

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

На правах рукописи

Лаврушкин Сергей Валерьевич

**Разработка нейросетевых методов оценки искажений
стереоскопических видео**

Специальность 2.3.5

Математическое и программное обеспечение вычислительных систем,
комплексов и компьютерных сетей

ДИССЕРТАЦИЯ

на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
кандидат физико-математических наук
Ватолин Дмитрий Сергеевич

Москва — 2025

Оглавление

	Стр.
Введение	5
Глава 1. Нейросетевой метод одновременной оценки	
искажений цвета и резкости между ракурсами	
стереовидео	19
1.1 Постановка задачи	19
1.2 Обзор существующих алгоритмов	23
1.2.1 Моноскопические методы	23
1.2.2 Стереоскопические методы	26
1.3 Общая модель искажений цвета и резкости стереоскопических видео	28
1.4 Описание предложенного алгоритма	31
1.4.1 Оценка карты диспаратности и вычисление карты доверия	33
1.4.2 Нейросетевая оценка искажений	35
1.5 Обучение нейронных сетей	45
1.6 Экспериментальная оценка	47
1.7 Программная реализация	49
1.8 Анализ видео в формате VR180	52
1.8.1 Подготовка набора видео в формате VR180	52
1.8.2 Предобработка собранных видео в формате VR180	54
1.8.3 Результаты оценки искажений цвета и резкости для видео в формате VR180	54
1.9 Заключение	56
Глава 2. Нейросетевой метод оценки геометрических	
искажений между ракурсами стереовидео	59

2.1	Постановка задачи	59
2.2	Обзор существующих алгоритмов	61
2.2.1	Методы ректификации стереовидео	62
2.2.2	Методы, оценивающие параметры геометрических искажений	63
2.3	Модель геометрических искажений	65
2.4	Описание предложенного алгоритма	68
2.5	Обучение нейронных сетей	70
2.6	Экспериментальная оценка	74
2.6.1	Выбор оптимизируемого функционала	74
2.6.2	Оценка модели на тестовом наборе данных	76
2.7	Исправление геометрических искажений	76
2.8	Программная реализация	79
2.9	Анализ видео в формате VR180	80
2.10	Заключение	83

Глава 3. Нейросетевой метод поиска перепутанных ракурсов в

	стереовидео	85
3.1	Постановка задачи	85
3.2	Обзор существующих алгоритмов	88
3.2.1	Методы упорядочивания глубины	89
3.2.2	Метод, основанный на проверке предположения о распределении диспаратности	90
3.2.3	Метод на основе анализа положения областей открытия в стереопаре	91
3.2.4	Композиционные методы	92
3.3	Описание предложенного алгоритма	95
3.3.1	Определение пригодности кадра для анализа	96
3.3.2	Вычисление областей открытия/закрытия по движению	99

3.3.3	Архитектура нейронной сети	100
3.3.4	Оценка пригодности сцены для анализа	101
3.3.5	Определение смены сцены	102
3.4	Обучение нейронных сетей	103
3.5	Экспериментальная оценка	104
3.6	Программная реализация	106
3.7	Анализ видео в формате VR180	107
3.8	Заключение	108
 Приложение А. Дополнительные результаты анализа видео в		
	формате VR180	109
A.1	Методология	109
A.2	Анализ параллакса	110
A.3	Примеры найденных стереоскопических искажений в VR180 . . .	115
 Заключение		128
 Список литературы		130
 Список рисунков		142
 Список таблиц		149

Введение

На сегодняшний день видео в стереоскопическом формате очень широко распространены и знакомы практически каждому. Но все чаще зрители предпочитают смотреть в кинотеатрах 2D-версии полнометражных фильмов вместо 3D-версий. Большинство таких зрителей после просмотра фильмов в стереоскопическом формате ощущает определенного рода дискомфорт: усталость, напряжение и боль в глазах, а также другие симптомы, вплоть до головной боли [1], из-за чего у них пропадает интерес к просмотру стереофильмов. Появление дискомфорта при просмотре 3D-фильмов может быть связано с несколькими факторами:

1. особенности зрительной системы;
2. условия показа;
3. техническое качество производимого контента.

Исследованию восприятия зрителями стереоскопических фильмов посвящен ряд научных работ. В работе [2] обсуждаются трудности восприятия фильмов в форматах 2D и 3D, а также их влияние на зрительную систему. Авторы отмечают, что восприятие фильмов в 2D и 3D форматах связано с различными физиологическими и психофизиологическими особенностями. При грамотной реализации стереоконтента нагрузка на зрительную систему может быть меньше, чем при просмотре в 2D. Однако технические ошибки и несоответствия между ракурсами создают трудности для адаптации, что вызывает зрительный дискомфорт. Ключевая проблема заключается в несовпадении естественных условий восприятия и искусственных параметров 3D-фильмов, таких как несоответствие аккомодации и конвергенции. В работе [3] подчеркивается, что зрительная система вынуждена перестраиваться при переходе от 2D к 3D восприятию. У людей, редко использующих бинокулярное зрение в повседневной жизни, такие изменения вызывают значительное напряжение. Трудности могут быть преодолены через тренировки, направленные на развитие бинокулярных

функций, что снижает ощущение дискомфорта. Также влияние наклона головы на восприятие стереоизображений играет важную роль, согласно работе [4]. Наклон головы ухудшает качество восприятия за счет изменения геометрии зрительного поля. Индивидуальные различия в адаптации связаны с физиологическими характеристиками, такими как торсионные движения глаз, что требует учета при проектировании оборудования и контента.

Кроме того, в работе [5] анализируется влияние современных технологий создания стереофильмов на зрителей с различными зрительными функциями. Грамотно выполненный стереоконтент не вызывает дискомфорта у зрителей с нормальным бинокулярным зрением. Однако недостатки в создании контента или условиях показа могут усиливать зрительный дискомфорт, вызывая усталость, головную боль и даже тошноту. Особое внимание уделено группе зрителей с нарушениями бинокулярных функций, для которых просмотр стереофильмов может быть наиболее проблематичным. Для улучшения восприятия авторы предлагают разрабатывать тесты для диагностики бинокулярных функций и тренировать их перед просмотром стереофильмов. Наконец, в работе [6] анализируются искажения пространственных образов в стереоскопических фильмах. Авторы рассматривают эффекты кулисности, миниатюризации и гигантизма, которые возникают из-за противоречий между монокулярными и бинокулярными факторами зрительного восприятия. Основные причины данных искажений включают несоответствие параметров стереосъемки и особенностей пространственного восприятия зрителя. Для устранения таких эффектов рекомендуется оптимизировать параметры стереосъемки, включая базис и фокусное расстояние, а также учитывать зрительное восприятие при разработке контента.

Таким образом, несмотря на то что особенности зрительского восприятия, такие как индивидуальные зрительные функции и физиологические реакции, играют важную роль в возникновении дискомфорта при просмотре 3D-фильмов, эти трудности не являются непреодолимыми. Со временем и при достаточной практике зрители способны адаптироваться к условиям восприятия стереоскопического контента, что снижает влияние физиологических факторов.

Поэтому по мере привыкания зрителей к 3D-формату на первый план выходят проблемы, связанные с условиями показа и техническим качеством контента.

Недостатки оборудования, такие как несовершенство проекторов, экранов или системы проекции, могут существенно повлиять на восприятие глубины, контрастность изображения и комфорт зрителя. Так в работах [7; 8] рассматривается проблема конфликта между аккомодацией и конвергенцией при просмотре стереоскопического контента на 3D-устройствах, что часто становится причиной визуального утомления и дискомфорта. Исследования показывают, что подобные конфликты приводят к искажению восприятия глубины и увеличению времени идентификации стереоскопических стимулов. Авторы продемонстрировали улучшение зрительного восприятия при минимизации указанных конфликтов на специализированных дисплеях, способных корректно воспроизводить аккомодационные и конвергенционные подсказки. В работе [9] рассматриваются артефакты и ограничения стереоскопических мониторов, возникающие при использовании временного и пространственного интерлейсинга. Данные методы приводят к таким проблемам, как искажения глубины, мерцание, снижение разрешения и неудобства, связанные с слиянием изображений для обоих глаз. Особое внимание уделяется влиянию данных искажений на комфорт зрителей и возможным способам их минимизации, включая использование гибридных методов интерлейсинга. В работе [10] рассматривается влияние размера экрана на восприятие стереоскопического контента. Авторы показывают, что конфликты аккомодации и конвергенции, усиливаемые особенностями дисплеев, влияют на визуальный комфорт и впечатление от 3D. Результаты подчёркивают важность выбора оптимального размера экрана и технологии для минимизации зрительного дискомфорта. Работа [11] систематизирует основные причины зрительного дискомфорта при использовании стереоскопических дисплеев. Помимо затронутых в других работах факторов, таких как конфликт аккомодации и конвергенции, также отмечаются недостатки стереоскопического контента, включая чрезмерную диспаратность стереоскопические искажения. Авторы подчеркивают необходимость создания субъективных и объективных

методов оценки дискомфорта для разработки стандартов, обеспечивающих комфортное восприятие стереовидео.

В работе [12] анализируется влияние различных условий просмотра на зрительное утомление при просмотре стереоскопического телевидения (3DTV) в домашних условиях. В эксперименте с 500 участниками оценивалась степень усталости при просмотре 3D-программ на телевизорах с активными затворными очками. Результаты показали, что использование 3D-очков вызывает значительный дискомфорт, включая усталость глаз, головную боль и общее ухудшение состояния, особенно при близком расстоянии просмотра. Также выявлено, что такие эффекты могут сохраняться на следующий день. В работе [13] анализируется влияние изменений цветовой насыщенности на восприятие стереоскопических изображений. Эксперименты показали, что при увеличении насыщенности восприятие глубины и качество изображения улучшается, однако это может сопровождаться снижением естественности изображения. Также исследование выявило, что стереоскопические изображения требуют более высокой контрастности для достижения той же естественности, что и 2D, что подчеркивает важность сбалансированного подхода к настройке параметров отображения для достижения комфортного и качественного восприятия.

Таким образом, дискомфорт зрителей при просмотре стереоскопического контента обусловлен как технологическими ограничениями устройств показа, так и качеством самого контента. Но несмотря на развитие технологий, включая улучшение дисплеев и оптимизацию способов показа стереоскопического контента, качество 3D-видео продолжает играть ключевую роль. Даже при совершенствовании технологий показа, недостатки самого стереоскопического контента могут свести на нет достигнутые технологические улучшения. И если для первых двух рассмотренных факторов появления дискомфорта при просмотре стереовидео зритель может решить эту проблему самостоятельно (путем тренировки и привыкания зрительной системы к стереоскопическому формату в случае первого фактора, и путем выбора для просмотра хорошего оборудования в случае второго фактора), то для третьего фактора решение

проблемы возникаемого дискомфорта ложится на плечи создателей стереоскопического контента. Следовательно, в стереокинематографе на первый план выходит качество производимого контента, и становится актуальным создание инструментов обеспечения контроля качества создаваемых стереофильмов.

При производстве 3D-фильмов возникает набор новых проблем, не характерных для обычных фильмов. При этом для различных методов производства стереоскопических фильмов характерны различные типы искажений (или артефакты). Существуют следующие методы производства стереофильмов:

- Съемка 3D-фильмов. При съемке фильмов в стереоскопическом формате используются две видеокамеры, находящиеся на небольшом расстоянии друг от друга, имитирующие тем самым зрительную систему человека. Каждая камера при этом записывает отдельную видеопоследовательность (ракурс), предназначенную для одного (левого или правого, в зависимости от положения камеры) глаза.
- Конвертация из 2D в 3D. При конвертации «плоские» (2D) фильмы преобразуются в 3D-форму путем генерации левого и правого ракурса на основе исходной видеопоследовательности и соответствующих данной видеопоследовательности карт глубины (глубина — оценка расстояния от камеры до рассматриваемой точки).
- Компьютерная графика. При рендеринге CGI-фильмов в 3D-формате используются две виртуальные камеры, аналогично обычной съемке 3D-фильмов.

Рассмотрим артефакты, характерные для каждого метода производства стереовидео. Так, при съемке 3D-фильмов достаточно часто появляются геометрические несоответствия между ракурсами, а также несоответствия по цвету, яркости и резкости [14]. Данные проблемы появляются в случае, когда используемые для съемки камеры по-разному настроены и/или какая-то компонента одной из камер вышла из строя. Таких проблем лишены стереофильмы, при производстве которых была использована конвертация или компьютерная графика (хотя описанные выше артефакты можно получить и при использовании

компьютерной графики, проблем с настройкой виртуальных камер обычно не возникает), из-за чего на сегодняшний день большинство высокобюджетных полнометражных фильмов создается при помощи конвертации. Но это не значит, что при таком способе создания стереоконтента удастся избавиться от всех характерных для 3D проблем: для конвертации характерны искажения, связанные с некачественной обработкой границ и карт глубины при генерации ракурсов, например, эффект «кулисности», искажения резкости на границах объектов, их деформация и другие [15].

Другим форматом видео, еще более погружающим чем 3D-видео, являются 360° видео или сферические видео. При этом наибольшее погружение достигается за счет использования наголовного дисплея или шлема виртуальной реальности (VR-шлема). При просмотре сферических видео в каждый момент времени зритель наблюдает только часть сферы, соответствующей текущему направлению взгляда. Дополнительно, сферические видео также могут быть стереоскопическими, при показе отдельных последовательностей для каждого глаза – левого и правого ракурсов, аналогично обычному стереоскопическому формату, что возможно при использовании VR-шлема для просмотра таких видео.

Для записи сферических видео обычно используется специальная система из нескольких камер, одновременно снимающих различные, но пересекающиеся ракурсы вокруг фиксированной позиции, которые далее объединяются в одно 360° видео с помощью алгоритмов склейки. Однако такой подход порождает целый ряд влияющих на итоговое качество видео проблем в зависимости от качества работы алгоритма склейки и калибровки камер. Более того, в сферических видео основное действие обычно происходит только на одной полусфере, в то время как на устройство передается весь видеопоток, что приводит к передаче и хранению избыточной информации. Для решения этих проблем Google анонсировала новый формат видео в VR – VR180, в котором видео проецируется на полусферу, в то время как другая часть сферы может быть использована для хранения второго ракурса, благодаря чему достигается стереоскопический эф-

фект. Вместо использования специального рига из нескольких камер для записи видео в формате VR180 используются только две камеры с объективами типа «рыбий глаз», аналогично обычной стереоскопической съемке. Такой подход значительно снижает стоимость конечного устройства. При этом значительно упрощается техника съемки, так как остаются актуальными все приемы работы с обычной камерой (только результат получается потенциально более зрелищным). В целом, видео в формате VR180 обеспечивает еще большее погружение, чем сферические видео, дешевле и проще в производстве, а также не обладает проблемами, возникающими при использовании алгоритмов склейки. Но, как и в случае с обычным стереоскопическим форматом, VR180-видео также страдают от стереоскопических артефактов, характерных для 3D-съемки. А так как съемкой видео в формате VR180 в основном занимаются любители, контролю качества уделяется не так много времени, и большое количество стереоскопических артефактов попадает в итоговые видео.

Целью данной работы является исследование и разработка нейросетевых алгоритмов объективной оценки качества стереоскопических видео, применимых на практике при анализе стереоскопических фильмов и видео в формате VR180. В данной работе рассматриваются наиболее часто встречающиеся артефакты, характерные для стереоскопической съемки: искажения цвета, резкости и геометрии, а также один из наиболее болезненных для зрителей артефактов – перепутанные ракурсы.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработка новых нейросетевых алгоритмов для:
 - одновременной оценки искажений цвета и резкости;
 - оценки геометрических искажений, включающих в себя сдвиг, масштабирование и поворот одного ракурса относительно другого;
 - поиска перепутанных ракурсов в стереоскопических видео.

2. Исследование объективного качества видео в формате VR180 с помощью предложенных методов оценки качества.

Основные положения, выносимые на защиту:

1. Нейросетевой метод оценки цветовых искажений и искажений резкости между ракурсами стереоскопического видео значительно сокращает число ложноположительных срабатываний за счет одновременного учета рассматриваемых искажений и по результатам объективного сравнения превосходит аналоги, ранее использовавшиеся при анализе полнометражных стереоскопических фильмов.
2. Нейросетевой метод оценки геометрических искажений между ракурсами стереоскопического видео по результатам объективного сравнения уменьшает ошибку оценки угла поворота более чем на 14%, а коэффициента масштабирования и вертикального сдвига на 2 порядка по сравнению с аналогами, ранее использовавшимися при анализе полнометражных стереоскопических фильмов. Метод также обеспечивает возможность автоматического исправления выявленных геометрических искажений.
3. Нейросетевой метод поиска перепутанных ракурсов в стереоскопических видео по результатам объективного сравнения улучшает точность классификации более чем на 8% по сравнению с применявшимися ранее при анализе полнометражных стереофильмов подходами.
4. Результаты исследования объективного качества 1000 VR180-видео, выполненного с использованием разработанных методов, показывают наличие по меньшей мере одного вида стереоскопического искажения в каждом из проанализированных материалов.

Все предложенные алгоритмы были реализованы и прошли экспериментальную апробацию.

Научная новизна:

1. Впервые предложен нейросетевой алгоритм для одновременной оценки цветовых искажений и искажений резкости в стереоскопических видео,

что значительно понизило число ложноположительных срабатываний по сравнению с отдельными методами, ранее использовавшимися при анализе полнометражных стереофильмов.

2. Предложен новый оптимизируемый функционал для обучения нейросетевых методов оценки геометрических искажений в стереоскопических видео, а также описан метод исправления геометрических искажений для VR180-видео на основе предложенного метода поиска искажений.
3. Предложен новый нейросетевой метод поиска перепутанных ракурсов, предсказывающий вероятность наличия перепутанных ракурсов на основе исходного ракурса, построенной карты диспаратности и областей открытия/закрытия по движению.
4. Проведен первый в мире масштабный анализ видео в формате VR180 на предмет технического качества стереоскопического контента.

Практическая значимость. Все предложенные в данной работе методы были реализованы в виде программного инструмента, позволяющего осуществлять оценку соответствующих артефактов по двум видеопоследовательностям — для левого и правого ракурсов. С помощью данного программного инструмента было проведено масштабное тестирование видео в формате VR180, показавшее, что практически в каждом видео из собранной выборки встречается как минимум один стереоскопический артефакт. Разработанный инструмент может быть применен на этапе контроля качества при производстве стереоскопических видео, как профессиональными студиями, так и любителями, что значительно сократит число нежелательных искажений в итоговом продукте и заметно сократит число зрителей, испытывающих дискомфорт при его просмотре.

Проведенное исследование объективного качества стереоскопического видео в формате VR180 оформлено в виде отдельного отчета. Данный отчет доступен в сети Интернет по адресу https://videoprocessing.ai/stereo_quality/report12.html.

Методология и методы исследования. В работе применялись методы линейной алгебры, теории алгоритмов, а также методы машинного обучения.

Апробация работы. Основные результаты работы докладывались на:

- International Conference on 3D Immersion (Бельгия, 13-15 декабря 2016)
- 3DTV Conference 2018 (Стокгольм, Швеция, 3-5 июня 2018)
- International Conference on 3D Immersion (дистанционно, 15 декабря 2020)
- Stereoscopic Displays and Applications XXXII (дистанционно, 18 января 2021)
- Семинар кафедры интеллектуальных информационных технологий ВМК МГУ (весна 2021)
- Семинар центра визуализации и спутниковых информационных технологий НИИСИ РАН (10 октября 2022)
- Семинар института информационных технологий, математики и механики ННГУ им. Н.И.Лобачевского (13 октября 2022)
- Семинар им. М.Р. Шура-Бура ИПМ им. М.В. Келдыша (10 ноября 2022)
- Семинар кафедры информатики и программного обеспечения БГТУ (17 марта 2023)
- Семинар подразделения интеллектуального анализа данных и технического зрения ГосНИИАС (13 апреля 2023)

Личный вклад автора заключается в выполнении основного объема теоретических и экспериментальных исследований, изложенных в диссертационной работе, включая разработку теоретических моделей, методик, разработку и реализацию алгоритмов, анализ и оформление результатов в виде публикаций и научных докладов. В работах [1; 2] Д.С. Ватолину принадлежит постановка задачи и обсуждение результатов ее решения.

Научные исследования, представленные в диссертации, были поддержаны грантом СТАРТ Фонда содействия инновациям в рамках проекта “Разработка системы автоматической объективной оценки качества и исправления стереоскопического видео и видео в формате VR180”, а также частично поддержаны грантами РФФИ 15-01-08632 а “Автоматизация создания и контроля качества

стерео видео” и РФФИ 19-01-00785 а “Разработка нейросетевых алгоритмов обработки и сжатия видеопоследовательностей”.

Публикации. Основные результаты по теме диссертации изложены в 3 публикациях [1—3], изданных в рецензируемых научных изданиях, определенных в п. 2.3 Положения о присуждении ученых степеней в Московском государственном университете имени М. В. Ломоносова.

Объем и структура работы. Диссертация состоит из введения, трех глав, приложения и заключения. Полный объем диссертации составляет 149 страниц, включая 81 рисунок и 8 таблиц. Список литературы содержит 95 наименований.

Дополнительно результаты работы были опубликованы в трудах международных конференций и рекомендованных ВАК журналах [16—24]. По итогам разработки получено свидетельство о государственной регистрации программы для ЭВМ "Система для обучения и тестирования нейросетевых методов объективной оценки качества и исправления стереоскопических видео" [25]. В работах [16—24] Д.С. Ватолину принадлежит постановка задачи и обсуждение результатов ее решения. В работе [16] вклад А.А. Бокова, М.В. Ерофеева и А.А. Федорова заключается в подготовке текста статьи. В работе [22] вклад К.А. Кожемякова состоит в проведении экспериментальных сравнений разработанного метода оценки геометрических искажений. В работе [24] вклад К.А. Кожемякова и И.А. Молодецких заключается в участии в подготовке отчета по результатам анализа объективного качества видео в формате VR180. В работе [17] вклад В.А. Людвиченко состоит в разработке метода исправления цветовых искажений в стереоскопических видео. В работах [18; 21; 23] практическая составляющая была выполнена основными авторами статей – М.С. Великановым, К.В. Малышевым и С.Д. Грохольским соответственно – автору диссертации принадлежит постановка задачи и курирование проектов.

Как было упомянуто выше, в данной диссертационной работе основное внимание уделяется артефактам стереосъемки. Из-за проблем с артефактами, характерными для натуральной съемки фильмов в 3D, в последние годы боль-

шинство стереофильмов производится с помощью конвертации отснятого в 2D контента в стереоскопический формат (рисунок 1). Хотя этот факт уменьшает востребованность поиска искажений стереосъемки в конвертированных полнометражных фильмах, данная проблема как никогда актуальна для стереоскопических VR-видео, так как съемка в 3D является единственным способом создания стереоскопического контента в VR помимо компьютерной графики. При этом рынок VR продолжает активно развиваться и расти (рисунок 2), и формат стереоскопических видео в VR – VR180 – может вскоре получить большую огласку. Однако это может не произойти из-за плохого технического качества стереоскопического контента в VR, в результате чего заинтересованность зрителей стремительно спадет из-за возникаемого при просмотре дискомфорта аналогично полнометражным стереоскопическим фильмам, где интерес поддерживается в основном в азиатском регионе. Например, сборы 3D версии фильма «Мир юрского периода» в США составили 48% от общих сборов в этом регионе, в то время как в Китае – 95% от общих сборов. Поэтому для продвижения стереоскопического формата в VR важен контроль качества создаваемого стереоскопического контента на этапе его производства для сокращения потенциального дискомфорта.

Решению задачи оценки цветовых искажений в стереоскопических видео, а также задачи оценки искажений резкости посвящено множество научных работ, обзор которых представлен в разделе 1.2. Проведенный обзор показал, что на текущий момент не существует метода, осуществляющего оценку данных искажений одновременно, в результате чего существующие методы будут давать неверные оценки на стереокадрах с другим искажением. В разделе 1.4 приводится описание нового нейросетевого метода для одновременной оценки искажений цвета и резкости между ракурсами стереовидео на основе модели искажений, представленной в разделе 1.3. В разделе 1.5 описывается процесс обучения нейросетевой части метода. Как показано в разделе 1.6, предложенный метод превосходит аналоги, ранее использовавшиеся при анализе полнометражных стереоскопических фильмов. В разделе 1.8 приводятся результаты



Рисунок 1 — Количество полнометражных стереоскопических фильмов по годам, разделенных по методу производства. Информация по фильмам получена со страницы

[https://en.wikipedia.org/wiki/List_of_3D_films_\(2005_onwards\)](https://en.wikipedia.org/wiki/List_of_3D_films_(2005_onwards)).

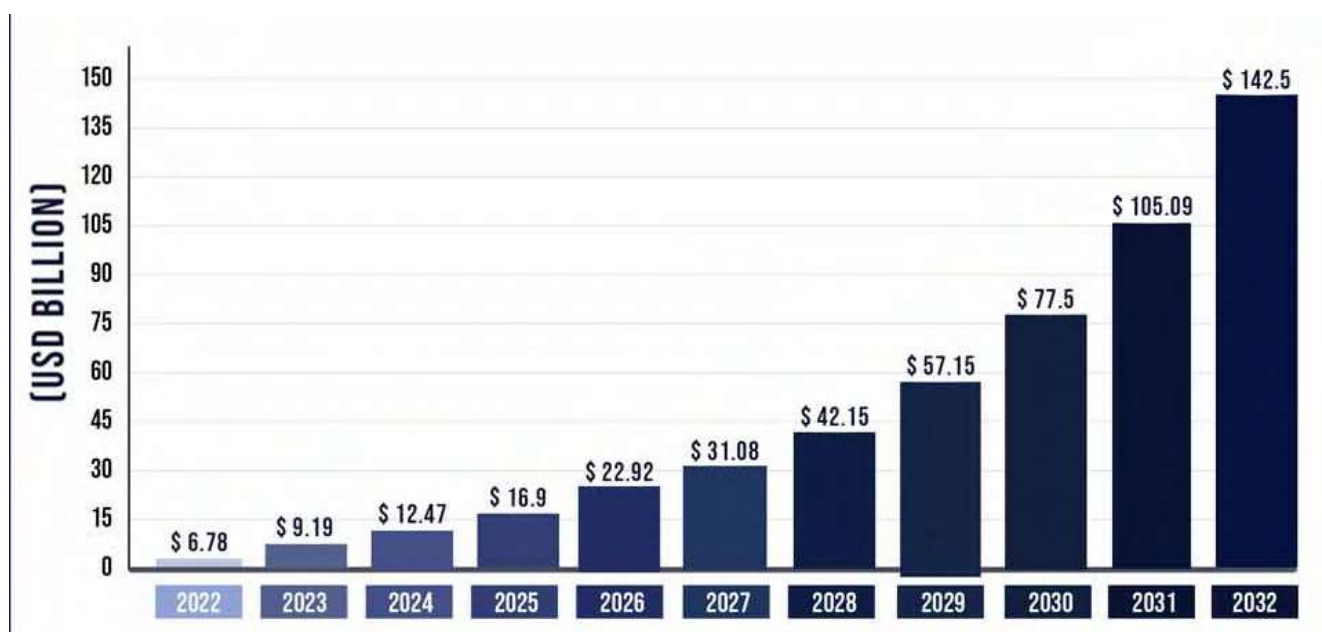


Рисунок 2 — Прогнозы объема рынка AR- и VR-шлемов во всем мире с 2022 по 2032. Данные получены с сайта <https://www.precedenceresearch.com/>

проведенного анализа объективного качества 1000 VR180-видео, собранных с платформы YouTube.

Аналогичным образом построены главы 2 и 3, посвященные новым предложенным методам по оценке геометрических искажений между ракурсами стереоскопического видео и поиску перепутанных ракурсов соответственно. Также в приложении А приведены дополнительные материалы по результатам исследования стереоскопического качества видео в формате VR180. В разделе А.1 описывается методология проводимой объективной оценки качества в целом, в разделе А.2 – результаты оценки диапазона параллаксов для VR180-видео, а в разделе А.3 приведены найденные кадры с стереоскопическими артефактами в проанализированных видео.

В Заключение сформулированы основные результаты диссертационного исследования, рассмотрены возможные варианты их применения и обозначены перспективы дальнейших исследований.

Глава 1. Нейросетевой метод одновременной оценки искажений цвета и резкости между ракурсами стереовидео

1.1 Постановка задачи

Цветовые искажения ракурсов и искажения ракурсов по резкости – одни из самых распространенных типов искажений стереоскопического видео, полученных в ходе нативной съемки в 3D, что справедливо как для классических стереоскопических видео, так и видео в формате VR180, так как в обоих случаях используются одни и те же методы съемки: данные искажения характерны для систем, состоящих из двух камер. Малейшие несоответствия в настройке и/или неполадки одной из них и приводят к возникновению названных выше артефактов.

Под цветовыми искажениями ракурсов стереоскопического видео понимается сильное несоответствие яркости и/или цвета одного и того же объекта кадра в левом и правом ракурсе или всего кадра, что наиболее заметно при переключении между ракурсами. На рисунке 3 представлен схематичный пример цветовых искажений в стереовидео. При этом часто в полнометражных стереофильмах встречается ситуация, когда лишь часть кадра отличается по цвету между ракурсами, как показано на рисунке 4. Эти несоответствия могут возникать из-за различий в матрицах камер (различия могут появиться непосредственно во время съемки стереоскопического видео, например, при неравномерном прогреве матриц), особенностей освещения (например, возникновение различных бликов в ракурсах из-за разного угла падения световых лучей на объективы камер), а также некорректном использовании светофильтров и/или их дефекте.

Под искажениями ракурсов по резкости понимается сильное несоответствие в детализации и/или размытии одного и того же объекта кадра в левом

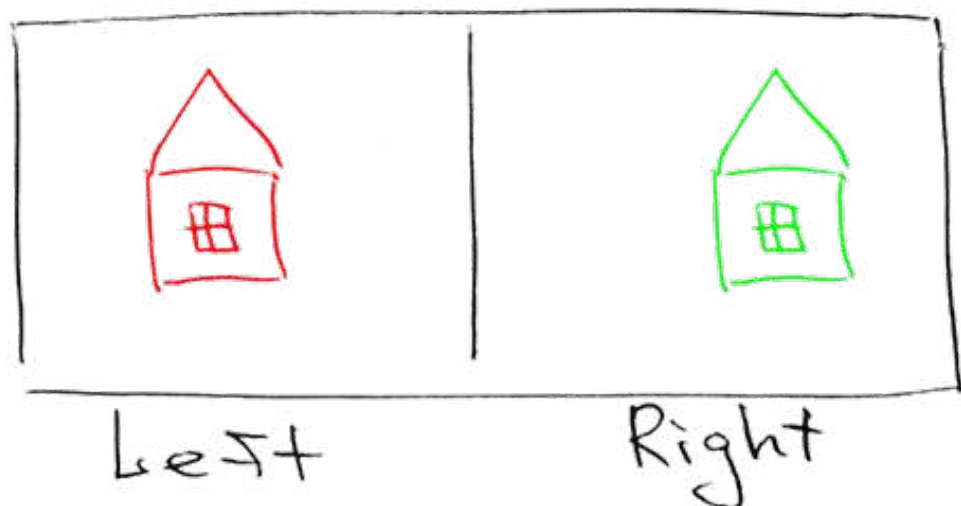


Рисунок 3 — Схематичный пример стереопары с цветовыми различиями между ракурсами.



Рисунок 4 — Пример локальных цветовых расхождений между ракурсами стереофильма. В кадре присутствует засвет на стене здания в правом ракурсе, в то время как цвет героя сильно не меняется при смене ракурсов. Кадр из фильма “Шаг вперед 4”.

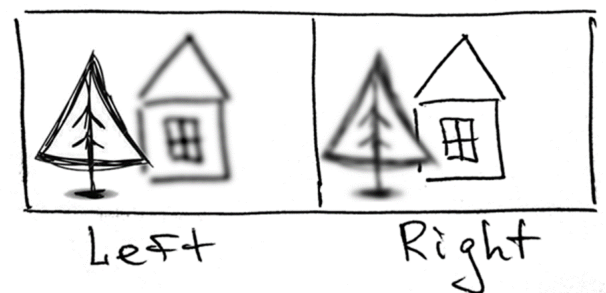
и правом ракурсе или всего кадра, что также наиболее заметно при переключении между ракурсами. Пример стереопары из полнометражного стереофильма с различиями по резкости представлен на рисунке 5. Обычно различия в резкости между ракурсами появляются из-за некорректной калибровки съемочного оборудования, а именно разной фокусировке камер, но также могут возникнуть из-за загрязнения объективов камер и их дефектов. В итоге данные артефакты стереовидео могут иметь различную природу и силу. В самом простом случае либо левый, либо правый ракурс резче другого во всех пикселях изображения, как показано на рисунке 6 (а). Более сложной является ситуация, при которой



Рисунок 5 — Увеличенный фрагмент стереопары, в котором различия по резкости максимальны. Кадр из фильма “Джек – покоритель великанов”.



(а) Левый ракурс находится полностью в фокусе в отличие от правого



(б) Различные объекты находятся в фокусе на разных ракурсах

Рисунок 6 — Схематичные примеры стереопар с различиями по резкости между ракурсами.

разные объекты находятся в фокусе в разных ракурсах стереовидео, как показано на рисунке 6 (б), поэтому для каждого ракурса будет справедливо, что часть объектов в нем окажется более четкой и детализированной по сравнению с теми же объектами в другом ракурсе. В данной работе рассматриваются оба варианта искажений резкости в стереовидео.

Также в данной работе предлагается осуществлять одновременный поиск кадров стереоскопического видео с различиями по цвету и резкости. Оба этих артефакта приводят к различиям в яркости и/или цвете между ракурсами стереовидео, поэтому при использовании отдельных алгоритмов для поиска дан-

ных артефактов может возникать большое количество ложноположительных срабатываний. Этой проблемой обладают предыдущие методы поиска данных искажений, которые ранее использовались для анализа полнометражных стереоскопических фильмов [26; 27].

Формально задачу одновременной оценки искажений цвета и резкости между ракурсами стереовидео можно поставить следующим образом. Пусть имеются две видеопоследовательности, являющиеся левым и правым ракурсами стереоскопического видео. Необходимо для каждого кадра стереовидео оценить различия по цвету и резкости между левым и правым ракурсом. Введем обозначения:

- Пиксель p – тройка (p_Y, p_U, p_V) целых чисел, $p_Y, p_U, p_V \in [0, 255]$, определяющих цветовые компоненты в модели YUV.
- Изображение I – матрица пикселей, имеющая размер $h \times w$, где h – высота изображения, w – ширина изображения.
- Видеопоследовательность $\{I_t\}_{t=1}^k$ – упорядоченный набор изображений, имеющих одинаковую высоту и ширину, длиной k .
- Стереовидео S – упорядоченная пара видеопоследовательностей $(\{I_t^L\}_{t=1}^k, \{I_t^R\}_{t=1}^k)$, имеющих одинаковую высоту, ширину и количество кадров. Данные видеопоследовательности называются левым и правым ракурсом соответственно.

На вход алгоритму подается стереовидео S . На выходе алгоритм должен предоставить оценку различий между левым и правым ракурсом по цвету $\{m_i^c\}_{i=1}^k$, $m_i^c \in \mathbb{R}^{[0, +\infty]}$ и по резкости $\{m_i^d\}_{i=1}^k$, $m_i^d \in \mathbb{R}^{[0, +\infty]}$ для каждого кадра стереовидео $t = \overline{1, k}$. Вид оценок напрямую зависит от рассматриваемой модели искажений. Подробное описание используемой в работе модели искажений представлено в разделе 1.3.

1.2 Обзор существующих алгоритмов

В общем случае для оценки стереоскопических искажений обычно необходимо сопоставить ракурсы (то есть найти для каждого пикселя одного ракурса соответствующий ему пиксель в другом ракурсе), после чего проводится анализ соответствующих друг другу пикселей. Для сопоставления ракурсов могут быть использованы различные методы: сопоставление особых точек [28—30], вычисление оптического потока [31—33] и блочные алгоритмы сопоставления [34]. Далее могут быть использованы стандартные метрики, например среднеквадратическое отклонение и средняя абсолютная ошибка, для поиска разницы между стереоскопическими изображениями для оценки цветовых искажений, либо для оценки искажений резкости, но в частотном диапазоне. Также возможны подходы, не использующие сопоставление ракурсов для оценки искажений между ними. Как правило, в них вычисляется некоторая характеристика для каждого ракурса, которая затем между ними сравнивается. Таким способом можно применять любые моноскопические методы оценки качества, не использующие эталон, для сравнения качества ракурсов стереоскопического видео.

1.2.1 Моноскопические методы

В научной литературе встречается большое количество методов для оценки уровня размытия изображения. Данные методы могут быть применены для оценки размытия между стереоскопическими ракурсами путем вычисления разницы уровня размытия между ними. Методы оценки уровня размытия изображения могут быть разделены на две основные группы: методы на основе анализа границ и методы на основе анализа областей. Также за последние несколько

лет появился целый ряд нейросетевых методов оценки уровня размытия в изображении.

Методы на основе анализа границ обычно следуют похожей стратегии: вычисляется уровень размытия границ изображения, в результате чего получается разреженная карта размытия, а далее эти оценки размытия распространяются на все изображение (что является необязательным шагом) с помощью методов интерполяции и/или экстраполяции для получения полной карты размытия. В работе [35] размытая граница моделируется путем свертки четкой границы с ядром фильтра Гаусса. Далее оценивается стандартное отклонение в ядре фильтра Гаусса (неизвестный уровень размытия) путем оценки силы изменения яркости вдоль границ. В работе [36] предлагается одновременный метод поиска границ и метод оценки размытия. Рассматриваемый метод вычисляет уровень размытия, измеряя места обнуления производных гауссианы третьего порядка по направлению градиента за счет использования поворачивающихся фильтров. В обоих рассмотренных методах вычисляется лишь разреженная карта размытия. В методе [37] было предложено использовать соотношение модулей градиента между исходным изображением и его дополнительно размытой версией для локальной оценки силы размытия в точках границ. Далее полученные оценки фильтровались с помощью билатерального фильтра и распространялись на все изображение с помощью матирующего лапласиана [38]. В дальнейшем проводились дополнительные исследования использования разницы модулей градиента и в других работах. Например, в работах [39; 40] предлагается использовать более одного параметра для дополнительного размытия изображения для уменьшения влияния шума, а в работе [41] предлагается иерархический подход, чтобы избежать неоднозначностей в определении границ при различных масштабах.

