2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

doi 10.26089/NumMet.v26r323 УДК 04.932.4

Программная система для автоматической оценки стереоскопических искажений видео в формате VR180

С. В. Лаврушкин

Московский государственный университет имени М. В. Ломоносова,
Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем,
Москва, Российская Федерация
Институт системного программирования имени В. П. Иванникова РАН,
Исследовательский центр доверенного искусственного интеллекта,
Москва, Российская Федерация
ORCID: 0009-0003-8544-2923, e-mail: sergey.lavrushkin@graphics.cs.msu.ru

Аннотация: В статье представлена комплексная программная система для автоматической оценки стереоскопических искажений видео в формате VR180. Предложенный подход учитывает наиболее распространенные типы артефактов: несоответствия по цвету, различия по резкости, геометрические искажения (вертикальный сдвиг, поворот, масштабирование), а также перепутанный порядок ракурсов. Для каждого из перечисленных искажений разработаны специализированные алгоритмы, основанные на вычислении карт диспаратности, векторов движения, карт доверия к ним и нейросетевых методов регрессии или классификации. Предложенные решения успешно прошли проверку на нескольких тестовых наборах данных и продемонстрировали высокую точность в обнаружении искажений разных типов для VR180-видео. Система может быть интегрирована в стандартные пайплайны постобработки и обеспечивает автоматизированную генерацию подробных отчетов, что может позволить создателям стереоконтента оперативно выявлять и устранять причины стереоскопических искажений до вывода продукции на массового зрителя.

Ключевые слова: стереоскопические искажения, стереоскопические видео, VR180, цветовые искажения, искажения резкости, геометрические искажения, перепутанные ракурсы, глубокое обучение.

Для цитирования: Лаврушкин С.В. Программная система для автоматической оценки стереоскопических искажений видео в формате VR180 // Вычислительные методы и программирование. 2025. **26**, № 3. 340–365. doi 10.26089/NumMet.v26r323.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

A software system for automated assessment of stereoscopic distortions in VR180 videos

Sergey V. Lavrushkin
Lomonosov Moscow State University, Institute for Artificial Intelligence, Moscow, Russia
Ivannikov Institute for System Programming of RAS, Research Center for Trusted Artificial Intelligence, Moscow, Russia
ORCID: 0009-0003-8544-2923, e-mail: sergey.lavrushkin@graphics.cs.msu.ru

Abstract: The paper presents a comprehensive software system for automatically evaluating stereoscopic distortions in VR180 video. The proposed approach takes into account the most common types of artifacts: color mismatches, differences in sharpness, geometric distortions (vertical shift, rotation, scaling), and channel mismatches. Specialized algorithms have been developed for each type of distortion, based on disparity maps, motion vectors and their confidence maps, as well as neural network regression or classification methods. The proposed solutions have been successfully tested on several datasets, demonstrating high accuracy in detecting various types of distortions in VR180 video. The system can be integrated into standard post-processing pipelines and provides automated generation of detailed reports, allowing stereoscopic content creators to quickly identify and eliminate stereoscopic distortions before releasing their products to a wide audience.

Keywords: stereoscopic distortions, stereoscopic video, VR180, color mismatch, sharpness mismatch, geometric distortions, channel mismatch, deep learning.

For citation: S. V. Lavrushkin, "A software system for automated assessment of stereoscopic distortions in VR180 videos," Numerical Methods and Programming. **26** (3), 340–365 (2025). doi 10.26089/NumMet.v26r323.

