

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

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

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2025

Диссертация подготовлена на кафедре интеллектуальных информационных технологий факультета вычислительной математики и кибернетики
МГУ имени М.В. Ломоносова.

Научный руководитель: **Ватолин Дмитрий Сергеевич**
кандидат физико-математических наук

Официальные оппоненты: **Ершов Егор Иванович**,
кандидат физико-математических наук, доцент,
Институт проблем передачи информации
им. А.А. Харкевича РАН, старший научный
сотрудник

Турлапов Вадим Евгеньевич,
доктор технических наук, доцент,
Институт информационных технологий
математики и механики Нижегородского
государственного университета
им. Н.И. Лобачевского, кафедра
высокопроизводительных вычислений и системного
программирования, профессор

Дьяконов Александр Геннадьевич,
доктор физико-математических наук, доцент,
профессор РАН,
ООО «ВК», руководитель отдела

Защита состоится «05» марта 2026 г. в 13 часов 00 минут на заседании диссертационного совета МГУ.012.2 Московского государственного университета имени М.В. Ломоносова по адресу: 119991, Москва, ГСП-1, Ленинские горы, МГУ, д. 1 строение 52, факультет Вычислительной математики и кибернетики, аудитория №238.

E-mail: ds012.2@cs.msu.ru.

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский проспект, д.27) и на портале:
<https://dissovet.msu.ru/dissertation/3736>.

Автореферат разослан «____» _____ 20____ г.

Ученый секретарь
диссертационного совета МГУ.012.2,
кандидат физико-математических наук



А.С. Антонов

Общая характеристика работы

Актуальность темы. В настоящее время стереоскопическое видео широко распространено и знакомо практически каждому. Однако всё чаще зрители предпочитают 2D-версии фильмов вместо 3D, поскольку просмотр стереоскопических фильмов нередко сопровождается дискомфортом: усталостью, напряжением и болью в глазах, а также головной болью. Причинами такого дискомфорта могут быть особенности зрительной системы, условия показа и техническое качество производимого контента.

Хотя зрители со временем могут адаптироваться к стереоскопическому восприятию и выбирать качественное оборудование для просмотра, проблемы, связанные с техническим качеством самого контента, остаются значительными. Именно качество производимого стереоскопического контента выходит на первый план, что делает актуальным создание инструментов для его контроля.

Производство 3D-фильмов сопряжено с новыми проблемами, не характерными для обычных фильмов. Различные методы создания стереоскопических фильмов — съёмка, конвертация из 2D в 3D и использование компьютерной графики — могут приводить к появлению разнообразных искажений или артефактов, таких как геометрические несоответствия, различия в цвете, яркости и резкости. Даже при совершенствовании технологий показа, недостатки самого стереоскопического контента могут свести на нет достигнутые технологические улучшения.

Дополнительно, новые форматы видео, такие как 360-градусные видео и VR180, обеспечивают ещё более глубокое погружение, но также подвержены стереоскопическим артефактам. Особенно это актуально для VR180-видео, производство которого зачастую осуществляется любителями без должного контроля качества, что приводит к появлению множества искажений в итоговых материалах.

Таким образом, учитывая дискомфорт зрителей, вызванный техническими недостатками стереоскопического контента, и растущее распространение новых видеоформатов, разработка методов обнаружения искажений в стереоскопических видео является важной и актуальной задачей. Создание эффективных инструментов для контроля качества стереоконтента позволит улучшить зрительский опыт и повысить интерес к просмотру 3D-фильмов.

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 раздельными 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.

Основное содержание работы

Во **введении** обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана их практическая значимость, представлены выносимые на защиту научные положения.

Первая глава посвящена разработке нейросетевого метода одновременной оценки искажений цвета и резкости. Цветовые искажения ракурсов и искажения ракурсов по резкости – одни из самых распространенных типов искажений стереоскопического видео, полученных в ходе нативной



Рисунок 1 — Схематичные примеры стереопар с искажениями цвета и резкости.

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. 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 1 (а) представлен схематичный пример цветовых искажений в стереовидео. Под искажениями ракурсов по резкости понимается сильное несоответствие в детализации и/или размытии одного и того же объекта кадра в левом и правом ракурсе или всего кадра, что также наиболее заметно при переключении между ракурсами. В самом простом случае либо левый, либо правый ракурс резче другого во всех пикселях изображения, как показано на рисунке 1 (б). Более сложной является ситуация, при которой разные объекты находятся в фокусе в разных ракурсах стереовидео, как показано на рисунке 1 (в), 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. Подробнее описание данных искажений приведено в [разделе 1.1](#).

