACT

The report contains 60 pages, 19 figures.

Keywords: stereo vision, machine learning, unmanned transport, optical flow.

This paper describes the issue of detecting obstacles in self-driving vehicles. Emphasis is placed on the use of stereo vision algorithms. The advantages and disadvantages of these approaches will be described. To compensate for the disadvantages, it is proposed to use some approaches that change the architecture of the obstacle detection unit and include the use of classical and deep machine learning methods in order to increase the robustness of the system and the frequency of data acquisition

СОДЕРЖАНИЕ

1.	Введение.....	11
2.	Обзор решений	14
2.1.	Датчики для беспилотного транспорта	14
2.1.1.	Камеры	14
2.1.1.1.	Особенности	14
	Работы в видимом диапазоне	14
	Теоретическая достаточность	14
	Дешевизна и распространенность	15
2.1.1.2.	Категории камер	15
	Видимый спектр	15
	Тепловизоры	15
	ИК	16
	SWIR.....	16
	TOF	16
	Structured light.....	17
2.1.2.	Лидары	17
2.1.3.	Радары	17
2.2.	Методы получения данных о трехмерной природе среды 18	
2.2.1.	Монокюлярное зрение	19
2.2.1.1.	Structure from motion	19
2.2.1.2.	Deep learning	20
2.2.2.	Стереозрение	21

2.2.2.1.	Алгоритмы стереосопоставления.....	21
2.2.2.2.	Deep learning.....	21
2.3.	Зависимость требований к системе от скорости и массы транспортного средства.....	21
2.4.	Компании и их подходы к решению	23
2.4.1.	Tesla.....	23
2.4.2.	AiMotive	24
2.4.3.	НИИАС	24
2.5.	Проблемы при использования стереокамер	26
2.5.1.	Сложность алгоритмов	26
2.5.2.	Снижение частоты.....	26
2.5.3.	Сопровождение объектов.....	27
2.5.4.	Шумы источников данных.....	27
2.6.	Постановка цели	28
3.	Концепция решения.....	29
3.1.	Пример архитектуры, использующей комбинирование методов машинного обучения со стереозрением	30
3.2.	Используемые подходы	31
3.2.1.	Переход к карте глубины как основному источнику данных	31
3.2.2.	Переход к задаче обнаружения проходимого пути по карте глубины	31
3.2.3.	Переход к задаче оптического потока в 2D	33
3.2.4.	Использование синтетического датасета для анализа проходимости по карте глубины	35

3.2.5.	Получение объектной сегментации из семантической сегментации и карты глубины	36
3.2.5.1.	Подход ассоциации данных детектора с облаком точек	37
3.2.5.2.	Подход ассоциации данных сегментации объектов с облаком точек	37
3.2.5.3.	Подход с классификаций препятствий, найденных по реконструкции	38
	Поиск препятствий по данным от стереозрения	38
	Подход к предположению о поверхности.....	38
	Подход анализа карты диспаратности.....	39
	Подход анализа облака точек.....	40
4.	Реализация	43
4.1.	Архитектура.....	43
4.1.1.	Алгоритм стереозрения	44
4.1.2.	Семантическая сегментация.....	46
4.1.3.	Коллектор кадров и оптический поток.....	46
4.1.4.	Кластеризация	46
4.1.5.	Сегментация проходимости	51
5.	Результаты работы алгоритмов.....	56
6.	Заключение	58
7.	Список литературы	59

1. ВВЕДЕНИЕ

Одной из наиболее быстро развивающихся технологий за последние годы стала система автопилотирования для различных видов транспортных средств, начиная с небольших роботизированных платформ для доставки мелких грузов, заканчивая грузовыми автомобилями и поездами. Современные достижения в машинном обучении совершили скачок в задачах распознавания образов, результаты развития аппаратного обеспечения позволяют параллельно вычислять большое количество примитивных математических операций используя ядра графических процессоров, а теория классического компьютерного развития позволяет извлечь из датчиков видимого спектра разнообразную информацию более высокого уровня абстракции, чем цветовая информация.

Технология беспилотного транспорта декомпозируется на несколько частей, одной из которых является восприятие транспортным средством окружающей среды, сенсорики. Среди наиболее часто используемых сенсоров можно назвать камеры, лидары, радары, ультразвуковые сканеры, инерциальные системы, гиростабилизаторы, системы спутниковой навигации. Но наибольшее количество информации можно извлечь именно из камер. Этот вид датчиков в большей степени обладает свойством антропоморфности и передаёт информацию, достаточную для принятия человеком решения о дальнейшем управлении. При работе программного обеспечения с потоком данных от видекамеры каждый кадр представляется в виде матрицы, способной быть

обработанной различными алгоритмами компьютерного зрения.

Количество различных алгоритмов компьютерного зрения на текущий момент велико, поделить их можно разными способами на различные категории. С точки зрения задач, компьютерное зрение делится на фильтрацию данных, геометрические преобразования, поиск и анализ характерных особенностей на кадрах, анализ видеопотока, классификацию изображений, детекцию объектов, улучшение качества и многие другие.

Одним из разделов классического компьютерного зрения, применяемым в отрасли беспилотных транспортных средств, является стереозрение. Оно позволяет получать огромное количество информации об окружающем пространстве роботизированных устройств. Так, система из двух камер позволяет дополнить информацию видимого спектра данными о трехмерном представлении объектов, за счет сопоставления отображения объектов на разных кадрах и вычисления глубины до них по разнице координат проекций на матрицах двух камер с помощью метода триангуляции. На протяжении многих лет появлялись новые, более совершенные алгоритмы стереозрения. Во время нового расцвета машинного обучения, стали появляться нейронные сети^[1], решающие задачу стерео сопоставления. Современные методы глубокого машинного обучения в области стереозрения, как правило, предоставляют решения в виде “черного ящика” для получения карты

диспаратности, но сама карта диспаратности при этом не является конечным результатом цикла работы всего программного обеспечения, а точность получаемых данных напрямую зависит от обучающей выборки.

Предлагается подход, при котором методы машинного обучения в совокупности с методами классического компьютерного зрения применяются в виде каскада операций, выполняющих преобразование одного типа входных данных, представляющих из себя ценную информацию, способную быть проанализированной человеком, в другой, решая конечную задачу обнаружения препятствий на пути, поскольку подход обмена данными между методами может существенно улучшить качество отдельных элементов системы^[2]. Кроме того, задача обнаружения препятствий решается как в своем классическом варианте, когда за цикл работы алгоритмов мы получаем данные и информацию о различных объектах в зоне интересов устройства, так и в инвертированном варианте, когда вместо определения препятствий происходит анализ проходимого пути. Будет показано, как такой подход позволит добиться повышения точности, частоты и количества получаемых полезных данных.

2. ОБЗОР РЕШЕНИЙ

Говоря об обнаружения препятствий, стоит начать с рассмотрения датчиков, используемых для решения данной задачи, чтобы подчеркнуть преимущества, особенности и недостатки работы с камерами.

2.1. Датчики для беспилотного транспорта

Для получения информации, пригодной для принятия решений об управляющем воздействии на транспортное средство, используется ряд датчиков, различных по принципу работы.

2.1.1. Камеры

Камеры являются одним из наиболее распространенных датчиков, используемых в беспилотном транспорте, по нескольким причинам.