1. Введение. Стереоскопические видео в наши дни стали неотъемлемой частью массовой культуры, однако многие зрители продолжают испытывать дискомфорт (усталость, напряжение, боль в глазах, вплоть до головной боли [1]) при просмотре 3D-фильмов и нередко предпочитают 2D-версии. Исследования [2–6] указывают, что такой дискомфорт может возникать и усиливаться не только из-за особенностей зрительной системы и условиями показа, но прежде всего из-за технических ошибок при создании стереофильмов, приводящих к возникновению искажений и различий между стереоскопическими ракурсами. По мере привыкания зрителя к 3D все более критичной становится именно техническая сторона: неточные настройки камер, несовпадение ракурсов и другие артефакты способны нивелировать усилия по улучшению оборудования и привести к дискомфорту, отталкивая потенциальную аудиторию.

Таким образом, в стереокинематографе выходит на первый план задача обеспечения качества производимого контента. Ответственность за устранение артефактов ложится на создателей 3D-фильмов, поскольку выбор зрителя в пользу более качественного оборудования не решит проблемы изначально некорректной стереосъемки. 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. При таком способе создания стереовидео достаточно часто появляются геометрические несоответствия между ракурсами, а также несоответствия по цвету и резкости [7]. Данные проблемы появляются в случае, когда используемые для съемки камеры по-разному настроены и/или какая-то компонента одной из камер вышла из строя. Также возможно появление дополнительных искажений, внесенных при постобработке уже отснятого материала. Примером такого артефакта являются перепутанные ракурсы, где левый и правый ракурсы меняются местами, либо неправильно накладывается компьютерная графика и спецэффекты на изначально правильные ракурсы, что приводит к конфликтам восприятия.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

Отдельного внимания заслуживает стереоскопический формат VR180, ориентированный на использование в шлемах виртуальной реальности. Вместо использования специального рига из нескольких камер для записи традиционного 360° видео, VR180 снимается двумя камерами с объективами типа "рыбий глаз", аналогично обычной стереоскопической съемке. Хотя такой подход охватывает только полусферу, это упрощает технику съемки, снижает стоимость оборудования и устраняет проблемы склейки, характерные для сферических видео. При этом сохраняется возможность стереоскопического эффекта: каждый глаз получает собственный ракурс, аналогично классической 3D-съемке. Однако и в VR180 форматах встречаются артефакты, обусловленные неточностью настройки парных камер или искажениями, связанными с объективами, что возникает из-за попытки упростить устройства съемки для широкого использования обычными потребителями, т.е. сделать дешевыми и простыми в использовании. Поскольку этот формат активно осваивают энтузиасты без достаточного опыта, вопросы контроля качества становятся особенно актуальными и требуют разработки методик оценки получаемых стереоматериалов.

Таким образом, несмотря на значительное развитие технологий отображения, именно качество 3D-контента является решающим фактором комфорта зрителя. Создание и внедрение систем контроля качества стереофильмов, включая VR180-видео, позволит повысить достоверность передачи глубины, снизить зрительный дискомфорт и расширить аудиторию, готовую выбирать стереоформат.

Данная работа является обобщением серии представленных на конференциях работ [8–12], посвященной оценке качества стереоскопических видео. В работе впервые представлена итоговая программная система для оценки стереоскопических искажений, включая 3D-формат для сферических видео — VR180, приведены последние версии методов по поиску стереоскопических искажений, а также описаны дополнительные результаты анализа стереоскопического качества видео в формате VR180. Используемые в системе методы позволяют надежно находить большинство присутствующих в анализируемом видео стереоскопических искажений 3D-съемки, что дает возможность исправлять их на этапе производства. Эффективность работы системы продемонстрирована на масштабном исследовании стереоскопического качества 1000 видео в формате VR180.

2. Постановка задачи. В рамках предложенной системы оценки стереоскопических искажений видео в формате VR180 рассматриваются характерные для стереоскопической съемки артефакты: искажения цвета, искажения резкости, геометрические искажения, а также перепутанные ракурсы.

Под цветовыми искажениями ракурсов стереоскопического видео понимается сильное несоответствие яркости и/или цвета одного и того же объекта кадра в левом и правом ракурсах или всего кадра, что наиболее заметно при переключении между ракурсами. При этом часто в стереовидео встречается ситуация, когда лишь часть кадра отличается по цвету между ракурсами. 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 при 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 и правом ракурсах или всего кадра, что также наиболее заметно при переключении между ракурсами. 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. Предложенная система осуществляет 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.