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 данных артефактов может возникать большое количество ложноположительных срабатываний. Этой проблемой обладают предыдущие методы поиска данных искажений, которые ранее использовались для анализа полнометражных стереоскопических фильмов. [Раздел 1.2](#) содержит обзор существующих методов оценки искажений цвета и резкости как для моноскопических видео, так и для стереоскопических.

В разделе 1.3 описана предложенная общая модель искажений цвета и резкости. Модель цветовых искажений строится на линейной модели, а модель искажений резкости – на основе фильтров Гаусса. Пусть I_{gt}^L и I_{gt}^R – соответственно левый и правый ракурсы стереопары, не содержащей различия по цвету и резкости между ними. Для моделирования рассматриваемых стереоскопических артефактов кадры без искажений модифицируются следующим образом:

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

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

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. Даный 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:

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

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

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.

На основе данных моделей был сгенерирован набор данных путем преобразования 9488 различных стереопар без искажений в разрешении 960×540 из 16 стереофильмов. Итоговый набор данных был разделен на

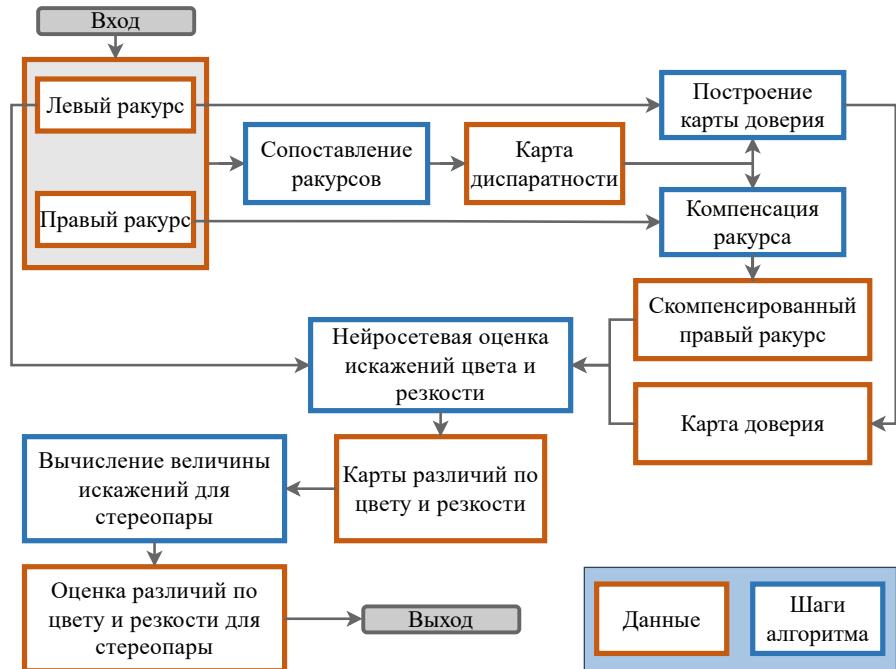


Рисунок 2 — Общая схема предложенного метода поиска различий по цвету и резкости между ракурсами стереоскопического видео.

обучающую и валидационную части в соотношении 95% и 5% соответственно.

Раздел 1.4 содержит описание предлагаемого 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.

Каждый кадр стереоскопического видео обрабатывается независимо друг от друга. Общая схема метода представлена на рисунке 2.

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

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

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

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. Используемый метод оценки карты диспаратности описывается в [разделе 1.4.1](#). Для каждой карты диспаратности также вычисляется карта доверия на основе двух показателей: меры достоверности сопоставления и блочной дисперсии цветовых значений ракурса.

В [разделе 1.4.2](#) представлены использованные в методе нейросетевые архитектуры для предсказания карт различий по цвету и резкости. В рамках работы рассмотрено два типа нейросетевых архитектур: кодировщик-декодировщик по типу U-net (рисунок 3) и кодировщик-декодировщик по типу GridNet (рисунок 4). 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) – фильтр, использующийся для распространения данных предсказанных карт искажений по маске карты доверия с учетом границ исходного изображения. В работе показано, что данный фильтр можно использовать в качестве нейросетевого блока, а также для него возможна эффективная параллельная реализация:

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

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.