Методы на основе анализа областей обрабатывают локальные блоки изображения для оценки уровня размытия и в большинстве случаев используют пострегуляризацию для получения согласованных результатов. В работе [42] задача оценки уровня размытия решается в частотном диапазоне с использованием теоремы свертки. В данном методе вычисляется функция правдоподобия

для заданного кандидата функции рассеяния точки, формулируя на ее основе разложение изображения на поддиапазоны и модель гауссовой смеси на различных масштабах. Рассматриваемый метод получил развитие в работе [43], в которой исследовалась непрерывная вероятностная функция для оценки уровня размытия в каждом пикселе на основе анализа локализованного преобразования Фурье. Дополнительно использовалась информация о цветовых границах, а также накладывались ограничения на гладкость для получения согласованных результатов. В работе [44] также рассматривается локализованное преобразование Фурье, но задача оценки силы размытия формулируется как задача разметки пикселей изображения. В частности, каждому пикселю изображения присваивается дискретное значение уровня размытия, используя методы машинного обучения, что гарантирует глобальную консистентность вычисленной карты размытия. В методе [45] анализируются признаковые векторы, построенные на основе высокочастотных коэффициентов дискретного косинусного преобразования блоков изображения для каждого пикселя. При этом для одного пикселя рассматривается несколько блоков разного размера для учета различных масштабов. В работе [46] предложен метод оценки размытия изображений, основанный на модели человеческого восприятия размытия при различных уровнях контрастности. Метод использует вероятностную модель для обнаружения размытия на каждой границе изображения, а для получения итогового значения вероятности для границ объединяются путем вычисления кумулятивной вероятности размытия. В работе [47] также предлагается метод оценки размытия всего изображения. Ключевым понятием в нем является разница разниц значений яркости изображения после обработки медианными фильтром. Авторы [47] утверждают, что данная характеристика оценивает ширину границ, которая соответствует силе размытия.

Также для оценки уровня размытия появилось несколько методов на основе нейронных сетей. В работе [48] предложена архитектура сверточной нейронной сети для вычисления значимых локальных признаков на уровне суперпикселей для оценки размытия в областях изображения. В работе [49] комбиниру-

ются классические признаки с нейросетевыми, что увеличило точность работы метода для оценки размытия границ. В методе [50] используется сверточная нейронная сеть для непосредственного предсказания уровня размытия по изображению. В рассматриваемом методе также дополнительно используется метод адаптации доменов наборов данных, который преобразует признаки натурально размытых изображений в признаки искусственно размытых изображений. В работе [51] предлагается еще одна нейросетевая архитектура для определения размытых областей изображения на основе последовательности остаточных уточнений. В нейросетевом методе [52] оценка уровня размытия осуществляется за счет дистилляции глубины для определения областей, находящихся в фокусе. Для оценки карты глубины в данном методе используется отдельная нейросетевая модель. В работе [53] предлагается использовать две сверточные нейронные сети: одна классифицирует границы изображения на границы по глубине, а вторая осуществляет непосредственную оценку размытия. При этом авторам [53] удалось добиться улучшения точности оценки как по сравнению с методами на основе анализа границ, так и с методами на основе анализа областей.

Таким образом, нейросетевые подходы являются многообещающими в задаче оценки уровня размытия изображения. Однако использовать существующие методы оценки уровня размытия в изображениях для оценки несоответствий резкости между ракурсами стереовидео нецелесообразно: данные методы не учитывают другие возможные стереоскопические искажения и в случае их наличия будут выдавать некорректные результаты. Наиболее распространенным будет ошибка в случае наличия цветовых искажений.

1.2.2 Стереоскопические методы

Все стереоскопические методы оценки искажений цвета и резкости можно разделить на глобальные (не использующие результаты сопоставления ракур-

сов) и локальные (использующие результаты сопоставления ракурсов). Так, для оценки цветовых искажений было предложено два локальных метода. В методе [54] вычисляется корреляция Пирсона между цветовыми гистограммами двух ракурсов в цветовом пространстве HSV. В работе [55] рассматривается аналогичный подход, в котором предложена простая глобальная мера цветовых искажений. Хотя эти методы достаточно просты, они не позволяют локализовать цветовые искажения и часто работают неточно, так как не учитывают различия двух ракурсов, например, области открытия/закрытия.

В работе [56] исследуется задача оценки искажений резкости, вызванных различиями в фокусных расстояниях камер. Авторы данной работы предполагают, что входная стереопара ректифицирована, и поэтому для вычисления карты диспаратности рассматриваются только горизонтальные смещения пикселей. Рассматриваемый метод использует сумму модифицированных лапласианов для вычисления уровня размытия в каждом пикселе изображения и на их основе строит модель, которая сопоставляет различия резкости со значениями диспаратности. Другой подход предлагается в работе [57] на основе анализа различий в ширине соответствующих границ в разных ракурсах. Однако эти подходы не измеряют напрямую силу размытия одного ракурса относительно другого. Так, метод [56] классифицирует кадр по 5 возможным уровням размытия, а метод [57] вычисляет вероятность наличия размытия одного ракурса относительно другого.

Предложенный в данной работе метод строится на идеях локальных методов оценки искажений цвета и резкости, используя стандартную схему сопоставления ракурсов и дальнейшем оценки стереоскопических артефактов. За основу берутся метод оценки искажений цвета [27] и метод оценки искажений резкости [26]. Метод оценки цветовых искажений вычисляет локальную цветовую разницу между соответствующими пикселями, а метод оценки различий резкости – локальную разницу размытия в частотном диапазоне. Данные методы были использованы при анализе 105 полнометражных стереоскопических фильмов [58].

Однако в ходе анализа было получено большое число ложных срабатываний данных методов, в первую очередь из-за присутствия другого типа искажения в кадре. Поэтому логичным дальнейшим шагом по улучшению точности работы этих методов является создание общего метода для одновременной оценки рассматриваемых стереоскопических артефактов. Также для более точной оценки силы искажений предлагается использовать нейросетевой подход.

1.3 Общая модель искажений цвета и резкости стереоскопических видео

Для обучения нейросетевого метода оценки стереоскопических искажений необходимо в первую очередь подготовить соответствующие наборы данных, для чего, в свою очередь, необходимо ввести модель рассматриваемых искажений для генерации элементов выборки. Пусть I_{gt}^L и I_{gt}^R – соответственно левый и правый ракурсы стереопары, не содержащей различия по цвету и резкости между ними. Все изображения рассматриваются в цветовом пространстве YUV. Для моделирования рассматриваемых стереоскопических артефактов кадры без искажений модифицируются следующим образом:

$$I^L(x, y, c) = a(x, y, c) \left(G(\sigma^{pos}(x, y)) * I_{gt}^L \right)(x, y, c) + b(x, y, c), \quad (1)$$

$$I^R(x, y, c) = \left(G(\sigma^{neg}(x, y)) * I_{gt}^R \right)(x, y, c), \quad (2)$$

где I^L и I^R – получаемые в результате применения модели левый и правый ракурсы с искажениями, c – один из цветовых каналов цветового пространства YUV, $a(x, y, c)$, $b(x, y, c)$ – линейный и константный коэффициенты для моделирования цветовых искажений, генерирующиеся с помощью шума Перлина для каждого пикселя с координатами (x, y) и каждого цветового канала c , $G(\sigma(x, y))$ – ядро фильтра Гаусса размера 11×11 , в котором сила размытия задается параметром $\sigma(x, y)$ – стандартным отклонением распределения Гаусса,

генерирующийся с помощью шума Перлина для каждого пикселя с координатами (x, y) , σ^{pos} и σ^{neg} – матрицы сгенерированных стандартных отклонений, содержащие положительные значения и модули отрицательных значений матрицы стандартных отклонений σ соответственно, и нули в остальных пикселях, $*$ – операция свертки. Здесь используется линейная модель цветовых искажений, которые добавляются только в левый ракурс, чего достаточно для получения разницы в цвете между ракурсами стереовидео, а для моделирования искажений резкости используется Гауссово размытие с изменяющейся силой размытия в каждом пикселе, применяющееся либо в левом, либо в правом ракурсе. Для генерации коэффициентов линейной модели и матрицы стандартных отклонений Гауссова размытия используется градиентный шум Перлина, позволяющий задавать непрерывное изменение силы искажений в зависимости от координат пикселя. Данный подход к генерации рассматриваемых стереоскопических искажений позволяет получить сложные примеры, соответствующие неравномерному прогреву матриц камер для цветовых искажений и нахождению различных объектов кадра в фокусе на разных ракурсах для искажений резкости.

Дополнительно также рассматривалась простая константная модель искажений, изменяющая исходную стереопару одинаково для каждого пикселя:

$$I^L(c) = a_c \left(G(\sigma^{pos}) * I_{gt}^L \right)(c) + b_c, \quad (3)$$

$$I^R = \left(G(\sigma^{neg}) * I_{gt}^R \right), \quad (4)$$

где параметры a_c , b_c , σ^{pos} , σ^{neg} – константы для стереопары и не зависят от координат пикселя, a_c , b_c индивидуальны для каждого цветового канала, а один из параметров σ^{pos} , σ^{neg} равен 0, из-за чего размытие добавляется в один из двух ракурсов, $I^L(c)$ – один из каналов изображения в пространстве YUV. Данная модель соответствует более простым искажениям цвета и резкости, которые также встречаются при анализе полнометражных стереоскопических фильмов.

Для генерации наборов данных на основе описанных выше моделей было вырезано 9488 различных стереопар в разрешении 960×540 без цветовых искажений и искажений резкости из полнометражных стереоскопических филь-

мов. При сборе кадров стереовидео рассматривались только стереопары, у которых были близкие к нулевым значения оценки соответствующих артефактов с помощью методов [26; 27], ранее применявшихся для анализа полнометражных стереоскопических фильмов. У данных методов практически отсутствуют ложноотрицательные срабатывания, тем самым в выборку не попадали стереопары с искажениями. Кадры были получены из следующих фильмов:

1. “Вольт”;
2. “Судья Дредд 3D”;
3. “Первый мститель”;
4. “Призрачный гонщик 2”;
5. “Гравитация”;
6. “Великий Гэтсби”;
7. “Гарри Поттер и Дары Смерти: Часть II”;
8. “Хоббит: Нежданное путешествие”;
9. “Железный человек 3”;
10. “Человек из стали”;
11. “Люди в черном 3”;
12. “Тихоокеанский рубеж”;
13. “Пастырь”;
14. “Призрачный патруль”;
15. “47 ронинов”;
16. “Сталинград”.

В этот список вошли фильмы, произведенные как по методу съемки в 3D, так и с помощью конвертации из 2D в 3D и компьютерной графики.

При подготовке набора данных рассматривались случаи без/с добавлением цветовых искажений, без/с добавлением искажений резкости. При добавлении искажений результирующие кадры генерировались с помощью одной из двух представленных выше моделей. Наряду с этим были использованы смешанные модели, в которых один вид искажений добавлялся по сложной модели, а другой – по упрощенной. Пример стереопары со сгенерированными иска-



Рисунок 7 — Пример левого ракурса со сгенерированными искажениями цвета и размытия и приведенного к нему правого ракурса. На данной стереопаре присутствуют локальные цветовые искажения, а также глобальное искажение резкости. Кадр из фильма “Первый мститель”.

жениями представлен на рисунке 7. Итоговый набор данных был разделен на обучающую и валидационную части в соотношении 95% и 5% соответственно.

1.4 Описание предложенного алгоритма

Для оценки расхождений ракурсов по цвету и резкости с помощью нейросети предлагается подход, состоящий из следующих шагов:

1. Вычисление карт диспаратности между левым и правым ракурсами, а также построение карт доверия к вычисленным картам, характеризующих точность оцененных значений диспаратности.
2. Интерполяция правого ракурса к левому ракурсу по вычисленной карте диспаратности.
3. Оценка карт различий по резкости и размытию с помощью нейронной сети на основе левого ракурса, интерполированного к нему правого ракурса, а также карте доверия, соответствующей карте диспаратности.
4. Вычисление величины искажения цвета и резкости между входными ракурсами по предсказанным нейросетью картам различий.

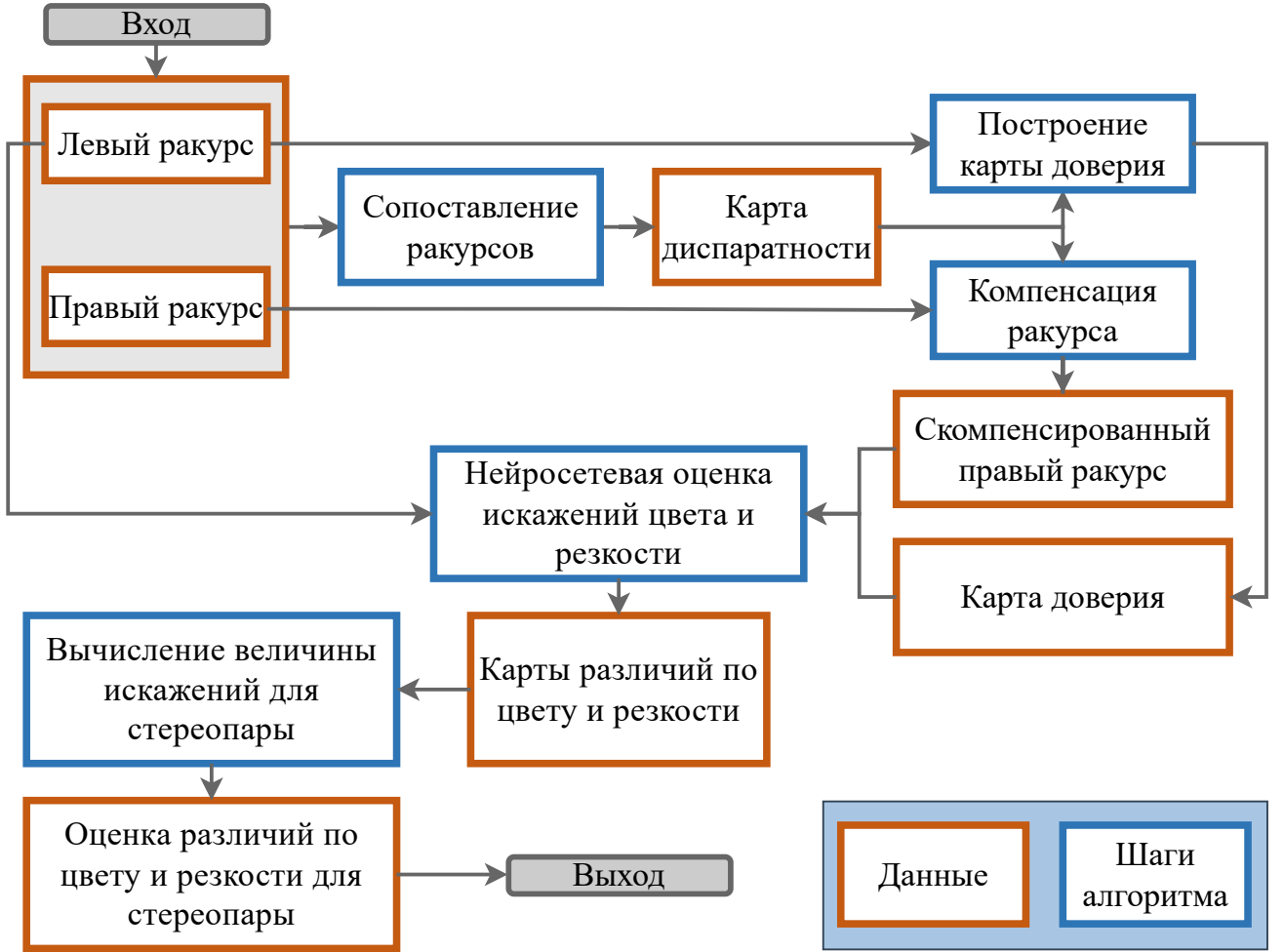


Рисунок 8 — Общая схема предложенного метода поиска различий по цвету и резкости между ракурсами стереоскопического видео.

Каждый кадр стереоскопического видео обрабатывается независимо друг от друга. Общая схема метода представлена на рисунке 8.

Итоговая оценка искажений по цвету m^c и резкости m^d в стереопаре формируется на основе предсказанных карт различий следующим образом:

$$m^c = \frac{\sum_{i=1}^n \text{conf}_i (\hat{c}_i^Y + \hat{c}_i^U + \hat{c}_i^V)}{3 \sum_{i=1}^n \text{conf}_i}, \quad (5)$$

$$m^d = \frac{\sum_{i=1}^n \text{conf}_i \hat{d}_i}{\sum_{i=1}^n \text{conf}_i}, \quad (6)$$

где \hat{c} – предсказанная карта различий по цвету для каждого цветового канала YUV, \hat{d} – предсказанная карта размытия, conf – карта доверия к диспаратности, используемая в качестве входной карты доверия для нейронной сети, n – количество пикселей в изображении.



Рисунок 9 — Пример карты диспаратности и соответствующей ей карты доверия, построенной для правого ракурса кадра стереофильма “Мстители”.

Далее описывается метод построения карт диспаратности и соответствующих им карт доверия, а также используемые нейросетевые архитектуры для оценки карт различий по цвету и резкости между входными ракурсами.

1.4.1 Оценка карты диспаратности и вычисление карты доверия

Оценка карты диспаратности осуществляется по исходным ракурсам стереоскопического видео в цветовом пространстве RGB с помощью блочного ме-

тогда сопоставления [34]. Так как при сопоставлении блоков возможны ошибки, для построенных карт диспаратности строятся карты доверия. Значения карты доверия характеризуют точность вычисленных векторов диспаратности. При вычислении значений доверия учитываются следующие показатели:

- Мера достоверности сопоставления LRC (left-right consistency) [59]. Поскольку левый и правый ракурсы являются изображениями одной сцены, то значение диспаратности пикселя в левом ракурсе должно быть равно по модулю и иметь противоположный знак по сравнению со значением диспаратности соответствующего ему пикселя в правом ракурсе. Более формально, мера достоверности сопоставления LRC вычисляется следующим образом: если пиксель с координатами $x = (x_1, x_2)$ одного ракурса соответствует пикселю с координатами $x' = (x'_1, x'_2) = x + v_x$ другого ракурса, то мера достоверности сопоставления LRC для него равна:

$$\text{lrc} = \frac{\text{dif}_1^2}{h} + \frac{\text{dif}_2^2}{w}, \quad (7)$$

$$\text{dif} = (\text{dif}_1, \text{dif}_2) = v'_{x'} + v_x, \quad (8)$$

где v_x – вектор диспаратности пикселя с координатами x в первом ракурсе, $v'_{x'}$ – вектор диспаратности пикселя с координатами x' во втором ракурсе.

- Блочная дисперсия цветовых значений ракурса, соответствующего карте диспаратности. Дисперсия вычисляется для каждого блока ракурса как сумма значений дисперсий каждой цветовой компоненты в блоке:

$$\text{var} = \text{var}^R + \text{var}^G + \text{var}^B, \quad (9)$$

$$\text{var}^i = \frac{1}{s} \sum_{p \in \text{block}} p_i^2 - \left(\frac{1}{s} \sum_{p \in \text{block}} p_i \right)^2, \quad (10)$$

где p – значение цветовой компоненты пикселя в блоке изображения размером 9×9 , i – один из каналов цветовой модели RGB, s – количество пикселей в блоке.

Итоговое значение доверия к значению диспаратности в пикселе, учитывающее две описанные характеристики, строится по следующей формуле:

$$\text{conf}_i = \min(1 - \min(1, a \text{ lrc}_i), \min(1, b \text{ var}_i)), \quad (11)$$

где $a = 40$, $b = 0.5$, i – индекс пикселя. Итоговое значение доверия лежит в диапазоне $[0, 1]$: $\text{conf}_i \in \mathbb{R}^{[0,1]}$. Пример построенной карты диспаратности и соответствующей ей карты доверия представлен на рисунке 9.

1.4.2 Нейросетевая оценка искажений

На вход нейронной сети подается исходный левый ракурс и интерполированный к нему по вычисленной карте диспаратности правый ракурс в цветовом пространстве YUV, а также соответствующая карте диспаратности карта доверия. По этим входным данным нейронная сеть одновременно предсказывает карты различий по цвету между ракурсами, а также карту размытия, которая соответствует матрице стандартных отклонений, используемой в модели искажений в фильтре Гаусса для задания силы искажения при генерации примеров набора данных.

Для предсказания карт различий по цвету и резкости были использованы следующие нейросетевые архитектуры. Первой была использована архитектура сети кодировщик-декодировщик по типу U-net, впервые примененная в задаче сегментации медицинских изображений [60] и хорошо себя зарекомендовавшая в дальнейшем в других задачах обработки изображений и видео. Данная архитектура состоит из кодировщика, постепенно уменьшающего разрешение карт признаков, и декодировщика, постепенно увеличивающего разрешение карт признаков и восстанавливающего по ним целевое изображение, которые и придают ей U-форму. В данной работе был использован кодировщик, представляющий из себя сверточную сеть с последовательным применением сверточных слоев и операций субдискретизации с выбором среднего. В качестве декодировщика

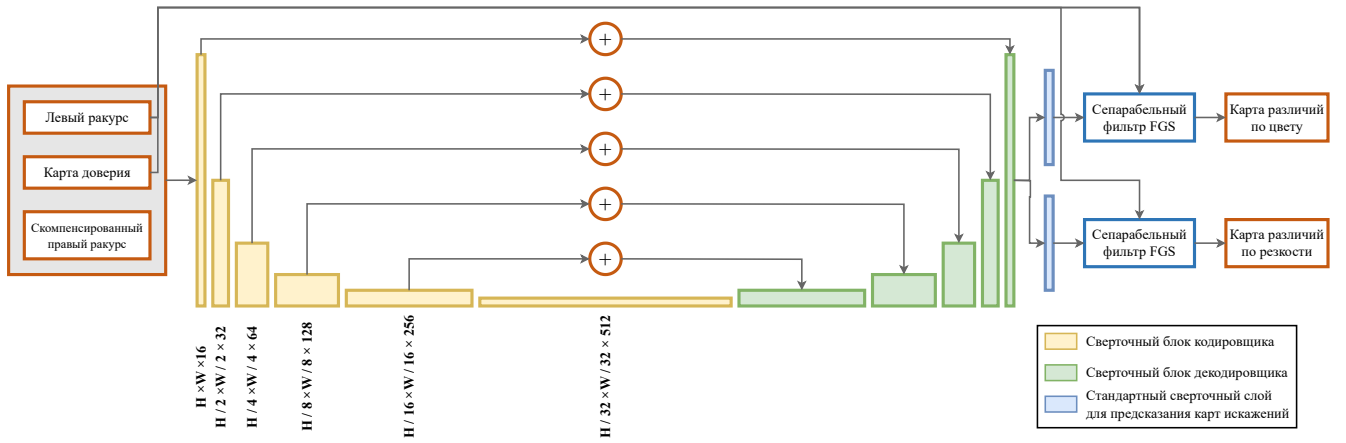


Рисунок 10 — Общая архитектура сети типа U-net.

была выбрана сверточная сеть с последовательным применением сверточных слоев и билинейной интерполяции для пространственного увеличения разрешения. Конфигурация блоков кодировщика и декодировщика представлена в таблице 1. Каждый сверточный блок состоит из нескольких сверточных слоев с функцией активации ReLU: $f(x) = \max(0, x)$, за которыми идет слой субдискретизации с выбором среднего в случае кодировщика, либо блок повышения размерности в случае декодировщика. Блок повышения размерности представляет из себя билинейную интерполяцию карты признаков, за которой следует сверточный слой с функцией активации ReLU. Между сверточными блоками, находящимися в разных частях сети, с одинаковой пространственной размерностью добавляются дополнительные соединения в виде сложения результатов работы сверточного слоя перед субдискретизацией в кодировщике с результатом работы сверточного слоя после восстановления разрешения в декодировщике для восстановления пространственной информации, утраченной после слоев субдискретизации с выбором среднего в кодировщике. Шаг и коэффициент дилатации каждого сверточного слоя в сети равен 1. Исключением является последний блок декодировщика, в котором последний сверточный слой заменен на два параллельных слоя, один из которых предсказывает карту различий по цвету, а другой — карту размытия. Общая архитектура сети представлена на рисунке 10.

Таблица 1 — Конфигурация блоков сверточной нейронной сети типа U-net.

Слой	Размерность	Шаг	Коэффициент дилатации	Функция активации
Сверточный блок кодировщика				
Сверточный	3×3	1	1	ReLU
Сверточный	3×3	1	1	ReLU
Сверточный	3×3	1	1	ReLU
Субдискретизация с выбором среднего	2×2	2	1	—
Сверточный блок декодировщика				
Сверточный	3×3	1	1	ReLU
Сверточный	3×3	1	1	ReLU
Сверточный	3×3	1	1	ReLU
Билинейной интерполяции	—	—	—	—
Сверточный	3×3	1	1	ReLU

В качестве второй архитектуры сети для предсказания карт различий по цвету и резкости была использована сверточная нейронная сеть типа GridNet [61], представляющая из себя модификацию архитектуры кодировщик-декодировщик и ранее применявшаяся для задачи семантической сегментации. Вместо использования последовательности сверточных слоев, как в типичном кодировщике-декодировщике, данная архитектура обрабатывает карты признаков в виде решетки из строк и столбцов. Слои в каждой строке образуют поток признаков, в котором их разрешение остается постоянным. Каждый поток обрабатывает карты признаков на разных масштабах, а столбцы соединяют потоки для обмена информацией между вышестоящим и нижестоящим потоком. Такая архитектура обобщает архитектуру кодировщик-декодировщик, в котором карты признаков обрабатываются лишь по одному потоку. Данный подход позволяет значительно сократить размеры сети по сравнению со стандартным кодировщиком-декодировщиком, а также увеличивает качество работы за счет использования потока карт признаков с полным пространственным разрешением. Сеть состоит из набора сверточных блоков: горизонтального для обработки

Таблица 2 — Конфигурация блоков сверточной нейронной сети типа GridNet.

Слой	Размерность	Шаг	Коэффициент дилатации
Горизонтальный блок			
Сверточный	1×1	1	1
Функция активации PReLU	—	—	—
Сверточный	3×3	1	1
Функция активации PReLU	—	—	—
Сверточный	3×3	1	1
Skip-соединение с первым сверточным блоком	—	—	—
Вертикальный блок для уменьшения разрешения			
Функция активации PReLU	—	—	—
Сверточный	3×3	2	1
Функция активации PReLU	—	—	—
Сверточный	3×3	1	1
Вертикальный блок для увеличения разрешения			
Билинейной интерполяции	—	—	—
Функция активации PReLU	—	—	—
Сверточный	3×3	1	1
Функция активации PReLU	—	—	—
Сверточный	3×3	1	1

карт признаков в одном потоке и двух типов вертикальных блоков для уменьшения и увеличения разрешения. Конфигурация блоков сети представлена в таблице 2. Каждый блок состоит из нескольких сверточных блоков с функциями активации PReLU [62]. Коэффициент дилатации каждого сверточного слоя в сети равен 1, а шаг изменяется с 1 до 2 в сверточных слоях, в которых осуществляется уменьшение размерности в 2 раза. После последнего горизонтального блока также используются два параллельных сверточных слоя для предсказания карт различий по цвету и резкости аналогично предыдущей архитектуре. Общая архитектура сети представлена на рисунке 11.

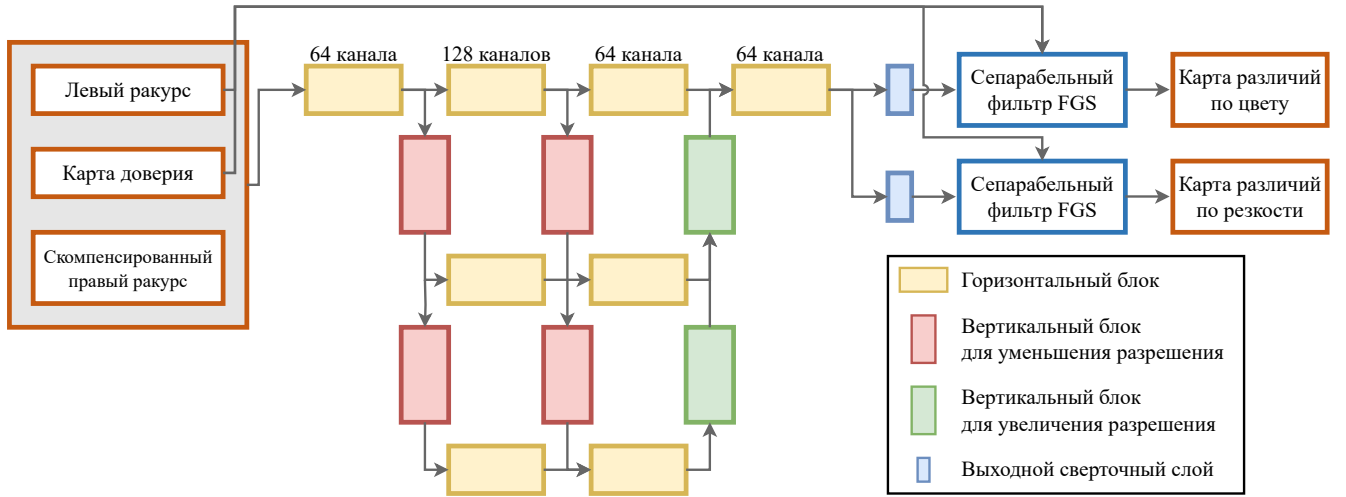


Рисунок 11 — Общая архитектура сети типа GridNet.

Для улучшения качества оценки и удаления эффектов блочности в итоговых картах искажений из-за блочного сопоставления ракурсов в рассматриваемые сверточные нейронные сети в качестве последних блоков был добавлен fast global smoother (FGS) [63] – фильтр, использующийся для распространения данных предсказанных карт искажений по маске карты доверия с учетом границ исходного изображения. Рассмотрим метод фильтрации FGS.

Задача фильтрации входного изображения $f \in \mathbb{R}^{H \times W \times 3}$ с учетом направляющего изображения $g \in \mathbb{R}^{H \times W \times 3} \mid \mathbb{R}^{H \times W}$ в фильтре FGS формулируется как задача минимизации функции энергии взвешенного метода наименьших квадратов:

$$\arg \min_u \sum_x ((u_x - f_x)^2 + \lambda \sum_{i \in N(x)} \omega_{x,i}(g)(u_x - u_i)^2), \quad (12)$$

где $u \in \mathbb{R}^{H \times W \times 3}$ – результат фильтрации, $N(x)$ – набор соседних пикселей к пикселю x , $\lambda > 0$ – параметр, контролирующий баланс между двумя компонентами функции энергии, $\omega_{x,i}(g)$ – веса, построенные на направляющем изображении g и представляющие из себя меру похожести пикселей x и i . Авторы FGS предлагают аппроксимировать решение данной задачи путем ее декомпозиции на серию одномерных задач для каждого пространственного измерения. В случае одномерной задачи минимизации набор соседних пикселей $N(x) = \{x-1, x+1\}$ и решение для нее получается в рекурсивной манере за два прохода:

1. Прямой проход:

$$\tilde{c}_x = \frac{c_x}{b_x - \tilde{c}_{x-1}a_x}, \tilde{c}_0 = \frac{c_0}{b_0}, \tilde{f}_x = \frac{f_x - \tilde{f}_{x-1}a_x}{b_x - \tilde{c}_{x-1}a_x}, \tilde{f}_0 = \frac{f_0}{b_0}, \quad (13)$$

$$a_x = -\lambda\omega_{x,x-1}, b_x = 1 + \lambda(\omega_{x,x-1} + \omega_{x+1,x}), c_x = -\lambda\omega_{x+1,x}, \\ a_0 = 0, c_{N-1} = 0,$$

где N – размерность одномерного сигнала, \tilde{c}_x и \tilde{f}_x – скрытые переменные, использующиеся для вычисления результата фильтрации на обратном проходе.

2. Обратный проход:

$$u_x = f_x - \tilde{c}_x u_{x+1}, u_{N-1} = f_{N-1}. \quad (14)$$

Сложность такого одномерного фильтра составляет $\mathcal{O}(N)$. Для использования данного одномерного фильтра для фильтрации двумерных изображений он применяется сначала к каждой строке, затем к каждому столбцу изображения в течение нескольких итераций T для достижения более качественного результата. Так как использование в течение нескольких итераций одного и того же параметра λ приводит к куда более значительному сглаживанию изображения, чем при решении исходной двумерной задачи сглаживания с тем же параметром, то для каждой итерации $t = \overline{1, T}$ вычисляется свой параметр $\lambda_t = \frac{3}{2} \frac{4^{T-t}}{4^T - 1} \lambda$, гарантирующий тот же уровень сглаживания, что и в исходной задаче.

Для применения сепарабельного двумерного фильтра FGS в качестве вычислительного блока в обучаемой нейронной сети необходимо, чтобы этот фильтр был дифференцируемым и его можно было использовать в методе обратного распространения ошибки, а также обладал некоторым ресурсом параллелизма для эффективного распараллеливания вычислений. Покажем, что это справедливо, в следующих теоремах.

Теорема 1. *Одномерный фильтр FGS дифференцируем и применим в методе обратного распространения ошибки.*

Доказательство. Покажем дифференцируемость обратного прохода одномерного фильтра FGS. При обратном распространении ошибки с последующих слоев нейронной сети приходит градиент $\frac{\partial L}{\partial u_x}, x = \overline{0, N-1}$. Тогда для обратного прохода получаем следующие производные:

$$\frac{\partial L}{\partial \tilde{f}_x} = \frac{\partial L}{\partial u_x} \frac{\partial u_x}{\partial \tilde{f}_x} = \frac{\partial L}{\partial u_x}, \quad (15)$$

$$\frac{\partial L}{\partial \tilde{c}_x} = \frac{\partial L}{\partial u_x} \frac{\partial u_x}{\partial \tilde{c}_x} = -u_{x+1} \frac{\partial L}{\partial u_x}, \quad (16)$$

$$\frac{\partial L}{\partial u_{x+1}} = \frac{\partial L}{\partial u_{x+1}} + \frac{\partial L}{\partial u_x} \frac{\partial u_x}{\partial u_{x+1}} = \frac{\partial L}{\partial u_{x+1}} - \tilde{c}_x \frac{\partial L}{\partial u_x}. \quad (17)$$

Аналогично для прямого прохода одномерного фильтра FGS. С обратного прохода приходят градиенты $\frac{\partial L}{\partial f_x}$ и $\frac{\partial L}{\partial \tilde{c}_x}$. Тогда получаем следующие производные для прямого прохода:

$$\frac{\partial L}{\partial f_x} = \frac{\partial L}{\partial \tilde{f}_x} \frac{\partial \tilde{f}_x}{\partial f_x} = \frac{1}{b_x - \tilde{c}_{x-1}a_x} \frac{\partial L}{\partial \tilde{f}_x}, \quad (18)$$

$$\frac{\partial L}{\partial \tilde{c}_{x-1}} = \frac{\partial L}{\partial \tilde{c}_{x-1}} + \frac{\partial L}{\partial \tilde{f}_x} \frac{\partial \tilde{f}_x}{\partial \tilde{c}_{x-1}} + \frac{\partial L}{\partial \tilde{c}_x} \frac{\partial \tilde{c}_x}{\partial \tilde{c}_{x-1}} = \quad (19)$$

$$= \frac{\partial L}{\partial \tilde{c}_{x-1}} + \frac{a_x(f_x - \tilde{f}_{x-1}a_x)}{(b_x - \tilde{c}_{x-1}a_x)^2} \frac{\partial L}{\partial \tilde{f}_x} + \frac{a_x c_x}{(b_x - \tilde{c}_{x-1}a_x)^2} \frac{\partial L}{\partial \tilde{c}_x},$$

$$\frac{\partial L}{\partial \tilde{f}_{x-1}} = \frac{\partial L}{\partial \tilde{f}_{x-1}} + \frac{\partial L}{\partial \tilde{f}_x} \frac{\partial \tilde{f}_x}{\partial \tilde{f}_{x-1}} = \frac{\partial L}{\partial \tilde{f}_{x-1}} + \frac{-a_x}{b_x - \tilde{c}_{x-1}a_x} \frac{\partial L}{\partial \tilde{f}_x}. \quad (20)$$

При этом также можно посчитать производные относительно весовых коэффициентов $\omega_{x,x-1}, x = \overline{1, N-1}$, что позволяет пропускать градиенты к направляющему изображению g и делает возможным обучение сетей для генерации и обработки не только фильтруемого изображения, но и направляющего изображения:

$$\frac{\partial L}{\partial a_x} = \frac{\partial L}{\partial \tilde{f}_x} \frac{\partial \tilde{f}_x}{\partial a_x} + \frac{\partial L}{\partial \tilde{c}_x} \frac{\partial \tilde{c}_x}{\partial a_x} = \frac{f_x \tilde{c}_{x-1} - \tilde{f}_{x-1}b_x}{(b_x - \tilde{c}_{x-1}a_x)^2} \frac{\partial L}{\partial \tilde{f}_x} + \frac{c_x \tilde{c}_{x-1}}{(b_x - \tilde{c}_{x-1}a_x)^2} \frac{\partial L}{\partial \tilde{c}_x}, \quad (21)$$

$$\frac{\partial L}{\partial b_x} = \frac{\partial L}{\partial \tilde{f}_x} \frac{\partial \tilde{f}_x}{\partial b_x} + \frac{\partial L}{\partial \tilde{c}_x} \frac{\partial \tilde{c}_x}{\partial b_x} = \frac{\tilde{f}_{x-1}a_x - f_x}{(b_x - \tilde{c}_{x-1}a_x)^2} \frac{\partial L}{\partial \tilde{f}_x} + \frac{c_x}{(b_x - \tilde{c}_{x-1}a_x)^2} \frac{\partial L}{\partial \tilde{c}_x}, \quad (22)$$

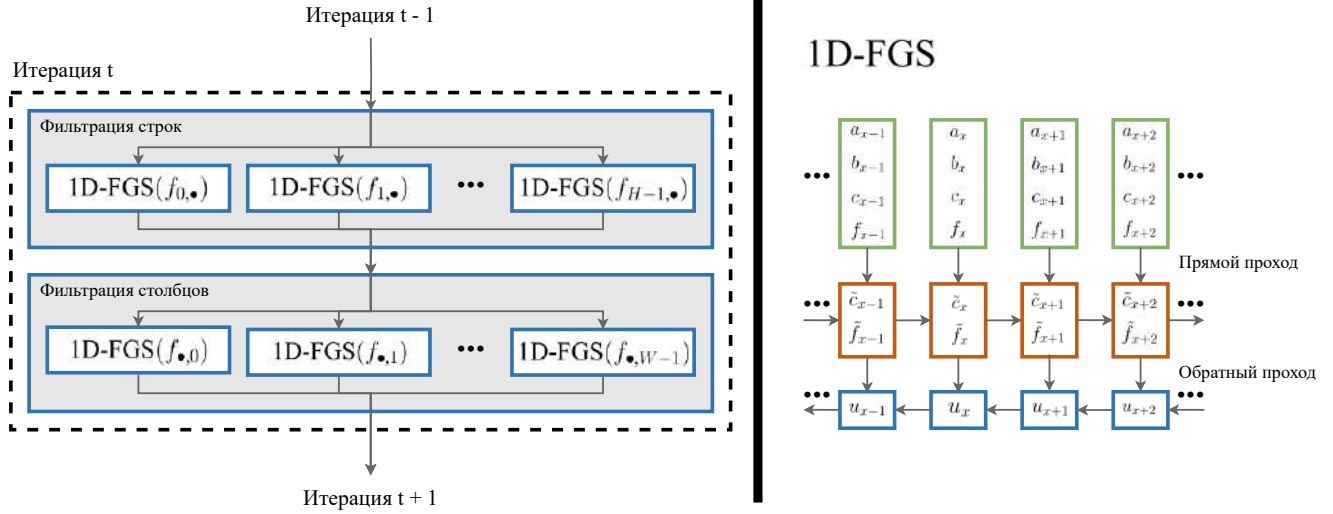


Рисунок 12 — Вычислительный граф двумерного сепарабельного фильтра FGS.

$$\frac{\partial L}{\partial c_x} = \frac{\partial L}{\partial \tilde{c}_x} \frac{\partial \tilde{c}_x}{\partial c_x} = \frac{1}{b_x - \tilde{c}_{x-1} a_x} \frac{\partial L}{\partial \tilde{c}_x}, \quad (23)$$

$$\frac{\partial L}{\partial \omega_{x,x-1}} = \frac{\partial L}{\partial a_x} \frac{\partial a_x}{\partial \omega_{x,x-1}} + \frac{\partial L}{\partial b_x} \frac{\partial b_x}{\partial \omega_{x,x-1}} = \lambda_t \left(\frac{\partial L}{\partial b_x} - \frac{\partial L}{\partial a_x} \right), \quad (24)$$

$$\frac{\partial L}{\partial \omega_{x+1,x}} = \frac{\partial L}{\partial \omega_{x+1,x}} + \frac{\partial L}{\partial b_x} \frac{\partial b_x}{\partial \omega_{x+1,x}} + \frac{\partial L}{\partial c_x} \frac{\partial c_x}{\partial \omega_{x+1,x}} = \frac{\partial L}{\partial \omega_{x+1,x}} + \lambda_t \left(\frac{\partial L}{\partial b_x} - \frac{\partial L}{\partial c_x} \right). \quad (25)$$

Таким образом, одномерный фильтр FGS является дифференцируемым фильтром относительно всех входных сигналов f и w . А представленные в данном доказательстве производные могут быть использованы в методе обратного распространения ошибки. \square

Теорема 2. Пусть фильтруемое изображение $f \in \mathbb{R}^{H \times W \times 3}$. Тогда последовательная сложность двумерного сепарабельного фильтра FGS с T итерациями составляет $\mathcal{O}(T \times H \times W)$, а его параллельная сложность — $\mathcal{O}(T \times \max(H, W))$. Аналогичная оценка на сложности справедлива и для метода обратного распространения ошибки по данному фильтру.