2.1.1.1. Особенности

Работы в видимом диапазоне

Камера является единственным датчиком, работающем в диапазоне видимого электромагнитного спектра, принцип работы которого аналогичен принципу работы человеческого глаза. Это позволяет ей решать классификационные задачи, определяя, к примеру, разновидность дорожных знаков, разметку, сигналы светофоров, тип динамического препятствия.

Теоретическая достаточность

Отличительной особенностью камеры является ее теоретическая достаточность для управления

транспортным средством. Иными словами, поскольку человек способен удалённо управлять транспортным средством, используя лишь визуальную информацию, можно сделать вывод, что информации, получаемой с камер, достаточно для выбора правильных действий управления.

Дешевизна и распространённость

Поскольку камеры являются массовым и устоявшимся на рынке устройством, компоненты камер со временем удешевлялись, а их использование в различных прикладных отраслях дало толчок улучшению различных показателей камер, таких как частота работы, минимальное необходимое количество света, разрешающая способность. По итогу мы имеем большое разнообразие устройств с различными характеристиками, которые мы можем выбирать исходя из конкретной задачи.

2.1.1.2. Категории камер

В свою очередь камеры можно разделить на категории по диапазону частот электромагнитных волн, с которыми они работают

Видимый спектр

Наиболее распространённый вид камер, с наибольшим разнообразием матриц. Позволяет получить информацию, воспринимаемую человеческим глазом, что делает этот датчик незаменимым.

Тепловизоры

Данный вид камер работает в среднем инфракрасном диапазоне и позволяет узнать

температуру объектов, что является, в большинстве случаев, хорошим признаком для определения людей и животных в зоне видимости. Основными недостатками этой категории камер является высокая стоимость и низкая разрешающая способность, что уменьшает пользу работы на дальних дистанциях.

ИК

Наиболее частым примером этой категории можно считать камеры видеонаблюдения для охраняемых зон, работающих в условиях плохой освещенности. Как правило данный вид камер захватывает часть видимого спектра излучения и ближний инфракрасный. За счёт этого, с помощью инфракрасной подсветки, данные камеры могут получить одноканальное изображение среды в условиях освещенности (ик), без помех для человека.

SWIR

Применяется для наблюдения за объектами в круглосуточном режиме при низких уровнях освещенности и в экстремальных климатических условиях в составе оптико-электронных систем на стационарных и подвижных комплексах. Подобные камеры имеют очень ограниченное распространение в гражданском секторе и высокую стоимость, из-за чего их применение становится затруднительным.

TOF

Камеры дальнего радиуса действия, в которой используются методы времени пролета для определения расстояния между камерой и объектом для каждой точки изображения путем измерения времени

прохождения сигнала искусственного света, излучаемого лазером или светодиодом, в оба конца. Имеет ограниченные для задачи автопилотирования характеристики, так как излучение, необходимое для работы на дальних дистанциях, опасно для человека.

Structured light

Камера структурированного света работают, как правило, частично в инфракрасном диапазоне. Специальный проектор отображает на видимой сцене определённый паттерн. В зависимости от деформации этого паттерна принимаются решения о том, какой трёхмерной координате соответствует данная точка кадра. Данный сканер имеет очень хорошую плотность реконструкции, однако, из-за ограниченной мощности проектора, датчик применяется лишь на ближней дистанции.

2.1.2. Лидары

Лидары являются одним из наиболее популярных датчиков на рынке, поскольку они с высокой частотой дают облако точек, с большой точностью измерений и практически отсутствием шумов.

2.1.3. Радары

Радар работает по принципу анализа отраженного радиосигнала. Из плюсов можно выделить работы практически в любых погодных условиях, высокую дальность работы, возможность анализа скорости и направления движения объектов. Главным

минусом является сложность обработки данных и практическая невозможность применения в условиях сильной радиозашумлённости, как, к примеру, в условиях применения на железных дорогах.

Как видно, камеры видимого спектра являются тем видом датчиков, который способен давать представление о среде на дальних дистанциях, сохраняя при этом приемлемую для массового применения.

Далее речь будет идти о системах, работающих с камерами в видимом диапазоне, поскольку данная система датчиков наиболее гибкая, масштабируемая и переносимая. Кроме того, данная система камер необходима и достаточна для решения классификационных задач (обнаружение и распознавание классов символов и дорожных знаков), определения расстояния до препятствий (мультизрение^[3] и т. п.) и выбора дальнейшего действия транспортного средства (удалённое управление ТС человеком).

2.2. Методы получения данных о трехмерной природе среды

Для обнаружения препятствий используются различные методы на основе трехмерного представления данных.

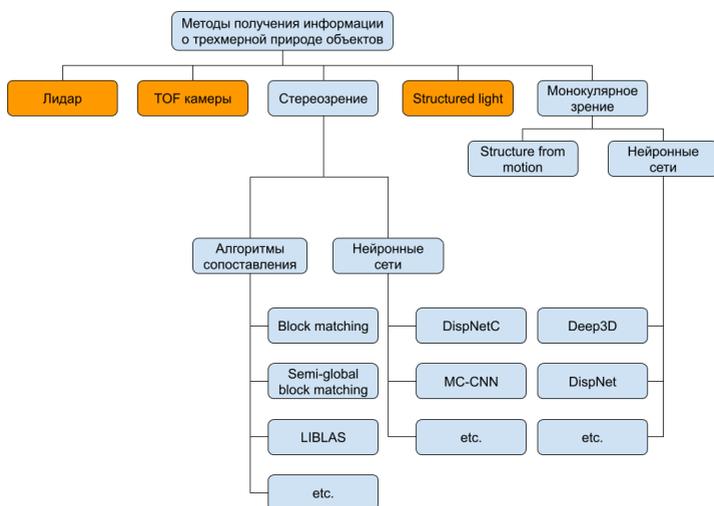


Рисунок 1. Различные подходы к получению информации о препятствиях

На приведённой схеме наибольший интерес представляют алгоритмы, работающие с камерами в видимом диапазоне частот. Рассмотрим их поподробнее.

2.2.1. Монокулярное зрение

2.2.1.1. Structure from motion

Алгоритмы класса structure from motion решают задачи реконструкции пространства по серии изображений, где объект или среда попадают в объектив камеры под разным углом. Как правило, алгоритмы данного класса в результате дают облако точек невысокой плотности (относительно разрешения кадра). Плотность точек, пожертвовав точностью, можно

увеличить с помощью методов оптического потока, вместо сопоставления признаков. Координаты определяются за счет триангуляции характерных признаков между кадрами. Главным недостатком данного алгоритма является невозможность реконструкции движущихся объектов. Поэтому имеет смысл рассматривать это семейство алгоритмов как вспомогательный источник информации, метод диагностирующий проблемы с калибровочными параметрами камер, уточнение показателей датчиков одометрии и запасной источник данных в случае сбоя камеры в системе стереозрения.

2.2.1.2. Deep learning

Методы реконструкции изображения по одному кадру, основанные на глубоком обучении, появились сравнительно недавно. Принцип работы этих сетей строится на умении сети запоминать паттерны объектов и их реконструкцию в зависимости от масштаба и участка изображения. Поскольку данные алгоритмы являются разновидностью алгоритмов черного ящика, а точность сильно зависит от датасета, применение данных методов как основного источника информации не отличается надежностью, однако можно использовать эту информацию как дополнительный канал данных.