В [разделе 1.5](#) описывается процесс обучения нейросетевой части метода. В нем вводится предложенный оптимизируемый функционал:

$$L(\hat{c}, c, \hat{d}, d, \theta) = L_c(\hat{c}, c) + L_d(\hat{d}, d) + L_2(\theta). \quad (7)$$

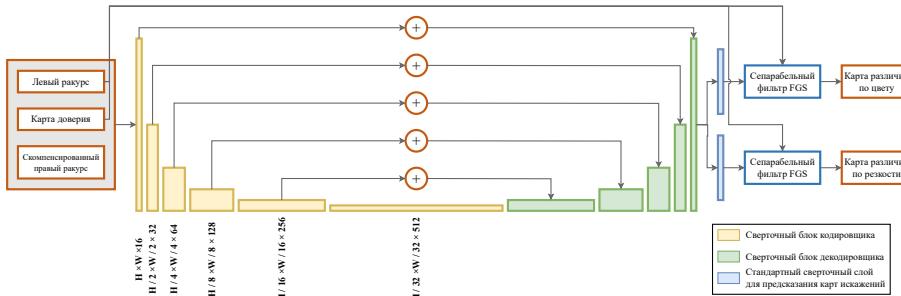


Рисунок 3 — Общая архитектура сети типа U-net.

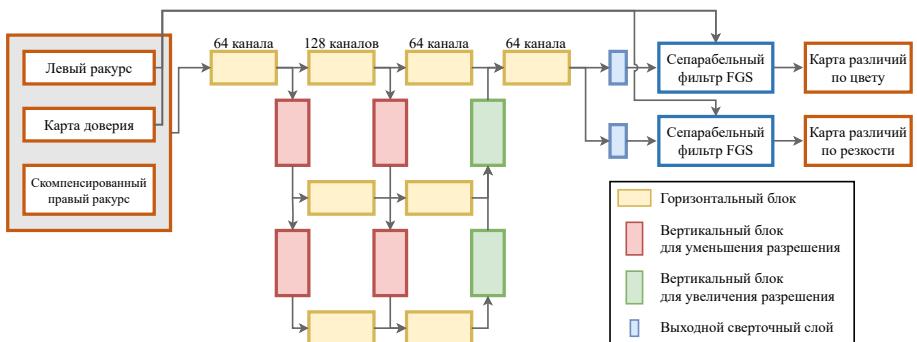


Рисунок 4 — Общая архитектура сети типа GridNet.

$$L_c(\hat{c}, c) = \frac{\sum_{i=1}^n \text{conf}_i \times \left((\hat{c}_i^Y - c_i^Y)^2 + (\hat{c}_i^U - c_i^U)^2 + (\hat{c}_i^V - c_i^V)^2 \right)}{3 \sum_{i=1}^n \text{conf}_i}, \quad (8)$$

$$L_d(\hat{d}, d) = \frac{\sum_{i=1}^n \text{conf}_i \times (\hat{d}_i - d_i)^2}{\sum_{i=1}^n \text{conf}_i}, \quad (9)$$

$$L_2(\theta) = \lambda \sum_{i=1}^k \theta_i^2, \quad (10)$$

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, θ – веса обучаемой нейросети, $\lambda = 10^{-2}$ – параметр регуляризации, k – общее количество весов в сети.

В разделе 1.6 преимущества предложенного алгоритма перед аналогами демонстрируются путем вычисления корреляций Пирсона и Спирмена

на с эталонными значениями искажений на отдельной тестовой выборке, подготовленной на основе набора данных Sintel. Предложенный метод на основе архитектуры GridNet в целом показал более высокие корреляции как по сравнению с аналогами, так и по сравнению с методом на основе U-net. Также была проведена экспертная оценка предложенного метода на наборе из 100 видео в формате VR180 с методами, применявшимися для анализа полнометражных стереоскопических фильмов, показавшая значительное сокращение числа ложноположительных параметров. Раздел 1.7 посвящен программной реализации предложенного метода.

В разделе 1.8 представлены результаты анализа 1000 видео в формате VR180, собранных с платформы YouTube, на предмет наличия искажений цвета и резкости между стереоскопическими ракурсами. В рамках работы рассматривались зависимости средних значений искажений относительно количества просмотров на YouTube, даты публикации и длительности каждого видео. Однако, ни один из рассматриваемых стереоскопических артефактов не демонстрирует какой-либо существенной тенденции по отношению к любой статистике видео, и вероятность встретить рассматриваемые артефакты в видео от них также не зависит.