Доказательство. Рассмотрим одну итерацию двумерного сепарабельного фильтра FGS по изображению $f \in \mathbb{R}^{H \times W \times 3}$. Каждая итерация состоит из горизонтального и вертикального проходов, в которых осуществляется применение одномерного фильтра FGS к каждой строке и столбцу изображения

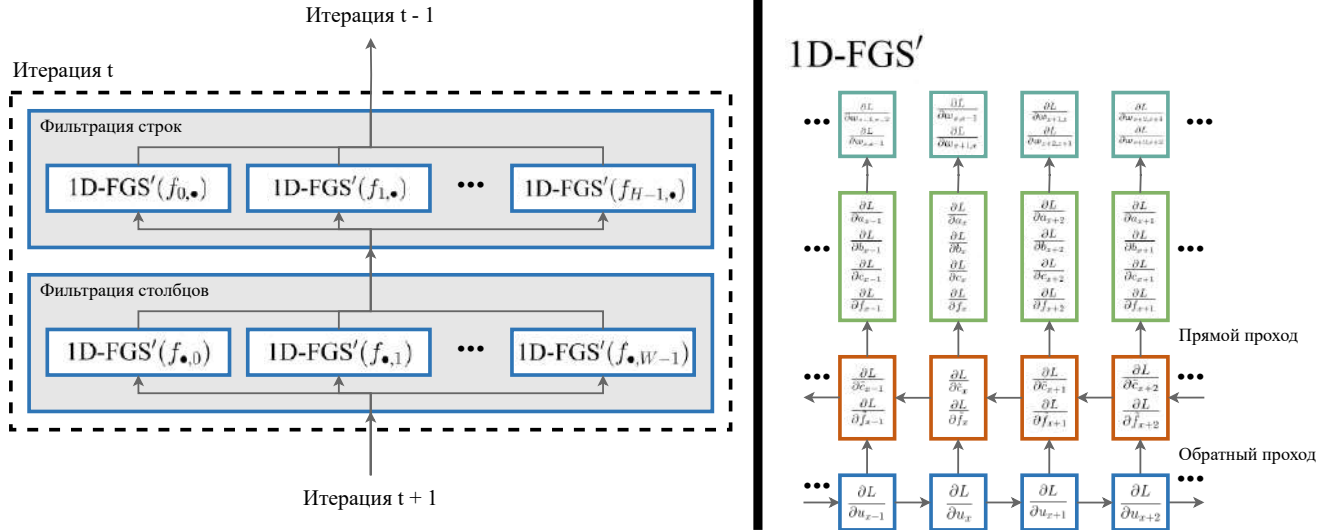


Рисунок 13 — Вычислительный граф метода обратного распространения ошибки для двумерного сепарабельного фильтра FGS.

соответственно. Тогда как показано в [63], сложность фильтрации одной строки составляет $\mathcal{O}(W)$, а сложность фильтрации одного столбца — $\mathcal{O}(H)$. Так как в изображении f всего H строк и W столбцов, то общая сложность одной итерации фильтра составляет $\mathcal{O}(H \times W)$, а следовательно общая последовательная сложность алгоритма — $\mathcal{O}(T \times H \times W)$. Граф вычислений одной итерации двумерного сепарабельного фильтра FGS показан на рисунке 12.

По графу 12 можно заметить, что фильтрация каждой отдельной строки/столбца осуществляется независимо от других строк/столбцов. Тогда, при условии наличия неограниченного числа процессоров эти операции можно распараллелить по числу строк в случае фильтрации строк и числу столбцов в случае фильтрации столбцов. Следовательно параллельная сложность фильтрации всех строк в изображении будет составлять $\mathcal{O}(W)$, а параллельная сложность фильтрации всех столбцов — $\mathcal{O}(H)$. Значит параллельная сложность одной итерации фильтра равна $\mathcal{O}(W) + \mathcal{O}(H) = \mathcal{O}(\max(H, W))$ и общая параллельная сложность алгоритма — $\mathcal{O}(T \times \max(H, W))$.

Рассмотрим последовательную сложность метода обратного распространения ошибки по одномерному фильтру FGS для сигнала размерности N . Аналогично самому фильтру, вычисление его производных осуществляется в два прохода:

1. Обратный проход – сложность $\mathcal{O}(N)$:

- $\frac{\partial L}{\partial \tilde{f}_x}$ не требует вычислений и равен приходящему с последующих слоев градиенту $\frac{\partial L}{\partial u_x}$;
- Вычисление $N - 1$ градиента $\frac{\partial L}{\partial \tilde{c}_x}$ по $\frac{\partial L}{\partial u_x}$;
- Рекурсивное обновление $N - 1$ градиента $\frac{\partial L}{\partial u_{x+1}}$ по предыдущему элементу.

2. Прямой проход – сложность $\mathcal{O}(N)$:

- Вычисление $N - 1$ градиента $\frac{\partial L}{\partial f_x}$ по $\frac{\partial L}{\partial \tilde{f}_x}$;
- Рекурсивное обновление $2(N - 1)$ градиентов $\frac{\partial L}{\partial \tilde{c}_{x-1}}, \frac{\partial L}{\partial \tilde{f}_{x-1}}$ по следующему элементу;
- При необходимости распространения ошибки на направляющее изображение – вычисление $3(N - 1)$ градиентов вспомогательных переменных $\frac{\partial L}{\partial a_x}, \frac{\partial L}{\partial b_x}, \frac{\partial L}{\partial c_x}$ по $\frac{\partial L}{\partial \tilde{f}_x}, \frac{\partial L}{\partial \tilde{c}_x}$;
- При необходимости распространения ошибки на направляющее изображение – вычисление $2(N - 1)$ обновлений градиентов весов $\frac{\partial L}{\partial \omega_{x+1,x}}, \frac{\partial L}{\partial \omega_{x,x-1}}$ по градиентам вспомогательных переменных.

Таким образом, последовательная сложность метода обратного распространения ошибки по одномерному фильтру FGS для сигнала размерности N также составляет $\mathcal{O}(N)$. Тогда, исходя из тех же соображений, что и при оценке сложностей двумерного сепарабельного фильтра FGS, последовательная сложность метода обратного распространения ошибки для него равна $\mathcal{O}(T \times H \times W)$, а параллельная сложность – $\mathcal{O}(T \times \max(H, W))$. Граф вычислений одной итерации метода обратного распространения ошибки для двумерного сепарабельного фильтра FGS показан на рисунке 13. □

Данные теоремы показывают, что двумерный сепарабельный фильтр FGS можно реализовать в качестве вычислительного блока нейронной сети. А также его параллельная сложность позволяет эффективно распараллелить реализацию алгоритма по строкам и столбцам изображения.

В качестве направляющего изображения для данного фильтра в рассматриваемых архитектурах нейронных сетей была использована карта границ исходного изображения. Для вычисления карт границ был использован фильтр

$$\text{Собеля: } g = \sqrt{g_x^2 + g_y^2}, \text{ где } g_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * f \text{ и } g_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * f. \text{ Тогда}$$

веса для оценки похожести соседних пикселей можно задать как $\omega_{x+1,x}(g) = e^{-\sqrt{2}\frac{gx}{\sigma_c}}$, где $\sigma_c > 0$ – параметр, контролирующий степень фильтрации.

Так как с помощью данного фильтра предполагается осуществлять распространение данных по маске карты доверия диспаратности, чтобы исключить из рассмотрения области с низким доверием и распространить на них значения предсказанных карт различий в области высокого доверия, то двумерный сепарабельный фильтр FGS применяется к маскированному входному изображению и самой маске, после чего результат фильтрации по маске вычисляется следующим образом:

$$u_{\text{conf}} = \frac{\text{FGS}(\text{conf} \times f)}{\text{FGS}(\text{conf})}, \quad (26)$$

где conf – карта доверия к диспаратности.

1.5 Обучение нейронных сетей

Для обучения нейросетей был использован набор данных, сгенерированный на основе описанной ранее модели искажений. В качестве оптимизируемого функционала была использована сумма квадратов разности предсказанных и истинных значений, взвешенных на доверие к карте диспаратности, как для карты различий по цвету, так и для карты различий по резкости:

$$L_c(\hat{c}, c) = \frac{\sum_{i=1}^n \text{conf}_i \left((\hat{c}_i^Y - c_i^Y)^2 + (\hat{c}_i^U - c_i^U)^2 + (\hat{c}_i^V - c_i^V)^2 \right)}{3 \sum_{i=1}^n \text{conf}_i}, \quad (27)$$

$$L_d(\hat{d}, d) = \frac{\sum_{i=1}^n \text{conf}_i (\hat{d}_i - d_i)^2}{\sum_{i=1}^n \text{conf}_i}, \quad (28)$$

где \hat{c} , c – предсказанная и истинная карты различий по цвету для каждого цветового канала YUV, \hat{d} , d – предсказанная и истинная карты размытия, conf – карта доверия к диспаратности, используемая в качестве входной карты доверия для нейронной сети, n – количество пикселей в изображении. Дополнительно была использована L_2 -регуляризация для уменьшения эффекта переобучения:

$$L_2(\theta) = \lambda \sum_{i=1}^k \theta_i^2, \quad (29)$$

где θ – веса обучаемой нейросети, $\lambda = 10^{-2}$ – параметр регуляризации, k – общее количество весов в сети. Итоговый оптимизируемый функционал выглядит следующим образом:

$$L(\hat{c}, c, \hat{d}, d, \theta) = L_c(\hat{c}, c) + L_d(\hat{d}, d) + L_2(\theta). \quad (30)$$

Для инициализации весов сверточных слоев в начале обучения был использован метод инициализации Xavier [64]. В качестве метода оптимизации был выбран алгоритм Adam [65]. Нейронные сети обучались в течение 100 эпох. Коэффициент скорости обучения составлял 10^{-4} с уменьшением в 10 раз каждые 40 эпох. Количество примеров из набора данных, используемых на одной итерации обучения, было равно 8, а разрешение используемых при обучении примеров составляло 256×256 . Участки изображений данного размера вырезались случайно во время обучения. Также для дополнительной аугментации данных осуществлялось случайное отражение изображения относительно горизонтальной или вертикальной оси и добавление шума к ракурсам по нормальному распределению с максимальным стандартным отклонением 0.02 и нулевым средним значением. Сходимость предложенных методов продемонстрирована на рисунке 14. Как можно видеть, метод на основе архитектуры GridNet показывает меньшую ошибку на валидационной выборке по сравнению с методом на основе архитектуры U-net.

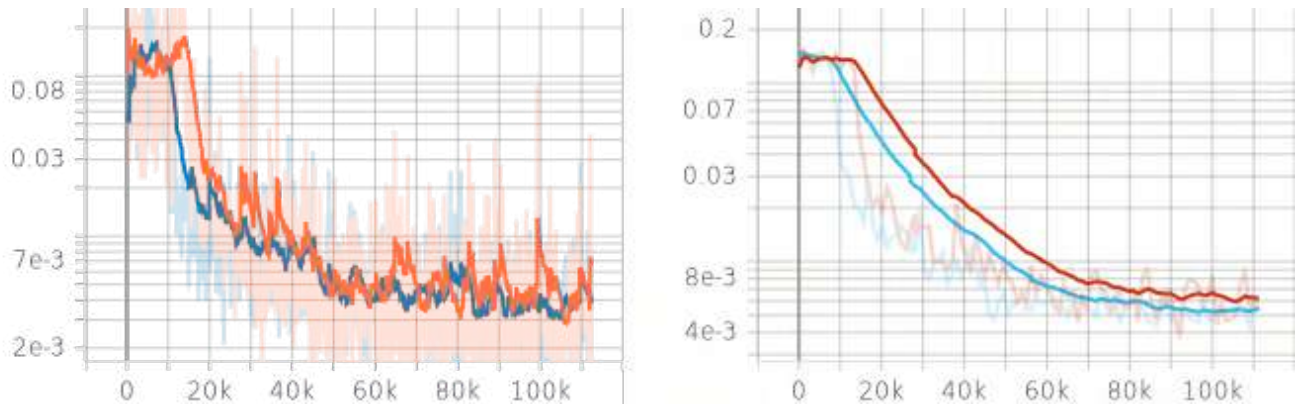


Рисунок 14 — Сглаженные графики оптимизируемого во время обучения функционала на обучающей (слева) и валидационной (справа) выборках. Оранжевая и красная линии соответствуют методу на основе архитектуры U-net, синяя и голубая — методу на основе архитектуры GridNet.

1.6 Экспериментальная оценка

Для тестирования предложенного метода была подготовлена тестовая выборка на основе набора данных Sintel [66]. Sintel содержит в себе 23 стереоскопические видеопоследовательности с разрешением 1024×436 , а также истинные значения оптического потока и диспаратности для каждого кадра. В исходных последовательностях отсутствуют искажения ракурсов по цвету и резкости, так как данные последовательности получены с помощью компьютерной графики. Для подготовки тестовой выборки на основе набора данных Sintel к каждой последовательности добавлялись искусственные искажения на основе представленной ранее общей модели искажений. Каждая последовательность преобразовывалась 3 раза с добавлением искажений разного типа и/или силы. На подготовленном наборе данных были протестированы предложенные нейросетевые методы, а также несколько аналогов, включая методы, ранее применявшиеся для анализа полнометражных стереоскопических фильмов. Результаты тестирования представлены в таблице 3. Предложенные нейросетевые методы превосходят по качеству другие методы как по корреляции Пирсона, так и по корреляции Спирмена. При этом метод на основе архитектуры GridNet превосходит по качеству работы метод на основе архитектуры U-net, хотя проигрывают ему

Таблица 3 — Результаты тестирования методов оценки различий по цвету и резкости между ракурсами стереовидео на искусственном наборе данных Sintel.

Метод	Корреляция Пирсона	Корреляция Спирмена
Искажения цвета		
MAE	0.1254	0.1626
MAE с компенсацией правого ракурса	0.1338	0.2039
Метод [54]	-0.4430	-0.4093
Метод [26]	0.8136	0.8760
Предложенный метод на основе U-net	0.9701	0.9477
Предложенный метод на основе GridNet	0.9696	0.9602
Искажения резкости		
Метод [46]	0.1310	0.0692
Метод [47]	0.9564	0.8047
Метод [45]	0.5176	0.3152
Метод [27]	0.7686	0.6815
Метод [52]	0.8151	0.4488
Предложенный метод на основе U-net	0.9482	0.8318
Предложенный метод на основе GridNet	0.9762	0.9078

по скорости: скорость работы первого метода составляет 0.1649 секунд на кадр, в то время как второго – 0.0549 секунд на кадр.

Также было проведено сравнение качества работы предложенного метода на основе архитектуры GridNet (как лучшего на сравнении на искусственном наборе данных) с методами, применявшимися для анализа полнометражных стереоскопических фильмов, на 100 случайных VR180-видео, скачанных с платформы YouTube. Процесс создания набора VR180-видео представлен в разделе 1.8.1. Каждый метод запускался на центральной части кубической проекции в разрешении 960×960 пикселей, вычисляемой для каждого видео в формате VR180 во время запуска. Для каждого метода были заданы аналогичные пороговые значения на их результаты оценки и по заданным порогам отобраны предполагаемые кадры с искажениями. Далее была осуществлена экспертная

Таблица 4 — Результаты тестирования методов оценки различий по цвету и резкости между ракурсами VR180-видео.

Метод	Количество истинных срабатываний	Количество ложноположительных срабатываний
Искажения цвета		
Метод [26]	601	213
Предложенный метод	600	51
Искажения резкости		
Метод [27]	320	779
Предложенный метод	339	180

оценка полученных стереопар, которые были размечены на две категории: содержащие и не содержащие рассматриваемые артефакты. Результаты оценки представлены в таблице 4. Предложенному нейросетевому методу удалось значительно сократить количество ложноположительных срабатываний, при этом не потеряв кадры, действительно содержащие рассматриваемые артефакты.

1.7 Программная реализация

Предложенный метод для одновременной оценки искажений цвета и резкости между ракурсами стереоскопических видео реализован на языке Python 3 в виде консольного приложения. В данном приложении используются следующие модули-зависимости:

- tensorflow – открытая программная библиотека для машинного обучения, направленная на решение задач построения и тренировки нейронных сетей;
- opencv, skimage – открытые библиотеки, использовавшиеся для работы с изображениями и видео;

- numpy, scipy, matplotlib, noise – многофункциональные открытые библиотеки, использовавшиеся для быстрой обработки, генерации и визуализации данных.
- MSU Motion Estimation для сопоставления ракурсов и оценки движения.

Реализованное консольное приложение включает в себя три основные компоненты для:

- подготовки наборов данных для обучения нейросетевых моделей на основе общей модели искажений цвета и резкости;
- обучения нейросетевых моделей одновременной оценки несоответствий по цвету и резкости между ракурсами стереоскопических видео;
- тестирования и использования на практике обученных моделей для одновременной оценки несоответствий по цвету и резкости.

Также предложенный метод реализован в виде подключаемого модуля к системе оценки качества стереовидео VQMT3D [67]. Основная часть данной системы – хост – реализует задачи по чтению входных ракурсов стереоскопических и VR180-видео; вычислению для считанных ракурсов общих данных, необходимых для работы методов оценки качества; записи результатов работы методов оценки артефактов стереоскопических видео, подключаемых к хосту в виде динамических модулей.

Для чтения входных ракурсов стереоскопических видео в хосте используется OpenCV, тем самым поддерживаются все форматы видео, которые поддерживаются в OpenCV. Также в системе поддерживаются в качестве входных данных скрипты в формате AviSynth и VapourSynth, позволяющие перед анализом качества входного стереоскопического видео провести предобработку ракурсов с помощью одной из поддерживаемых программ для обработки видеоматериалов. В рамках данной работы в хост добавлена дополнительная предобработка входных данных в случае анализа видео в формате VR180: применение кубической проекции к исходному видео для генерации центральной части кадра, пригодной для анализа стереоскопических артефактов.

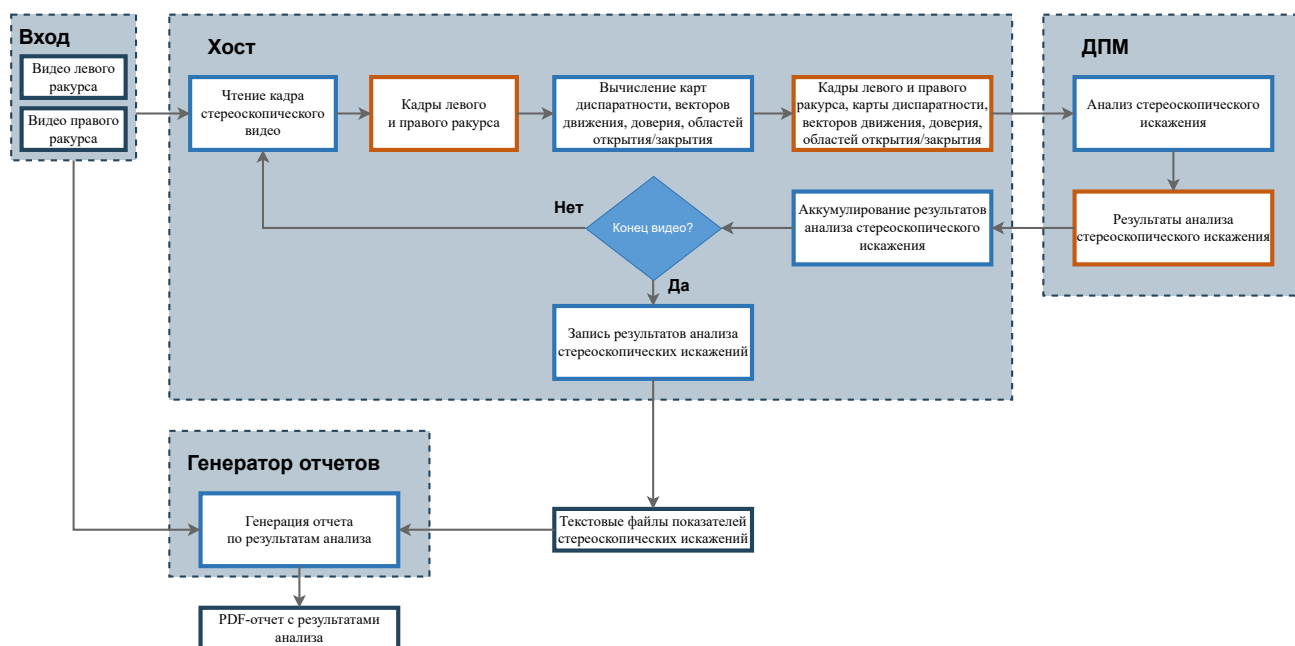


Рисунок 15 — Схема работы системы объективной оценки качества стереоскопических видео VQMT3D.

При анализе искажений стереоскопического видео в хосте осуществляется покадровое чтение левого и правого ракурсов стереоскопического видео и для каждого кадра вычисляются карты диспаратности, карты векторов движения, соответствующие карты доверия, карты областей открытия/закрытия. Далее считанные кадры, а также вычисленные промежуточные данные передаются методам по анализу стереоскопических артефактов, реализованных в виде динамически подключаемых модулей, которые возвращают значения анализируемых показателей.

Полученные результаты оценки значений стереоскопических искажений аккумулируются в хосте, которые далее записываются в отдельные текстовые файлы с покадровыми значениями метрик. Результаты анализа стереоскопических видео, записанные в текстовых файлах, далее используются в системе генерации отчетов. Выпущенные в рамках проекта стереоскопические отчеты представлены на странице https://videoprocessing.ai/stereo_quality/reports/.

Схема работы описанной системы представлена на рисунке 15.

Одной из важных характеристик реализации является ее скорость работы, что влияет на возможность практической применимости. Для кадров с разрешением 1024×436 из тестовой последовательности на основе набора данных Sintel средняя скорость работы предложенного метода составила 0.1649 секунд на кадр. Время работы реализации замерялось на стационарном компьютере с восьмиядерным процессором серии AMD Ryzen 7 1700 и графическим ускорителем GeForce GTX 1070.

1.8 Анализ видео в формате VR180

1.8.1 Подготовка набора видео в формате VR180

Для проведения масштабного анализа видео в формате VR180 было собрано 1000 видео с платформы YouTube. Для увеличения разнообразия выборки сбор осуществлялся по 36 запросам: по запросу на каждую английскую букву и на каждую цифру от 0 до 9. Для сбора видео только в формате VR180 устанавливался соответствующий фильтр на выдачу в YouTube. Для каждого запроса отбирались видео с первых 5-10 страниц результатов поиска. При этом исключались видео, недоступные для загрузки, обладающие низким разрешением и/или являющиеся не стереоскопическими.

На рисунках 16, 17 представлены распределения собранных видео по числу просмотров на YouTube и по длительности (в секундах) соответственно. Ось x на обоих графиках логарифмическая. У большинства отобранных видео от 10000 до 100000 просмотров, но также встречаются видео с несколькими миллионами просмотров. При этом длительность большинства видео находится в диапазоне от 5 до 10 минут. Также на рисунках 18, 19 можно увидеть распределения видео относительно соотношений “лайков” и “дизлайков”. Сбор видео

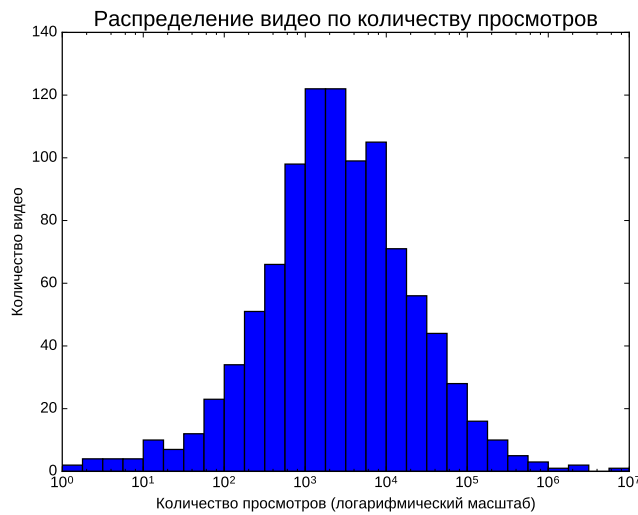


Рисунок 16 — Распределение отобранных видео по числу просмотров.

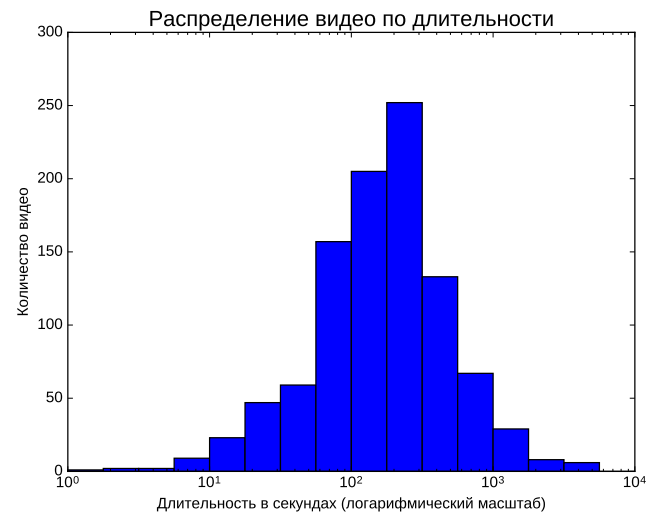


Рисунок 17 — Распределение отобранных видео по их длительности.

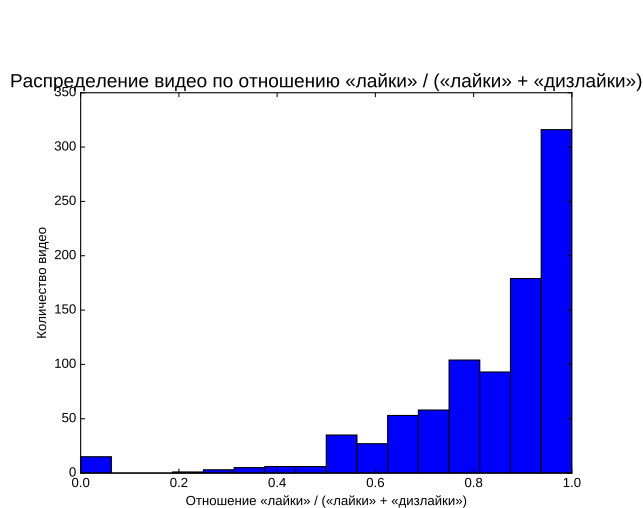


Рисунок 18 — Распределение отобранных видео по соотношению $\frac{likes}{likes+dislikes}$.

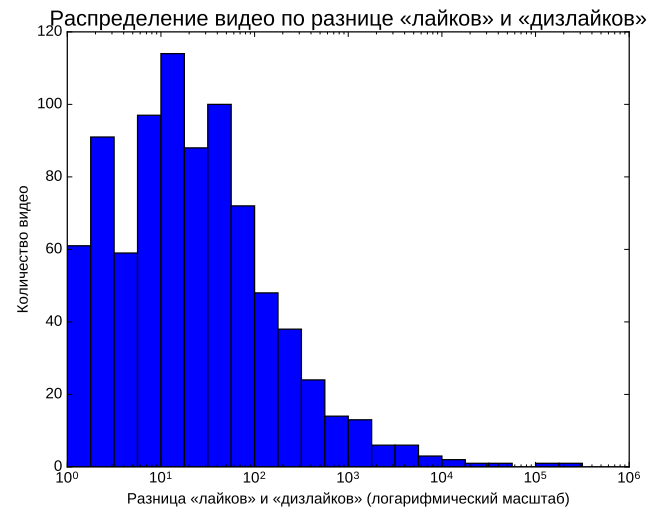


Рисунок 19 — Распределение отобранных видео по соотношению $likes - dislikes$.

осуществлялся до середины 2020 года, а в дальнейшем производилась работа по их обработке и оценке качества.

1.8.2 Предобработка собранных видео в формате VR180

Все кадры видео в формате VR180 изначально представлены в равнопромежуточной проекции. При анализе кадров непосредственно в этой проекции могут возникнуть проблемы как с сопоставлением ракурсов, так и с оценкой геометрических искажений, так как данная проекция вносит дополнительные нелинейные геометрические искажения, усиливающиеся при движении от центра кадра к его краям. Поэтому для корректного анализа видео в формате VR180 все кадры преобразуются к кубической проекции. Пример такого преобразования представлен на рисунке 20. Так как поле зрения исходного видео составляет 180 градусов, грани кубической проекции кадра заполнены следующим образом: верхняя, нижняя и боковые – только на половину, фронтальная – полностью, а задняя не заполнена совсем. Для дальнейшего анализа на предмет стереоскопического качества отбирается только фронтальная грань кубической проекции, так как она заполнена полностью и содержит большинство информации из исходного кадра, а также лишена геометрических искажений. Все эти факторы значительно упрощают применение методов оценки стереоскопического качества и дают возможность получить правильные результаты анализа. Далее предобработанные кадры V180 видео обрабатываются таким же способом как обычные стереоскопические кадры.

1.8.3 Результаты оценки искажений цвета и резкости для видео в формате VR180

Для всех 1000 видео в формате VR180 было проведено измерение силы цветовых искажений (рисунок 21) и искажений резкости (рисунок 22). Результаты анализа продемонстрированы относительно количества просмотров на YouTube



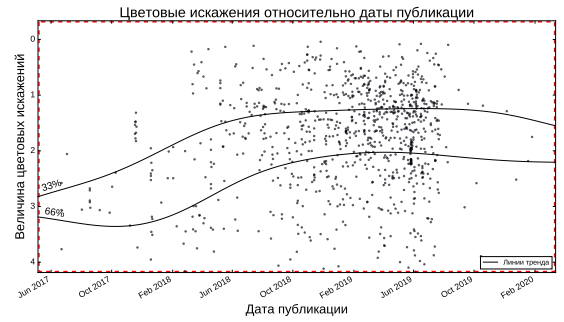
Рисунок 20 — Предобработка кадров видео в формате V180. Красным квадратом выделена фронтальная грань кубической проекции, непосредственно используемая при анализе качества VR180-видео.

(а), даты публикации (б) и длительности (в) каждого видео. Ось x на этих графиках соответствует конкретной статистике видео, а ось y – оцененной величине стереоскопического искажения. Синими точками изображены отдельные видео. Также графики включают в себя две линии тренда: верхняя линия соответствует 33 перцентилю, а нижняя – 66 перцентилю. Ни один из рассматриваемых стереоскопических артефактов не демонстрирует какой-либо существенной тенденции по отношению к любой статистике видео: на некоторых графиках присутствуют небольшие увеличивающиеся либо уменьшающиеся тренды, однако средние вычисленные значения искажений изменяются незначительно. Внезапные спуски и подъемы появляются слева и справа на некоторых графиках, но они в основном связаны с небольшим количеством видео с соответствующими статистиками. Данные графики позволяют сделать следующие заключения:

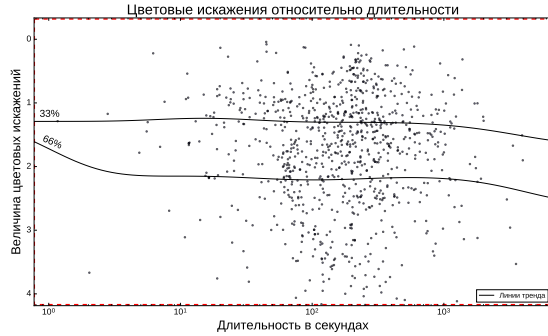
- Вычисленные значения искажений для видео с большим количеством просмотров на YouTube в среднем такие же как и у видео с небольшим количеством просмотров;



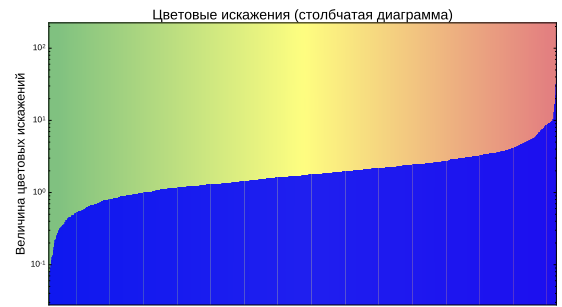
(а) Общие результаты анализа относительно количества просмотров на YouTube.



(б) Общие результаты анализа относительно даты публикации видео.



(в) Общие результаты анализа относительно длительности видео.



(г) Средние значения искажений для каждого видео.

Рисунок 21 — Результаты анализа цветowych искажений в VR180-видео.

- В целом, ситуация не изменялась со временем, так как видео, опубликованные позже, обладают в среднем теми же оценками стереоскопических артефактов, что и видео, опубликованные гораздо раньше;
- Средние значения искажений в видео независимы от их длительности.

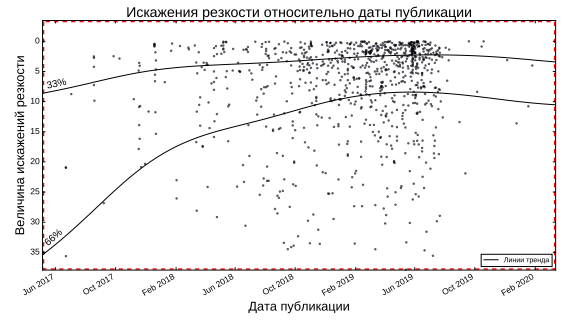
В приложении на рисунках 55-58 представлены найденные в VR180 примеры искажений цвета, на рисунках 59-62 – искажений резкости.

1.9 Заключение

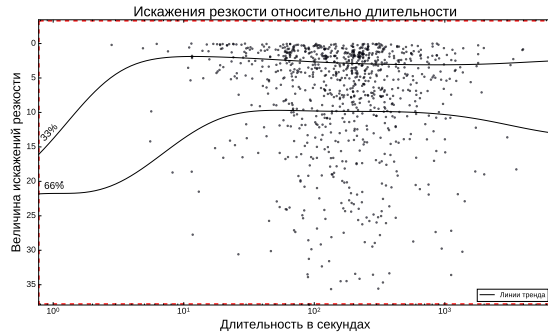
В данной главе была предложена общая модель искажений цвета и резкости, на основе которой был подготовлен набор данных для обучения и валидации нейросетевого метода оценки искажений. Рассматривались как простые, глобальные случаи искажений, так и более сложные, локальные случаи. Всего



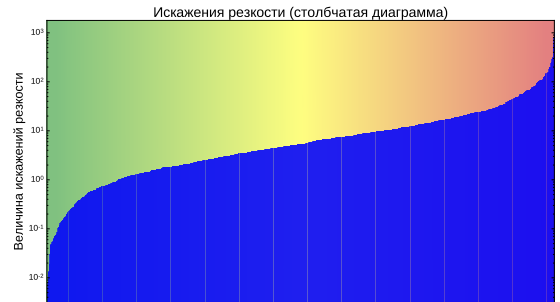
(а) Общие результаты анализа относительно количества просмотров на YouTube.



(б) Общие результаты анализа относительно даты публикации видео.



(в) Общие результаты анализа относительно длительности видео.



(г) Средние значения искажений для каждого видео.

Рисунок 22 — Результаты анализа искажений резкости в VR180-видео.

в набор данных вошло 9488 различных стереопар, вырезанных из 16 стереоскопических фильмов. Для оценки цветовых искажений и искажений резкости была использована модифицированная модель GridNet, предсказывающая карты различий по цвету и резкости на основе левого ракурса, интерполированного к нему по карте диспаратности правого ракурса и соответствующей карте доверия. Итоговый метод аккумулирует предсказанные значения карт искажений для общей оценки силы искажений в кадре. Разработанный нейросетевой алгоритм оценки различий по цвету и резкости между ракурсами стереоскопического видео был протестирован на искусственной тестовой выборке, показав значительное улучшение точности работы по сравнению с аналогами, ранее использовавшимися при анализе полнометражных стереоскопических фильмов. Так для оценки различий по цвету было достигнуто увеличение значений корреляций Пирсона на 15% и Спирмена на 10%, а для оценки различий по резкости корреляции в среднем были улучшены на 20%.