В качестве геометрических искажений в системе анализируются постоянный вертикальный сдвиг, поворот и масштабирование одного ракурса относительно другого. Геометрические искажения в первую очередь возникают из-за неправильной калибровки камер. Небольшие несоответствия в вертикальных положениях камер или небольшие наклоны приводят к появлению данных артефактов.

Перепутанный порядок ракурсов — артефакт, при котором в сцене стереовидео на месте левого ракурса оказывается правый и наоборот. Данное искажение встречается достаточно редко в стереофильмах, но наличие даже одной сцены с перепутанными ракурсами может вызвать серьезный дискомфорт у зрителей при ее просмотре. Эффект перепутанных ракурсов может возникнуть как при простом изменении порядка левого и правого ракурсов, так и при неправильном редактировании готового видеоматериала:

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

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.

Формально задачу оценки данных искажений можно поставить следующим образом. На вход системе подается стереоскопическое видео S — упорядоченная пара видеопоследовательностей (левый и правый ракурсы) $\left(\{I_i^{\mathrm{L}}\}_{i=1}^{n}, \{I_i^{\mathrm{R}}\}_{i=1}^{n}\right)$, $I_i^{\mathrm{L,R}} \in \mathbb{R}^{h\times w\times 3}$, где h — высота, w — ширина кадров. На выходе необходимо получить:

- Для искажений цвета и резкости: оценку различий между левым и правым ракурсами по цвету $\{m_i^{\mathrm{c}}\}_{i=1}^{n}$, $m_i^{\mathrm{c}} \in [0, +\infty)$, и по резкости $\{m_i^{\mathrm{d}}\}_{i=1}^{n}$, $m_i^{\mathrm{d}} \in [0, +\infty)$, для каждого кадра стереовидео.
- Для геометрических искажений: оценку трех геометрических искажений между левым и правым ракурсами $\{\alpha_i, k_i, t_i\}_{i=1}^{n}$, $\alpha_i,\ k_i,\ t_i \in \mathbb{R}$, для каждого кадра стереовидео, где α_i — угол поворота, k_i — коэффициент масштабирования, t_i — вертикальный сдвиг.
- Для перепутанных ракурсов: оценку вероятности наличия перепутанных ракурсов $\{p_j\}_{j=1}^{m}$, $p_j \in [0, 1]$, для каждой сцены Sc_j^S стереовидео S, где под сценой понимается упорядоченная пара непрерывных фрагментов видеопоследовательностей $\left(\{I_i^{\mathrm{L}}\}_{i=n_1}^{n_2}, \{I_i^{\mathrm{R}}\}_{i=n_1}^{n_2}\right)$, $n_2 > n_1$.

3. Обзор методов оценки стереоскопических искажений. В общем случае для оценки стереоскопических искажений необходимо построить карту диспаратности, т.е. сопоставить ракурсы и найти для каждого пикселя одного ракурса соответствующий ему пиксель в другом ракурсе, тем самым оценив смещение, после чего проводится анализ соответствующих друг другу пикселей. Далее могут быть использованы стандартные метрики, например среднеквадратическое отклонение и средняя абсолютная ошибка, для поиска разницы между стереоскопическими изображениями, например для оценки цветовых искажений либо для оценки искажений резкости, но в частотном диапазоне. Дополнительно, построенные карты смещений могут быть использованы для оценки геометрических различий, так как геометрические искажения между ракурсами стереоскопических видео порождают вертикальные смещения пикселей в одном из ракурсов относительно другого, и для проверки порядка объектов в кадре для оценки перепутанности ракурсов. 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. Рассмотрим различные подходы для оценки стереоскопических искажений.