Вторая глава посвящена разработке метода оценки геометрических искажений в стереоскопических видео. 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 5, слева), поворот (рисунок 5, посередине) и масштабирование (рисунок 5, справа) одного ракурса относительно другого. Рассматриваемые геометрические искажения в первую очередь возникают из-за неправильной калибровки камер. При просмотре стереоскопических сцен с данными артефактами, зритель часто испытывает дискомфорт, вплоть до головных болей, тошноты и головокружения, аналогично искажениям цвета и резкости. Подробнее геометрические искажения описаны в разделе 2.1. В обзоре существующих методов, приведенного в разделе 2.2, показано, что существующие методы оценки геометрических искажений либо требуют калибровочные параметры камер для точной оценки искажений, либо обладают нестабильными результатами предсказаний. Поэтому в данной работе предлагается использовать нейросетевой регрессор для непосредственной оценки параметров геометрических искажений.

В разделе 2.3 приводится используемая модель геометрических искажений. Для моделирования данных искажений достаточно применить аффинное преобразование к одному из ракурсов стереоскопического видео. Пусть I_{gt}^L и I_{gt}^R – левый и правый ракурсы стереопары, не содержащей геометрических искажений, $p = [x \ y \ 1]^T$ и $p' = [x' \ y' \ 1]^T$ – однородные координаты двух точек до и после применения преобразования соответственно. Тогда геометрические несоответствия между ракурсами стереовидео мож-

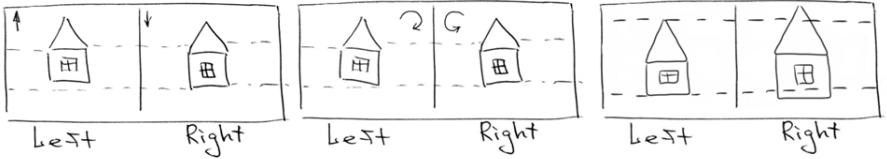


Рисунок 5 — Схематичная иллюстрация типов геометрических искажений, возникаемых при съемке 3D-видео.

но промоделировать следующим аффинным преобразованием:

$$p' = A \times p, \quad (11)$$

$$A = \begin{bmatrix} (1+k) \cos(\alpha) & -(1+k) \sin(\alpha) & 0 \\ (1+k) \sin(\alpha) & (1+k) \cos(\alpha) & t \\ 0 & 0 & 1 \end{bmatrix}, \quad (12)$$

где α – угол поворота, k – коэффициент масштабирования, t – вертикальный сдвиг. Примеры с геометрическими искажениями получаются путем интерполяции одного из ракурсов по преобразованной координатной сетке.

На основе данной модели был сгенерирован набор данных путем преобразования 22800 различных стереопар без геометрических искажений из 29 стереофильмов. Итоговый набор данных был разделен на 3 части: обучающую (15500 стереопар), валидационную (3600 стереопар) и тестовую (3700 стереопар). Каждая стереопара из обучающего набора данных иска-жалась 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.

Предложенный алгоритм для оценки геометрических искажений описан в разделе 2.4. Данный алгоритм состоит 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.
2. Оценка параметров геометрических искажений с помощью нейронной сети. Для оценки параметров геометрических искажений используется нейросетевая архитектура, аналогичная ResNet-18. В качестве входных данных используется нормированная карта диспаратности, а также соответствующая карта доверия. На выходе сеть предсказывает вектор $\theta \in \mathbb{R}^3$, содержащий параметры предсказанных геометрических искажений.

Общая схема метода представлена на рисунке 6.

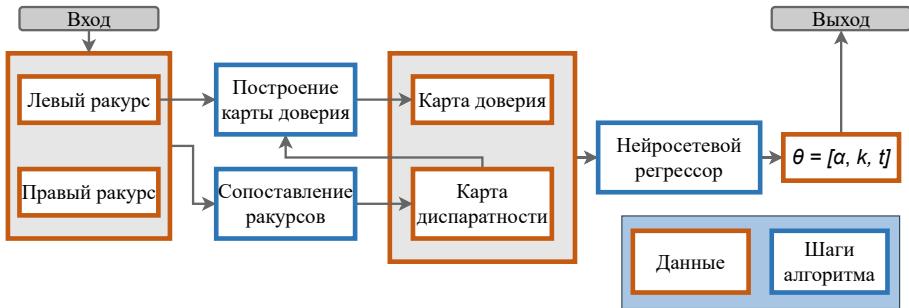


Рисунок 6 — Общая схема предложенного метода оценки геометрических искажений между ракурсами стереоскопического видео.