Также в главе представлены результаты анализа искажений цвета и резкости для 1000 видео в формате VR180, собранных с платформы YouTube. Ни полученные показатели искажений цвета, ни полученные показатели искажений резкости не показали существенных трендов по отношению к таким статистикам видео, как количество просмотров, дата публикации и длительность.

Результаты главы опубликованы автором в [2; 3; 17; 22; 24].

Глава 2. Нейросетевой метод оценки геометрических искажений между ракурсами стереовидео

2.1 Постановка задачи

Аналогично искажениям цвета и резкости при съемке стереоскопических видео и видео в формате VR180 также достаточно часто возникают геометрические искажения между ракурсами. Среди них можно выделить постоянный вертикальный сдвиг (рисунок 23, слева), поворот (рисунок 23, посередине) и масштабирование (рисунок 23, справа) одного ракурса относительно другого. Пример вертикального сдвига одного из ракурсов полнометражного стереоскопического фильма представлен на рисунке 24. Помимо данных артефактов также могут возникать и другие геометрические артефакты, например, перспективные искажения и конвергированные оптические оси [21], которые также приводят к возникновению вертикальных смещений пикселей в стереопаре, однако для их точного определения необходимо знать внутренние параметры калибровки камер. Доступом к съемочной аппаратуре обладают операторы фильма, и оценку параметров можно произвести во время съемки стереоскопического контента, но в других ситуациях точная оценка данных параметров невозможна. В данной работе источником данных являются только ракурсы отснятого стереоскопического видео. Поэтому упомянутые выше артефакты не рассматриваются.

Рассматриваемые геометрические искажения (рисунок 23) в первую очередь возникают из-за неправильной калибровки камер. Небольшие несоответствия в вертикальных положениях камер или небольшие наклоны приводят к появлению данных артефактов. Если при производстве стереоскопического контента контролю технического качества уделяется недостаточно внимания, такие артефакты зачастую попадают в финальный продукт. При просмотре стерео-

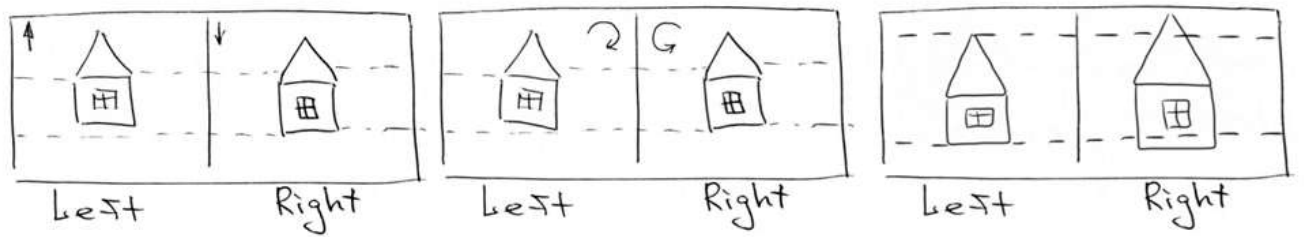


Рисунок 23 — Схематичная иллюстрация типов геометрических искажений, возникаемых при съемке 3D-видео.



Рисунок 24 — Пример вертикального сдвига между ракурсами стереофильма.
Кадр из фильма “Убойное Рождество Гарольда и Кумара”.

скопических сцен с данными артефактами, зритель часто испытывает дискомфорт [1], вплоть до головных болей, тошноты и головокружения, аналогично искажениям цвета и резкости.

Формально задачу оценки геометрических искажений можно поставить следующим образом. Пусть имеются две видеопоследовательности, являющиеся левым и правым ракурсами стереоскопического видео. Необходимо для каждого кадра стереовидео оценить 3 параметра геометрических искажений – константный вертикальный сдвиг, поворот и масштабирование правого ракурса относительно левого. Используя обозначения из раздела 1.1, пусть на вход алгоритму подается стереовидео S . На выходе алгоритм должен предоставить оценки трех геометрических искажений между левым и правым ракурсом

$\{\alpha_i, k_i, t_i\}_{i=1}^k$, $\alpha_i, k_i, t_i \in \mathbb{R}$. Вид оценок также зависит от рассматриваемой модели искажений, представленной в разделе 2.3.

2.2 Обзор существующих алгоритмов

Геометрические искажения между ракурсами стереоскопических видео порождают вертикальные смещения пикселей в одном из ракурсов относительно другого. Поэтому для оценки геометрических искажений необходимо осуществить анализ данных смещений, а алгоритмы оценки геометрических искажений можно разделить на две явные части: сопоставление ракурсов для оценки возникаемых вертикальных смещений и оценку некоторой модели геометрических искажений. Для сопоставления ракурсов могут использоваться различные методы: поиск и сопоставление особых точек в ракурсах [28—30], оценка оптического потока между ракурсами [31—33], блочное сопоставление ракурсов [34]. Однако эти методы должны быть устойчивыми к искажениям цвета и резкости между ракурсами стереовидео, которые часто возникают при съемке в 3D.

Непосредственно геометрические искажения могут оцениваться как с использованием общей модели, так и отдельно друг от друга, что может улучшить распознавание конкретного искажения, однако не будет учитывать влияние других артефактов на рассматриваемое искажение. Также в отдельный класс алгоритмов, оценивающих параметры геометрических искажений, выносятся алгоритмы ректификации ракурсов стереовидео, направленные на устранение всех геометрических искажений в стереопаре. В ходе ректификации стереопары оцениваются параметры преобразования ракурсов, из которых также можно получить оценку исходных геометрических искажений.

2.2.1 Методы ректификации стереовидео

Для ректификации стереопары необходимо привести ракурсы друг к другу таким образом, чтобы выполнялось эпполярное ограничение. Другими словами, два ракурса стереопары считаются ректифицированными, если все соответствующие друг другу точки в обоих изображениях лежат на одной и той же горизонтальной оси. Задача ректификации стереопары с геометрическими искажениями сводится к оценке матриц камер для каждого ракурса, составляющих искомую фундаментальную матрицу, и преобразованию ракурсов с использованием оцененных матриц.

В работе [68] оценка матриц камер осуществляется путем минимизации функции энергии с шестью степенями свободы. В методе [69] предлагается декомпозировать искомое преобразование на две части: проективное преобразование и аффинное преобразование. Проективное преобразование отображает эпполярные линии на бесконечность, в то время как аффинное преобразование вращает и переносит эпполярные линии так, чтобы они были параллельны оси x . В работе [70] рассматривается проблема ректификации стереопары при наличии различных фокусных расстояний и решается задача нелинейной условной оптимизации. Большинство методов ректификации основано на использовании соответствий, полученных с помощью методов поиска особых точек, таких как SIFT [28], SURF [29] и ORB [30], что сильно ограничивает возможности ректификации в виду сильной разреженности полученных соответствий для определенных типов кадров. Для устранения этого недостатка в последнее время начали появляться нейросетевые методы ректификации [71], в которых поиск соответствий для процесса ректификации осуществляется за счет использования нейронных сетей, обучаемых непосредственно на данной задаче.

Однако для полноценной работы методов ректификации стереовидео требуются калибровочные параметры камер. Также данные методы имеют тенденцию к ухудшению результатов работы при наличии шума в данных. Данные

факторы серьезно ограничивают применимость методов ректификации на практике, как для исправления стереоскопических видео, так и для простой оценки параметров геометрических искажений.

2.2.2 Методы, оценивающие параметры геометрических искажений

Для оценки параметров геометрических искажений в научной литературе был предложен ряд методов. В работе [72] осуществляется одновременное вычисление двух параметров геометрических искажений: вертикального сдвига и относительного масштабирования. Для этого осуществляется оценка параметров следующей модели с помощью метода наименьших квадратов на основе сопоставлений, полученных с помощью метода SIFT [28]:

$$y_R = y_L k + t, \quad (31)$$

где y_L , y_R – вертикальные координаты соответствующих друг другу точек в левом и правом ракурсе соответственно, k – параметр масштабирования, t – параметр сдвига. Аналогичный метод используется в работе [73], где вместо поиска особых точек используется блочное иерархическое сопоставление ракурсов и оценивается сразу три геометрических искажения: сдвиг, масштабирование и поворот:

$$y_R = [\alpha \quad k] \begin{bmatrix} x_L \\ y_L \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}, \quad (32)$$

где x_L , y_L – координаты точки в левом ракурсе, y_R – вертикальная координата соответствующей точки в правом ракурсе, α – параметр поворота, k – параметр масштабирования, t_x, t_y – параметры сдвига. Однако использование метода наименьших квадратов для оценки параметров модели неустойчиво к шуму в исходных данных, что снижает практичность данных методов.

В некоторых работах предлагается оценивать геометрические искажения независимо друг от друга. Так, в работе [74] оценивается вертикальный сдвиг и поворот одного ракурса относительно другого путем многоступенчатой медианной фильтрации на результатах работы алгоритма блочного сопоставления ракурсов. Аналогичный подход применяется в работе [55] на основе результатов сопоставления особых точек SIFT. Для оценки вертикального сдвига в рассматриваемом методе анализируется гистограмма вертикальных составляющих векторов сопоставлений; для оценки масштабирования используются параметр масштабирования сопоставленных точек из SIFT; для оценки поворота выбирается такой угол поворота, который минимизирует разницу между повернутым левым ракурсом и исходным правым ракурсом. Однако данные подходы также малопрактичны, так как в присутствии других искажений либо одновременном присутствии нескольких рассматриваемых искажений результаты оценки параметров геометрических искажений будут недостоверны.

В целом, для оценки параметров заданной модели геометрических искажений можно использовать любой оптимизационный метод при условии его устойчивости к шуму и выбросам в исходных данных, которые часто встречаются при сопоставлении ракурсов. Например, для этих целей подходит метод RANSAC [75], а также его модификации, которые в последнее время берут за основу нейросетевые подходы [76; 77]. Также нейронные сети начинают использовать и на других этапах методов оценки параметров. Так, в работах [78; 79] предлагается вычислять дополнительные веса с помощью нейронной сети для полученных сопоставлений перед оценкой модели. Данный шаг может быть использован как дополнительный этап для лучшей фильтрации сопоставлений между ракурсами стереовидео. Также в работе [80] предлагается заменить этап сопоставления стереоскопических ракурсов на вычисление полных корреляций между двумя картами признаков, полученных с помощью нейронной сети. После полного сопоставления карт признаков авторы работы добавляют регрессионную нейронную сеть, которая вычисляет матрицу аффинного преобразования для сведения левого ракурса к правому. Модификация данного метода подразумевает обуче-

ние регрессионной нейронной сети для предсказания матрицы произвольного геометрического преобразования. В своей следующей работе [81] авторы дополнительно предлагают оценивать матрицу проективного преобразования.

Данная работа базируется на методе [26], который ранее использовался для анализа 105 полнометражных стереоскопических фильмов. Рассматриваемый метод оценивает параметры аффинного преобразования с помощью метода RANSAC на основе результатов блочного сопоставления ракурсов. Но в виду случайной природы работы метода RANSAC, результаты вычислений параметров искажений могут быть нестабильны. Поэтому в данной работе предлагается использовать нейросетевой регрессор для непосредственной оценки параметров геометрических искажений.

2.3 Модель геометрических искажений

Пусть I_{gt}^L и I_{gt}^R – соответственно левый и правый ракурсы стереопары, не содержащей геометрических искажений. Для моделирования описанных искажений достаточно применить аффинное преобразование к одному из ракурсов. Пусть $p = [x \ y \ 1]^T$ и $p' = [x' \ y' \ 1]^T$ – однородные координаты двух точек до и после применения преобразования соответственно. Тогда геометрические несоответствия между ракурсами стереовидео можно промоделировать следующим аффинным преобразованием:

$$p' = A p, \quad (33)$$

$$A = \begin{bmatrix} (1+k)\cos(\alpha) & -(1+k)\sin(\alpha) & 0 \\ (1+k)\sin(\alpha) & (1+k)\cos(\alpha) & t \\ 0 & 0 & 1 \end{bmatrix}, \quad (34)$$

где α – угол поворота, k – коэффициент масштабирования, t – вертикальный сдвиг. Пусть $G \in \mathbb{R}^{h \times w \times 3}$ – нормированные однородные координаты пикселей ракурса. Предполагается, что $\forall p = [x \ y \ 1]^T \in G$, $x \in [-1, 1)$, $y \in [-1, 1)$,

то есть нормированная координата $(-1, -1, 1)$ соответствует координате пикселя $(0, 0)$ в изображении, а $(1, 1, 1)$ – координате пикселя (w, h) . Для получения нормированной координатной сетки из ненормированной достаточно применить

преобразование $G = G_p \times N^T$, где $N = \begin{bmatrix} 2/W & 0 & -1 \\ 0 & 2/H & -1 \\ 0 & 0 & 1 \end{bmatrix}$, а для обратного пре-

образования – $G_p = G \times N_r^T$, где $N_r = \begin{bmatrix} W/2 & 0 & W/2 \\ 0 & H/2 & H/2 \\ 0 & 0 & 1 \end{bmatrix}$. Тогда $G' = G \times A^T$ –

нормированные однородные координаты пикселей трансформированного ракурса. Следовательно, для получения стереопары с геометрическими искажениями достаточно интерполировать пиксели исходного правого ракурса по трансформированным координатам:

$$I^R = \text{Warp} (I_{gt}^R, G'), \quad (35)$$

где Warp – операция интерполяции изображения по координатной сетке. Таким образом, в работе рассматривается задача оценки параметров $\theta = \begin{bmatrix} \alpha & k & t \end{bmatrix}$, задающих моделирующее аффинное преобразование.

На основе предложенной модели был подготовлен набор данных для обучения, валидации и тестирования нейросетевых моделей. При создании набора данных использовались кадры из 29 полнометражных стереоскопических фильмов. Список из 16 стереофильмов, представленных в разделе 1.3, был расширен следующими фильмами:

1. “Врата дракона”;
2. “Фантом”;
3. “Сумасшедшая езда”;
4. “Другой мир: Пробуждение”;
5. “Дети шпионов 4D”;
6. “Челюсти 3D”;
7. “История дельфина”;

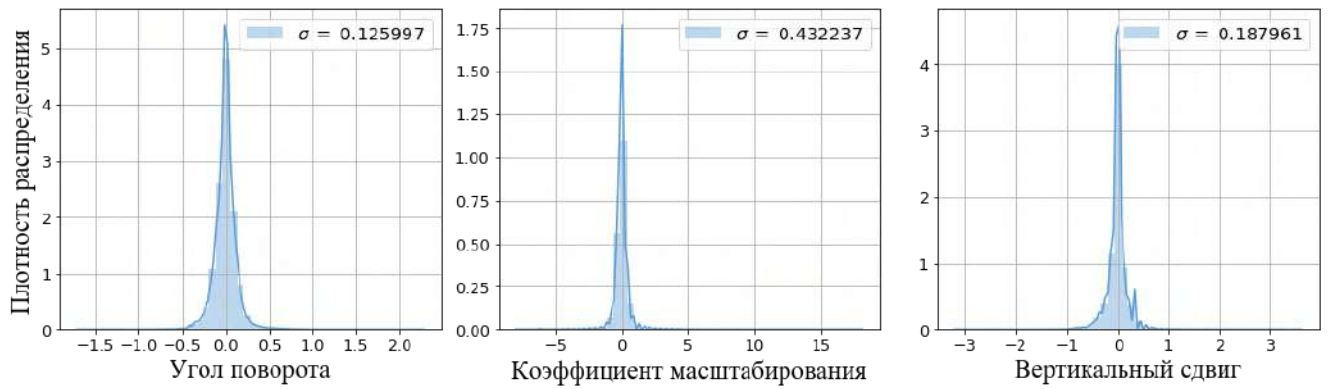


Рисунок 25 — Распределения значений вычисленных геометрических искажений в проанализированных 3D-фильмах с соответствующими стандартными отклонениями.

8. “Пираты Карибского моря: На странных берегах”;
9. “Хранитель времени”;
10. “Повелитель стихий”;
11. “Мушкетёры”;
12. “Санктум”;
13. “Тор”.

Так как эталонные значения геометрических искажений для полнометражных стереоскопических фильмов неизвестны, для их оценки был использован метод [26], ранее применявшийся при анализе стереофильмов. Данный метод также оценивает геометрические искажения по трем параметрам: поворота, масштабирования и вертикального сдвига одного ракурса относительно другого. На рисунке 25 представлены распределения значений данных искажений, вычисленные по всем проанализированным фильмам. Для каждого рассматриваемого искажения было вычислено стандартное отклонение его распределения σ , и в набор данных добавлялись только те стереопары, у которых все значения искажений имели абсолютное значение меньше чем $\frac{\sigma}{10}$. Для избежания дублирования сцен в выборке кадры выбирались с определенным шагом. Всего было собрано 22800 стереопар.

Весь подготовленный набор данных был разделен на три части: обучающую (15500 стереопар), валидационную (3600 стереопар) и тестовую (3700

стереопар, фильмы в тестовой части не пересекаются с фильмами из обучающей и валидационной частей). Далее каждая отобранная стереопара искажалась путем применения аффинного преобразования к правому ракурсу согласно выражению 35. Параметры аффинного преобразования генерировались по нормальному распределению с нулевым средним и стандартным отклонением в 5 раз больше вычисленного выборочного стандартного отклонения. Такой выбор стандартного отклонения расширяет охват возможных значений искажений, встречающихся в полнометражных стереофильмах. Каждая стереопара из обучающего набора данных искажалась 9 раз для расширения выборки, а также в выборку добавлялись неискаженные варианты стереопар. Стереопары из валидационного и тестового наборов данных добавлялись ровно один раз, в 20% случаев из которых искажения не применялись.

2.4 Описание предложенного алгоритма

Для оценки геометрических искажений между ракурсами стереоскопического видео предлагается подход, состоящий из следующих шагов:

1. Вычисление карт диспаратности между левым и правым ракурсами, а также построение карт доверия к вычисленным картам, характеризующих точность оцененных значений диспаратности. Данный шаг полностью совпадает с первым шагом предложенного нейросетевого метода оценки искажений цвета и резкости, описанном в разделе 1.4.1.
2. Оценка параметров геометрических искажений с помощью нейронной сети на основе карты диспаратности, а также соответствующей ей карте доверия.

Общая схема метода представлена на рисунке 26. Вычисленные параметры геометрических искажений могут далее быть использованы для исправления ракурсов путем применения обратного аффинного преобразования к правому ра-

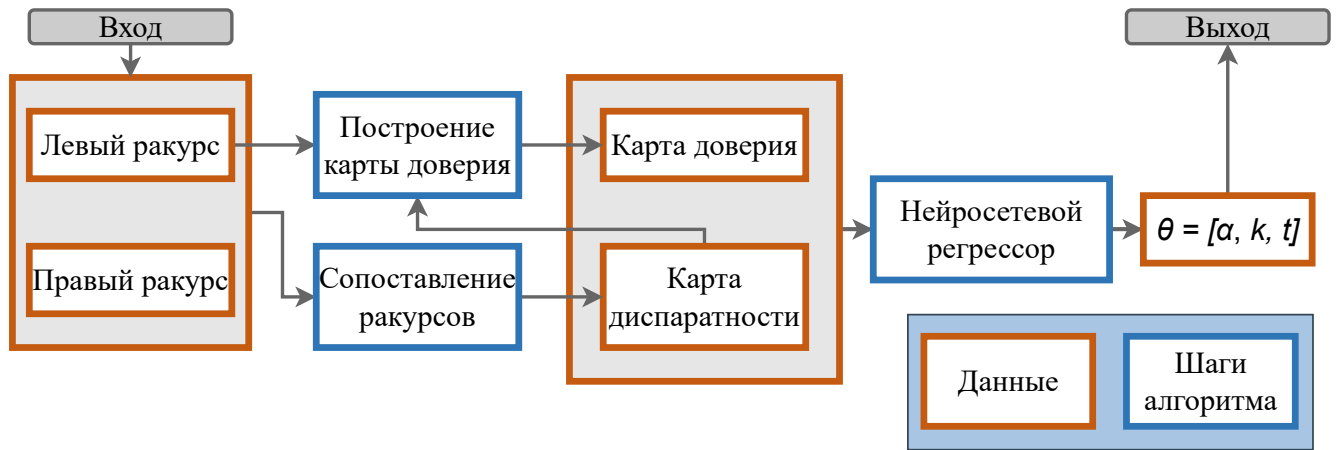


Рисунок 26 — Общая схема предложенного метода оценки геометрических искажений между ракурсами стереоскопического видео.

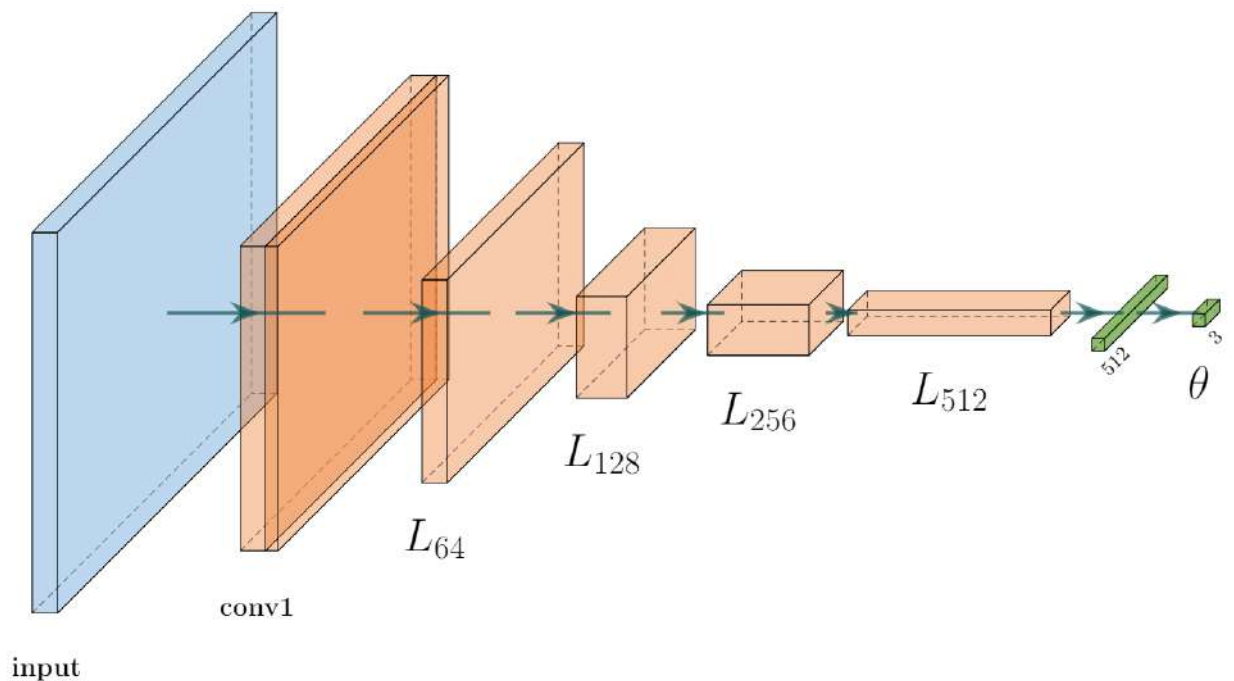


Рисунок 27 — Предложенная архитектура нейронной сети для предсказания геометрических искажений.

курсу. Данная процедура усложняется в случае исправления геометрических искажений в VR180-видео. Подробнее процесс исправления описан в разделе 2.7.

Для оценки параметров геометрических искажений используется нейросетевая архитектура, аналогичная ResNet-18 [82]. Данная архитектура проиллюстрирована на рисунке 27. В начале входной тензор обрабатывается сверточным слоем размера 7×7 с шагом 2. Далее следуют четыре последовательных

остаточных блоков L_k , где $k \in \{64, 128, 256, 512\}$ задает количество карт признаков, вычисляемых в сверточных слоях рассматриваемого блока. На рисунке 28 показана детальная архитектура каждого блока. Количество подблоков в каждом блоке было выбрано $M = 4$. Уменьшение пространственного разрешения осуществляется за счет использования сверточного слоя с шагом 2 в первом подблоке каждого блока. Два последних слоя в сети – полносвязные, при этом последний слой вычисляет вектор $\theta \in \mathbb{R}^3$, содержащий параметры предсказанных геометрических искажений. В качестве входных данных используется нормированная карта диспаратности, а также соответствующая карта доверия. Каждое значение смещения в карте диспаратности $(\Delta x_i, \Delta y_i)$ нормируется следующим образом:

$$(\Delta x'_i, \Delta y'_i) = \left(\frac{2\Delta x_i}{w}, \frac{2\Delta y_i}{h} \right), i = \overline{1, n}, n = h \times w, \quad (36)$$

где h – высота изображения, w – ширина изображения. Пространственные размеры входного тензора могут быть произвольными – перед полносвязными слоями используется глобальный слой субдискретизации с выбором среднего. Также в отличие от исходной архитектуры в используемой модели не используется батч-нормализация [83]. Использование батч-нормализации приводило к ухудшению сходимости модели и замедлению скорости обучения.

2.5 Обучение нейронных сетей

Для обучения нейронных сетей, оценивающих геометрические искажения, предлагается оптимизировать следующий функционал:

$$L(\theta, \theta_{gt}, I^R, I_{gt}^R, \theta_b) = L_{SE}(\theta, \theta_{gt}) + L_{Grid}(\theta, \theta_{gt}) + L_{Warp}(\theta, I^R, I_{gt}^R) + L_{Siam}(\theta, \theta_b), \quad (37)$$

где θ – вычисленные нейросетью значения геометрических искажений по картам диспаратности и доверия для левого ракурса, θ_{gt} – эталонные значения геометрических искажений, I^R и I_{gt}^R – правый ракурс стереопары, содержащий

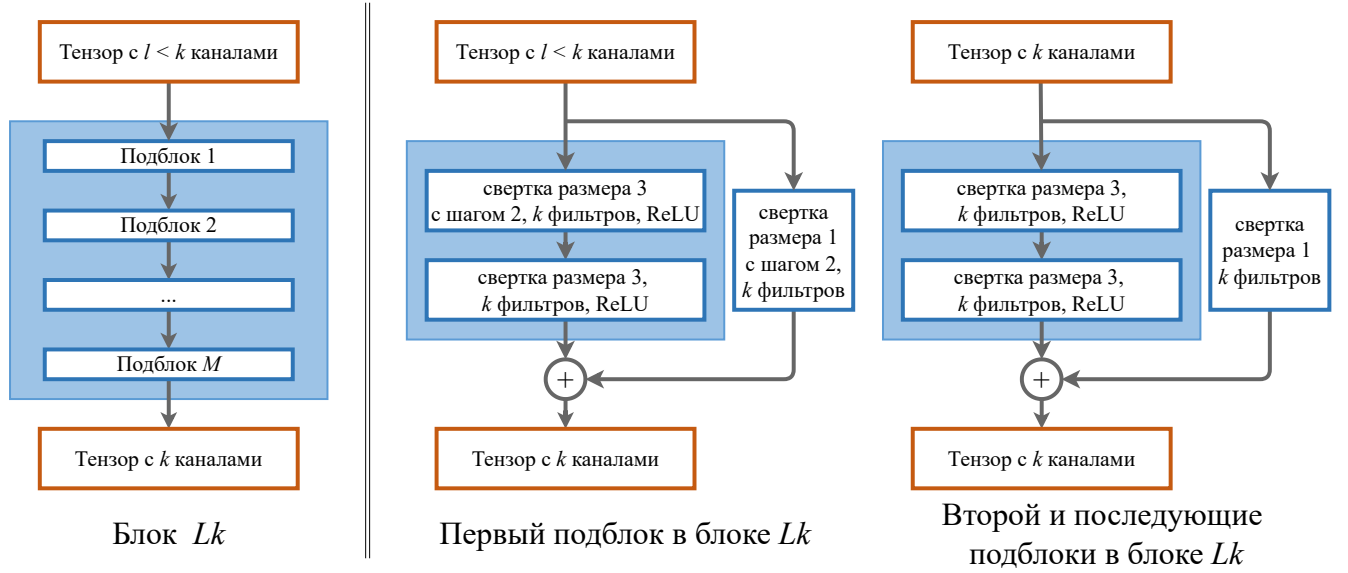


Рисунок 28 — Конфигурация блоков нейронной сети для предсказания геометрических искажений.

и несодержащий геометрические искажения соответственно, θ_b — вычисленные нейросетью значения геометрических искажений по картам диспаратности и доверия для правого ракурса. Данный функционал состоит из двух основных компонент (первые две компоненты) для обучения модели по эталонным значениям геометрических искажений, а также из двух регуляризационных компонент (последние две компоненты), для которых не требуются эталонные значения искажений.

Первая компонента оптимизируемого функционала, L_{SE} , представляет собой взвешенную сумму квадратичных разниц между вычисленными и эталонными значениями геометрических искажений с эмпирически подобранными весами для каждого типа искажений:

$$L_{SE}(\theta, \theta_{gt}) = w_\alpha (\alpha - \alpha_{gt})^2 + w_k (k - k_{gt})^2 + w_t (t - t_{gt})^2, \quad (38)$$

где $w_\alpha = 1$, $w_k = 10^4$ и $w_t = 10^4$.

Вторая компонента, L_{Grid} , вычисляет функцию потерь между двумя сетками, преобразованными с помощью аффинных преобразований, построенных по вычисленным и эталонным значениям геометрических искажений. Пусть $G \in \mathbb{R}^{H \times W \times 3}$ — однородные координаты точек на плоскости. Для вычисления данной компоненты были выбраны равноудаленные координаты на плоскости

$[-1; 1] \times [-1; 1]$ с шагом 0.1, таким образом $H = W = 21$. При вычислении L_{Grid} вектор параметров геометрических искажений разбивается на три различных вектора: $\theta^{rotate} = [\alpha \ 0 \ 0]$, $\theta^{scale} = [0 \ k \ 0]$, $\theta^{shift} = [0 \ 0 \ t]$. Далее последовательно применяется каждое аффинное преобразование к исходной сетке G как по вычисленным с помощью нейросетевого регрессора значениям, так и по эталонным значениям для генерации новых сеток, соответствующих одному из геометрических искажений:

$$G^\alpha = T(G, \theta^\alpha), G_{gt}^\alpha = T(G, \theta_{gt}^\alpha), \quad (39)$$

$$G^k = T(G^\alpha, \theta^k), G_{gt}^k = T(G_{gt}^\alpha, \theta_{gt}^k), \quad (40)$$

$$G^t = T(G^k, \theta^t), G_{gt}^t = T(G_{gt}^k, \theta_{gt}^t), \quad (41)$$

где $T(G, \theta) = G \times A^T(\theta)$ – операция применения аффинного преобразования A с параметрами θ к сетке однородных координат точек на плоскости G . Взвешенная сумма среднеквадратичных ошибок между соответствующими сетками и формирует вторую компоненту в оптимизируемом функционале:

$$L_{Grid}(\theta, \theta_{gt}) = w_{Grid}^\alpha MSE(G^\alpha, G_{gt}^\alpha) + w_{Grid}^k MSE(G^k, G_{gt}^k) + w_{Grid}^t MSE(G^t, G_{gt}^t), \quad (42)$$

где $MSE(G^1, G^2) = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W \left(G_{\{i,j\}}^1 \sim G_{\{i,j\}}^2 \right)^2$, $w_{Grid}^\alpha = 5000$, $w_{Grid}^k = 3000$ и $w_{Grid}^t = 3000$.

Первая регуляризационная компонента L_{Warp} оценивает качество восстановления правого ракурса с внесенными геометрическими искажениями I^R из исходного правого ракурса I_{gt}^R по вычисленным параметрам геометрических искажений. Для этого аналогично описанной модели искажений исходный правый ракурс I_{gt}^R интерполируется по трансформированным координатам, полученным после применения к ним аффинного преобразования с параметрами θ : $I_{Warp}^R = \text{Warp}(I_{gt}^R, G')$, $G' = T(G, \theta)$. Итоговое значение компоненты функции стоимости вычисляется как среднеквадратичная ошибка между входным правым ракурсом I^R и реконструированным по исходному правому ракурсу I_{gt}^R и вычисленным значениям геометрических искажений ракурсом I_{Warp}^R :

$$L_{Warp}(\theta, I^R, I_{gt}^R) = MSE(I^R, I_{Warp}^R). \quad (43)$$

Последняя регуляризационная компонента, L_{Siam} , оценивает консистентность между нейросетевыми предсказаниями на основе входных данных как для левого, так и для правого ракурсов. Если вычисление значений геометрических искажений корректно, то при подаче на вход карты диспаратности и соответствующей карты доверия правого ракурса нейронная сеть должна выдавать такие же по модулю параметры искажений, как и для левого ракурса, но с противоположным знаком. Другими словами, $\theta = -1 \cdot \theta_b$. Таким образом, четвертая компонента оптимизируемого функционала штрафует разницу между предсказанными векторами параметров для левого и правого ракурсов:

$$L_{Siam}(\theta, \theta_b) = L_{SE}(\theta, -\theta_b). \quad (44)$$

Для вычисления данной компоненты дополнительно вычисляются параметры геометрических искажений на основе данных для правого ракурса стереопары во время обучения. Однако при использовании обученной сети для вычисления геометрических искажений достаточно карты диспаратности и карты доверия, построенных только для левого ракурса.

При обучении нейронных сетей был использован метод инициализации весов He [62], а также оптимизационный метод Adam [65], для которого использовались стандартные параметры, за исключением коэффициента скорости обучения, составившего 10^{-4} , изменявшимся по косинусному правилу с уменьшающейся амплитудой (рисунок 29). Модели обучались в течение 120 эпох. Сходимость предложенного метода продемонстрирована на рисунке 30.

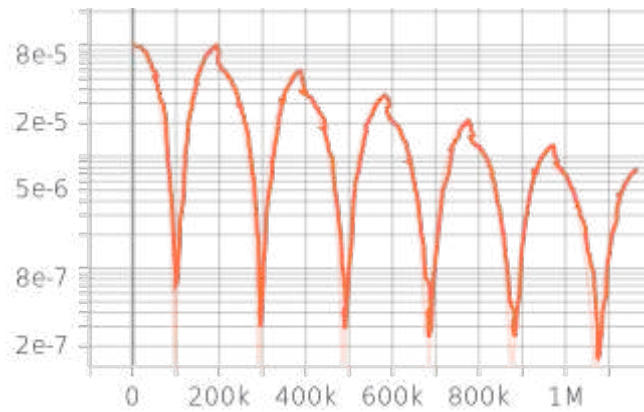


Рисунок 29 — График изменения коэффициента скорости обучения в зависимости от номера итерации.

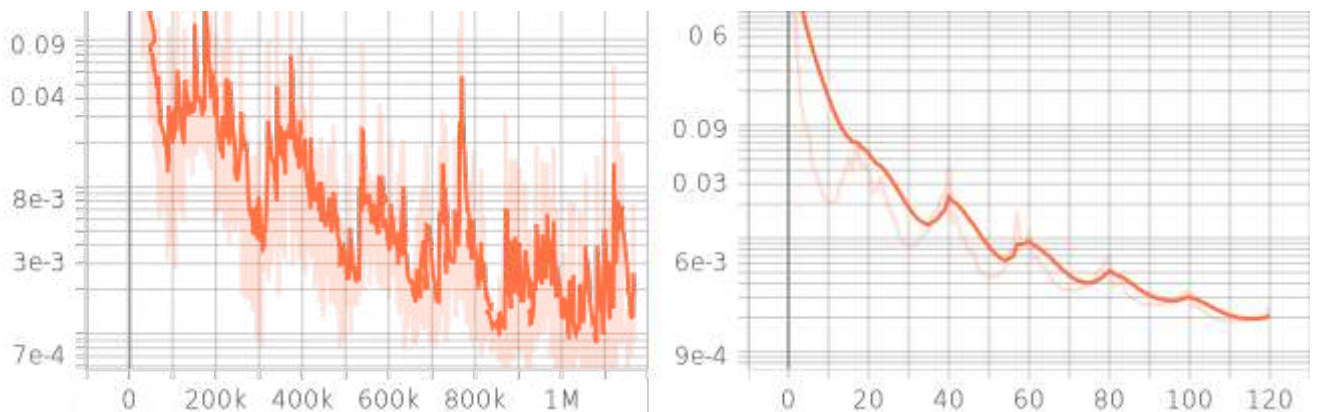


Рисунок 30 — Сглаженные графики оптимизируемого во время обучения функционала на обучающей (слева) и валидационной (справа) выборках.

2.6 Экспериментальная оценка

2.6.1 Выбор оптимизируемого функционала

При разработке метода были обучены различные модели на основе разных оптимизируемых функционалов. Во-первых, рассматривались разные комбинации базовых компонент оптимизируемого функционала, использующих эталонные значения искажений. В таблице 5 представлены результаты работы модели при обучении на основе компонент L_{SE} , L_{Grid} и $L_{SE} + L_{Grid}$. Как можно видеть из результатов, одновременное использование двух базовых компонент улучшает точность вычисленных параметров геометрических искажений. Во-вторых,

Таблица 5 — Результаты сравнения базовых компонент оптимизируемого функционала. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению.

Компоненты	Угол поворота (градусы)	Коэффициент масштабирования (процент от начального разрешения кадра)	Вертикальный сдвиг (процент от высоты кадра)
L_{SE}	0.02063	0.0004925	0.0002258
L_{Grid}	0.01935	0.0004951	0.0002519
$L_{SE} + L_{Grid}$	0.01925	0.0004862	0.0002237

Таблица 6 — Результаты сравнения регуляризационных компонент оптимизируемого функционала. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению.

Компоненты	Угол поворота (градусы)	Коэффициент масштабирования (процент от начального разрешения кадра)	Вертикальный сдвиг (процент от высоты кадра)
Без регуляризации	0.01925	0.0004862	0.0002237
L_{Warp}	0.01914	0.0004792	0.0002019
L_{Siam}	0.01879	0.0004582	0.0002225
$L_{Warp} + L_{Siam}$	0.01832	0.0004525	0.0002001

рассматривались разные комбинации регуляризационных компонент оптимизируемого функционала при обучении с использованием двух базовых компонент L_{SE} и L_{Grid} . Таблица 6 демонстрирует результаты работы модели при использовании в качестве регуляризационной части L_{Warp} , L_{Siam} и $L_{Warp} + L_{Siam}$. Аналогично одновременное использование обеих регуляризационных компонент повышает точность работы нейросетевого регрессора.