3.1. Искажения цвета и резкости. Существует широкий набор *моноскопических* методов для оценки размытия и резкости, которые условно делятся на два класса. Методы на основе анализа границ используют разреженные карты уровня размытия [13–18], а методы на основе анализа областей анализируют локальные блоки, в том числе в частотном диапазоне [19–24]. В последние годы активно развиваются нейросетевые методы, которые могут напрямую предсказывать уровень размытия или выделять размытые области с использованием различных архитектур сверточных нейронных сетей [25–32]. Однако все эти методы были разработаны для одиночных изображений и не учитывают дополнительные стереоскопические искажения, в частности цветовые, что может приводить к некорректным результатам.

Стереоскопические методы оценки искажений цвета и резкости обычно делятся на глобальные и локальные. Простейшие оценки цветовых искажений базируются на сравнении цветовых гистограмм двух ракурсов [33, 34], однако часто не учитывают различия в них (например, области открытия/закрытия). В задачах оценки стереоскопических искажений резкости анализируют либо суммарные показатели размытости, сопоставляя их с картой диспаратности [35], либо ширину соответствующих границ [36].

Используемый в предложенной системе метод строится на идеях локальных методов оценки искажений цвета и резкости с применением сопоставления ракурсов и дальнейшей оценки стереоскопических артефактов. За основу берутся метод оценки искажений цвета [37] и метод оценки искажений резкости [38]. Метод оценки цветовых искажений вычисляет локальную цветовую разницу между соответствующими пикселями, а метод оценки различий резкости — локальную разницу размытия в частотном диапазоне. Данные методы применялись при анализе полнометражных стереоскопических фильмов. 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 рассматри-

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

ваемых стереоскопических артефактов. Также для более точной оценки силы искажений предлагается использовать нейросетевой подход.

3.2. Геометрические искажения. Для оценки параметров геометрических искажений в научной литературе был предложен ряд методов. В работе [39] осуществляется одновременное вычисление двух параметров геометрических искажений: вертикального сдвига и относительного масштабирования. Для этого осуществляется оценка параметров искажений с помощью метода наименьших квадратов на основе сопоставлений, полученных с помощью метода SIFT [40]. Аналогичный метод используется в работе [41], где вместо поиска особых точек используется блочное иерархическое сопоставление ракурсов и оценивается сразу три геометрических искажения. 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 [42] 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 [34] на основе результатов сопоставления особых точек SIFT. Для оценки вертикального сдвига в рассматриваемом методе анализируется гистограмма вертикальных составляющих векторов сопоставлений; для оценки масштабирования используется 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 [43], а также его модификации, которые в последнее время берут за основу нейросетевые подходы [44–46]. Также нейронные сети начинают использовать и на других этапах методов оценки параметров. Так, в работах [47, 48] предлагается вычислять дополнительные веса с помощью нейронной сети для полученных сопоставлений перед оценкой модели. Также в работе [49] 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 обучение регрессионной нейронной сети для предсказания матрицы произвольного геометрического преобразования. В своей следующей работе [50] авторы дополнительно предлагают оценивать матрицу проективного преобразования.

Используемый в предложенной системе метод базируется на методе [38], который ранее использовался для анализа полнометражных стереоскопических фильмов. Рассматриваемый метод оценивает параметры аффинного преобразования с помощью метода RANSAC на основе результатов блочного сопоставления ракурсов. Но ввиду 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.

3.3. Перепутанные ракурсы. Для определения порядка ракурсов в стереопаре можно использовать различные *методы упорядочивания глубины*. Данные методы используют монокулярные признаки для построения карты диспаратности по одному ракурсу. Например, в работе [51] карта внимания помогает делить кадр на передний и задний планы, и по разнице координат особых точек SIFT определяется порядок ракурсов. Подобные идеи развивают и другие подходы [52–54], однако их точность или вычислительная сложность затрудняют применение при анализе стереовидео. Более перспективны нейросетевые методы для предсказания карт глубины по одному изображению [55–59], так как они показывают высокую обобщающую способность при обучении сразу на нескольких наборах данных эталонных карт глубины.