2.2.2. Стереозрение

2.2.2.1. Алгоритмы стереосопоставления

Классический вид алгоритмов, использующихся в стереозрении. Принцип работы алгоритмов основывается на поиске одинаковых паттернов на двух кадрах и вычислению расстояния между проекциями. Являются наиболее надежными из имеющихся алгоритмов, однако имеют ряд недостатков: зашумленность данных, требования к жесткой синхронизации при работе с динамическими объектами, подбор параметров фильтрации, проблемы с мало текстурированными объектами.

2.2.2.2. Deep learning

Глубокие нейронные сети, позволяющие получать данные о глубине, работают по разнообразным принципам. Обычно в архитектурах данных решений используются сиамские сети, а сама задача сводит к семантической сегментации, где класс пикселя является значением карты диспаратности в соответствующей координате. Данный подход требует значительно больше вычислительных ресурсов при сопоставимых результатах.

2.3. Зависимость требований к системе от скорости и массы транспортного средства

Определяя требования к сенсорам и работе системы беспилотного транспортного средства одним из

главных критериев можно назвать тормозной путь, который допускается при экстренном торможении. В зависимости от данной характеристики меняется как необходимая дальность видимой области, так и частота обработки данных от сенсоров.

Расчет тормозного пути колесного транспортного средства проводится по следующей формуле^[4]:

$$t_o = t_p + t_c + 0,5 \cdot t_H + \frac{V_0 \cdot K_{\text{э}}}{g \cdot \varphi_{\chi}}, \text{ с}$$

где t_p - время реакции водителя (с) в нашем случае данный параметр принимает значение времени реакции системы;

t_c - время срабатывания тормозного привода (с);

t_H - время нарастания тормозных сил, с;

V_0 - начальная скорость торможения (м/с);

g - ускорение свободного падения (м/с²);

φ_{χ} - коэффициент продольного сцепления колёс с дорожным покрытием;

$K_{\text{э}}$ - коэффициент эффективности торможения.

$K_{\text{э}}$ увеличивается при увеличении нагрузочной массы транспортного средства

Время нарастания тормозных сил рассчитывается по формуле:

$$t_H = \frac{G \cdot (b + \varphi \cdot h_{\text{ц}})}{K_1 \cdot L} \phi \quad \chi, \text{ с};$$

где G - вес транспортного средства с данной нагрузкой, Н;

b – расстояние от центра тяжести до заднего моста, м;

$h_{\text{ц}}$ - высота центра тяжести от поверхности дороги, м;

K_1 - скорость нарастания тормозных сил на колесах переднего моста

реакция тормозной системы

L - колесная база (м).

2.4. Компании и их подходы к решению

Рассмотрим подходы различных компаний к использованию камер как основного датчика.

2.4.1. Tesla

Тесла является одной из наиболее крупных автомобильных компаний, которая отказалась от датчиков лидарного типа в пользу комбинации камер и радара. За счёт визуальной информации и её сопоставления с данными от радара.

2.4.2. AiMotive

Ещё одна компания, создающая технологии беспилотных транспортных средств с акцентом на камеры является AiMotive. В статье журнала рассказывается о том, как использование камер позволяет решить задачи беспилотного управления, а комбинирование методов классического зрения с машинным обучением позволяет перейти к более точным результатам^[5].

2.4.3. НИИАС

Российская научно-исследовательский институт автоматизации и связи разрабатывает ряд беспилотных локомотивов. Условия эксплуатации данных транспортных средств определяет также необходимость работы с большой зоной интересов, поскольку тормозной путь локомотива значительно превышает тормозной путь легковых автомобилей. Для этого решение включает в себя несколько систем камер, составляющих стереопары с различным фокусным расстоянием, что обеспечивает покрытие и перекрытие зон на необходимой дальности.



Рисунок 2. Фронтальный вид беспилотного локомотива

[1]

2.5. Проблемы при использования стереокамер

Несмотря на все преимущества работы с камерами, работа с ними, в частности со стереозрением, имеет свои недостатки:

2.5.1. Сложность алгоритмов

Говоря об алгоритмах стереозрения стоит упомянуть тот факт, что большинство из них требуют большого количества вычислений и проходов по матрицам, соответствующим кадрам. Таким образом приходится выбирать между алгоритмами, дающими точный результат за продолжительное количество времени, либо быстрыми алгоритмами, зачастую предоставляющими склонный к зашумлению результат.

2.5.2. Снижение частоты

Из предыдущего пункта вытекает недостаток связанный с низкой частотой работы. Поскольку время обработки двух синхронизированных кадров превышает время получения кадра, система начинает получать данные со сниженной частотой. Частично данная проблема решается добавлением графических ускорителей или использованием программируемой логики. Зачастую второй вариант может позволить практически перейти в режим реального времени, но при этом значительно увеличивается стоимость аппаратного обеспечения и сложность его поддержки.

2.5.3. Сопровождение объектов

Низкая частота получения данных и временная задержка между получением оригинальных данных (кадр с камеры) и результатом работы алгоритмов (данными о препятствиях) сказывается на алгоритмах сопровождения объектов. Поскольку мы всегда получаем несколько устаревшие данные, беспилотный транспорт требует отдельного модуля, отвечающего за выравнивание данных во времени и пространстве и предсказании дальнейших координат объектов, чтобы мы могли основывать свои предположения на текущем состоянии окружающей среды. Очевидно, чем больше задержка между получением данных и их обработкой, тем сложнее нам предсказывать точное местоположение объектов, из-за чего нам приходится вводить вероятностную модель для описания объектов на карте, достоверность данных от которой соразмерно снижается. Аналогичный негативный эффект оказывает низкая частота работы алгоритмов.

2.5.4. Шумы источников данных

Если рассмотреть задачу стереозрения в обратном порядке, то местоположение объекта определяется по точкам объекта, полученным во время реконструкции, которая в свою очередь высчитывается из карты диспаратности. Карта диспаратности представляет из себя матрицу значений расстояний между проекциями точек на матрицы объективов в системе координат кадра. Низкая текстурированность,

проблемы с освещением, собственные шумы камеры могут ухудшить качество карты диспаратности, из-за чего в процессе реконструкции можно получить массу точек в пространстве, не соответствующую действительности. Большинство алгоритмов решают эту задачу с помощью различных штрафных функций и постфильтрации, однако это требует определенного подбора параметров и, зачастую, не лучшим образом отражается на точности реконструкции объектов.

2.6. Постановка цели

Из описанных ранее особенностей, недостатках и текущем состоянии подходов, связанных с использованием камер и стереозрения, вытекают цели и задачи, решение которых можно предоставить различным алгоритмам машинного обучения, что будет более подробно освещено ниже.

Цель: обеспечить увеличение частоты и устойчивость к зашумлению данных, в задаче обнаружения препятствий с помощью камер.

3. КОНЦЕПЦИЯ РЕШЕНИЯ

Данное решение предлагает систему на основе нескольких камер и программного обеспечения, позволяющих получать данные о расположении и классе различных объектов, позволяющих сделать оценку обстановки дорожной ситуации, и, с более высокой частотой, получать данные о проходимом пути, для моментального реагирования в случае внезапного, с точки зрения системы, возникновения препятствия на пути следования.

Для выполнение данных условий в программном обеспечении происходит комбинирование классических алгоритмов, используемых в компьютерном зрении и беспилотных автомобилях, с алгоритмами машинного обучения.