2.6.2 Оценка модели на тестовом наборе данных

На тестовом наборе данных было проведено сравнение предложенного метода как с нейросетевыми аналогами [78; 80], так и с методом [26], ранее применявшимся при анализе полнометражных стереоскопических фильмов. Оба нейросетевых аналога были обучены на тренировочной части подготовленного набора данных. Точность работы всех методов была оценена на тестовой части подготовленного набора данных. Результаты представлены в таблице 7. Она содержит средние значения абсолютной ошибки между вычисленными и истинными значениями параметров для каждого из трех рассматриваемых геометрических искажений. “Нулевой вектор” – модель, предсказывающая отсутствие геометрических искажений для каждого примера. Предложенный метод работает значительно точнее нейросетевых аналогов. Также удалось добиться увеличения точности работы по сравнению с методом, ранее применявшимся на практике при анализе полнометражных стереоскопических фильмов, который также обладает лучшим качеством по сравнению с другими нейросетевыми методами. Так, удалось сократить ошибку вычислений на 14.43% при оценке угла поворота, а при оценке масштабирования и вертикального сдвига – более чем на 99%.

2.7 Исправление геометрических искажений

Утверждение 1. *Для исправления геометрических искажений между ракурсами стереоскопического видео достаточно применить аффинное преобразование с вычисленными параметрами геометрических искажений к левому ракурсу. Для исправления искажений в видео в формате VR180 необходимо сначала*

Таблица 7 — Результаты тестирования рассматриваемых методов. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению.

Метод	Угол поворота (градусы)	Коэффициент масштабирования (процент от начального разрешения кадра)	Вертикальный сдвиг (процент от высоты кадра)
Нулевой вектор	0.63406	0.6507	0.57497
Метод [80]	0.43735	1.23582	0.82534
Метод [78]	0.05115	0.10810	0.19109
Метод [26]	0.01158	0.02622	0.02004
Предложенный метод	0.009909	0.0001952	9.118e-05

перейти к трехмерным координатам на сфере, а только затем применить найденное аффинное преобразование к левому ракурсу.

Доказательство. Подробности применения аффинного преобразования к одному из ракурсов стереопары представлены в разделе 2.3. Также для исправления стереопары можно применить аффинное преобразование с параметрами $\theta' = -\theta$ к правому ракурсу. Выбор ракурса для преобразования должен осуществляться оператором или монтажером в зависимости от искомого желаемого положения камер.

Исправление геометрических искажений усложняется в случае видео в формате VR180. Для корректного исправления геометрических артефактов в данном случае необходимо трансформировать сетку координат в трехмерном пространстве по вычисленным параметрам, вместо двухмерного, чтобы не вносить дополнительных искажений ближе к краям кадра. Искомое видео в формате VR180 хранится в равнопромежуточной проекции, то есть каждой точке на сфере фиксированного радиуса с широтой φ и долготой λ ставится в соответствие точка на изображении с координатами x и y : $y = \varphi$, $x = \lambda$ с фиксированной базисной точкой $(\varphi_0, \lambda_0) = (0, 0)$. Пусть $G \in \mathbb{R}^{h \times w \times 3}$ – нормированные однородные координаты пикселей ракурса видео в формате VR180. Для корректного перехода к трехмерным координатам необходимо перевести нормированные ко-

ординаты к угловым координатам, то есть отобразить исходный диапазон значений $[-1, 1)$ на $[-\pi/2, \pi/2)$, для чего достаточно применить преобразование

$$G_s = G \times F_s^T, \text{ где } F_s = \begin{bmatrix} \pi/2 & 0 & 0 \\ 0 & \pi/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \text{ Далее полученные угловые координаты}$$

переводятся в трехмерные координаты, считая что радиус сферы равен 1:

$$G_{vr, \{i,j\}} = (\cos \varphi_{\{i,j\}} \cdot \sin \lambda_{\{i,j\}}, \sin \varphi_{\{i,j\}}, -\cos \varphi_{\{i,j\}} \cdot \cos \lambda_{\{i,j\}}, 1) \in \mathbb{R}^{h \times w \times 4}.$$

После данного преобразования для исправления геометрических искажений необходимо произвести поворот вокруг оси Oz , вертикальный сдвиг по оси Oy , масштабирование по всем координатным плоскостям по вычисленным параметрам.

В такой постановке матрица аффинного преобразования будет выглядеть следующим образом:

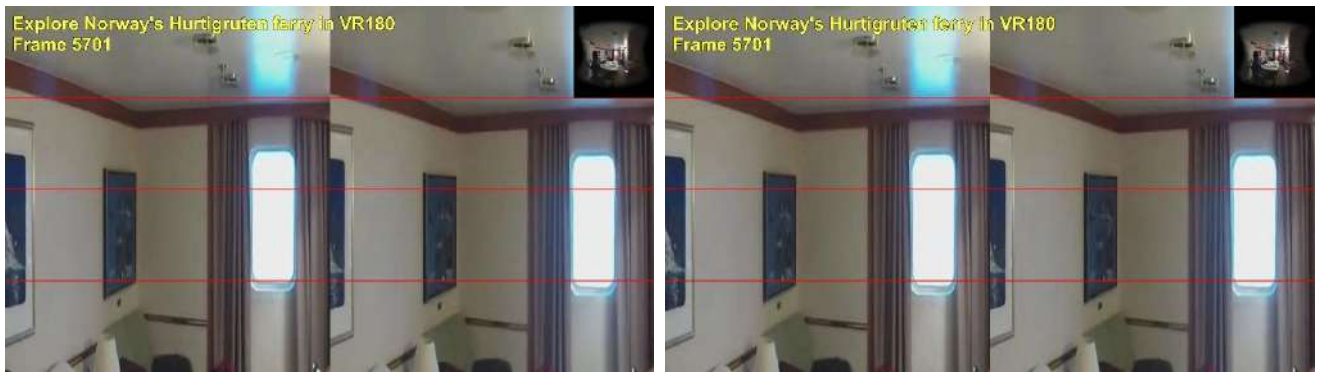
$$A = \begin{bmatrix} (1+k) \cos(\alpha) & -(1+k) \sin(\alpha) & 0 & 0 \\ (1+k) \sin(\alpha) & (1+k) \cos(\alpha) & 0 & t \\ 0 & 0 & (1+k) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (45)$$

Тогда для получения трансформированной сетки достаточно применить данное преобразование $G'_{vr} = G_{vr} \times A^T$, перейти обратно к угловым координатам

$$G'_{s, \{i,j\}} = \left(\arcsin \frac{x_{\{i,j\}}}{\sqrt{1 - y_{\{i,j\}}^2}}, \arcsin y_{\{i,j\}}, 1 \right) \in \mathbb{R}^{h \times w \times 3}, \text{ а затем обратно к нормиро-}$$

$$\text{ванным координатам } G' = G'_s \times B_s^T, \text{ где } F_s = \begin{bmatrix} 2/\pi & 0 & 0 \\ 0 & 2/\pi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \text{ Далее кадр в рав-}$$

нопромежуточной проекции интерполируется по полученным нормированным координатам, как и в случае с обычной стереопарой. Примеры исправления геометрических искажений для кадров видео в формате VR180 представлены на рисунках 31 и 32. □



До

После

Рисунок 31 — Пример исправления вертикально сдвига в видео “Explore Norway’s Hurtigruten ferry in VR180”.Arteфакт наиболее заметен на окне и рамках картины.



До

После

Рисунок 32 — Пример исправления поворота одного ракурса относительно другого в видео “Charlie Puth In London (VR180 Experience)”.Arteфакт наиболее заметен на центральной красной линии.

2.8 Программная реализация

Предложенный метод для оценки геометрических искажений между ракурсами стереоскопических видео реализован на языке Python 3 в виде консольного приложения. В данном приложении используются следующие модули-зависимости:

- pytorch – открытая программная библиотека для машинного обучения, направленная на решение задач построения и тренировки нейронных сетей;

- `opencv`, `skimage` – открытые библиотеки, использовавшиеся для работы с изображениями и видео;
- `numpy`, `scipy`, `matplotlib` – многофункциональные открытые библиотеки, использовавшиеся для быстрой обработки, генерации и визуализации данных.
- MSU Motion Estimation для сопоставления ракурсов и оценки движения.

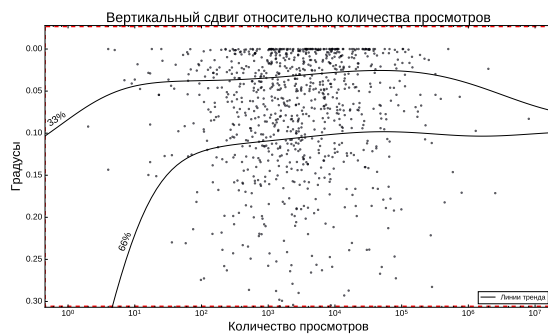
Реализованное консольное приложение аналогично приложению, описанному в разделе 1.7, включает в себя три основные компоненты для:

- подготовки наборов данных для обучения нейросетевых моделей на основе предложенной модели геометрических искажений;
- обучения нейросетевых моделей для оценки геометрических искажений между ракурсами стереоскопических видео;
- тестирования и использования на практике обученных моделей для оценки геометрических искажений.

Также предложенный метод реализован в виде подключаемого модуля к системе оценки качества стереовидео VQMT3D [67], описание которой приведено в разделе 1.7. Средняя скорость работы предложенного метода на кадрах с разрешением 940×544 составила 0.2027 секунд на кадр. Время работы реализации замерялось на стационарном компьютере с восьмиядерным процессором серии AMD Ryzen 7 1700 и графическим ускорителем GeForce GTX 1070.

2.9 Анализ видео в формате VR180

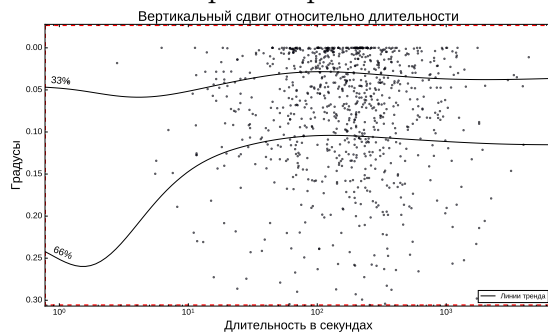
На подготовленном наборе данных из 1000 видео в формате VR180, описанном в разделе 1.8.1, также измерялись значения геометрических искажений. Для вертикального сдвига результаты продемонстрированы на рисунке 33, для относительного масштабирования – на рисунке 34, для относительного поворота



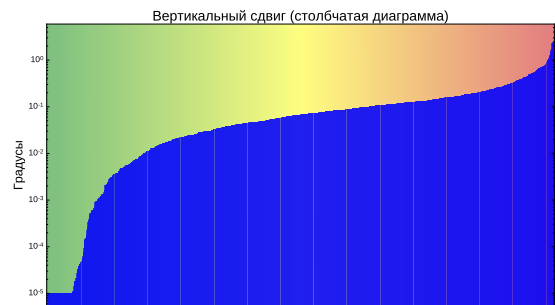
(а) Общие результаты анализа относительно количества просмотров на YouTube.



(б) Общие результаты анализа относительно даты публикации видео.

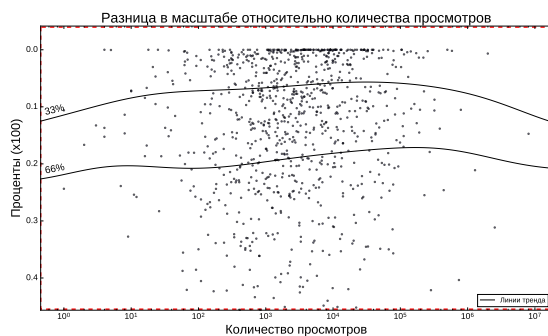


(в) Общие результаты анализа относительно длительности видео.

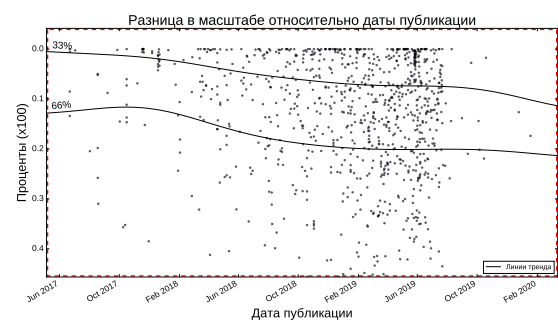


(г) Средние значения искажений для каждого видео.

Рисунок 33 — Результаты анализа вертикального сдвига в VR180-видео.



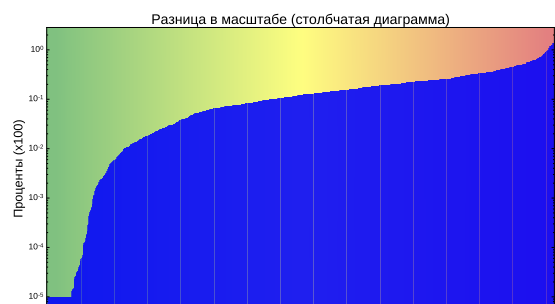
(а) Общие результаты анализа относительно количества просмотров на YouTube.



(б) Общие результаты анализа относительно даты публикации видео.



(в) Общие результаты анализа относительно длительности видео.

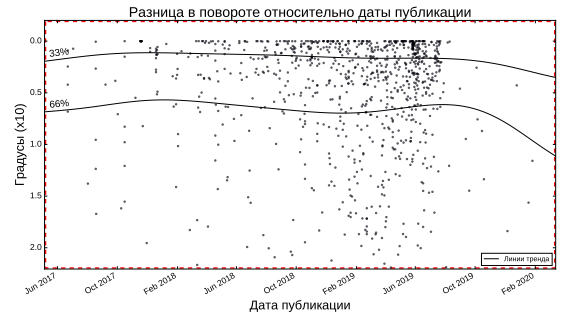


(г) Средние значения искажений для каждого видео.

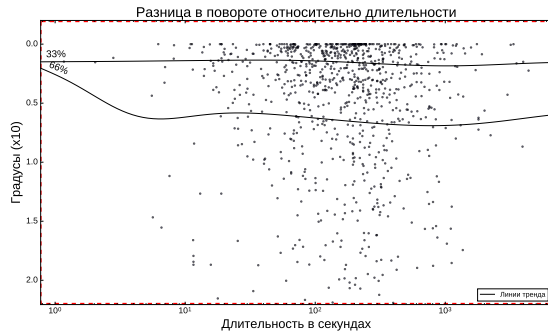
Рисунок 34 — Результаты анализа относительного масштабирования в VR180-видео.



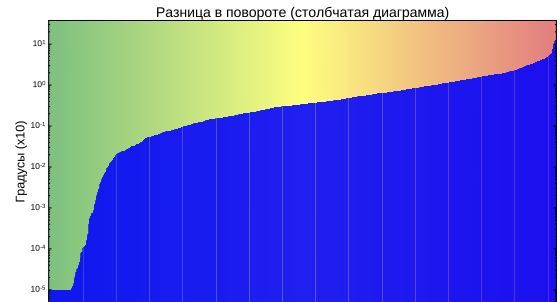
(а) Общие результаты анализа относительно количества просмотров на YouTube.



(б) Общие результаты анализа относительно даты публикации видео.



(в) Общие результаты анализа относительно длительности видео.



(г) Средние значения искажений для каждого видео.

Рисунок 35 — Результаты анализа относительного поворота в VR180-видео.

– на рисунке 35. Результаты анализа продемонстрированы относительно количества просмотров на YouTube (а), даты публикации (б) и длительности (в) каждого видео. Аналогично искажениям цвета и резкости геометрические искажения не демонстрируют какие-либо существенные тенденции по отношению к рассматриваемым статистикам, и для них также справедливы все выводы, сделанные в разделе 1.8.3.

При этом значительное количество проанализированных видео в формате VR180 демонстрирует наличие по крайней мере одного стереоскопического артефакта из рассмотренной группы искажений. Рисунки (г) показывают средние значения оцененных искажений (ось y) для каждого видео (ось x). Небольшие стереоскопические искажения встречаются во многих видео, однако также есть случаи с внушительными значениями артефактов. В левой части графиков для геометрических искажений также присутствуют плоские области, указывающие на отсутствие геометрических артефактов в них. Эти области соответ-

ствуют либо “плоским” видео с одинаковыми ракурсами, либо видео на основе компьютерной графики.

В приложении на рисунках 63-66 представлены найденные в VR180 примеры вертикального сдвига одного ракурса относительно другого, на рисунках 67-70 – масштабирования одного ракурса относительно другого, на рисунках 71-74 – поворота одного ракурса относительно другого.

2.10 Заключение

В рамках главы предложен новый метод для автоматической оценки геометрических искажений между ракурсами стереовидео. Ключевой особенностью метода является использование нейросетевого регрессора для непосредственного предсказания параметров геометрических искажений. Для обучения и тестирования нейросети подготовлен набор данных из 22800 стереопар, преобразованных по предложенной модели геометрических искажений. Данная модель поддерживает три геометрических искажения: вертикальный сдвиг, масштабирование и поворот одного ракурса относительно другого. Предложенный метод показал высокую точность работы по сравнению с аналогами, включая метод, ранее использовавшийся для анализа полнометражных стереоскопических фильмов. Удалось добиться уменьшения погрешности вычислений метода при оценке угла поворота более чем на 14%, а при оценке относительного масштабирования и вертикального сдвига более чем на 99%. Результаты работы предложенного метода можно непосредственно использовать для исправления геометрических искажений между ракурсами стереовидео, в том числе и для видео в формате VR180, для которых в данной главе был описан метод их исправления.

Также в главе представлены результаты анализа геометрических искажений для 1000 видео в формате VR180, собранных с платформы YouTube. Подво-

для итог, ни один из оцененных технических параметров не показал существенных трендов по отношению к таким статистикам видео, как количество просмотров, дата публикации и длительность. Большинство проанализированных видео в формате VR180 содержит как минимум одно серьезное стереоскопическое искажение, что может привести к возникновению у зрителей дискомфорта после просмотра нескольких таких видео. Данная ситуация свидетельствует о необходимости разработки инструментов контроля качества и исправления стереоскопических искажений, которые бы помогли как профессионалам, так и любителям создавать более качественный стереоскопический контент.

Основные результаты главы опубликованы автором в работе [3; 22; 24].

Глава 3. Нейросетевой метод поиска перепутанных ракурсов в стереовидео

3.1 Постановка задачи

Перепутанный порядок ракурсов – артефакт, при котором в сцене стереовидео на месте левого ракурса оказывается правый и наоборот (рисунок 36). Данное искажение встречается достаточно редко в стереофильмах, но наличие даже одной сцены с перепутанными ракурсами может вызвать серьезный дискомфорт у зрителей при ее просмотре [1]. При этом этот артефакт встречается в стереофильмах вне зависимости от способа их создания. На рисунке 37 представлены примеры перепутанных ракурсов в полнометражных стереофильмах для разного способа производства. Эффект перепутанных ракурсов может возникнуть как при простом изменении порядка левого и правого ракурса, так и при неправильном редактировании готового видеоматериала: неправильной конвертации из 2D в 3D, например из-за неточной карты глубины или некачественного метода конвертации, а также добавлением титров и элементов компьютерной графики на неправильную глубину.

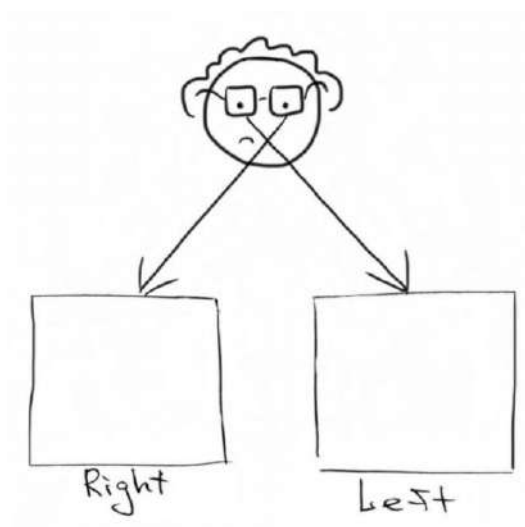


Рисунок 36 — Схематичный пример перепутанных ракурсов в стереовидео.

Левый ракурс



Карта диспаратности



Дальше

Ближе

Случай, в котором левый и правый ракурс перепутаны полностью, что чаще всего случается при съемке 3D-фильмов (кадр из фильма “Сталинград”).



Дальше

Ближе

Случай некорректного наложения компьютерной графики на исходные кадры (кадр из фильма “Хроники Нарнии: Покоритель Зари”).



Дальше

Ближе

Случай эффекта перепутанных ракурсов, полученный в результате некачественной конвертации из 2D в 3D (кадр из фильма “Щелкунчик и Крысиный король”).

Рисунок 37 — Примеры кадров с перепутанными ракурсами, найденных в полнометражных стереофильмах, для различных методов производства

При просмотре сцены с неправильным порядком ракурсов зритель наблюдает невозможную в реальности картину: из-за смены порядка ракурсов меняются знаки значений диспаратности для каждой точки изображения, что приводит к инвертированию значений расстояний до объектов. То есть ближние точки пространства переходят в дальние и наоборот. При этом человеческий мозг помимо бинокулярных признаков для определения расстояния до объектов использует также монокулярные, основанные на опыте, что приводит к конфликту восприятия сцены и порождает дискомфорт. И хотя сила дискомфорта зависит от набора параметров сцены, например, величины диапазона параллакса, характера движения объектов, освещенность сцены, перепутанные ракурсы оказывают негативное влияние на зрителей, из-за чего необходим строгий контроль при производстве стереоскопических видео, не допускающий появления рассматриваемого искажения в финальном продукте. При этом можно достаточно тривиально исправить данное искажение при его обнаружении, для чего нужно изменить подачу ракурсов с одного глаза на другой.

Задача поиска перепутанных ракурсов в стереовидео является задачей бинарной классификации сцен 3D-видео на 2 класса:

- сцены с правильным порядком ракурсов;
- сцены с перепутанным порядком ракурсов.

Рассмотрим формальную постановку данной задачи, используя обозначения из раздела 1.1. Введем дополнительно обозначение сцены стереоскопического видео. Сцена $Sc_i^I \subset I$ – непрерывный фрагмент видеопоследовательности $\{I_t\}_{t=k_1}^{k_2}$, $1 \leq k_1 \leq k_2 \leq n$, с похожим содержанием. Сцены одной видеопоследовательности не пересекаются между собой. Предполагается, что левый и правый ракурсы стереовидео имеют одинаковое разбиение на сцены. Тогда под сценой стереовидео Sc_i^S подразумевается упорядоченная пара непрерывных фрагментов видеопоследовательностей $(\{I_t^L\}_{t=k_1}^{k_2}, \{I_t^R\}_{t=k_1}^{k_2})$, соответствующих левому и правому ракурсу. На вход алгоритму подается стереовидео S . На выходе алгоритм должен предоставить оценку вероятности наличия перепутанных ракурсов $\{p_i\}_{i=1}^m$, $p_i \in \mathbb{R}^{[0,1]}$ для каждой сцены стереовидео Sc_i^S , $i = \overline{1, m}$.

3.2 Обзор существующих алгоритмов

Для определения порядка ракурсов необходимо при помощи некоторых признаков произвести оценку карты диспаратности и сравнить ее с полученной при использовании бинокулярных признаков (то есть при сопоставлении ракурсов стереопары) картой диспаратности, в которой при наличии перепутанных ракурсов значения диспаратности будут иметь неправильный знак. Используемые для оценки диспаратности признаки можно классифицировать следующим образом:

- Использование методов упорядочивания глубины. Данные алгоритмы используют монокулярные признаки для оценки карты диспаратности, то есть строят эти карты по одному из ракурсов.
- Проверка предположений о распределении диспаратности. Большинство сцен стереофильмов обладают похожей структурой (например, увеличение расстояния до объектов при переходе от нижней части кадра к верхней), так что они имеют схожие карты диспаратности. В методах, использующих данные признаки, проводится проверка карт диспаратности, полученных в ходе сопоставления двух ракурсов, на соответствие таким предположениям.
- Анализ областей открытия в стереопаре. Области открытия в стереопаре – это области одного из ракурсов, видимые в нем и невидимые в другом ракурсе. Для таких областей справедлив следующий факт: области открытия в левом ракурсе находятся слева от объектов переднего плана, а в правом ракурсе – справа. Методы, основанные на анализе областей открытия, проверяют справедливость этого факта для каждой стереопары.
- Анализ областей открытия/закрытия по движению. Области открытия/закрытия по движению – это области текущего кадра, видимые в нем и невидимые в предыдущем/следующем кадре соответственно. В боль-

шинстве сцен стереовидео присутствует движение объектов переднего плана, которое и порождает области открытия/закрытия, по которым можно определить местоположения перекрытия объектов заднего плана объектами переднего плана и выяснить порядок этих объектов относительно друг друга.

3.2.1 Методы упорядочивания глубины

Метод [84] основан на использовании карты внимания для осуществления сегментации кадра на передний и задний план, а для определения порядка ракурсов вычисляются средние значения диспаратности у пикселей соответствующих областей. Под вниманием понимается свойство объекта выделяться на фоне соседних объектов, а карта внимания содержит вероятность того, что человек во время просмотра видео будет смотреть на конкретную область кадра. В работе [84] предполагается, что объекты с высоким вниманием являются объектами переднего плана. В качестве заднего плана берутся пиксели, не принадлежащие объектам переднего плана. Для сопоставления ракурсов используется метод поиска особых точек SIFT. В качестве значений диспаратности для найденных особых точек берется разница горизонтальных координат соответствующих друг другу особых точек левого и правого ракурса. Предложенный в работе [84] подход был протестирован на 40 стереопарах, в половине которых порядок ракурсов был изменен. В ходе тестирования было показано, что данный метод превосходит субъективную оценку респондентов, которые часто затруднялись в определении порядка ракурсов при просмотре тестовых стереопоследовательностей. Тем не менее, качество работы данного метода существенно зависит от точности метода построения карты внимания, а также выполнения условия о нахождении объектов переднего плана в салиентной зоне, что делает этот метод не применимым при анализе полнометражных фильмов.

Также не для всех сцен стереофильмов будет справедлив тот факт, что салиентные объекты находятся на переднем плане, что зависит, в первую очередь, от структурной сложности сцены. Для оценки карты диспаратности по одному из ракурсов также могут быть использованы и другие методы упорядочивания глубины, например [85—87]. Данные методы позволяют по заданной видеопоследовательности восстановить относительный порядок объектов в сцене. Однако большинство этих алгоритмов не применимо на практике при анализе полнометражных стереофильмов, так как они обладают низкой точностью и высокой вычислительной сложностью. Исключением являются нейросетевые методы для предсказания карт диспаратности по одному изображению, например [88; 89]. В данных работах были получены относительно точные результаты при предсказании карт диспаратности по изображениям из наборов данных NYUDepth и KITTI, что демонстрирует возможность обучения сверточных нейронных сетей монокулярным признакам упорядочивания глубины. Также в некоторых работах [90; 91] обучение нейросетевых моделей осуществляется сразу на нескольких наборах данных. Такая стратегия позволяет повысить обобщаемость методов на потенциальные варианты использования и в целом увеличивает практичность. Таким образом, многообещающим шагом для улучшения методов поиска перепутанных ракурсов является применение сверточных нейронных сетей для решения данной задачи.

3.2.2 Метод, основанный на проверке предположения о распределении диспаратности

Описанный в работе [92] алгоритм поиска перепутанных ракурсов в стереовидео основан на проверке предположения о пространственном распределении диспаратности. Предполагается, что объекты, расположенные в нижней части кадра, находятся ближе к камере, чем объекты, находящиеся в верхней

части кадра, а объекты, расположенные в центре кадра, – ближе, чем объекты в левой и правой частях кадра. Для проверки этого предположения строится шаблонная карта диспаратности, следующая описанному выше условию, путем усреднения карт диспаратности по 6000 кадрам, а для определения порядка ракурсов достаточно посчитать корреляцию исследуемой карты диспаратности с шаблонной. Благодаря своей простоте рассматриваемый метод имеет низкую вычислительную сложность, так как заключается в вычислении корреляции между анализируемой и шаблонной картами диспаратности. Однако он также обладает низкой точностью, так как большое количество сцен в стереовидео не будут следовать указанному выше предположению о распределении диспаратности. Описанный метод можно надежно использовать при анализе стереовидео на наличие перепутанных ракурсов в некоторых случаях (например, при анализе спортивных телетрансляций), когда заведомо известны возможные положения камер и пространственная структура кадров, для которых возможно посчитать шаблонную карту диспаратности. Но в случае анализа полнометражных стереофильмов данный подход будет иметь большое число ошибок второго рода, поэтому он не применим на практике.

3.2.3 Метод на основе анализа положения областей открытия в стереопаре

Для проверки предположения об областях открытия в стереопаре в методе [93] анализируются центроиды данных карт: в случае выполнения этого предположения центроида областей открытия в левом ракурсе будет находится левее, чем соответствующая центроида в правом ракурсе. Поэтому для определения порядка ракурсов в данной работе предлагается вычислять разницу горизонтальных координат центроидов. Рассматриваемый метод был протестирован всего на 52 последовательностях, что не гарантирует практической примени-

мости. Также проверки данного предположения может быть недостаточно, так как его качество сильно зависит от ширины областей открытия в конкретной сцене.

3.2.4 Композиционные методы

В методе, описанном в работе [94], используется два критерия для определения порядка ракурсов в стереовидео:

- Анализ областей открытия. Для проверки утверждения о расположении областей открытия в анализируемой стереопаре осуществляется анализ границ рядом с областями открытия. Больше границ должно наблюдаться на той стороне областей открытия, которая примыкает к объекту переднего плана, поэтому таким способом можно определить положение областей открытия относительно объектов переднего плана.
- Второй критерий основан на проверке предположения о распределении диспаратности. Данный критерий используется в случае, когда вычисленные области открытия имеют слишком маленькую ширину и непригодны для анализа либо когда области открытия вообще не обнаруживаются алгоритмом поиска областей открытия. Проверяемое предположение о распределении диспаратности основано на следующем факте: стереографы обычно организуют сцену таким образом, чтобы объекты, находящиеся перед плоскостью экрана, занимали треть пространства кадра, а объекты, находящиеся за плоскостью экрана, – две трети, в результате чего больше положительных значений будет у карты диспаратности левого ракурса, а у карты диспаратности правого ракурса – отрицательных.

Однако метод [94] показал большое число ложных срабатываний при его использовании для анализа полнометражных стереофильмов. Идеи метода [94]

были развиты в работе [16], в которой для поиска перепутанных ракурсов были предложены 3 дополнительных критерия: анализ перспективы; анализ «выпадающих» объектов; анализ областей открытия/закрытия по движению; а также был переработан критерий по анализу областей открытия в стереопаре.

Часто в сценах объекты с меньшей глубиной находятся внизу кадра, а объекты с большей глубиной – вверху кадра. Критерий по анализу перспективы основан на предположении, что чем ближе объект к низу кадра, тем меньше его глубина. Для проверки данного предположения вычисляется среднее значение диспаратности в каждой строке, взвешенное на значения карты доверия для карты диспаратности. Аналогичным образом вычисляется среднее значение доверия для каждой строки. В качестве результата описанный признак выдает сумму разностей средних значений диспаратности соседних строк, взвешенных на среднее доверие.

Критерий по анализу «выпадающих» объектов проверяет наличие в центре экрана выпуклого объекта, окруженного объектами заднего плана, что часто справедливо для полнометражных фильмов, так как в центре обычно находится объект интереса зрителей. Для проверки этого в каждой строке и столбце карты диспаратности левого ракурса перебираются всевозможные пары точек, и вычисляется разница между значениями диспаратности точек, находящихся между рассматриваемыми точками, и средним значением диспаратности рассматриваемых точек. Тогда, если выбранные точки лежат на одном и том же уровне карты диспаратности, предполагается, что диспаратность точек, находящихся между рассматриваемыми, будет меньше или совпадает со средним значением диспаратности этих точек. Алгоритм аккумулирует вычисленную разницу диспаратности для всевозможных пар точек в каждой строке и столбце карты диспаратности.

Критерий по анализу областей открытия в стереопаре определяет положение отрезков вычисленных областей открытия по отношению к объекту переднего плана за счет анализа соседних областей треугольной формы, распо-

ложенных по горизонтали. Если цвета объектов переднего и заднего плана в анализируемом кадре отличаются, то можно определить, с какой стороны от рассматриваемого отрезка области открытия находится объект переднего плана. Для каждого отрезка области открытия вычисляются цветовые гистограммы соседних областей, разреженные до 16 градаций по каждому каналу. Далее для каждого пикселя отрезка вычисляется его координата в гистограмме, по которой вычисляется разница значений гистограмм, взвешенная на доверие.

В критерии по анализу областей открытия/закрытия по движению осуществляется определение объектов переднего плана за счет движения в сцене и анализу их глубины. Для обнаружения точек объектов переднего плана, предположительно перекрывающих рассматриваемые точки области открытия/закрытия, для них протягиваются вектора движения из карты векторов движения, построенной для кадров, в которых эти точки являются видимыми. Для областей открытия такими кадрами будут текущий и следующий кадры, а для областей закрытия – предыдущий и текущий кадры. Тогда при выполнении предположения о достаточно постоянном поле векторов движения переход по протянутому вектору движения в следующий кадр (для точек областей закрытия) или в предыдущий кадр (для точек областей открытия) приводит в точку, перекрывающую рассматриваемую точку области открытия/закрытия. Далее для полученной перекрывающей точки выполняется переход по обратной карте векторов движения в текущий кадр, в результате чего получается искомая пара точек: точка области открытия/закрытия и перекрывающая её точка. В данном критерии вычисляется взвешенная на доверие разница средних значений диспаратности в соседних для данных точек областях. Полученные для каждой пары точек значения суммируются.

Для получения финального результата метод [16] осуществляет простую классификацию кадра на наличие/отсутствие в нем перепутанных ракурсов путем определения знака взвешенной суммы результатов работы критериев. Для увеличения скорости работы метода вычисления производятся только для одного кадра сцены. Хотя метод [16] был успешно применен для анализа 105 пол-

нометражных фильмов, в ходе анализа все равно было получено большое число ложноположительных срабатываний.

3.3 Описание предложенного алгоритма

Предложенный метод осуществляет классификацию сцены на наличие/отсутствие перепутанных ракурсов по усредненному по сцене значению нейросетевого признака, предсказывающего вероятность наличия перепутанных ракурсов в кадре. Данный метод состоит из следующих шагов при обработке одного кадра:

1. Вычисление карт диспаратности между левым и правым ракурсами, карт векторов движения между текущим и предыдущим, текущим и следующим кадрами, а также построение карт доверия к вычисленным картам диспаратности, характеризующих точность вычисленных векторов. Вычисление карт диспаратности, векторов движения, а также их доверия, совпадает с аналогичным шагом для методов оценки искажений цвета и резкости, описанном в разделе 1.4.1.
2. Определение пригодности кадра для анализа на наличие/отсутствие перепутанных ракурсов. Из рассмотрения исключаются кадры с константной диспаратностью и/или с низкой яркостью.
3. Вычисление областей открытия/закрытия по движению.
4. Нейросетевая оценка вероятности наличия перепутанных ракурсов в кадре на основе левого ракурса, соответствующей карты диспаратности, карты доверия и карты областей открытия/закрытия.

На рисунке 38 представлена общая схема работы предложенного метода. Формально, результатом работы предложенного алгоритма для сцены является число:

$$\bar{x} = \frac{1}{n_A} \sum_{j \in A} x_j, \quad (46)$$

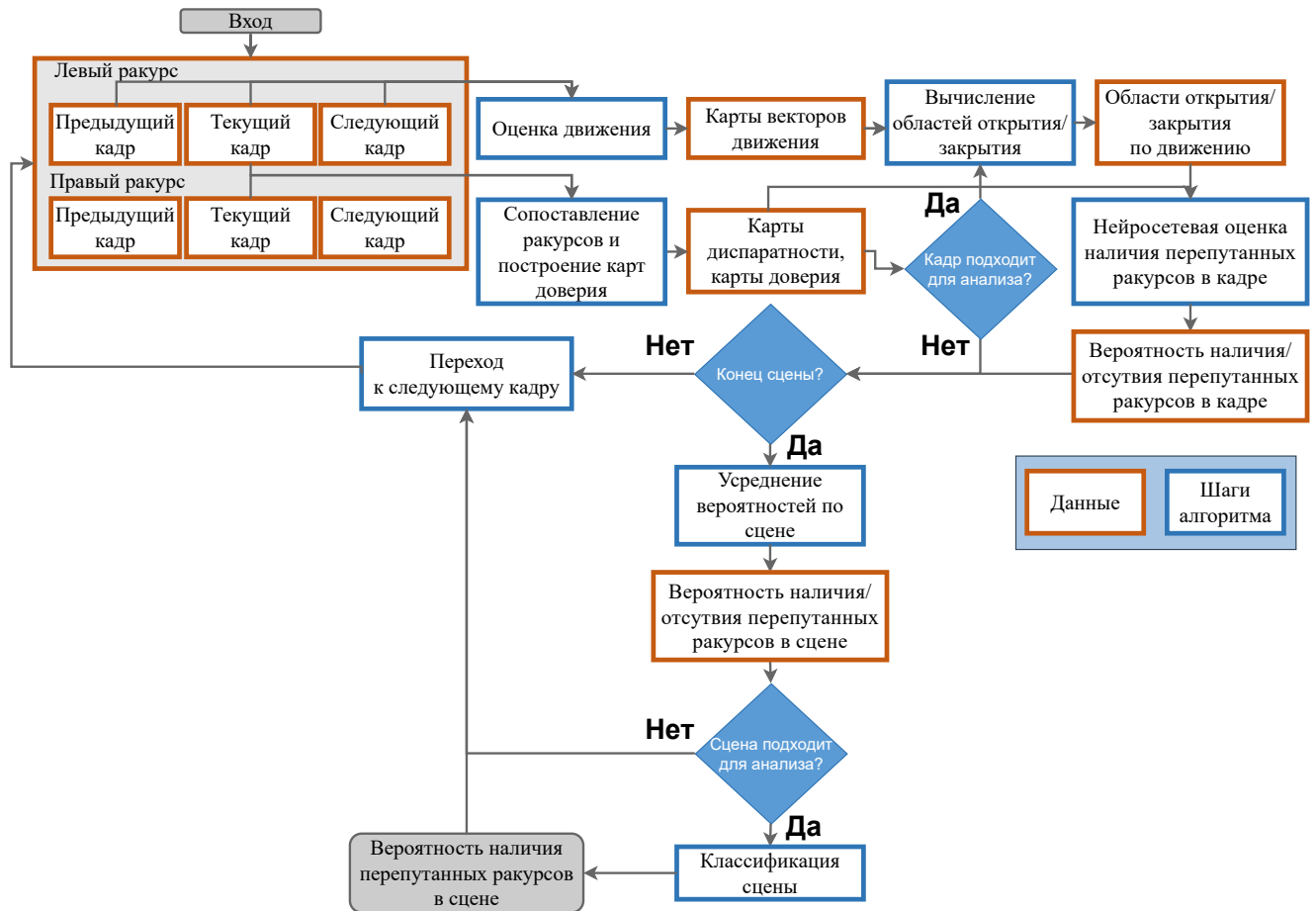
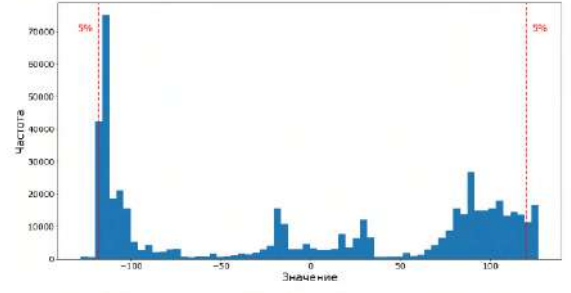


Рисунок 38 — Общая схема предложенного метода поиска перепутанных ракурсов в стереовидео.