Процесс обучения нейросетевого регрессора описан в [разделе 2.5](#). В нем представлен предложенный в работе оптимизируемый функционал:

$$L(\theta, \theta_{gt}, I^R, I_{gt}^R, \theta_b) = L_{SE}(\theta, \theta_{gt}) + L_{Grid}(\theta, \theta_{gt}) + L_{Warp}(\theta, I^R, I_{gt}^R) + L_{Siam}(\theta, \theta_b), \quad (13)$$

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 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. Вторая компонента, L_{Grid} , вычисляет функцию потерь между двумя сетками, преобразованными с помощью аффинных преобразований, построенных по вычисленным и эталонным значениям геометрических искажений. 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. Наконец, последняя регуляризационная компонента, L_{Siam} , оценивает консистентность между нейросетевыми предсказаниями на основе входных данных как для левого, так и для правого ракурсов.

В [разделе 2.6](#) приведена экспериментальная оценка предложенного метода. Вклад каждой компоненты оптимизируемого функционала в улуч-

шение точности работы нейросетевого регрессора продемонстрирован на валидационной выборке в разделе 2.6.1, а в разделе 2.6.2 представлены результаты сравнения предложенного метода с аналогами на тестовой выборке. Предложенному методу удалось сократить ошибку вычислений на 14.43% при оценке угла поворота, а при оценке масштабирования и вертикального сдвига – более чем на 99%.

Также предложенный метод для оценки геометрических искажений между ракурсами стереовидео можно использовать и для их исправления:

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

Процесс данного преобразования подробно показан в разделе 2.7. Программная реализация предложенных методов описана в разделе 2.8.

Результаты анализа 1000 видео в формате VR180 на предмет наличия в них геометрических искажений описаны в разделе 2.9. Аналогично искажениям цвета и резкости геометрические искажения не демонстрируют какие-либо существенные тенденции по отношению к рассматриваемым статистикам, и для них также справедливы все те же выводы. При этом значительное количество проанализированных видео в формате VR180 демонстрирует наличие по крайней мере одного стереоскопического артефакта из рассмотренной группы искажений, 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.

Третья глава посвящена разработке метода поиска перепутанных ракурсов в стереоскопических видео. Перепутанный порядок ракурсов – артефакт, при котором в сцене стереовидео на месте левого ракурса оказывается правый и наоборот (рисунок 7). Данное искажение встречается достаточно редко в стереофильмах, но наличие даже одной сцены с перепутанными ракурсами может вызвать серьезный дискомфорт у зрителей при ее просмотре. При этом этот артефакт встречается в стереофильмах вне зависимости от способа их создания. 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.

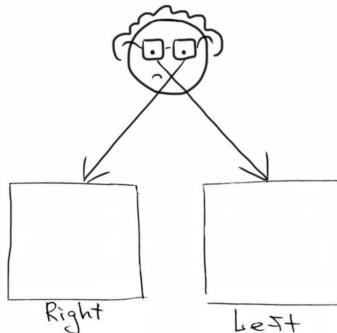


Рисунок 7 — Схематичный пример перепутанных ракурсов в стереовидео.

Подробнее постановка данной задачи приведена в разделе 3.1.

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. В разделе 3.2 приводится классификация существующих методов по используемым признакам оценки карты диспаратности и описывается каждый из классов методов: использование методов упорядочивания глубины; проверка предположений о распределении диспаратности; анализ областей открытия в стереопаре; анализ областей открытия/закрытия по движению. Из обзора стоит заметить, что нейросетевые методы получают достаточно точные результаты при предсказании карт диспаратности, что демонстрирует возможность обучения сверточных нейронных сетей монокулярным признакам упорядочивания глубины. Таким образом, применение сверточных нейронных сетей является многообещающим шагом для улучшения методов поиска перепутанных ракурсов.