Также большинство подходов будет использовать модель, при которой в первую очередь происходит обработка двумерных данных, то есть кадра, но результат содержит информацию о трехмерной природе. Такой класс решений позволяет упростить задачу с точки зрения реализации алгоритмов и надежности их работы.

3.1. Пример архитектуры, использующей комбинирование методов машинного обучения со стереозрением

Существует ряд подзадач, уже реализованных на основе методов машинного обучения в задаче обнаружения препятствий с помощью стереокамер.

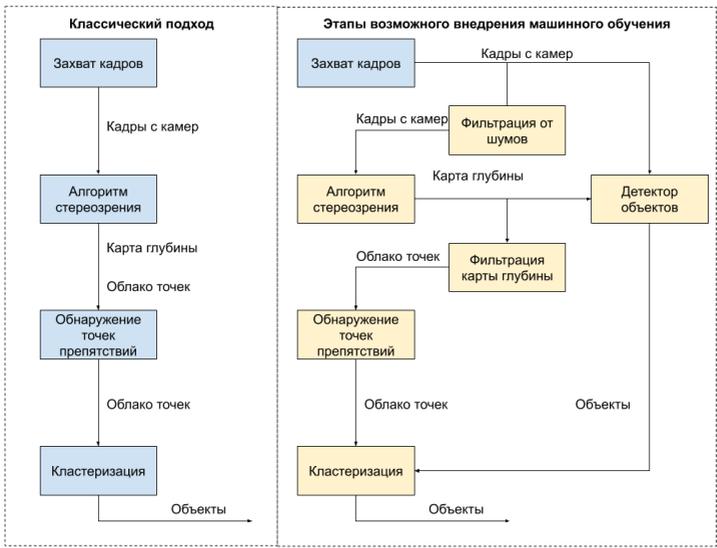


Рисунок 3. Классический подход и подход с использованием машинного обучения для определения препятствий с помощью стереозрения

Часть из этих подходов будет использована в данном решении, однако внедрение некоторых подходов к решению задач исключит часть методов и изменит общую архитектуру решения.

3.2. Используемые подходы

3.2.1. Переход к карте глубины как основному источнику данных

Важным элементом предложенного решения является переход к карте глубины как основному источнику данных. В отличие от классического трёхканального представления изображений, канал глубины, кодируемый одним байтом, представляет гораздо меньшее количество всевозможных вариантов, из-за чего обучение нейронных сетей должно проходить гораздо эффективнее при прочих равных условиях. Кроме того, канал глубины является достаточным для того, чтобы обнаружить на нём препятствия, особенности ландшафта, расстояния до объектов. Далее будут показаны несколько подходов, базирующиеся на описанном здесь.

3.2.2. Переход к задаче обнаружения проходимого пути по карте глубины

С точки зрения оценки условий возможного продолжения движения транспортных средств, стоит рассмотреть два подхода к определению свободного пути: обнаружение препятствий и оценка проходимости пути. С одной стороны, это абсолютно противоположные подходы, решающие одну и ту же задачу, но между ними есть несколько важных различий.

При анализе свободного пути мы не учитываем возможный динамический характер препятствий, поэтому частота получения данных должна быть очень высокой, так как время их жизни не велико. При подходе с обнаружением препятствий, мы способны предсказывать вектор движения различных объектов, что позволяет снизить требования к частоте данных, однако это усложняет комплексирование данных и создает в системе место возможных ошибок. Предлагается использовать два этих метода в комплексе, компенсируя недостатки друг друга.

Проходимый путь, фактически, определяется геометрией пространства видимой области. Поскольку в системе уже имеется блок получения данных о трёхмерной природе окружающей среды, логично использовать именно этот канал, чтобы научить алгоритмы распознавать допустимую геометрию проходимого пути. Для этого предлагается использовать сети семантической сегментации, решающие задачи бинарной сегментации на проходимый и непроходимый путь, принимая в качестве входных данных канал глубины.

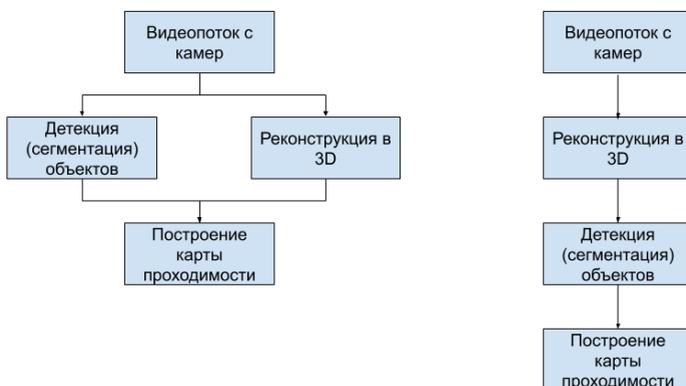


Рисунок 4. Подходы к обнаружению препятствий с помощью стереокамеры

Данный подход должен уменьшить склонность сети к перетренировке, ускорить время обучения сети и получить более робастное решение задачи. Однако для полноценного применения данного подхода необходимо обеспечить высокую частоту работы модуля.

3.2.3. Переход к задаче оптического потока в 2D

Говоря о карте проходимости, стоит принять во внимание тот факт, что проходимый путь является статической поверхностью, не обладающей собственным вектором движения с точки зрения глобальной системы координат. Следовательно, мы можем принять априори тот факт, что участок проходимого пути одного кадра, соответствующий участку изображения на другом, относится к одному и тому же физическому объекту, следовательно также не представляет из себя препятствие.

Таким образом, отслеживая перемещения пикселей, относящихся к проходимому пути между кадрами, можно продлевать время жизни маски проходимости, следовательно, повысить частоту практически к частоте работы камер.

Класс данных задач называется оптический поток. Он разделяется на две категории: разреженный оптический поток, когда ищется смещение отдельных точек и полный оптический поток, когда определяется смещение для всех пикселей кадра. Первый подход выполняется за короткое время, однако нас интересует результат второй категории методов. Для обеспечения необходимой частоты, предлагается использовать метод, на основе двух этих категорий, когда изначально оптический поток рассчитывается для отдельных точек, а затем интерполируется на оставшуюся часть кадра^[6]. Помимо увеличения частоты работы модуля, данный подход также относится к более робастным методам.

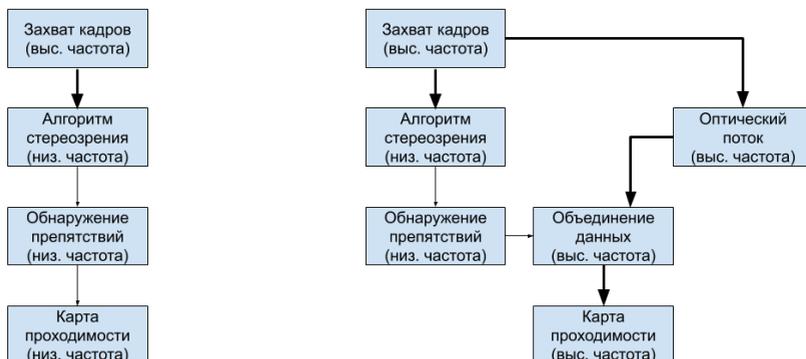


Рисунок 5. Повышение частоты данных о проходимости пути от стереокамеры с помощью методов оптического потока

3.2.4. Использование синтетического датасета для анализа проходимости по карте глубины

Переход с анализа трехканального цветного изображения к одноканальному описанию глубины снижает количество возможных вариантов входных данных, а также использует физическое описание объектов. Благодаря этому, снижается вероятность переобучения сетей, а данные разных датасетов становятся более близкими друг к другу. Это позволяет с большей уверенностью использовать симуляторы для создания тренировочного набора данных. Таким образом, мы можем автоматизировать генерацию большого количества данных для обучения, с минимальными затратами труда. Для данной работы были сгенерированы датасеты для тренировки, тестирования и валидации с помощью симулятора CARLA^[7]. Наборы данных применялись для решения задач фильтрации карты диспаратности, обнаружении

проходимого пути и тестирования семантической разметки по цветному изображению.