где x_j , $j = \overline{1, n}$ – значение нейросетевого признака j -го кадра анализируемой сцены, n – число кадров в сцене, $n_A = |A|$ – число подходящих для анализа кадров в сцене, $A = \{k_j | 1 \leq k_j \leq n\}$ – множество номеров кадров сцены, подходящих для анализа.

3.3.1 Определение пригодности кадра для анализа

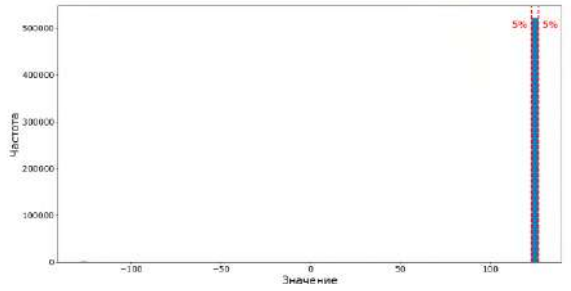
В предложенном методе при анализе сцены не учитываются кадры с постоянной диспаратностью и кадры с очень низкой яркостью, которые считаются непригодными для анализа. Для кадров с постоянной диспаратностью не имеет смысла проводить анализ на наличие перепутанных ракурсов, а при анализе кадров с очень низкой яркостью часто возникают ошибки при вычислении карт



Дальше Ближе

“Неплоская” карта диспаратности с построенной по ней гистограммой.

$$\sigma^{d^L} = 92.72.$$



Дальше Ближе

“Плоская” карта диспаратности с построенной по ней гистограммой. $\sigma^{d^L} = 0$.

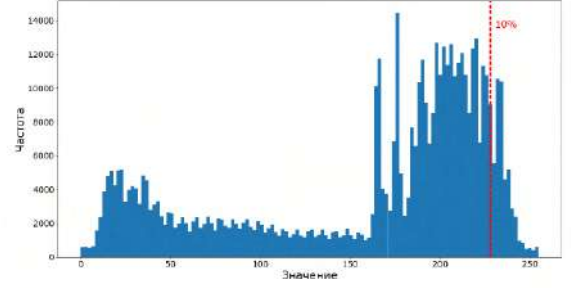
Рисунок 39 — Примеры вычисленных стандартных отклонений диспаратности для кадра с переменной диспаратностью и для кадра с постоянной диспаратностью.

диспаратности и векторов движения. При этом возникающий дискомфорт при просмотре «темных» кадров с перепутанными ракурсами значительно меньше, чем при просмотре «ярких» кадров [1], что в целом позволяет не учитывать такие кадры при анализе стереофильмов.

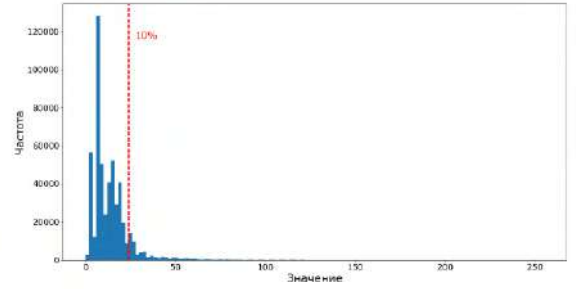
Для исключения из рассмотрения кадров с постоянной диспаратностью строится взвешенная гистограмма карты диспаратности левого ракурса d^L , где в качестве весов выступают значения доверия к рассматриваемой карте conf^{d^L} :

$$H_k^{d^L} = \sum_{\substack{1 \leq i \leq h \\ 1 \leq j \leq w \\ d_{ij}^L = k}} \text{conf}_{ij}^{d^L}, \quad (47)$$

где k — индекс гистограммы. В построенной гистограмме отбрасываются 5% максимальных и 5% минимальных значений для уменьшения влияния выбо-



“Яркий” кадр с построенной по нему гистограммой яркости. $\bar{Y}^L = 150.55$.



“Темный” кадр с построенной по нему гистограммой яркости. $\bar{Y}^L = 10.48$.

Рисунок 40 — Примеры вычисленных средних значений яркости для кадров с высокой и низкой яркостью. Кадры из фильма “Мстители”.

сов, и вычисляется стандартное отклонение диспаратности:

$$\sigma^{d^L} = \sqrt{\frac{\sum_{i=k_1}^{k_2} H_i^{d^L} i^2}{\sum_{i=k_1}^{k_2} H_i^{d^L}} - \left(\frac{\sum_{i=k_1}^{k_2} H_i^{d^L} i}{\sum_{i=k_1}^{k_2} H_i^{d^L}} \right)^2}, \quad (48)$$

где $k_1, k_2, k_1 < k_2$ – нижняя и верхняя границы индексов гистограммы. Если полученное стандартное отклонение диспаратности оказалось слишком малым, то кадр не используется для анализа. Примеры карт диспаратности, построенных по ним гистограмм и вычисленных значений σ^{d^L} представлены на рисунке 39.

Для исключения из рассмотрения кадров с очень низкой яркостью, вычисляется яркость левого ракурса Y^L и строится гистограмма яркости:

$$H_k^{Y^L} = \#\{y \in Y^L \mid y = k\}, \quad (49)$$

где $\#$ – оператор вычисления числа элементов. В построенной гистограмме отбрасываются 10% максимальных значений для уменьшения влияния небольших

ярких участков изображения, и вычисляется среднее значение яркости:

$$\bar{Y}^L = \frac{\sum_{i=0}^{255} H_i^{Y^L} i}{\sum_{i=0}^{255} H_i^{Y^L}}. \quad (50)$$

Если полученное среднее значение яркости оказалось слишком малым, то кадр не используется для анализа. Примеры кадров, построенных по ним гистограмм яркости и средних значений яркости представлены на рисунке 40.

3.3.2 Вычисление областей открытия/закрытия по движению

Для вычисления областей открытия/закрытия по движению используется подход на основе вычисления меры достоверности карт векторов движения FBC (forward-backward consistency), аналогичной мере достоверности сопоставления LRC для карт диспаратности, описанной в разделе 1.4.1. Пусть пиксель с координатами $x = (x_1, x_2)$ одного кадра соответствует пикселю с координатами $x' = (x'_1, x'_2) = x + v_x$ другого кадра. Тогда мера достоверности сопоставления для него равна:

$$\text{fbc} = 1 \wedge \max \left(\min \left(\frac{1}{255} \left(\frac{\text{dif}_1^2}{h} + \frac{\text{dif}_2^2}{w} \right), 0 \right), 1 \right), \quad (51)$$

$$\text{dif} = (\text{dif}_1, \text{dif}_2) = v'_{x'} + v_x, \quad (52)$$

где v_x – вектор движения пикселя с координатами x в первом кадре, $v'_{x'}$ – вектор движения пикселя с координатами x' во втором кадре. Пусть fbc_{prev} – вычисленная мера достоверности FBC для текущего кадра относительно предыдущего кадра, а fbc_{next} – вычисленная мера достоверности FBC для текущего кадра относительно следующего кадра. Тогда $\text{oss} = \text{fbc}_{prev} + \text{fbc}_{next}$ формирует карту областей открытия/закрытия для текущего кадра. Пример вычисленной карты областей открытия/закрытия представлен на рисунке 41



Рисунок 41 — Левый ракурс с вычисленной для него картой областей открытия/закрытия по движению, изображенной красным цветом. Большая часть областей открытия/закрытия обрамляет движущиеся объекты, находящиеся на переднем плане. Кадр из фильма “Гарри Поттер и Дары Смерти: Часть I”.

3.3.3 Архитектура нейронной сети

Для предсказания вероятности наличия перепутанных ракурсов в кадре была также использована архитектура нейронной сети, аналогичная ResNet-18 [82], как и в методе по оценке геометрических искажений в разделе 2.4. В модификации архитектуры не используется батч-нормализация [83], а также было использовано четыре остаточных блока перед каждым увеличением размера канала признаков. Помимо этого, последний слой сети предсказывает вектор из двух значений, представляющих из себя вероятность наличия и отсутствия перепутанных ракурсов в стереовидео соответственно после применения к этим значениям функции Softmax. Входными данными для сети являются яркость левого ракурса, соответствующая карта диспаратности и карта доверия к ней,

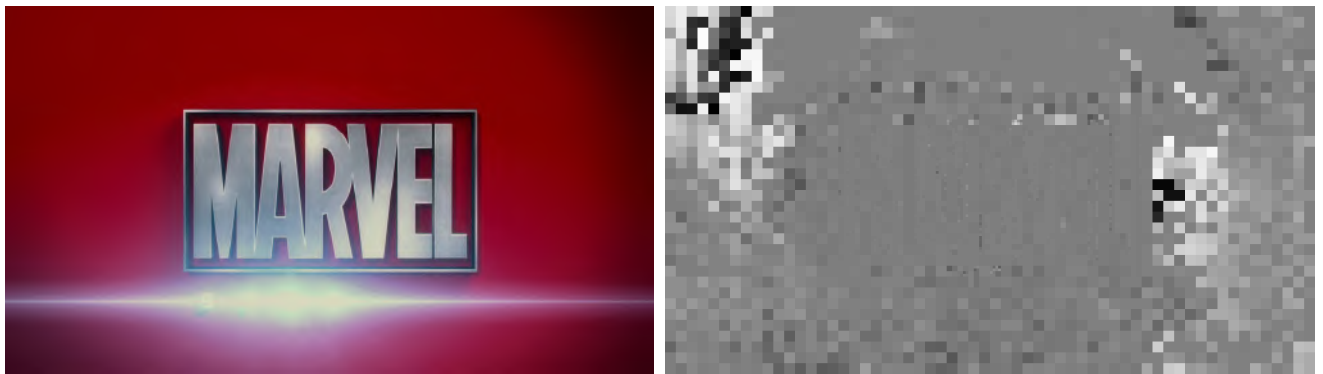


Рисунок 42 — Пример сцены (левый ракурс и карта диспаратности), непригодной для анализа перепутанных ракурсов, так как изменение порядка ракурсов в данной сцене не изменит ее восприятие зрителями. Кадр из фильма “Тор 2: Царство тьмы”.

а также карта областей открытия/закрытия по движению. Этой информации обычно достаточно для подготовленного человека, чтобы определить наличие перепутанных ракурсов. При этом пространственная размерность входных данных может быть произвольной благодаря использованию глобального слоя субдискретизации с выбором среднего перед финальным полносвязным слоем.

3.3.4 Оценка пригодности сцены для анализа

При использовании предложенной нейросетевой архитектуры для предсказания наличия перепутанных ракурсов достаточно просто осуществить проверку пригодности сцены для ее анализа. Для этого достаточно получить выход сети как для одного порядка входных ракурсов, так и для другого. Далее можно сравнить усредненные показатели по сцене для одного и другого порядка ракурсов. Если полученное значение будет одновременно больше или меньше 0.5, значит сеть одновременно для двух порядков ракурсов предсказывает либо перепутанность, либо неперепутанность. Такой исход может возникнуть при анализе плоских сцен, а также сцен, в которых изменение порядка ракурсов не влияет на восприятие сцены. Пример такой сцены представлен на рисунке 42.

3.3.5 Определение смены сцены

Определение последнего кадра сцены осуществляется за счет сравнения гистограмм яркости блоков левого ракурса для текущего и следующего кадра. Для каждого кадра по их яркости Y_t^L строятся гистограммы блоков размера 32×32 , разреженные до 64 градаций:

$$H_{ij,k}^{Y_t^L} = \# \left\{ y \in B_{ij}^{Y_t^L} \mid \left\lfloor \frac{y}{4} \right\rfloor = k \right\}, \quad (53)$$

где $i = \overline{1, h_b}$, $j = \overline{1, w_b}$ – вертикальная и горизонтальная координаты гистограмм блоков яркости, $h_b = \left\lfloor \frac{h}{32} \right\rfloor$, $w_b = \left\lfloor \frac{w}{32} \right\rfloor$ – число блоков по вертикали и горизонтали соответственно, t – номер кадра, $B_{ij}^{Y_t^L}$ – блок яркости Y_t^L с координатами $(32i, 32j)$. При сравнении текущего и следующего кадра вычисляется разность между полученными гистограммами блоков яркости:

$$\text{Dif}_{ij,k}^{Y_t^L, Y_{t+1}^L} = H_{ij,k}^{Y_t^L} - H_{ij,k}^{Y_{t+1}^L}. \quad (54)$$

Дополнительно осуществляется свертка полученных разностей гистограмм с ядром $\{0.05, 0.1, 0.2, 0.3, 0.2, 0.1, 0.05\}$ для сглаживания результата. Значения, необходимые для фильтрации левой и правой границы гистограммы, достраиваются путем зеркального отражения. Все вычисленные разности аккумулируются следующим образом:

$$\text{Dif}^{Y_t^L, Y_{t+1}^L} = \frac{1}{1024 h_b w_b} \sum_{i=1}^{h_b} \sum_{j=1}^{w_b} \sum_{k=0}^{63} |\text{Dif}_{ij,k}^{Y_t^L, Y_{t+1}^L}|. \quad (55)$$

Данное значение характеризует степень похожести текущего кадра на следующий кадр. Предполагается, что при смене сцены это число будет большим, а также оно будет сильно отличаться от аналогичных значений для предыдущих и следующих кадров. Для проверки последнего утверждения вычисляется следующая характеристика:

$$\text{Dis}^{Y_t^L} = 6 \text{Dif}^{Y_t^L, Y_{t+1}^L} - \sum_{k \in [-3, -1] \cup [1, 3]} \text{Dif}^{Y_{t+k}^L, Y_{t+k+1}^L}. \quad (56)$$



$$\text{Dif}^{Y_{t-1}^L, Y_t^L} = 0.0676 \quad \text{Dif}^{Y_t^L, Y_{t+1}^L} = \mathbf{1.3105} \quad \text{Dif}^{Y_{t+1}^L, Y_{t+2}^L} = 0.0178 \quad \text{Dif}^{Y_{t+2}^L, Y_{t+3}^L} = 0.0181$$

$$\text{Dis}^{Y_{t-1}^L} = -1.123 \quad \text{Dis}^{Y_t^L} = \mathbf{7.6222} \quad \text{Dis}^{Y_{t+1}^L} = -1.3888 \quad \text{Dis}^{Y_{t+2}^L} = -1.3436$$

Рисунок 43 — Пример смены сцены и соответствующих кадрам значений

$\text{Dif}^{Y_t^L, Y_{t+1}^L}$ и $\text{Dis}^{Y_t^L}$. Кадры из фильма “Новый Человек-паук”.

Если в ходе анализа яркости кадра Y_t^L значения $\text{Dif}^{Y_t^L, Y_{t+1}^L}$ и $\text{Dis}^{Y_t^L}$ превысили определенные пороги, то считается, что рассматриваемый кадр является последним в сцене. Пример определения смены сцены представлен на рисунке 43.

3.4 Обучение нейронных сетей

Задача поиска перепутанных ракурсов в стереовидео является задачей бинарной классификации сцен 3D-видео на 2 класса. Поэтому для обучения нейронной сети для определения порядка ракурсов достаточно использовать бинарную кросс-энтропию в качестве оптимизируемой функции:

$$L_{CE}(y, p) = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_{i1}) + (1 - y_i) \log(p_{i2})),$$

где N — число примеров, используемых на каждой итерации обучения, y_i — метка о наличии/отсутствии перепутанных ракурсов в примере i , p_{ij} — выходные значения сверточной нейронной сети для примера i . Дополнительно для предотвращения переобучения в оптимизируемой функции используется L_2 -регуляризация с коэффициентом 0.0005 для всех весов в сети.

Для обучения нейронной сети был подготовлен обучающий набор данных на основе кадров из полнометражных стереоскопических фильмов, ранее использованных для обучения методов оценки искажений цвета и резкости. Спи-

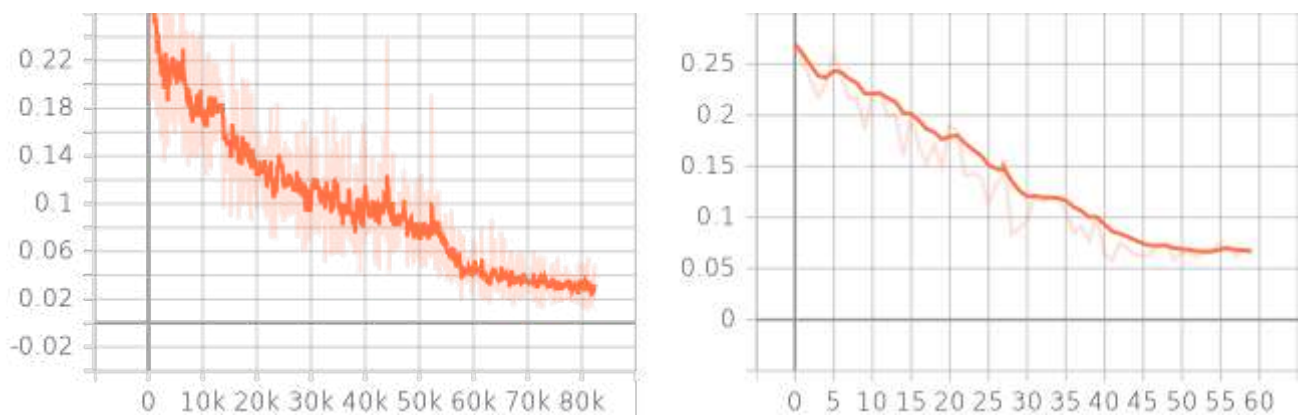


Рисунок 44 — Сглаженные графики оптимизируемого во время обучения функционала на обучающей (слева) и валидационной (справа) выборках.

сок фильмов представлен в разделе 1.3. При этом сам порядок ракурсов выбирался во время обучения случайно.

В качестве метода инициализации весов сети был использован метод Xavier [64], а для оптимизации был выбран алгоритм Adam [65]. Предложенная нейронная сеть обучалась в течение 60 эпох с коэффициентом скорости обучения 10^{-4} , который уменьшался в 10 раз каждые 40 эпох. Количество примеров из набора данных, используемых на одной итерации обучения, было равно 8. Размер входных данных при обучении – 928×512 . Сходимость рассматриваемого метода продемонстрирована на рисунке 44.

3.5 Экспериментальная оценка

На тестовой выборке, состоящей из 900 сцен длиной в 30 кадров, было проведено сравнение предложенного алгоритма поиска перепутанных ракурсов с аналогами [16; 92; 94], применявшимися на практике при анализе полнометражных стереофильмов, а также с современными нейросетевыми методами построения карт диспаратности по одному кадру [90; 91]. Для использования последних методов в задаче оценки порядка ракурсов вычислялась корреляция Пирсона между предсказанными картами диспаратности и картами, вычислен-

Таблица 8 — Результаты тестирования алгоритмов поиска перепутанных ракурсов в стереовидео.

Показатель Алгоритм	AUC	Точность	F-мера
Метод [92]	0.7223	0.6614	0.6683
Метод [94]	0.901	0.8378	0.8409
Метод [16]	0.957	0.8946	0.8928
Метод [90]	0.9913	0.8394	0.8613
Метод [91]	0.9899	0.8256	0.8515
Предложенный метод	0.9963	0.9784	0.9789

ными с помощью блочного метода компенсации движения [34]. Положительная корреляция указывает на совпадение порядка ракурсов, в то время как отрицательная – на их перепутанность. Во время тестирования для всех оцениваемых алгоритмов вычислялись следующие показатели:

- площадь под ROC-кривой;
- точность на тестовой выборке;
- F-мера.

Сравнение методов проводилось без использования дополнительных методов фильтрации результатов, описанных в разделах 3.3.1 и 3.3.4. Полученные показатели представлены в таблице 8. По результатам тестирования видно, что предложенный метод поиска перепутанных ракурсов в стереовидео превосходит существующие аналоги по качеству классификации. При этом удалось улучшить точность классификации более чем на 8% по сравнению с ранее применявшимися на практике методами. Можно заметить, что нейросетевые методы построения карт диспаратности в целом неплохо справляются с задачей определения порядка ракурсов и даже обгоняют специализированные методы поиска перепутанных ракурсов, что в очередной раз демонстрирует их высокий потенциал на данной задаче. Однако для определения порядка ракурсов достаточно использования небольших, легковесных классификационных моделей, как по-

казывает данное сравнение, а восстановление полной карты диспаратности с помощью моделей типа кодировщик-декодировщик является излишним. Также в предложенном методе за счет использования меньшего числа критериев для оценки перепутанных ракурсов удалось достичь ускорения предложенного метода по сравнению с предшественником. Так предложенный метод обрабатывает 4.68 тестовых кадра в секунду, в то время как метод [16] – 0.58 тестовых кадра в секунду.

3.6 Программная реализация

Предложенный метод поиска перепутанных ракурсов в стереоскопических видео реализован на языке Python 3 в виде консольного приложения. В данном приложении используются следующие модули-зависимости:

- `pytorch` – открытая программная библиотека для машинного обучения, направленная на решение задач построения и тренировки нейронных сетей;
- `opencv`, `skimage` – открытые библиотеки, использовавшиеся для работы с изображениями и видео;
- `numpy`, `scipy`, `matplotlib` – многофункциональные открытые библиотеки, использовавшиеся для быстрой обработки, генерации и визуализации данных.
- `MSU Motion Estimation` для сопоставления ракурсов и оценки движения.

Реализованное консольное приложение, аналогично приложению, описанному в разделе 1.7, включает в себя три основные компоненты для:

- подготовки наборов данных для обучения нейросетевых моделей для предсказания вероятности наличия перепутанных ракурсов в кадре;

- обучения нейросетевых моделей для предсказания вероятности наличия перепутанных ракурсов;
- тестирования и использования на практике обученных моделей для поиска перепутанных ракурсов в сценах стереоскопических видео.

Также предложенный метод реализован в виде подключаемого модуля к системе оценки качества стереовидео VQMT3D [67], описание которой приведено в разделе 1.7. Средняя скорость работы предложенного метода на кадрах с разрешением 940×544 составила 0.2137 секунд на кадр. Время работы реализации замерялось на стационарном компьютере с восьмиядерным процессором серии AMD Ryzen 7 1700 и графическим ускорителем GeForce GTX 1070.

3.7 Анализ видео в формате VR180

Для поиска перепутанных ракурсов в VR180 было проанализировано 50 наиболее просматриваемых видео в формате VR180. Перепутанные ракурсы встречаются редко в стереовидео, однако даже одна сцена с перепутанными ракурсами может вызвать серьезный дискомфорт у зрителя [1]. С помощью предложенного метода поиска перепутанных ракурсов в стереовидео была найдена 21 сцена с перепутанными ракурсами в 10 видео. Согласно данному результату вероятность встретить сцену с перепутанными ракурсами в VR180-видео составляет 20%. При этом в большинстве случаев перепутанные ракурсы возникают из-за неграмотного наложения элементов компьютерной графики и/или титров поверх отснятого материала, что неудивительно, так как съемкой видео в формате VR180 занимаются любители, не обладающие знаниями о композиции трехмерных сцен, а также из-за отсутствия необходимых инструментов для проверки диспаратности добавленных в видео объектов. В приложении продемонстрированы примеры найденных сцен с перепутанными ракурсами (рисунки 75-81).

3.8 Заключение

В данной главе предложен метод автоматического поиска перепутанных ракурсов в стереовидео. Разработанный метод для каждой сцены стереовидео вычисляет вероятность наличия в ней перепутанных ракурсов. Данный метод:

- Использует в качестве основного признака результаты работы нейронной сети, вычисляющей вероятность наличия перепутанных ракурсов в кадре на основе исходного левого ракурса, и соответствующих ему карт диспаратности, доверия и областей открытия/закрытия;
- Превзошел аналоги, использовавшиеся ранее для анализа полнометражных стереофильмов на предмет наличия/отсутствия перепутанных ракурсов, на подготовленной эталонной выборке, состоящей из 900 сцен по 30 кадров. При этом удалось добиться увеличения точности классификации на 8%.

С помощью предложенного метода проанализировано 50 наиболее просматриваемых VR180-видео, в 10 из которых была найдена 21 сцена с перепутанными ракурсами. Поэтому вероятность встретить данные искажения при просмотре видео в формате VR180 остается достаточно высокой.

Результаты главы опубликованы автором в [1; 3; 19; 20; 24].

Приложение А. Дополнительные результаты анализа видео в формате VR180

А.1 Методология

Для оценки технического качества видео в формате VR180 рассматривались следующие стереоскопические параметры:

1. Положительный параллакс;
2. Отрицательный параллакс;
3. Цветовые искажения;
4. Искажения резкости;
5. Вертикальный сдвиг одного ракурса относительно другого;
6. Поворот одного ракурса относительно другого;
7. Разница в масштабе между ракурсами;
8. Перепутанные ракурсы.

Таким образом, помимо рассматриваемых в данной работе стереоскопических искажений также оценивалась величина параллакса в видео. Для этого использовался модифицированный метод из работы [95] по поиску экстремальных положительных и отрицательных значений диспаратности. Основное отличие модификации заключается в вычислении взвешенной гистограммы значений диспаратности с использованием соответствующих значений доверия, вычисление которого описано в разделе 1.4.1, вместо использования гистограммы, не учитывающей доверия, в базовом методе. В остальном, используемый метод не отличается от базового, и построенная гистограмма значений диспаратности используется для нахождения положительного и отрицательного параллакса в кадре путем вычисления 5 и 95 перцентилей в ней.

А.2 Анализ параллакса

На рисунке 45 представлены средние значения диспаратности для наиболее близких и наиболее удаленных объектов в проанализированных видео. Положительные значения диспаратности соответствуют объектам, находящимся за плоскостью экрана, в то время как отрицательные значения – объектам, находящимся перед плоскостью экрана. На графике индивидуальные видео представлены в виде линий, соединяющих точки, представляющие значения наибольшего положительного и наименьшего отрицательного параллаксов в видео, измеренные в процентах от ширины экрана. Данные линии соответствуют приблизительному расстоянию между объектами, находящимися дальше всего от зрителей, и объектами, находящимися ближе всего к зрителям. Чем длиннее линия, тем больше диапазон параллакса в видео и более заметен 3D-эффект. Рисунок 45 демонстрирует ряд видео с несущественным диапазоном параллакса, однако среди проанализированных видео также встречаются видео с огромным диапазоном параллакса. На графике также изображены линии тренда для положительных и отрицательных значений диспаратности, показывающие средние диапазоны параллакса в зависимости от даты выпуска видео. Видео с диапазоном параллакса больше среднего, вероятно, могут вызвать дискомфорт при просмотре на некоторых шлемах виртуальной реальности, – неутешительный вывод, так как диапазон параллакса трудно изменить во время постобработки видео. Дополнительно ряд видео обладает значительным положительным параллаксом. Так как левый и правый ракурсы находятся непосредственно перед соответствующими глазами зрителя при просмотре видео в шлеме виртуальной реальности, нулевое значение параллакса (0%) соответствует объектам, находящимся на максимально удаленном от зрителя расстоянии, то есть на уровне “бесконечности”. Для сравнения, нулевое значение параллакса для обычного стереоскопического формата соответствует объектам, находящимся на уровне плоскости экрана. Поэтому значения положительного параллакса в VR180-видео

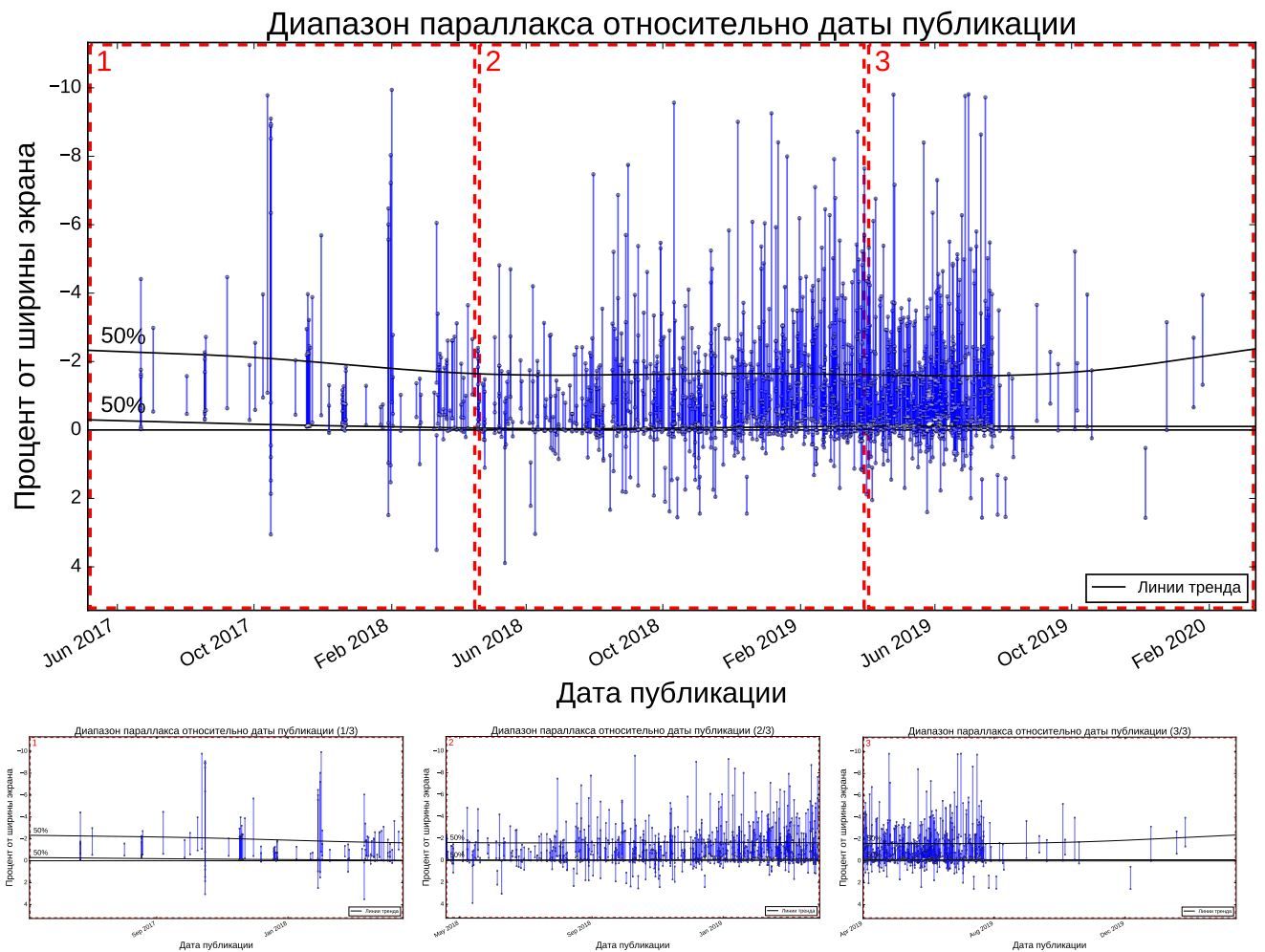


Рисунок 45 — Общие результаты анализа диапазонов параллакса в VR180-видео относительно даты их публикации. Для удобства просмотра также представлены увеличенные фрагменты для разных временных отрезков.

должны быть как можно меньше, так как они соответствуют объектам, находящимся за уровнем “бесконечности” – невозможная ситуация для мозга зрителя. То есть при просмотре сцены со значительным положительным параллаксом в формате VR180 с большой вероятностью зрители будут испытывать дискомфорт.

Рисунок 46 демонстрирует разбиение сцен каждого видео по соответствующим значениям диапазона параллакса. На оси x представлены различные видео, в то время как на оси y отложены столбцы для представления количества кадров в каждом видео с хорошими, промежуточными и плохими значениями диапазонов параллакса. Почти все кадры видео на левой части графика обла-

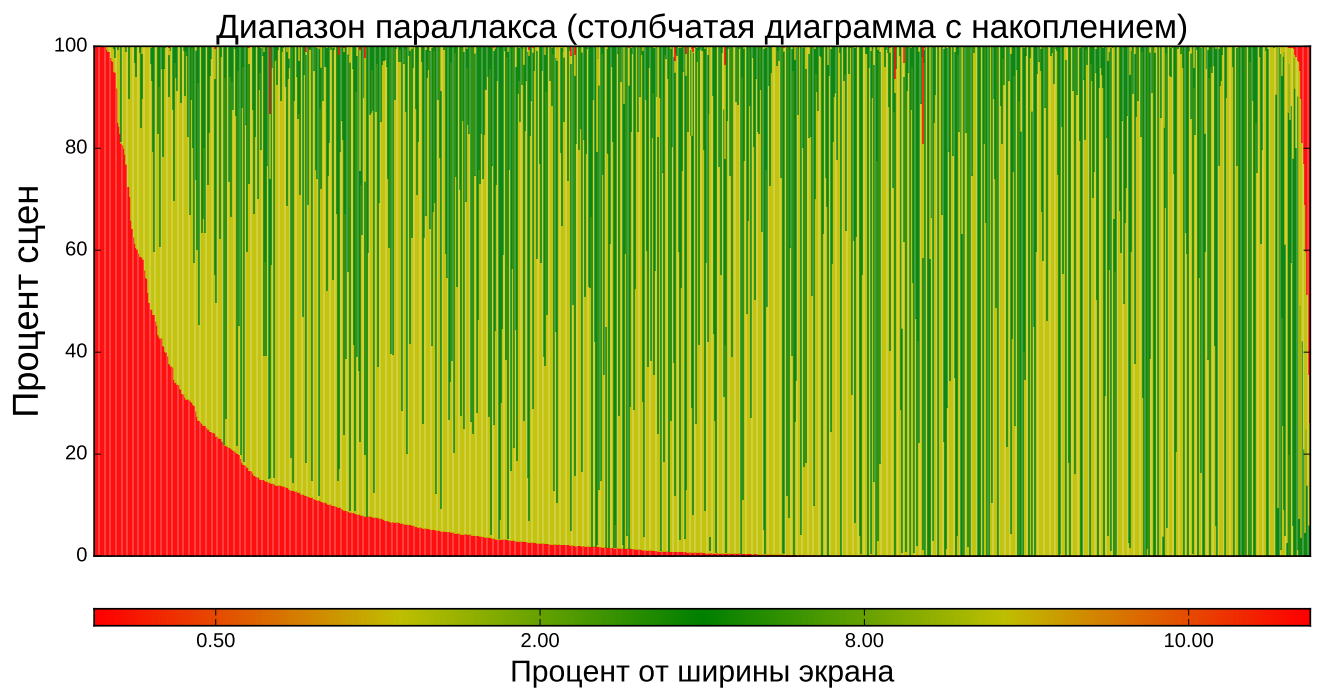


Рисунок 46 — Диапазон параллаксов для каждого кадра видео в формате VR180, отсортированные по количеству кадров с низким диапазоном параллакса.

дают маленьким диапазоном параллакса, то есть 3D-эффект в них может быть совсем не заметен. Видео в центре графика обладают средними значениями диапазонов параллакса, но в них все равно встречаются сцены с слишком маленьким, либо слишком высоким диапазоном параллаксов. Несколько видео в правой части графика в основном состоят из кадров с огромными значениями диапазонов параллакса. Согласно рисунку 46, общая ситуация значительно лучше по сравнению с предыдущим графиком: большинство кадров обладает средними значениями диспаратности, а огромные диапазоны параллакса встречаются довольно редко. Однако большое количество кадров слишком “плоские”, то есть обладают небольшими значениями диапазонов параллакса, и воспринимаемый стереоэффект от просмотра таких сцен будет незначительным.

На рисунках 47-50 продемонстрированы найденные примеры с большими положительными значениями диспаратности, а на рисунках 51-54 – с большими отрицательными значениями диспаратности.



Рисунок 47 — Визуализация кадра #12745 с большими положительными значениями диспаратности из видео youtu.be/AVTcC513iLQ.



Рисунок 48 — Визуализация кадра #3864 с большими положительными значениями диспаратности из видео youtu.be/KAoQwCVQFDU.

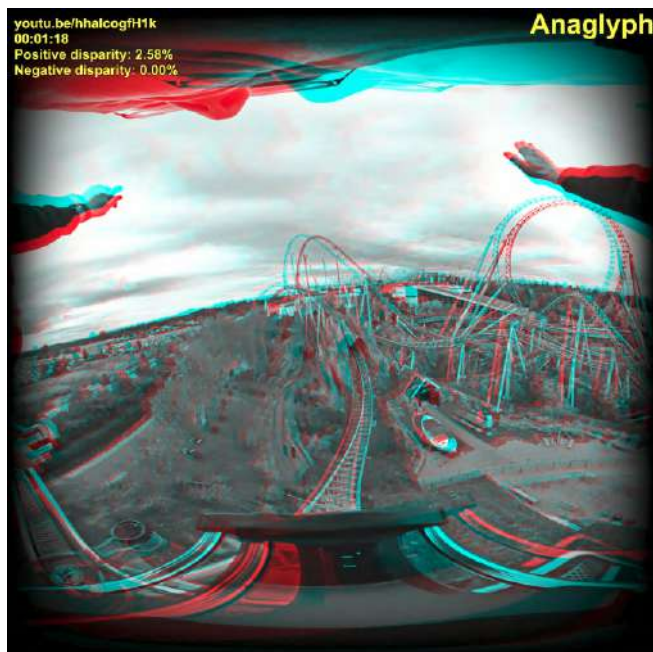


Рисунок 49 — Визуализация кадра #2346 с большими положительными значениями диспаратности из видео youtu.be/hhaIcogfH1k.