Также существует ряд специализированных методов для поиска перепутанных ракурсов в стереовидео. Метод [60] основан на проверке предположения о распределении диспаратности. Он полагается на шаблонную карту, где объекты в нижней части кадра ближе к камере, чем в верхней. Метод вычисляет корреляцию между анализируемой картой и шаблоном, но из-за упрощенного предположения об устройстве сцены дает много ошибок при анализе сцен со сложной компоновкой. Метод [61] основан

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

на анализе положения областей открытия путем сравнения центроидов данных областей для левого и правого ракурсов. Такой подход тестировался лишь на ограниченном наборе данных, и его надежность в масштабах стереовидео не была доказана. Методы [8, 62] являются композиционными и объединяют несколько критериев. Критерии включают анализ областей открытия, предположения о распределении таких характеристик, как диспаратность, перспектива (связь глубины с вертикальным положением объектов), а также наличие "выпадающих" объектов (часто находящихся в центре экрана) и движения для уточнения положения объектов переднего плана. Несмотря на более высокую гибкость, эти методы при анализе больших объемов данных все еще дают ощутимое количество ложноположительных результатов, требуя дальнейших улучшений, что предлагается сделать в рамках данной работы за счет использования нейросетевых признаков.

4. Архитектура предложенной системы для оценки стереоскопических искажений. Для эффективной оценки качества стереоскопических фильмов, включая формат VR180, была разработана система, позволяющая переиспользовать общие выходные данные для различных методов оценки стереоскопических искажений. Архитектура данной системы представлена на рис. 1. Основная часть данной системы — хост — реализует задачи по

1) чтению входных ракурсов стереоскопических и VR180-видео,
2) вычислению для считанных ракурсов общих данных, необходимых для работы методов оценки качества,
3) записи результатов работы методов оценки артефактов стереоскопических видео, подключаемых к хосту в виде динамических модулей.

Для чтения входных ракурсов стереоскопических видео в хосте используется библиотека 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. Для работы с видео в формате VR180 в системе осуществляется дополнительная предобработка входных данных: применение