3.2.5. Получение объектной сегментации из семантической сегментации и карты глубины

В большинстве случаев принадлежность объекта к какому-либо классу происходит с использованием данных от камеры, а именно кадру изображения в формате цветного изображения, а непосредственно расположение объекта определяется по координатам реконструируемых точек, найденных с помощью стереопары или иного датчика, дающего представление о его трехмерной природе. Имеет место быть и классификация объекта по данным реконструкции, но цветное изображение дает наиболее полную и достаточную информацию об объекте, к тому же размеры и количество датасетов с размеченными на классы объекты кардинально превышает датасеты реконструируемых объектов.

Существует два подхода к ассоциации данных между классами объектов и их физическом расположении. Первый подход заключается в использовании нейронных сетей, решающих задачу детекции объектов, то есть нахождении описывающего объект прямоугольника на кадре, либо, в более сложном варианте, сегментацию по объектам, где каждый отдельный объект получает маску, описывающего принадлежность отдельных пикселей кадра к нему, после чего, область кадра, содержащая объект, ассоциируется с облаком точек, за счёт которого

определяется местоположение и габариты препятствий. Второй подход изначально работает лишь с облаком точек, независимо от класса определяются точки, соответствующие препятствиям, после кластеризации рассчитываются их проекции на кадр, для каждой из которых решается более простая задача классификации.

Далее приведено описание достоинств и недостатков каждого из алгоритмов.

3.2.5.1. Подход ассоциации данных детектора с облаком точек

Преимущества данного подхода заключаются в простоте алгоритма ассоциации данных и возможности использования множества разнообразных решений для детекции объектов, начиная с выбора архитектур, заканчивая датасетом.

Негативная сторона данного подхода заключается в том, что детектор не может гарантировать нахождение объектов в любых условиях, а также не способен определять объекты препятствий, не входящие в список классов тренировочного датасета.

3.2.5.2. Подход ассоциации данных сегментации объектов с облаком точек

Преимущества и недостатки данного подхода повторяют ассоциацию данных детектора с облаком точек, но не ограничиваются ими. Так сети сегментации по объектам более требовательны к вычислительным ресурсам, что увеличивает задержку между принятием и обработкой данных, но при этом дают более точное описание контура объекта.

3.2.5.3. Подход с классификаций препятствий, найденных по реконструкции

Принцип данного подхода заключается в разделении облака точек на категории препятствий и поверхности движения, после чего дополнительно в облаке точек, относящихся лишь к препятствиям, производится кластеризация объектов. Задача проекции кластеров точек на кадр тривиальна, а задача классификации является одной из наиболее хорошо освещенных задач машинного обучения в области компьютерного зрения. Главное преимущество подхода - обнаружение препятствия вне зависимости от класса препятствия. При этом классификация объекта зачастую имеет более высокую вероятность правильного решения, чем задача детекции. Проблемы могут возникать в тех случаях, когда один кластер частично перекрывается на кадре другим. В этом случае мы можем получить неверные данные о природе объекта с высоким значением вероятности, что исказит нашу модель окружающей среды, в отличие от задач детекции и сегментации по объектам, которые способны решать подобные проблемы. Но главным недостатком можно назвать сложность определения препятствий по облаку точек. Решения этой задачи можно разделить на три категории.

Поиск препятствий по данным от стереозрения

Подход к предположению о поверхности

В данном случае мы принимаем как факт утверждение, что поверхность, по которой мы двигаемся, можно описать некоторой математической

моделью. Один из методов, входящих в данную категорию - v -disparity^[8]. Предполагая, что наша поверхность описывается плоскостью, мы строим гистограмму распределения глубины по строкам кадра, после чего находим уравнение прямой для этой гистограммы, таким образом находя зависимость распределения координат в пространстве, которая подходит для описания модели движения по идеальной плоскости. Данный метод является простым и быстрым с точки зрения работы системы, однако в случае движения по поверхности, не подходящей под предполагаемую модель, мы получаем неверные результаты.

Подход анализа карты диспаратности

Полезными с точки зрения внедрения в условиях ограниченных вычислительных ресурсов являются методы, основанные на анализе карты диспаратности, без реконструкции и анализа всего облака точек. К этой категории можно было бы отнести и предыдущий алгоритм, поскольку он также основан на анализе карты диспаратности, однако основная его суть сводилась к строгому определению модели поверхности.

Существует похожий подход, основанный на описанном выше, приобретший большую популярность за счёт своей простоты и надежности - uv -disparity^[9]. Данный метод также предполагает некоторую однородность поверхности, но за счёт построения нескольких гистограмм по разным измерениям, способен находить препятствия не только на поверхностях, близких к плоскостям. Однако, как и предыдущий метод, использование его в условиях, к примеру, пересечённой

местности или движения по насыпи очень затруднительно.

Подход анализа облака точек

Подход к анализу облака точек обладает наибольшей вычислительной сложностью, но при этом представляет наиболее универсальное решение. Примерами таких алгоритмов можно назвать сикстелизацию пространства^[10] и алгоритм инвертированных конусов^[11]. Каждый из этих подходов производит поиск препятствий в облаке точек, исходя из анализа распределения точек в виде некоторых объектов, обладающих высотой, которая по необходимому по пороговому значению отличается от окружающей обстановки.

Поскольку решение задачи определения проходимого пути уже включает в себя обнаружение любого класса препятствий, мы имеем достаточную надежность системы с точки зрения обнаружения препятствий, чтобы использовать подходы, основанные на анализе кадра на поиск объектов, интересующих нас классов. Данная информация позволит расширить представление о среде для системы в целом и повысить качество модуля объединения данных и модуля планирования, поскольку данные сущности смогут учитывать возможный динамический характер объектов.

Предлагается использовать подход с попиксельным выделением объектов, однако, поскольку карта глубины уже дает нам информацию о трехмерной структуре объектов и взаимном местоположении точек

относительно друг друга, предлагается использовать сети семантической сегментации с разбиением изображения на маски определенных классов, после чего внутри каждого класса решать задачу кластеризации данных с помощью плотностного алгоритма, поскольку данный вид алгоритмов кластеризации наиболее точно описывает распространение данных, анализируемых нами. Подобный подход позволит нам использовать сети семантической сегментации, тренировочные наборы данных для которых более распространены. В приведенной реализации мы будем использовать DBSCAN^[12].