В разделе 3.3 представлен предложенный метод для поиска перепутанных ракурсов в стереоскопических видео. 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.
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. Для предсказания вероятности наличия перепутанных ракурсов в кадре была также использована архитектура нейронной сети, аналогичная ResNet-18, как и в методе по оценке геометрических искажений.

На рисунке 8 представлена общая схема работы предложенного метода. Формально, результатом работы предложенного алгоритма для сцены является число:

$$\bar{x} = \frac{1}{n_A} \sum_{j \in A} x_j, \quad (14)$$

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.

Так как задача поиска перепутанных ракурсов в стереовидео является задачей бинарной классификации, для обучения нейронной сети для определения порядка ракурсов достаточно использовать бинарную кросс-энтропию в качестве оптимизируемой функции. При этом дополнительно для предотвращения переобучения в оптимизируемой функции используется L_2 -регуляризация. Подробнее процесс обучения нейронной сети описан в разделе 3.4. Обучающий набор данных был подготовлен на основе кадров из полнометражных стереоскопических фильмов, ранее использованных для обучения методов оценки искажений цвета и резкости. При этом сам порядок ракурсов выбирался во время обучения случайно.

Тестирование предложенного метода и сравнение с аналогами проводилось на выборке из 900 сцен длиной в 30 кадров. Во время тестирования для всех оцениваемых алгоритмов вычислялись следующие показатели:

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

Результаты тестирования, описанные в разделе 3.5, показали, 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

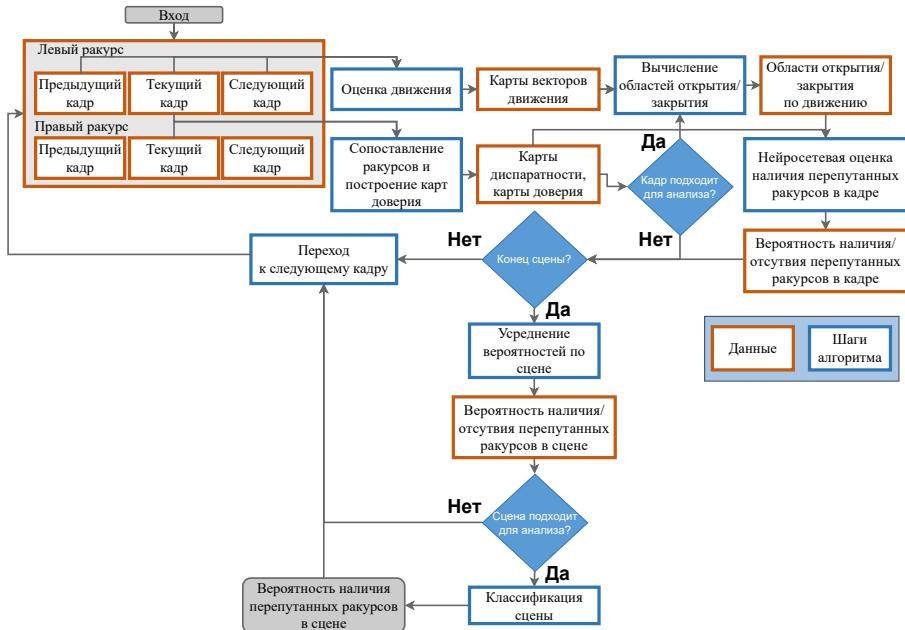


Рисунок 8 — Общая схема предложенного метода поиска перепутанных ракурсов в стереовидео.

применявшимися на практике методами. Программная реализация предложенного метода описана в разделе 3.6.

Также с помощью предложенного метода поиска перепутанных ракурсов было проанализировано 50 наиболее просматриваемых видео в формате VR180. 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 даже одна сцена с перепутанными ракурсами может вызвать серьезный дискомфорт у зрителя.

В Приложениях представлены дополнительные результаты анализа видео в формате VR180. Помимо рассматриваемых в данной работе стереоскопических артефактов для каждого видео измерялся диапазон параллаксов. Среди них были найдены видео со 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”. Поэтому значения положительного параллакса в VR180-видео 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. Также в **Приложении** представлены примеры, обнаруженные с помощью предложенных методов оценки искажений.

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

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. Предложенная 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 видео-последовательностей”. Обучение нейросетевых моделей производилось с использованием высокопроизводительного кластера IBM Polus факультета ВМК МГУ: <https://hpc.cmc.msu.ru/polus>.

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.