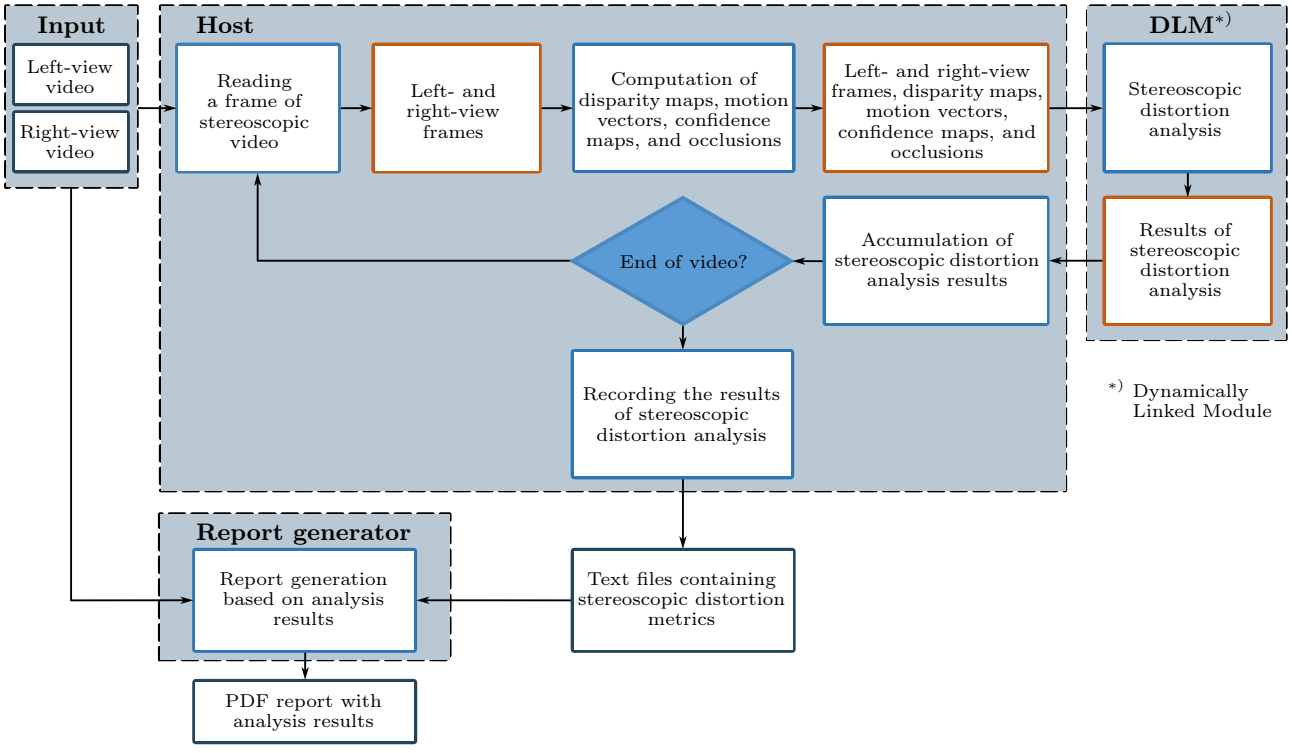


Рис. 1. Схема работы предложенной системы автоматической оценки стереоскопических искажений

Fig. 1. Scheme of the proposed system for automatic assessment of stereoscopic distortions

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

Рис. 2. Предобработка кадров видео в формате V180. Красным квадратом выделена фронтальная грань кубической проекции, непосредственно использующаяся при анализе качества VR180-видео

Fig. 2. Preprocessing of V180 video frames. The red square highlights the frontal face of the cubic projection, which is directly used in analyzing the quality of VR180 video

Рис. 3. Пример карты диспаратности и соответствующей ей карты доверия, построенной для правого ракурса кадра стереофильма “Мстители”

Fig. 3. An example of a disparity map and the corresponding confidence map constructed for the right view of a frame from the stereoscopic film “The Avengers”

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

кубической проекции к исходному видео для генерации центральной части кадра, пригодной для анализа стереоскопических артефактов.

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 рис. 2. Так как поле зрения исходного видео составляет 180°, грани кубической проекции кадра заполнены следующим образом: верхняя, нижняя и боковые — только наполовину, 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.

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 и 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/.

4.1. Оценка карт диспаратности, векторов движения и соответствующих им карт доверия и областей открытия/закрытия. Оценка карт диспаратности и карт векторов движения осуществляется по исходным ракурсам стереоскопического видео в цветовом пространстве RGB с помощью блочного метода сопоставления [63]. Так как при сопоставлении блоков возможны ошибки, для построенных карт строятся карты доверия. Значения карт доверия характеризуют точность вычисленных векторов. При вычислении значений доверия учитываются следующие показатели:

- Мера достоверности сопоставления LRC (left-right consistency) [64]. Поскольку левый и правый ракурсы (или соседние кадры) являются изображениями одной сцены, то значение диспаратности (вектора движения) пикселя в левом ракурсе (одном кадре) должно быть равно по модулю и иметь противоположный знак по сравнению со значением диспаратности (вектора движения) соответствующего ему пикселя в правом ракурсе (другом ракурсе). Более формально, мера достоверности сопоставления LRC вычисляется следующим образом: если пиксель с координатами $x = (x_1,\ x_2)$ одного кадра соответствует пикселю с координатами $x' = (x'_1,\ x'_2) = x + v_x$ другого кадра, то мера достоверности сопоставления LRC для него равна

$$\mathrm{lrc} = \frac{\mathrm{dif}_1^2}{h} + \frac{\mathrm{dif}_2^2}{w}, \tag{1}$$

$$\mathrm{dif} = (\mathrm{dif}_1, \mathrm{dif}_2) = v'_{x'} + v_x, \tag{2}$$

  где v_x — вектор пикселя с координатами x в первом кадре, $v'_{x'}$ — вектор пикселя с координатами x' во втором кадре.