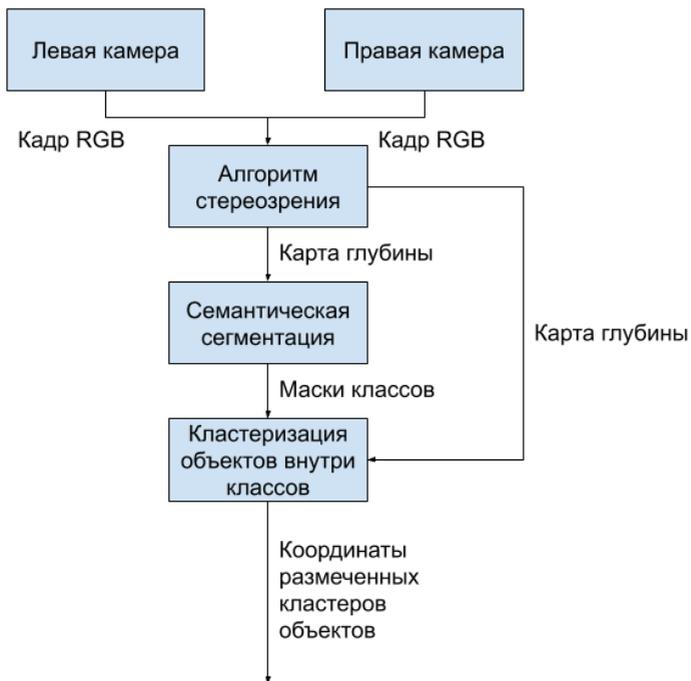


Рисунок 6. Объектная сегментация кадра с помощью комбинирования семантической сегментации и кластеризации облака точек от стереопары

4. РЕАЛИЗАЦИЯ

4.1. Архитектура



Рисунок 7. Итоговая архитектура решения для определения препятствий с помощью камер

Выше приведена итоговая архитектура предлагаемого решения. Далее опишем детали реализации каждого блока, помимо блока камер, поскольку работа с ними в большей степени зависит от деталей аппаратной реализации платформы. Будем считать, что левая и правая камеры дают синхронные, ректифицированные кадры с постоянной частотой, а все калибровочные параметры нам известны.

4.1.1. Алгоритм стереозрения

Одним из важнейших этапов данной системы является алгоритм стереозрения. Было рассмотрено несколько различных алгоритмов для решения задачи стереосопоставления, как классических вариантов, так и использования алгоритмов нахождения полного оптического потока между двумя кадрами стереопары.

Сравнение алгоритмов для решения задачи стереосопоставления		
Название	Средняя скорость работы	Средняя квадратичная ошибка
Block Matching	$6.77 \text{ ms} \pm 44 \mu\text{s}$	62.47994
Semi-global block matching	$50.9 \text{ ms} \pm 227 \mu\text{s}$	32.77910
DeepFlow	$869 \text{ ms} \pm 45.6 \text{ ms}$	11.25950
Farneback	$182 \text{ ms} \pm 2.54 \text{ ms}$	107.9442
OpticalFlowSparseToDense	$120 \text{ ms} \pm 4.32 \text{ ms}$	2.413946

Выше были приведены результаты сравнения алгоритмов с точки зрения их соответствия истинному значению глубины и с точки зрения их производительности.

Как можно заметить исходя из таблицы, сравнивающих точность и скорость работы различных алгоритмов, пожертвовав скоростью выполнения, можно значительно повысить точность и робастность нахождения карты диспаратности с помощью методов нахождения оптического потока, путём интерполяции точечного оптического потока между двумя кадрами.

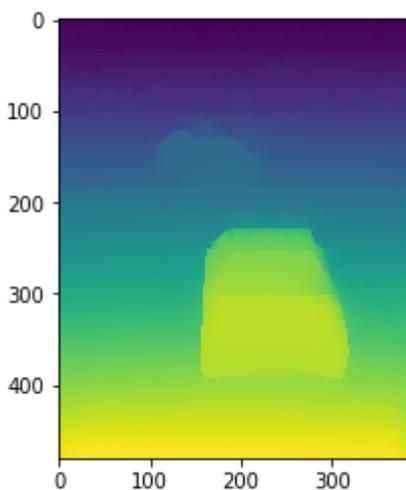


Рисунок 8. Карта диспаратности, полученная с помощью метода нахождения точечного оптического потока и его интерполяции

4.1.2. Семантическая сегментация

Для решения задачи семантической сегментации предлагается использовать претренированные сети. Многоклассовая сегментация требует большого количества обучающих данных, кроме того, для определения принадлежности различных объектов к классам, важно использовать цветовую составляющую, поскольку зачастую она позволяет повысить точность получаемых данных. Была использована архитектура FCN^[13] на основе классификаторов VGG16^[14]

4.1.3. Коллектор кадров и оптический поток

4.1.4. Кластеризация

В качестве алгоритма кластеризации используется алгоритм DBSCAN. Данный алгоритм относится к подвиду плотностных алгоритмов кластеризации, то есть решение о соответствии точек кластеру принимается на основе того, как близко расположены они относительно друг друга. Подобный подход наиболее точно описывает природу данных карты глубины. Далее можно видеть, как работает сегментация по объектам на примере класса машин.