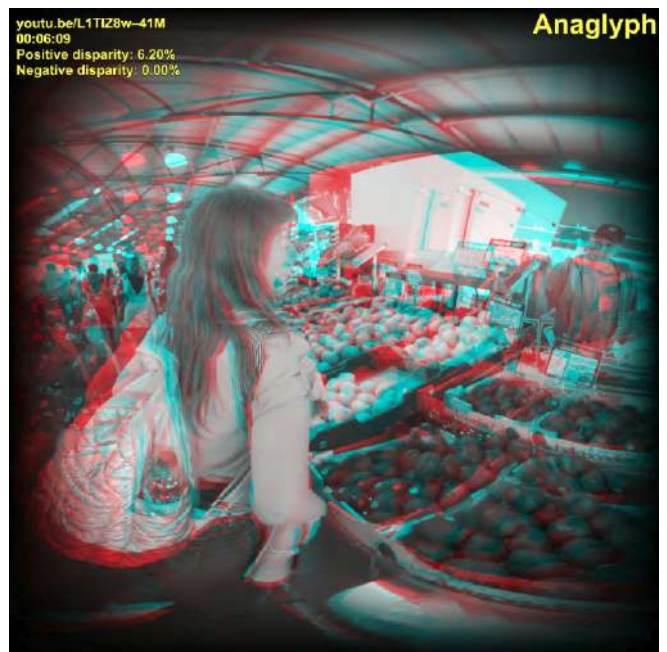


Рисунок 50 — Визуализация кадра #11084 с большими положительными значениями диспаратности из видео youtu.be/L1TIZ8w-41M.



Рисунок 51 — Визуализация кадра #7574 с большими отрицательными значениями диспаратности из видео youtu.be/vwKjoXx2Ic.



Рисунок 52 — Визуализация кадра #498 с большими отрицательными значениями диспаратности из видео youtu.be/8MY42p1_yh4.



Рисунок 53 — Визуализация кадра #9431 с большими отрицательными значениями диспаратности из видео youtu.be/AVTcC513iLQ.



Рисунок 54 — Визуализация кадра #6353 с большими отрицательными значениями диспаратности из видео youtu.be/CHh3pgJXhwM.

А.3 Примеры найденных стереоскопических искажений в VR180

В данном разделе на рисунках 55-81 приведены примеры кадров со стереоскопическими искажениями, найденные в ходе анализа 1000 видео в формате VR180.

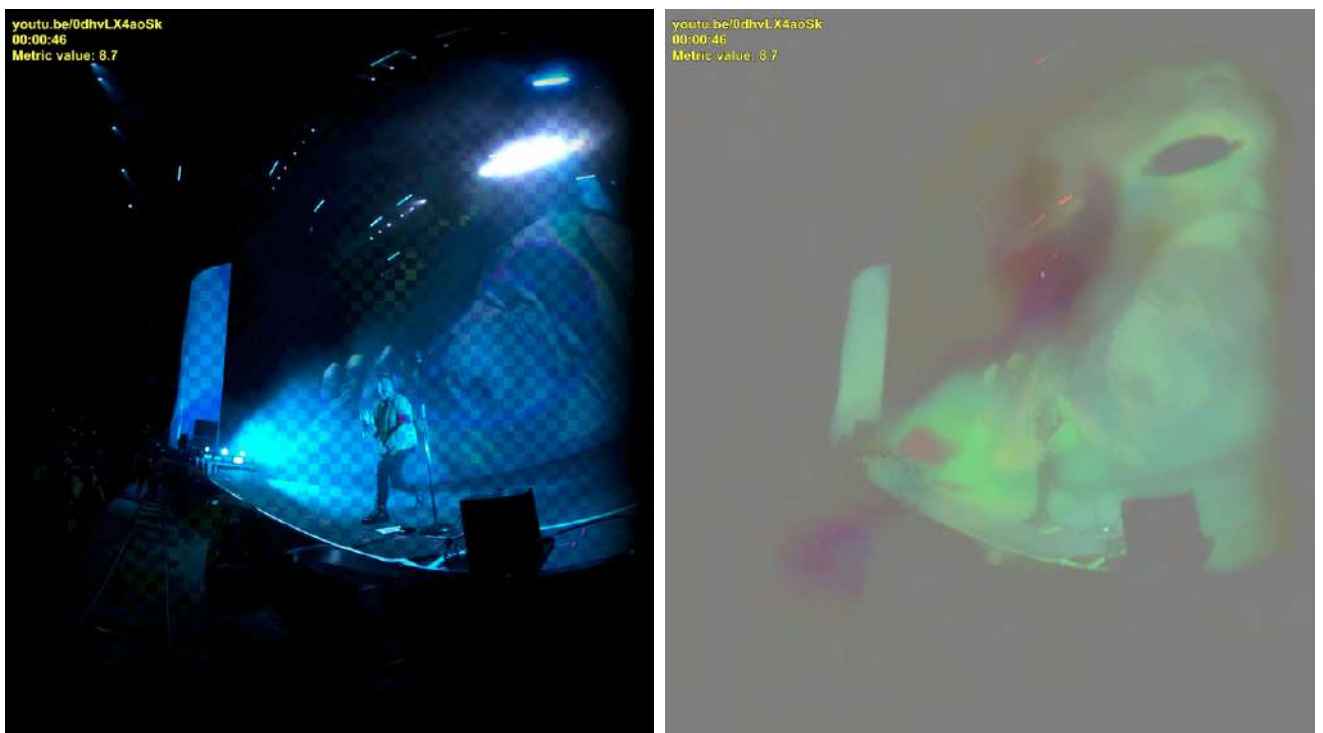


Рисунок 55 — Визуализация кадра #1382 с цветовыми искажениями из видео youtu.be/0dhvLX4aoSk. Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.



Рисунок 56 — Визуализация кадра #1267 с цветовыми искажениями из видео youtu.be/2vwKjoXx2Ic. Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.

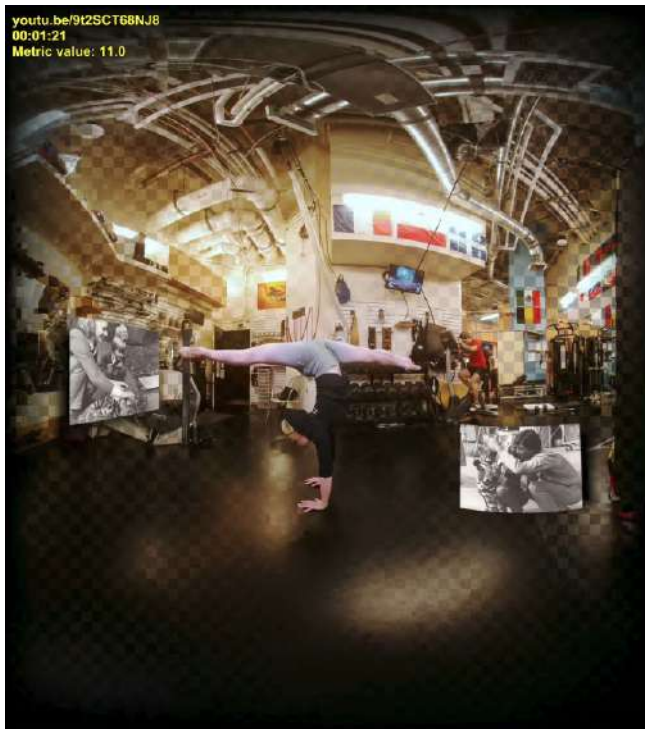


Рисунок 57 — Визуализация кадра #2436 с цветовыми искажениями из видео youtu.be/9t2SCT68NJ8. Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.



Рисунок 58 — Визуализация кадра #2259 с цветовыми искажениями из видео youtu.be/HBasBaz22dE. Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.



Рисунок 59 — Визуализация кадра #860 с искажениями резкости из видео youtu.be/2vwKjoXx2Ic.



Рисунок 60 — Визуализация кадра #3641 с искажениями резкости из видео youtu.be/8MY42p1_yh4.



Рисунок 61 — Визуализация кадра #6152 с искажениями резкости из видео youtu.be/AVTcC513iLQ.

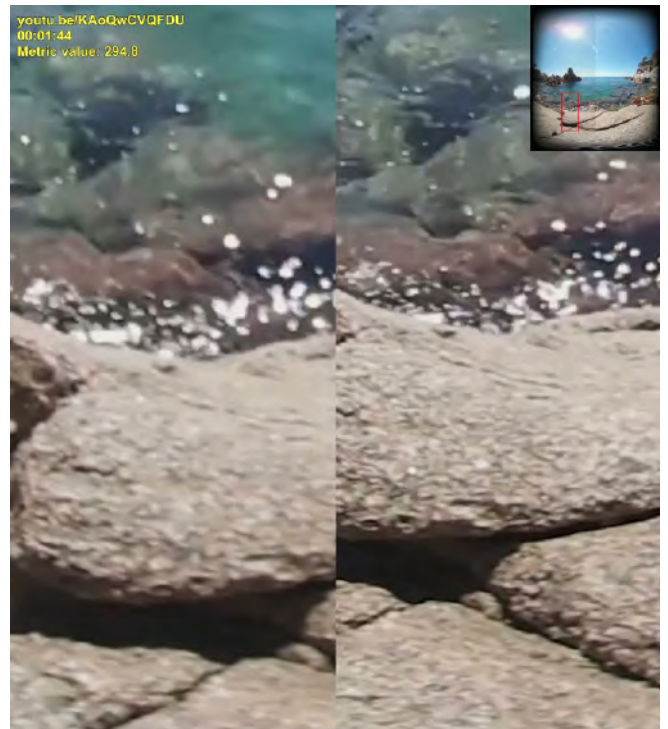


Рисунок 62 — Визуализация кадра #3129 с искажениями резкости из видео youtu.be/KAoQwCVQFDU.

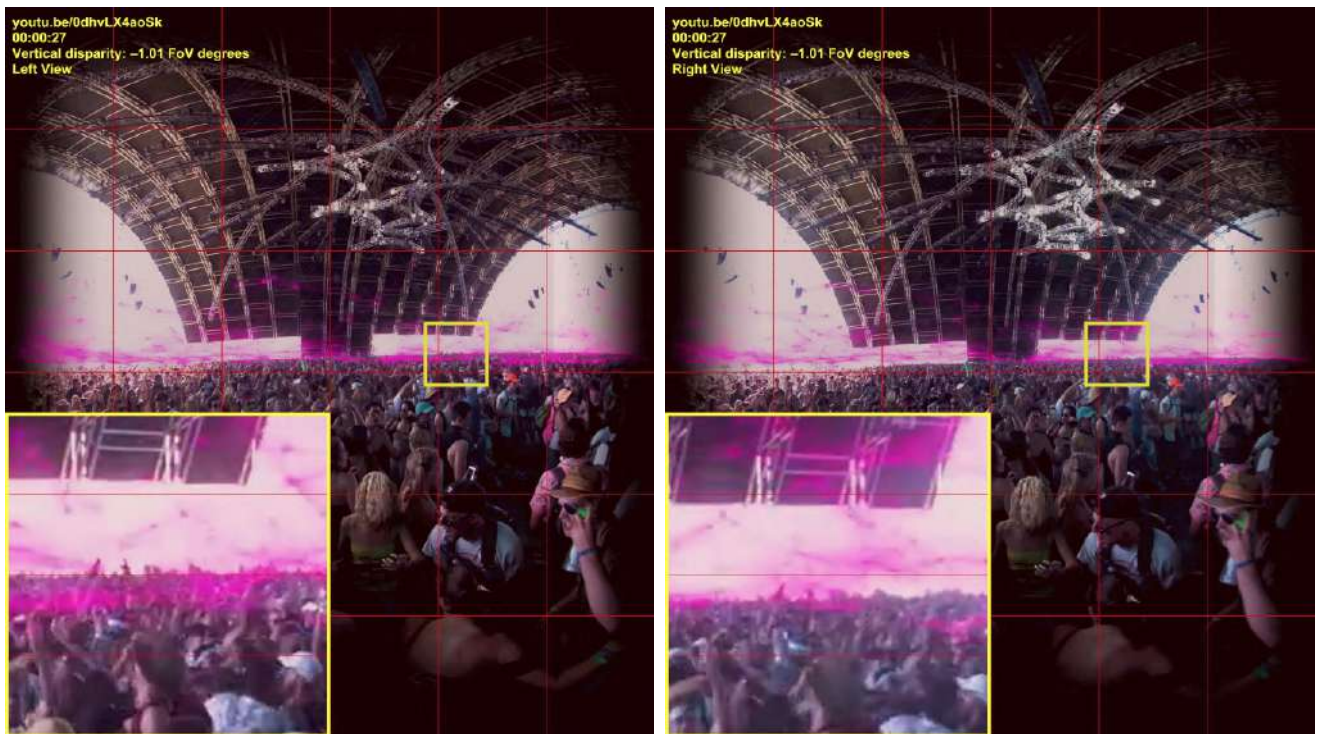


Рисунок 63 — Визуализация кадра #814 с вертикальным сдвигом из видео youtu.be/0dhvLX4aoSk.

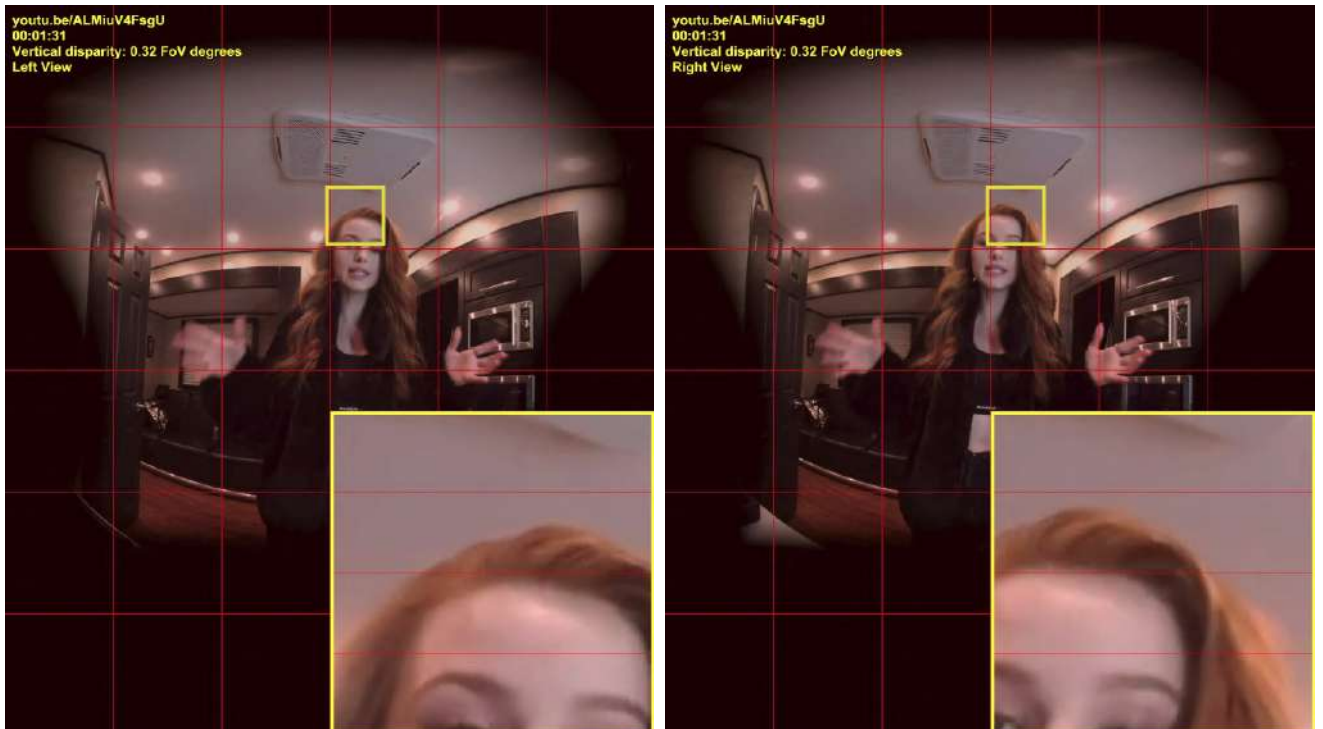


Рисунок 64 — Визуализация кадра #2749 с вертикальным сдвигом из видео youtu.be/ALMiuV4FsgU.

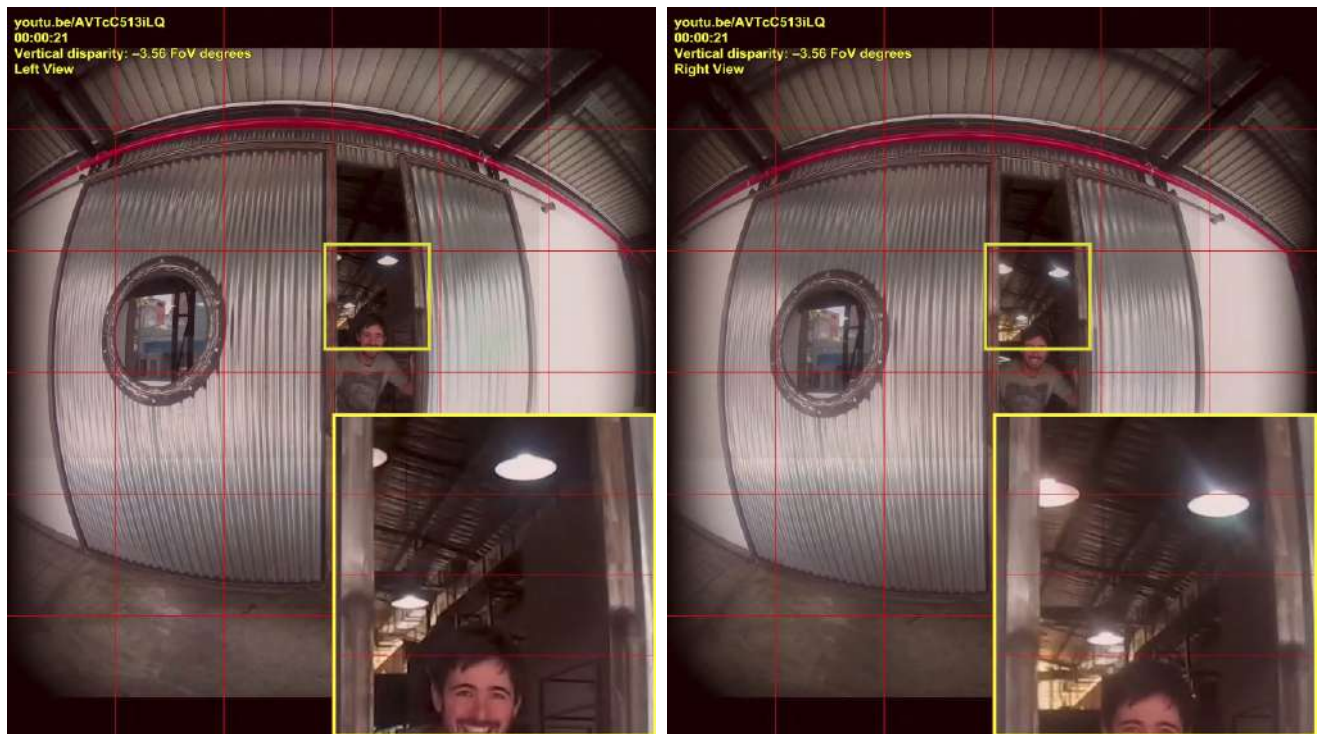


Рисунок 65 — Визуализация кадра #630 с вертикальным сдвигом из видео youtu.be/AVTcC513iLQ.

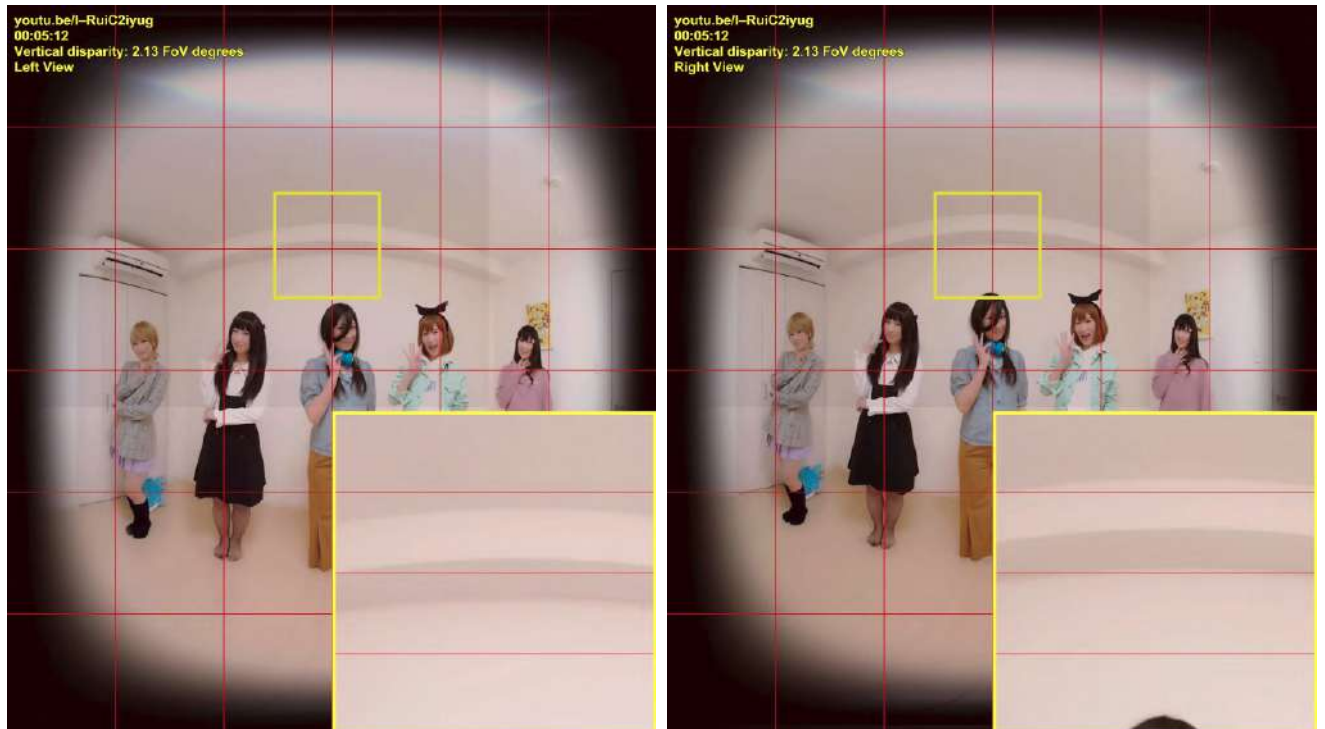


Рисунок 66 — Визуализация кадра #9361 с вертикальным сдвигом из видео youtu.be/l-RuiC2iyug.

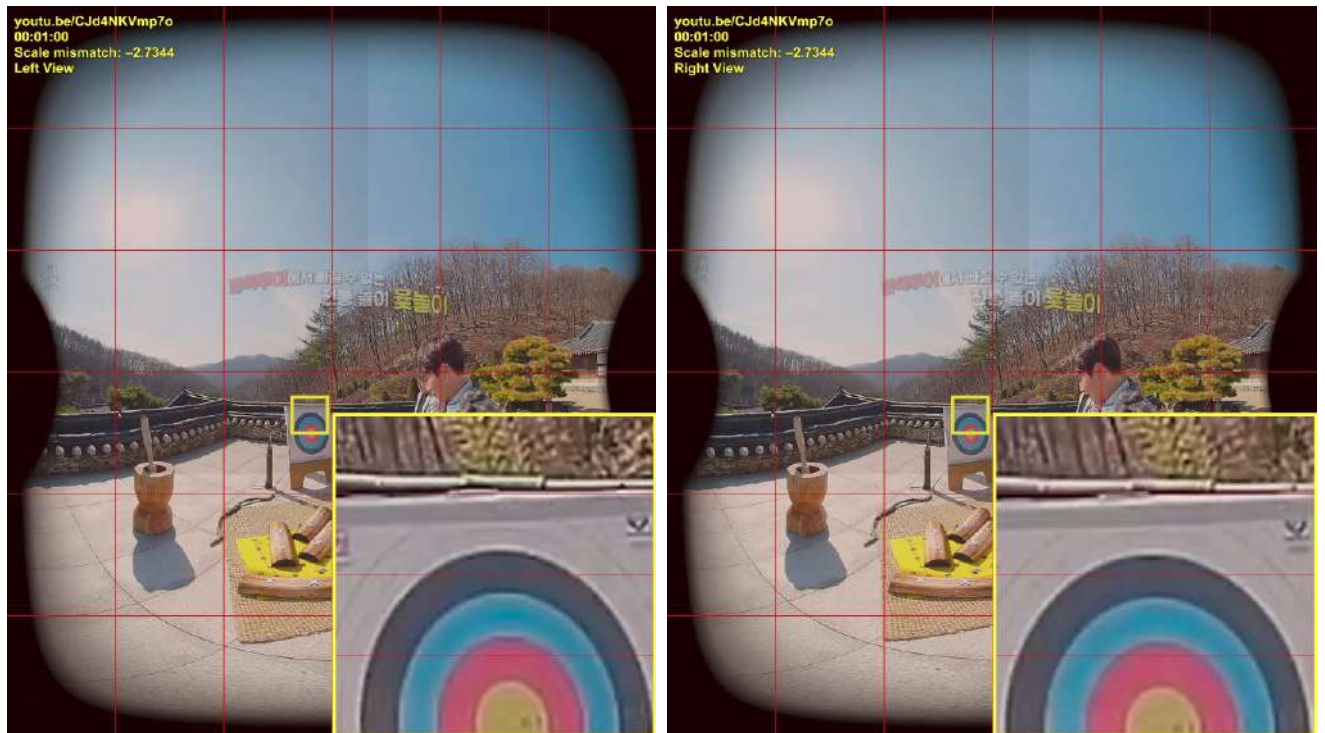


Рисунок 67 — Визуализация кадра #1824 с масштабированием из видео youtu.be/CJd4NKVmp7o.

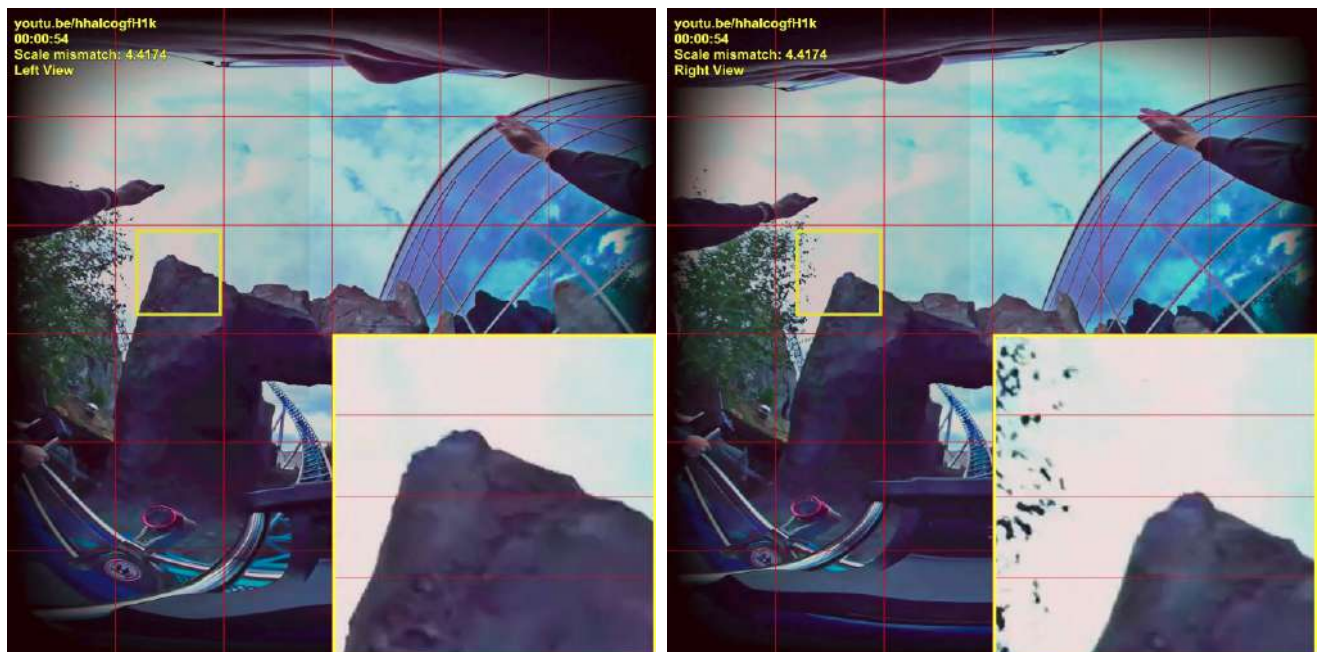


Рисунок 68 — Визуализация кадра #1630 с масштабированием из видео youtu.be/hhaIcogfH1k.

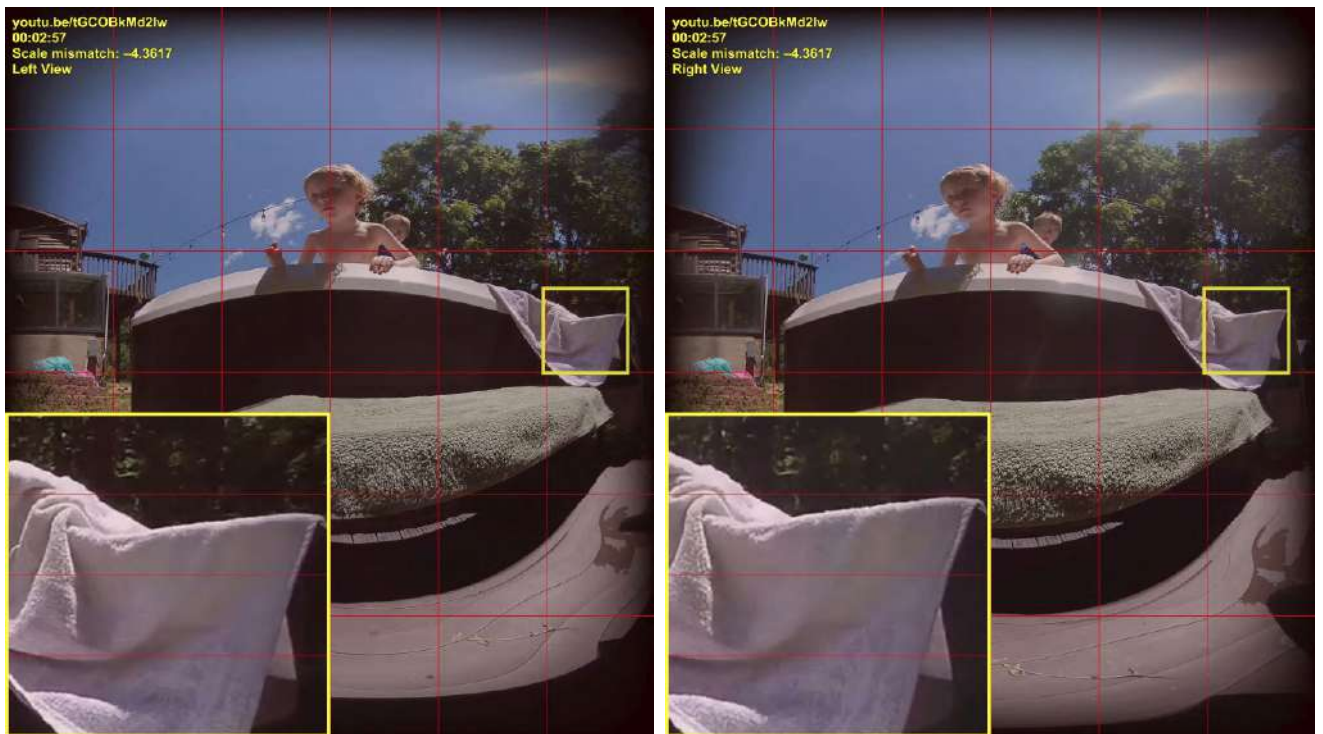


Рисунок 69 — Визуализация кадра #5311 с масштабированием из видео youtu.be/GCOBkMd2Iw.



Рисунок 70 — Визуализация кадра #768 с масштабированием из видео youtu.be/xfEX2de3N4U.

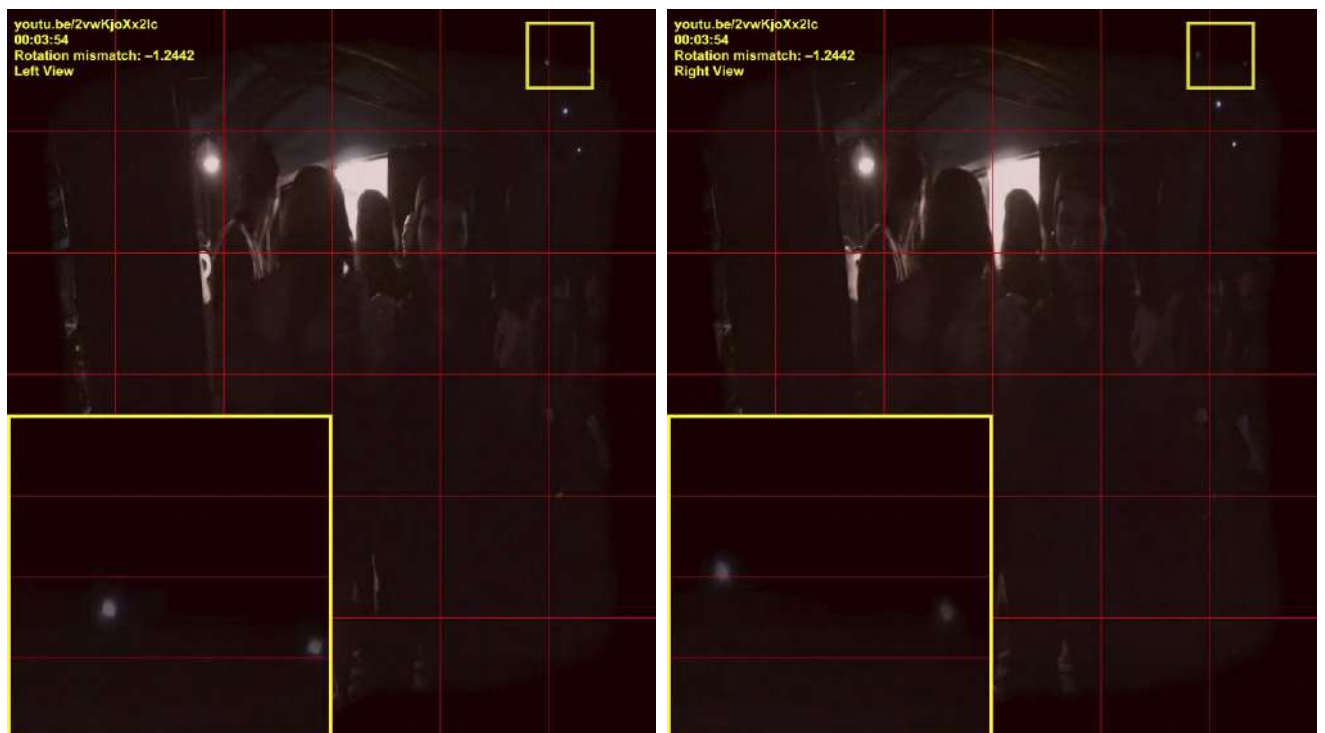


Рисунок 71 — Визуализация кадра #7022 с поворотом из видео youtu.be/2vwKjoXx2Ic.

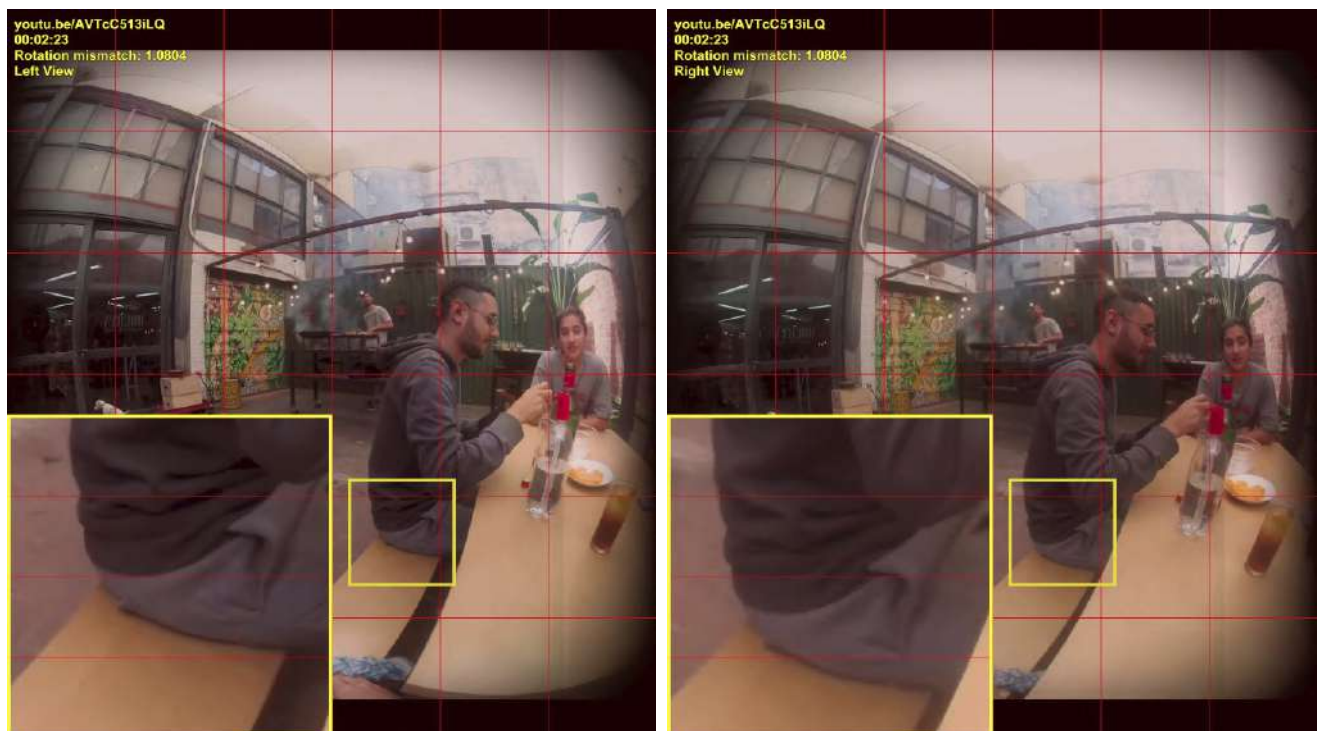


Рисунок 72 — Визуализация кадра #4302 с поворотом из видео youtu.be/AVTcC513iLQ.



Рисунок 73 — Визуализация кадра #2690 с поворотом из видео youtu.be/hhaIcogfH1k.