Рисунок 9. Оригинальный кадр левой камеры



Рисунок 10. Семантическая разметка кадра в формате CityScape

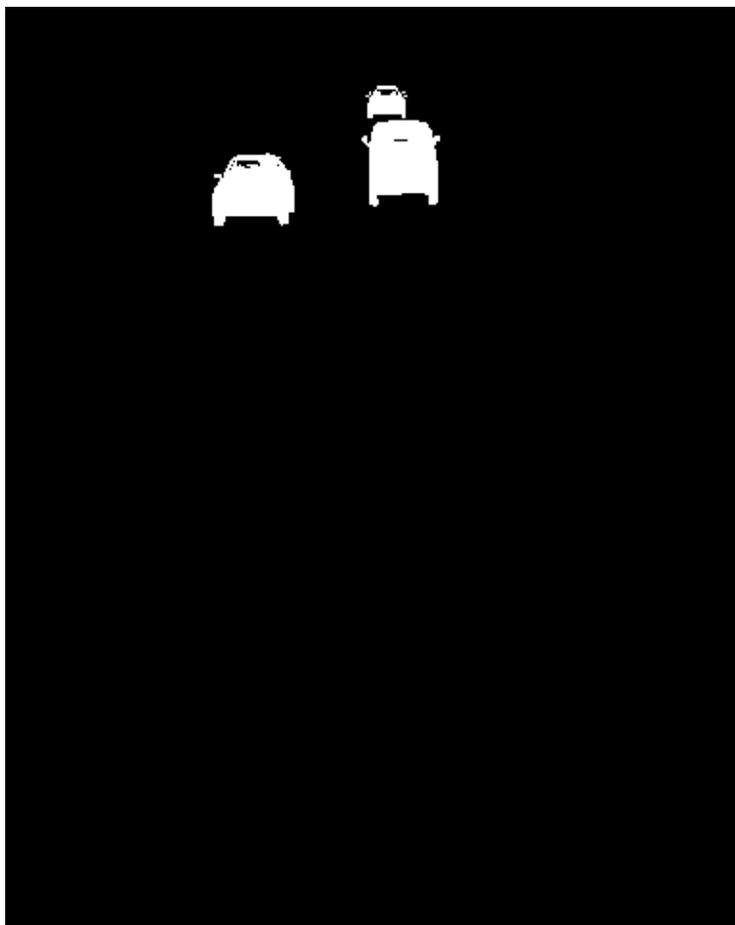


Рисунок 11. Маска с отображением машин на кадре

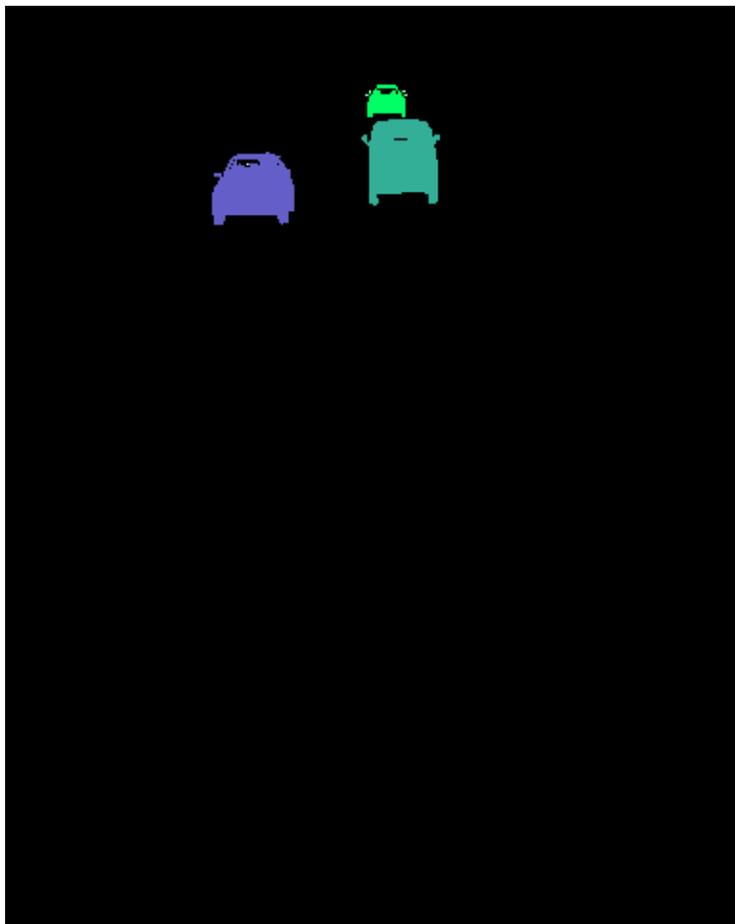


Рисунок 12. Кластеризация облака точек от стереопары, принадлежащих машинам и их проекция на кадр



Рисунок 13. Наложение маски объектной сегментации на кадр

4.1.5. Сегментация проходимости

Для сегментации проходимого пути была использована архитектура автокодера, на основе принципов архитектуры UNet^[15], однако в более упрощённом варианте. Так использован подход, при котором выход слоев верхнего уровня кодировщика

передаётся кроме следующего слоя кодировщика и слою автокодировщика. Было проведено исследования относительно формата входных данных, которые принимает нейронная сеть.

Среди анализируемых форматов были:

1. Карта глубины
2. Цветное изображения
3. Составляющие насыщенности и яркости HSV палитры изображения и глубины
4. Цветное изображение с дополнительным каналом глубины

Далее приведены графики тренировок сетей при различных входных данных. Коэффициент обучения подбирается автоматически, в качестве оптимизатора используется градиентный спуск.

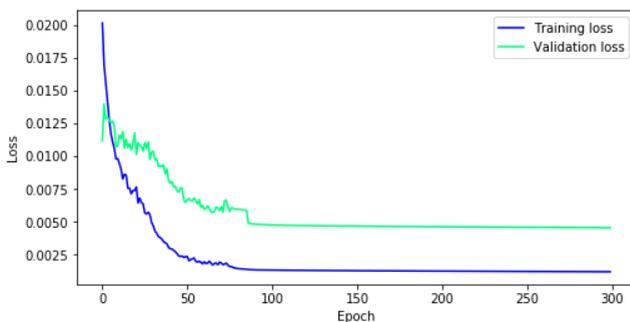


Рисунок 14. График обучения на каналах насыщенности, яркости и глубины

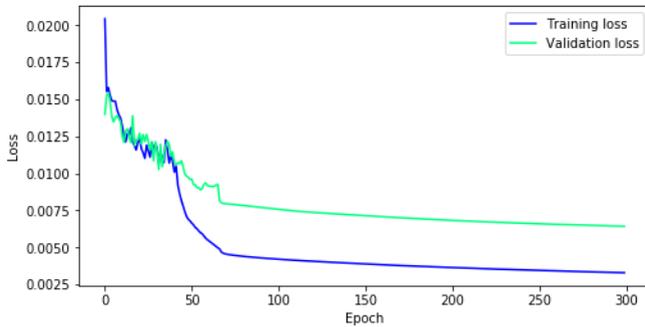


Рисунок 15. График обучения на канале глубины

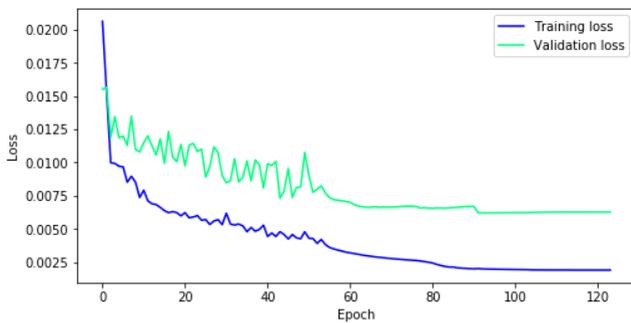


Рисунок 16. График обучения на каналах цветного изображения с каналом глубины

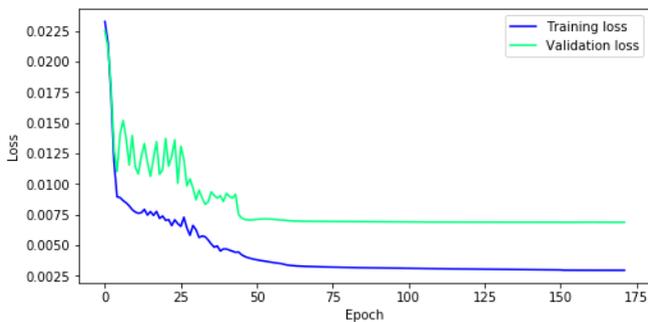


Рисунок 17. График обучения на каналах цветного изображения

Входные каналы	Корректный результат	ложно-положительные ошибки	ложно-негативных ошибки
Depth (Глубина)	97.38%	0.36%	2.26%
RGB (цветное изображение)	95.78%	1.10%	3.12%
SVD	98.30%	0.51%	1.19%
RGBD (цветное изображение + глубина)	96.90%	0.73%	2.37%

Как видно из графиков и таблицы, канал глубины повышает точность работы нейронной сети. А цветовые составляющие способствуют переобучению сети. В дальнейшем предлагается использовать сеть, работающую только с каналом глубины, поскольку она имеет наименьшее значение ложно-положительных ошибок, которые являются наиболее опасными при решении данной задачи.

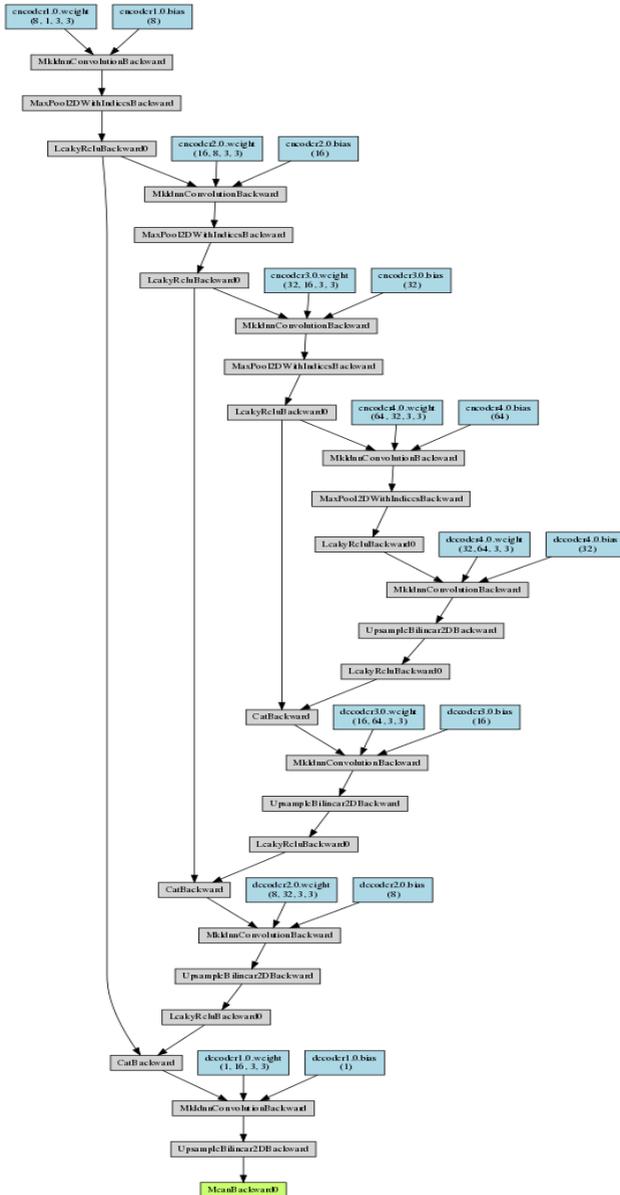


Рисунок 18. Архитектура нейронной сети для сегментации

5. РЕЗУЛЬТАТЫ РАБОТЫ АЛГОРИТМОВ

В ходе разработки данных методологий был решен следующий ряд задач:

1. Сегментация проходимого пути по данным о глубине
2. Решение объектной сегментации с помощью решения задачи семантической сегментации и кластеризации данных от стереопары
3. Решение задачи стереосопоставления с помощью оптического потока для уменьшения зашумления данных
4. Увеличение частоты данных о проходимом пути с помощью решения задачи оптического потока

Была смоделирована работа нейронной сети для прохождения карты проходимости с частотой 1 Гц, что в 5 раз меньше частоты получения данных от камер. Было показано успешное продление результата нейронной сети с помощью решения задачи оптического потока, тем самым обеспечивая результат работы с частотой 5Гц, то есть фактически в режиме реального времени с небольшой постоянной задержкой.

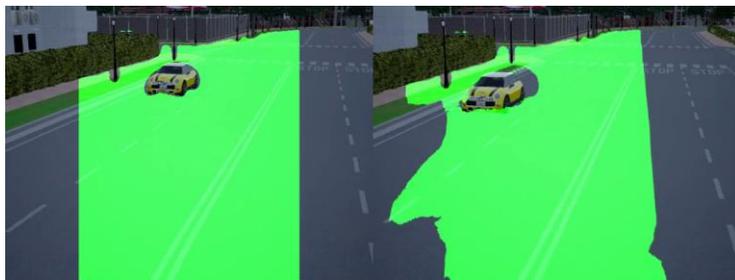


Рисунок 19. Продление маски проходимого пути по оптическому потоку

Переход к задаче объектной сегментации по семантической сегментации и карте глубины позволяет использовать более обширное разнообразие архитектур нейронных сетей, которые менее требовательны к обучающему датасету и имеют большую точность. Использование плотностного алгоритма кластеризации позволяет достоверно разделить объекты по их облакам точек.

Дальнейшее использование разработанных подходов и архитектуры предполагает реализацию блока временного выравнивания данных, сопоставления данных от других датчиков и разработки модуля планирования и принятия решений.

6. ЗАКЛЮЧЕНИЕ

В ходе данной работы была разработана архитектура совместного использования программных модулей для решения задачи обнаружения препятствий по данным от камер. Были проанализированы различные подходы решений задач компьютерного зрения, проведена оценка влияния выбора различных входных данных и готовых реализаций алгоритмов на работу системы, реализованы блоки, входящие в разработанную архитектуру.

Использование данных методологий позволяет увеличить частоту работы системы для определения проходимости пути, упростить обучение нейронных сетей за счёт использования канала глубины как основного источника данных в задачах обнаружения препятствий и их объектной сегментации, используя для этого синтетический датасет, полученный с применением открытых симуляторов. При этом использование предложенных подходов применимо также в комбинации с иными методами обнаружения препятствий и свободного пути, применяемых в системах беспилотного транспорта, как дополнительный источник данных, использующие данные от системы камер.

7. СПИСОК ЛИТЕРАТУРЫ

1. Joel Janai, Fatma Guney, Aseem Behl, Andreas Geiger - Computer Vision for Autonomous Vehicles: Problems, Datasets and State-of-the-Art
2. Amir R. Zamir, Alexander Sax, William Shen, Leonidas Guibas, Jitendra Malik, Silvio Savarese - Taskonomy: Disentangling Task Transfer Learning
3. M. Domínguez-Morales, A. Jiménez-Fernández, R. Paz-Vicente, A. Linares-Barranco and G. Jiménez-Moreno - Stereo Matching: From the Basis to Neuromorphic Engineering
4. Боровский Б.Е. «Безопасность движения автомобильного транспорта», Ленинград, Лениздат, 1984 г.
5. AiMotive // RSIP Vision magazine. — Spring 2018. — С. 20.
6. Jean-Yves Bouguet - Pyramidal Implementation of the Lucas Kanade Feature Tracker
7. Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio L, Vladlen Koltun - CARLA: An Open Urban Driving Simulator

8. Labayrade, R. and Aubert, D. and Tarel, J.P. [SEP]Real time obstacle detection in stereovision on non flat road geometry through v-disparity representation
9. Hu, Z. and Uchimura, K. [SEP]UV-disparity: an efficient algorithm for stereovision based scene analysis
10. Yanir Kleiman - introduction to rendering techniques
11. Manduchi, R. and Castano, A. and Talukder, A. and Matthies, L. [SEP]Obstacle detection and terrain classification for autonomous off-road navigation
12. Martin Ester, Hans-Peter Kriegel, Jiirg Sander, Xiaowei Xu - A Density-Based Algorithm for Discovering Clustersin Large Spatial Databases with Noise
13. Jonathan Long, Evan Shelhamer, Trevor Darrell - Fully Convolutional Networks for Semantic Segmentation
14. Karen Simonyan, Andrew Zisserman - very deep convolutional networks for large-scale image recognition
15. Olaf Ronneberger, Philipp Fischer, Thomas Brox - U-Net: Convolutional Networks for Biomedical Image Segmentation