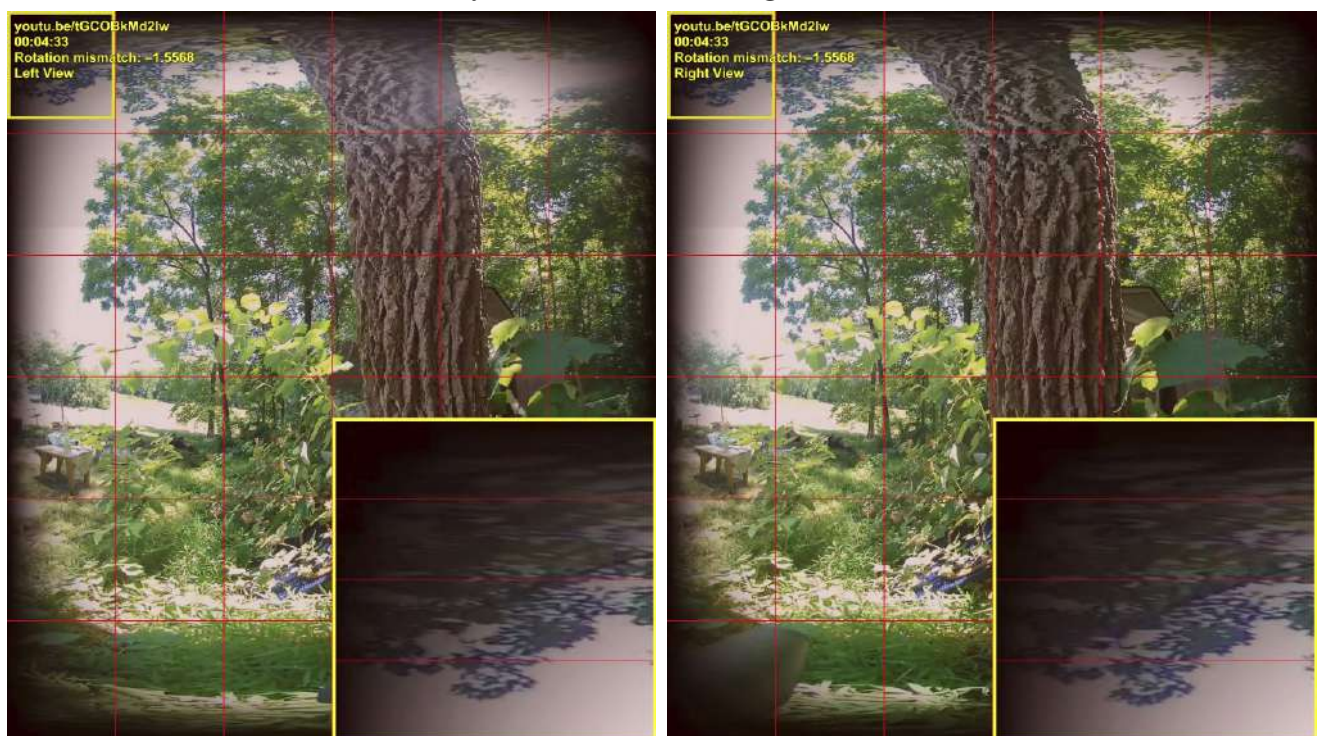


Рисунок 74 — Визуализация кадра #8193 с поворотом из видео youtu.be/tGCOBkMd2Iw.



Рисунок 75 — Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения компьютерной графики. Кадр из видео “The Aces”.



Рисунок 76 — Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения титров. Кадр из видео “All Time Low - Dressing Room Tour at Lollapalooza”.

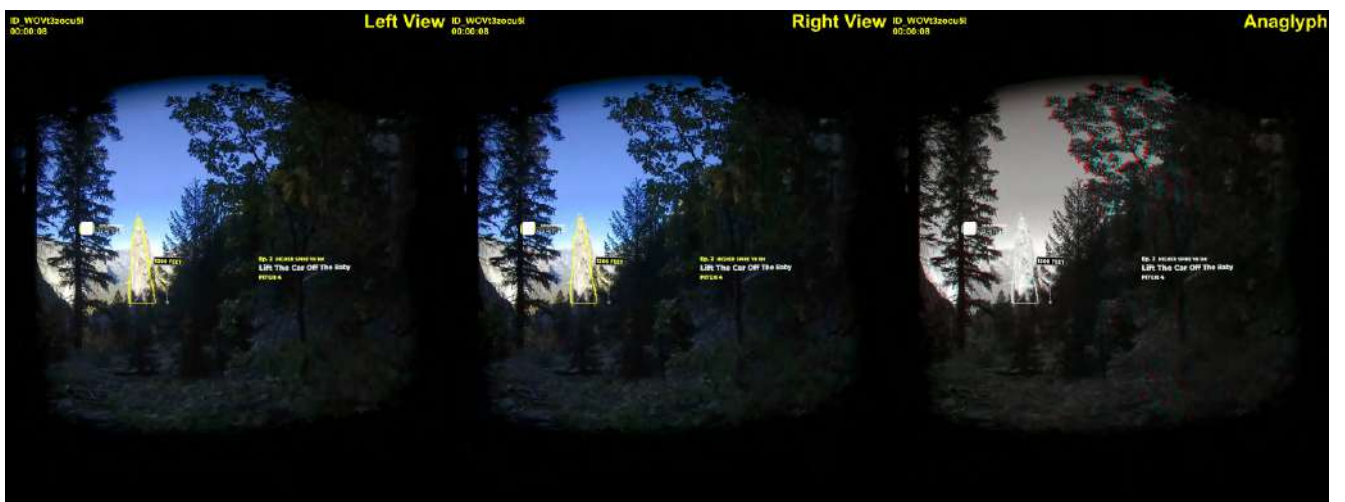


Рисунок 77 — Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения элементов компьютерной графики титров. Кадр из видео “Lift The Car Off The Baby | Yosemite Higher Spire Free”.



Рисунок 78 — Пример сцены с полностью перепутанными ракурсами. Кадр из видео “Hayley Kiyoko - VR180”.



Рисунок 79 — Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения титров. Кадр из видео “The Man - VR180”.



Рисунок 80 — Пример сцены с перепутанными ракурсами, возникшими из-за неправильного использования спецэффектов. Кадр из видео “Google Pixel 2 — Up-close in VR180”.



Рисунок 81 — Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения элементов компьютерной графики. Кадр из видео “Best Gaming Console (VR180)”.

Заключение

В ходе диссертационного исследования были получены следующие основные результаты:

1. Разработан новый нейросетевой метод одновременной оценки цветовых искажений и искажений резкости между ракурсами стереоскопического видео. Предложенный метод значительно сократил число ложноположительных срабатываний за счет одновременного учета рассматриваемых искажений и по результатам объективного сравнения превзошел аналоги, ранее использовавшиеся при анализе полнометражных стереоскопических фильмов.
2. Разработан новый нейросетевой метод оценки геометрических искажений между ракурсами стереоскопического видео. По результатам объективного сравнения предложенный метод превзошел аналоги, ранее использовавшиеся при анализе полнометражных стереоскопических фильмов. Предложенный метод также позволяет автоматически исправлять найденные геометрические искажения.
3. Разработан новый нейросетевой метод поиска перепутанных ракурсов в стереоскопических видео. По результатам объективного сравнения предложенный алгоритм превзошел аналоги, ранее использовавшиеся при анализе полнометражных стереоскопических фильмов.
4. Проведено исследование объективного качества 1000 VR180-видео с помощью разработанных методов. Исследование показало наличие по меньшей мере одного вида стереоскопического искажения в каждом из проанализированных материалов.

Предложенные алгоритмы могут быть использованы для разработки программных инструментов автоматического контроля качества стереоскопических видео, включая видео в формате виртуальной реальности, а также послужить основой для создания автоматических методов их исправления. Предло-

женная методология объективной оценки качества видео в формате VR180 может быть использована для анализа стереоскопического качества новых видео, а результаты представленного анализа могут послужить их отправной точкой.

Дальнейшее развитие темы исследования может включать в себя:

- Исследование нейросетевых методов исправления искажений цвета и резкости. Для успешного исправления данных артефактов необходимо также восстанавливать значения в областях, которые не были успешно сопоставлены методом построения карт диспаратности, что подразумевает исследование методов по восстановлению деталей и/или пикселей в неизвестных областях.
- Исследование метода поиска объектов с неправильной глубиной. Как показало проведенное сравнение видео в формате VR180, большинство найденных сцен с перепутанными ракурсами возникают из-за неправильной постобработки видео путем неверного наложения компьютерной графики и/или титров.
- Исследование нейросетевых методов оценки и исправления других стереоскопических артефактов. Например, оценка временного сдвига между ракурсами стереоскопического видео и его исправление, что актуально для 3D-съемки. Или оценка размытости границ и поиск плоских объектов и сцен, что актуально для конвертации из 2D в 3D.

Данная работа была поддержана грантом СТАРТ-19-1 по теме “Разработка системы автоматической объективной оценки качества и исправления стереоскопических видео и видео в формате VR180”. Также работа была частично поддержана грантами РФФИ 15-01-08632 а про теме “Автоматизация создания и контроля качества стерео видео” и РФФИ №19-01-00785 по теме “Разработка нейросетевых алгоритмов обработки и сжатия видеопоследовательностей”. Обучение нейросетевых моделей производилось с использованием высокопроизводительного кластера IBM Polus факультета ВМК МГУ: <https://hpc.cmc.msu.ru/polus>.

Список литературы

1. *Antsiferova A., Vatolin D.* The influence of 3D video artifacts on discomfort of 302 viewers // 2017 International Conference on 3D Immersion (IC3D). — IEEE. 2017. — С. 1—8.
2. *Рожкова Г., Васильева Н.* Сравнительные трудности восприятия фильмов в 2D и 3D форматах // Мир техники кино. — 2010. — Т. 4, № 2. — С. 12—18.
3. *Рожкова Г., Алексеенко С.* Зрительный дискомфорт при восприятии стереоскопических изображений как следствие непривычного распределения нагрузки на различные механизмы зрительной системы // Мир техники кино. — 2011. — Т. 5, № 3. — С. 12—21.
4. *Рожкова Г., Крутцова Е., Забалуева Н.* Влияние наклона головы к плечу при просмотре фильмов 3D формата: общие закономерности и индивидуальные особенности // Мир техники кино. — 2012. — Т. 6, № 3. — С. 17—21.
5. *Васильева Н., Рожкова Г., Рожков С.* О пользе и вреде современных технологий формирования стереокиноизображений для людей с различным состоянием зрительных функций // Мир техники кино. — 2011. — Т. 5, № 1. — С. 7—15.
6. *Рожков С., Рожкова Г.* Искажения пространственных образов в стереокино: иллюзии уменьшения, увеличения и уплощения объектов // Мир техники кино. — 2013. — Т. 7, № 3. — С. 13—20.
7. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue / D. M. Hoffman [и др.] // Journal of vision. — 2008. — Т. 8, № 3. — С. 33—33.

8. The zone of comfort: Predicting visual discomfort with stereo displays / T. Shibata [и др.] // Journal of vision. — 2011. — Т. 11, № 8. — С. 11—11.
9. *Johnson P. V., Kim J., Banks M. S.* Visible artifacts and limitations in stereoscopic 3D Displays // Information Display. — 2017. — Т. 33, № 1. — С. 12—17.
10. Display-size dependent effects of 3D viewing on subjective impressions / Y. Miyashita [и др.] // ACM Transactions on Applied Perceptions (TAP). — 2022. — Т. 19, № 2. — С. 1—15.
11. *Lambooij M. T., IJsselsteijn W. A., Heynderickx I.* Visual discomfort in stereoscopic displays: a review // Stereoscopic Displays and Virtual Reality Systems XIV. — 2007. — Т. 6490. — С. 183—195.
12. *Morita T., Ando H.* Effects of viewing conditions on fatigue caused by watching 3DTV // The 2012 Annual Technical Conference & Exhibition. — SMPTE. 2012. — С. 1—9.
13. Influence of chroma variations on naturalness and image quality of stereoscopic images / A. Kuijsters [и др.] // Human Vision and Electronic Imaging XIV. Т. 7240. — SPIE. 2009. — С. 461—469.
14. An objective method for 3D quality prediction using visual annoyance and acceptability level / D. Khaustova [и др.] // Stereoscopic Displays and Applications XXVI. Т. 9391. — International Society for Optics, Photonics. 2015. — 93910P.
15. Automatic detection of artifacts in converted S3D video / A. Bokov [и др.] // Stereoscopic Displays and Applications XXV. Т. 9011. — International Society for Optics, Photonics. 2014. — С. 901112.
16. Toward fully automatic channel-mismatch detection and discomfort prediction for S3D video / A. Bokov [et al.] // 2016 International Conference on 3D Imaging (IC3D). — Liege, 2016. — P. 1—7.

17. *Lavrushkin S., Lyudvichenko V., Vatolin D.* Local Method of Color-Difference Correction Between Stereoscopic-Video Views // Proceedings of the 2018 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON). — 2018. — P. 1–4.
18. Автоматическое обнаружение и оценка цветовых искажений с использованием карт доверия / С. Д. Грохольский [и др.] // International Journal of Open Information Technologies. — Москва, 2017. — Т. 5, № 5. — С. 1–8.
19. *Lavrushkin S., Vatolin D.* Channel-Mismatch Detection Algorithm for Stereoscopic Video Using Convolutional Neural Network // Proceedings of the 2018 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON). — 2018. — P. 1–4.
20. *Лаврушкин С. В., Ватолин Д. С.* Разработка метода поиска перепутанных ракурсов в стереофильмах // Мир техники кино. — 2018. — 2018—1(12). — С. 12–17.
21. *Malyshev K., Lavrushkin S., Vatolin D.* Stereoscopic Dataset from A Video Game: Detecting Converged Axes and Perspective Distortions in S3D Videos // 2020 International Conference on 3D Immersion (IC3D). — IEEE, 2020.
22. *Lavrushkin S., Kozhemyakov K., Vatolin D.* Neural-Network-Based Detection Methods for Color, Sharpness, and Geometry Artifacts in Stereoscopic and VR180 Videos // 2020 International Conference on 3D Immersion (IC3D). — IEEE, 2020.
23. Нейросетевой алгоритм поиска областей открытия/закрытия в видеопоследовательностях / М. С. Великанов [и др.] // International Journal of Open Information Technologies. — Москва, 2020. — Т. 8, № 3. — С. 1–7.
24. Stereoscopic quality assessment of 1,000 VR180 videos using 8 metrics / S. Lavrushkin [и др.] // Electronic Imaging. — 2021. — Т. 33. — С. 1–7.

25. *Свидетельство о гос. регистрации программы для ЭВМ. Система для обучения и тестирования нейросетевых методов объективной оценки качества и исправления стереоскопических видео* / С. В. Лаврушкин. — № 2021610246 ; заявл. 07.12.2020 ; опубл. 12.01.2021, 2020665915 (Рос. Федерация).
26. Trends in S3D-movie quality evaluated on 105 films using 10 metrics / D. Vatolin [и др.] // *Electronic Imaging*. — 2016. — Т. 2016, № 5. — С. 1—10.
27. *Vatolin D., Bokov A. Sharpness mismatch and 6 other stereoscopic artifacts measured on 10 Chinese S3D movies* // *Electronic Imaging*. — 2017. — Т. 2017, № 5. — С. 137—144.
28. *Lowe D. G. Distinctive image features from scale-invariant keypoints* // *International journal of computer vision*. — 2004. — Т. 60, № 2. — С. 91—110.
29. *Bay H., Tuytelaars T., Van Gool L. Surf: Speeded up robust features* // *European conference on computer vision*. — Springer. 2006. — С. 404—417.
30. ORB: An efficient alternative to SIFT or SURF / E. Rublee [и др.] // *2011 International conference on computer vision*. — Ieee. 2011. — С. 2564—2571.
31. *Teed Z., Deng J. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow* // *16th European Conference on Computer Vision, ECCV 2020*. — Springer Science, Business Media Deutschland GmbH. 2020. — С. 402—419.
32. *Bar-Haim A., Wolf L. Scopeflow: Dynamic scene scoping for optical flow* // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. — 2020. — С. 7998—8007.
33. Maskflownet: Asymmetric feature matching with learnable occlusion mask / S. Zhao [и др.] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. — 2020. — С. 6278—6287.

34. Fast video super-resolution via classification / K. Simonyan [и др.] // 2008 15th IEEE international conference on image processing. — IEEE. 2008. — C. 349—352.
35. *Pentland A. P.* A new sense for depth of field // IEEE transactions on pattern analysis and machine intelligence. — 1987. — № 4. — C. 523—531.
36. *Elder J. H., Zucker S. W.* Local scale control for edge detection and blur estimation // IEEE Transactions on pattern analysis and machine intelligence. — 1998. — T. 20, № 7. — C. 699—716.
37. *Zhuo S., Sim T.* Defocus map estimation from a single image // Pattern Recognition. — 2011. — T. 44, № 9. — C. 1852—1858.
38. *Levin A., Lischinski D., Weiss Y.* A closed-form solution to natural image matting // IEEE transactions on pattern analysis and machine intelligence. — 2007. — T. 30, № 2. — C. 228—242.
39. *Cao Y., Fang S., Wang Z.* Digital multi-focusing from a single photograph taken with an uncalibrated conventional camera // IEEE Transactions on image processing. — 2013. — T. 22, № 9. — C. 3703—3714.
40. *Karaali A., Jung C. R.* Adaptive scale selection for multiresolution defocus blur estimation // 2014 IEEE International Conference on Image Processing (ICIP). — IEEE. 2014. — C. 4597—4601.
41. *Karaali A., Jung C. R.* Edge-based defocus blur estimation with adaptive scale selection // IEEE Transactions on Image Processing. — 2017. — T. 27, № 3. — C. 1126—1137.
42. *Chakrabarti A., Zickler T., Freeman W. T.* Analyzing spatially-varying blur // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — IEEE. 2010. — C. 2512—2519.
43. Estimating spatially varying defocus blur from a single image / X. Zhu [и др.] // IEEE Transactions on Image Processing. — 2013. — T. 22, № 12. — C. 4879—4891.

44. Non-parametric blur map regression for depth of field extension / L. D'Andrès [и др.] // IEEE Transactions on Image Processing. — 2016. — Т. 25, № 4. — С. 1660—1673.
45. *Alireza Golestaneh S., Karam L. J.* Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — С. 5800—5809.
46. *Narvekar N. D., Karam L. J.* A no-reference image blur metric based on the cumulative probability of blur detection (CPBD) // IEEE Transactions on Image Processing. — 2011. — Т. 20, № 9. — С. 2678—2683.
47. *Kumar J., Chen F., Doermann D.* Sharpness estimation for document and scene images // Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). — IEEE. 2012. — С. 3292—3295.
48. A local metric for defocus blur detection based on CNN feature learning / K. Zeng [и др.] // IEEE Transactions on Image Processing. — 2018. — Т. 28, № 5. — С. 2107—2115.
49. A unified approach of multi-scale deep and hand-crafted features for defocus estimation / J. Park [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — С. 1736—1745.
50. Deep defocus map estimation using domain adaptation / J. Lee [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2019. — С. 12222—12230.
51. R²MRF: Defocus Blur Detection via Recurrently Refining Multi-Scale Residual Features / C. Tang [и др.] // Proceedings of the AAAI Conference on Artificial Intelligence. Т. 34. — 2020. — С. 12063—12070.
52. *Cun X., Pun C.-M.* Defocus blur detection via depth distillation // European Conference on Computer Vision. — Springer. 2020. — С. 747—763.

53. *Karaali A., Harte N., Jung C. R.* Deep Multi-Scale Feature Learning for Defocus Blur Estimation // arXiv preprint arXiv:2009.11939. — 2020.
54. *Winkler S.* Efficient measurement of stereoscopic 3D video content issues // Image Quality and System Performance XI. T. 9016. — International Society for Optics, Photonics. 2014. — 90160Q.
55. A stereo camera distortion detecting method for 3DTV video quality assessment / Q. Dong [и др.] // 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. — IEEE. 2013. — C. 1—4.
56. *Devernay F., Pujades S., AV V. C.* Focus mismatch detection in stereoscopic content // Stereoscopic Displays and Applications XXIII. T. 8288. — International Society for Optics, Photonics. 2012. — 82880E.
57. *Liu M., Müller K.* Automatic analysis of sharpness mismatch between stereoscopic views for stereo 3D videos // 2014 International Conference on 3D Imaging (IC3D). — IEEE. 2014. — C. 1—6.
58. VQMT3D Project: Report 10 on 3D-video Quality Analysis. — [Online]. https://videoprocessing.ai/stereo_quality/report10.html.
59. *Egnal G., Wildes R. P.* Detecting binocular half-occlusions: Empirical comparisons of five approaches // IEEE Transactions on pattern analysis and machine intelligence. — 2002. — T. 24, № 8. — C. 1127—1133.
60. *Ronneberger O., Fischer P., Brox T.* U-net: Convolutional networks for biomedical image segmentation // International Conference on Medical image computing and computer-assisted intervention. — Springer. 2015. — C. 234—241.
61. Residual Conv-Deconv Grid Network for Semantic Segmentation / D. Fourure [и др.] // BMVC 2017. — 2017.

62. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification / K. He [и др.] // Proceedings of the IEEE international conference on computer vision. — 2015. — С. 1026—1034.
63. Fast global image smoothing based on weighted least squares / D. Min [и др.] // IEEE Transactions on Image Processing. — 2014. — Т. 23, № 12. — С. 5638—5653.
64. *Glorot X., Bengio Y.* Understanding the difficulty of training deep feedforward neural networks // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. — JMLR Workshop, Conference Proceedings. 2010. — С. 249—256.
65. *Kingma D. P., Ba L. J.* Adam: A Method for Stochastic Optimization // International Conference on Learning Representations (ICLR). — 2015.
66. A naturalistic open source movie for optical flow evaluation / D. J. Butler [и др.] // European Conference on Computer Vision. — Springer. 2012. — С. 611—625.
67. Video Quality Measurement Tool 3D Project. — [Online]. [https : / / videoprocessing.ai/stereo_quality/](https://videoprocessing.ai/stereo_quality/).
68. *Al-Zahrani A., Ipson S. S., Haigh J.* Applications of a direct algorithm for the rectification of uncalibrated images // Information Sciences. — 2004. — Т. 160, № 1—4. — С. 53—71.
69. *Loop C., Zhang Z.* Computing rectifying homographies for stereo vision // Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). Т. 1. — IEEE. 1999. — С. 125—131.
70. Stereo rectification of uncalibrated and heterogeneous images / S. Kumar [и др.] // Pattern Recognition Letters. — 2010. — Т. 31, № 11. — С. 1445—1452.

71. *Nguyen T. P., Tran T. H.-P., Jeon J. W.* Multi-level Feature Pooling Network for Uncalibrated Stereo Rectification in Autonomous Vehicles // IEEE Transactions on Industrial Electronics. — 2020.
72. Correcting unsynchronized zoom in 3d video / C. Doutre [и др.] // Proceedings of 2010 IEEE International Symposium on Circuits and Systems. — IEEE. 2010. — С. 3244—3247.
73. *Pekkucuksen I. E., Batur A. U., Zhang B.* A real-time misalignment correction algorithm for stereoscopic 3D cameras // Stereoscopic Displays and Applications XXIII. Т. 8288. — International Society for Optics, Photonics. 2012. — 82880J.
74. *Voronov A., Borisov A., Vatolin D.* System for automatic detection of distorted scenes in stereo video // Proceedings of Sixth International Workshop on Video Processing and Quality Metrics (VPQM). — 2012.
75. *Fischler M. A., Bolles R. C.* Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Communications of the ACM. — 1981. — Т. 24, № 6. — С. 381—395.
76. Dsac-differentiable ransac for camera localization / E. Brachmann [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — С. 6684—6692.
77. *Brachmann E., Rother C.* Neural-guided RANSAC: Learning where to sample model hypotheses // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2019. — С. 4322—4331.
78. Learning to find good correspondences / K. M. Yi [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — С. 2666—2674.
79. Acne: Attentive context normalization for robust permutation-equivariant learning / W. Sun [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2020. — С. 11286—11295.

80. *Rocco I., Arandjelovic R., Sivic J.* Convolutional neural network architecture for geometric matching // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — С. 6148—6157.
81. *Rocco I., Arandjelović R., Sivic J.* End-to-end weakly-supervised semantic alignment // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — С. 6917—6925.
82. Deep residual learning for image recognition / К. Хе [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — С. 770—778.
83. *Ioffe S., Szegedy C.* Batch normalization: Accelerating deep network training by reducing internal covariate shift // International Conference on Machine Learning. — PMLR. 2015. — С. 448—456.
84. Content-based pseudoscopic view detection / J. Lee [и др.] // Journal of Signal Processing Systems. — 2012. — Т. 68, № 2. — С. 261—271.
85. *Palou G., Salembier P.* Monocular depth ordering using t-junctions and convexity occlusion cues // IEEE transactions on image processing. — 2013. — Т. 22, № 5. — С. 1926—1939.
86. *Lee H., Jung C., Kim C.* Depth map estimation based on geometric scene categorization // The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision. — IEEE. 2013. — С. 170—173.
87. *Hoiem D., Efros A. A., Hebert M.* Recovering occlusion boundaries from an image // International Journal of Computer Vision. — 2011. — Т. 91, № 3. — С. 328—346.
88. Deep ordinal regression network for monocular depth estimation / H. Fu [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — С. 2002—2011.

89. Digging into self-supervised monocular depth estimation / C. Godard [и др.] // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2019. — С. 3828—3838.
90. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer / R. Ranftl [и др.] // IEEE transactions on pattern analysis and machine intelligence. — 2020.
91. Learning to recover 3d scene shape from a single image / W. Yin [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2021. — С. 204—213.
92. *Knee M.* Getting machines to watch 3d for you // SMPTE Motion Imaging Journal. — 2012. — Т. 121, № 3. — С. 52—58.
93. *Bouchard J., Nazzar Y., Clark J. J.* Half-occluded regions and detection of pseudoscopy // 2015 International Conference on 3D Vision. — IEEE. 2015. — С. 215—223.
94. *Shestov A., Voronov A., Vatolin D.* Detection of swapped views in stereo image // 22st GraphiCon International Conference on Computer Graphics and Vision. — 2012. — С. 23—27.
95. Methodology for stereoscopic motion-picture quality assessment / A. Voronov [и др.] // Stereoscopic Displays and Applications XXIV. Т. 8648. — International Society for Optics, Photonics. 2013. — С. 864810.

Публикации автора по теме диссертации

Научные статьи, опубликованные в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности и отрасли наук:

1. *Ватолин Д. С., Лаврушкин С. В.* Исследование и предсказание заметности перепутанных ракурсов в стереовидео // Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика. — 2016. — № 4. — С. 40—46. — RSCI, Импакт-фактор 0,246 (РИНЦ), EDN: XDBUVL [0,4375 п.л. / 0,4155 п.л.].

Работа полностью выполнена автором самостоятельно. Автором разработаны метод поиска перепутанных ракурсов в стереовидео и метод оценки заметности перепутанных ракурсов, подготовлен набор данных с субъективными оценками заметности перепутанных ракурсов и проведены экспериментальные сравнения предложенных методов.

2. *Лаврушкин С. В., Ватолин Д. С.* Разработка нейросетевых методов распространения цвета в видео // Цифровая обработка сигналов. — 2020. — № 3. — С. 25—34. — RSCI, Импакт-фактор 0,520 (РИНЦ), EDN: HWWNUW [0,625 п.л. / 0,6 п.л.].

Работа полностью выполнена автором самостоятельно. Автором разработаны нейросетевые методы распространения цвета в видео, показана дифференцируемость фильтра FGS и продемонстрирована возможность его использования в качестве нейросетевого блока и проведены экспериментальные сравнения предложенных методов.

3. *Лаврушкин С. В.* Программная система для автоматической оценки стереоскопических искажений видео в формате VR180 // Вычислительные методы и программирование. — 2025. — Т. 26, № 3. — С. 340—365. — RSCI, Импакт-фактор 0,789 (РИНЦ), EDN: VEISIO [1,625 п.л. / 1,625 п.л.].

Работа полностью выполнена автором самостоятельно. Автором разработаны нейросетевые методы оценки стереоскопических искажений, проведены экспериментальные сравнения предложенных методов и реализована программная система для оценки искажений стереоскопических видео, включая стереоскопический формат VR180.

Список рисунков

1	Количество полнометражных стереоскопических фильмов по годам, разделенных по методу производства. Информация по фильмам получена со страницы https://en.wikipedia.org/wiki/List_of_3D_films_(2005_onwards)	17
2	Прогнозы объема рынка AR- и VR-шлемов во всем мире с 2022 по 2032. Данные получены с сайта https://www.precedenceresearch.com/	17
3	Схематичный пример стереопары с цветовыми различиями между ракурсами.	20
4	Пример локальных цветовых расхождений между ракурсами стереофильма. В кадре присутствует засвет на стене здания в правом ракурсе, в то время как цвет героя сильно не меняется при смене ракурсов. Кадр из фильма “Шаг вперед 4”.	20
5	Увеличенный фрагмент стереопары, в котором различия по резкости максимальны. Кадр из фильма “Джек – покоритель великанов”.	21
6	Схематичные примеры стереопар с различиями по резкости между ракурсами.	21
7	Пример левого ракурса со сгенерированными искажениями цвета и размытия и приведенного к нему правого ракурса. На данной стереопаре присутствуют локальные цветовые искажения, а также глобальное искажение резкости. Кадр из фильма “Первый мститель”.	31
8	Общая схема предложенного метода поиска различий по цвету и резкости между ракурсами стереоскопического видео.	32
9	Пример карты диспаратности и соответствующей ей карты доверия, построенной для правого ракурса кадра стереофильма “Мстители”.	33

10	Общая архитектура сети типа U-net.	36
11	Общая архитектура сети типа GridNet.	39
12	Вычислительный граф двумерного сепарабельного фильтра FGS. . .	42
13	Вычислительный граф метода обратного распространения ошибки для двумерного сепарабельного фильтра FGS.	43
14	Сглаженные графики оптимизируемого во время обучения функционала на обучающей (слева) и валидационной (справа) выборках. Оранжевая и красная линии соответствуют методу на основе архитектуры U-net, синяя и голубая — методу на основе архитектуры GridNet.	47
15	Схема работы системы объективной оценки качества стереоскопических видео VQMT3D.	51
16	Распределение отобранных видео по числу просмотров.	53
17	Распределение отобранных видео по их длительности.	53
18	Распределение отобранных видео по соотношению $\frac{likes}{likes+dislikes}$	53
19	Распределение отобранных видео по соотношению $likes - dislikes$. .	53
20	Предобработка кадров видео в формате V180. Красным квадратом выделена фронтальная грань кубической проекции, непосредственно используемая при анализе качества VR180-видео.	55
21	Результаты анализа цветовых искажений в VR180-видео.	56
22	Результаты анализа искажений резкости в VR180-видео.	57
23	Схематичная иллюстрация типов геометрических искажений, возникаемых при съемке 3D-видео.	60
24	Пример вертикального сдвига между ракурсами стереофильма. Кадр из фильма “Убойное Рождество Гарольда и Кумара”.	60
25	Распределения значений вычисленных геометрических искажений в проанализированных 3D-фильмах с соответствующими стандартными отклонениями.	67

26	Общая схема предложенного метода оценки геометрических искажений между ракурсами стереоскопического видео.	69
27	Предложенная архитектура нейронной сети для предсказания геометрических искажений.	69
28	Конфигурация блоков нейронной сети для предсказания геометрических искажений.	71
29	График изменения коэффициента скорости обучения в зависимости от номера итерации.	74
30	Сглаженные графики оптимизируемого во время обучения функционала на обучающей (слева) и валидационной (справа) выборках.	74
31	Пример исправления вертикально сдвига в видео “Explore Norway’s Hurtigruten ferry in VR180”. Артефакт наиболее заметен на окне и рамках картины.	79
32	Пример исправления поворота одного ракурса относительно другого в видео “Charlie Puth In London (VR180 Experience)”. Артефакт наиболее заметен на центральной красной линии.	79
33	Результаты анализа вертикального сдвига в VR180-видео.	81
34	Результаты анализа относительного масштабирования в VR180-видео.	81
35	Результаты анализа относительного поворота в VR180-видео.	82
36	Схематичный пример перепутанных ракурсов в стереовидео.	85
37	Примеры кадров с перепутанными ракурсами, найденных в полнометражных стереофильмах, для различных методов производства	86
38	Общая схема предложенного метода поиска перепутанных ракурсов в стереовидео.	96
39	Примеры вычисленных стандартных отклонений диспаратности для кадра с переменной диспаратностью и для кадра с постоянной диспаратностью.	97

40	Примеры вычисленных средних значений яркости для кадров с высокой и низкой яркостью. Кадры из фильма “Мстители”.	98
41	Левый ракурс с вычисленной для него картой областей открытия/закрытия по движению, изображенной красным цветом. Большая часть областей открытия/закрытия обрамляет движущиеся объекты, находящиеся на переднем плане. Кадр из фильма “Гарри Поттер и Дары Смерти: Часть I”.	100
42	Пример сцены (левый ракурс и карта диспаратности), непригодной для анализа перепутанных ракурсов, так как изменение порядка ракурсов в данной сцене не изменит ее восприятие зрителями. Кадр из фильма “Тор 2: Царство тьмы”.	101
43	Пример смены сцены и соответствующих кадрам значений $\text{Dif}^{Y_t^L, Y_{t+1}^L}$ и $\text{Dis}^{Y_t^L}$. Кадры из фильма “Новый Человек-паук”.	103
44	Сглаженные графики оптимизируемого во время обучения функционала на обучающей (слева) и валидационной (справа) выборках.	104
45	Общие результаты анализа диапазонов параллакса в VR180-видео относительно даты их публикации. Для удобства просмотра также представлены увеличенные фрагменты для разных временных отрезков.	111
46	Диапазон параллаксов для каждого кадра видео в формате VR180, отсортированные по количеству кадров с низким диапазоном параллакса.	112
47	Визуализация кадра #12745 с большими положительными значениями диспаратности из видео youtu.be/AVTcC513iLQ	113
48	Визуализация кадра #3864 с большими положительными значениями диспаратности из видео youtu.be/KAoQwCVQFDU	113
49	Визуализация кадра #2346 с большими положительными значениями диспаратности из видео youtu.be/hhaIcogfH1k	113

50	Визуализация кадра #11084 с большими положительными значениями диспаратности из видео youtu.be/L1TIZ8w-41M	113
51	Визуализация кадра #7574 с большими отрицательными значениями диспаратности из видео youtu.be/vwKjoXx2Ic	114
52	Визуализация кадра #498 с большими отрицательными значениями диспаратности из видео youtu.be/8MY42p1_yh4	114
53	Визуализация кадра #9431 с большими отрицательными значениями диспаратности из видео youtu.be/AVTcC513iLQ	114
54	Визуализация кадра #6353 с большими отрицательными значениями диспаратности из видео youtu.be/CHh3pgJXhwM	114
55	Визуализация кадра #1382 с цветовыми искажениями из видео youtu.be/0dhvLX4aoSk . Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.	115
56	Визуализация кадра #1267 с цветовыми искажениями из видео youtu.be/2vwKjoXx2Ic . Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.	116
57	Визуализация кадра #2436 с цветовыми искажениями из видео youtu.be/9t2SCT68NJ8 . Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.	116
58	Визуализация кадра #2259 с цветовыми искажениями из видео youtu.be/HBasBaz22dE . Слева представлена шахматная визуализация, а справа – цветовая разница между ракурсами.	117
59	Визуализация кадра #860 с искажениями резкости из видео youtu.be/2vwKjoXx2Ic	118
60	Визуализация кадра #3641 с искажениями резкости из видео youtu.be/8MY42p1_yh4	118
61	Визуализация кадра #6152 с искажениями резкости из видео youtu.be/AVTcC513iLQ	118
62	Визуализация кадра #3129 с искажениями резкости из видео youtu.be/KAoQwCVQFDU	118

63	Визуализация кадра #814 с вертикальным сдвигом из видео youtu.be/0dhvLX4aoSk.	119
64	Визуализация кадра #2749 с вертикальным сдвигом из видео youtu.be/ALMiuV4FsgU.	119
65	Визуализация кадра #630 с вертикальным сдвигом из видео youtu.be/AVTcC513iLQ.	120
66	Визуализация кадра #9361 с вертикальным сдвигом из видео youtu.be/l-RuiC2iyug.	120
67	Визуализация кадра #1824 с масштабированием из видео youtu.be/CJd4NKVmp7o.	121
68	Визуализация кадра #1630 с масштабированием из видео youtu.be/hhaIcogfH1k.	121
69	Визуализация кадра #5311 с масштабированием из видео youtu.be/GCOBkMd2Iw.	122
70	Визуализация кадра #768 с масштабированием из видео youtu.be/xfEX2de3N4U.	122
71	Визуализация кадра #7022 с поворотом из видео youtu.be/2vwKjoXx2Ic.	123
72	Визуализация кадра #4302 с поворотом из видео youtu.be/AVTcC513iLQ.	123
73	Визуализация кадра #2690 с поворотом из видео youtu.be/hhaIcogfH1k.	124
74	Визуализация кадра #8193 с поворотом из видео youtu.be/tGCOBkMd2Iw.	124
75	Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения компьютерной графики. Кадр из видео “The Aces”.	125
76	Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения титров. Кадр из видео “All Time Low - Dressing Room Tour at Lollapalooza”.	126

- 77 Пример сцены с перепутанными ракурсами, возникшими из-за
неправильного наложения элементов компьютерной графики
титров. Кадр из видео “Lift The Car Off The Baby | Yosemite Higher
Spire Free”. 126
- 78 Пример сцены с полностью перепутанными ракурсами. Кадр из
видео “Hayley Kiyoko - VR180”. 126
- 79 Пример сцены с перепутанными ракурсами, возникшими из-за
неправильного наложения титров. Кадр из видео “The Man - VR180”. 127
- 80 Пример сцены с перепутанными ракурсами, возникшими из-за
неправильного использования спецэффектов. Кадр из видео “Google
Pixel 2 — Up-close in VR180”. 127
- 81 Пример сцены с перепутанными ракурсами, возникшими из-за
неправильного наложения элементов компьютерной графики. Кадр
из видео “Best Gaming Console (VR180)”. 127

Список таблиц

1	Конфигурация блоков сверточной нейронной сети типа U-net.	37
2	Конфигурация блоков сверточной нейронной сети типа GridNet.	38
3	Результаты тестирования методов оценки различий по цвету и резкости между ракурсами стереовидео на искусственном наборе данных Sintel.	48
4	Результаты тестирования методов оценки различий по цвету и резкости между ракурсами VR180-видео.	49
5	Результаты сравнения базовых компонент оптимизируемого функционала. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению.	75
6	Результаты сравнения регуляризационных компонент оптимизируемого функционала. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению.	75
7	Результаты тестирования рассматриваемых методов. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению.	77
8	Результаты тестирования алгоритмов поиска перепутанных ракурсов в стереовидео.	105