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

$$\mathrm{var} = \mathrm{var}^{\mathrm{R}} + \mathrm{var}^{\mathrm{G}} + \mathrm{var}^{\mathrm{B}}, \tag{3}$$

$$\mathrm{var}^i = \frac{1}{s} \sum_{p \,\in\, \mathrm{block}} p_i^2 - \left(\frac{1}{s} \sum_{p \,\in\, \mathrm{block}} p_i \right)^2, \tag{4}$$

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

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

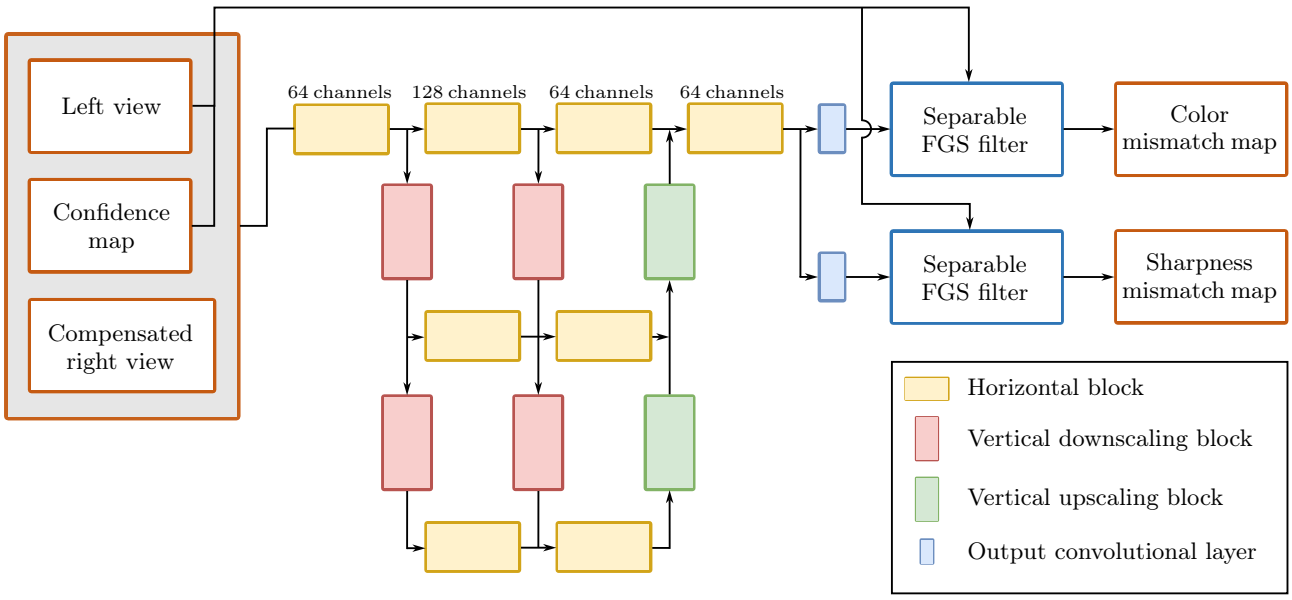


Рис. 4. Архитектура нейронной сети для одновременной оценки искажений цвета и резкости

Fig. 4. Neural network architecture for joint estimation of color and sharpness distortions

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

$$\mathrm{conf}_i = \min\left(1 - \min\left(1, a\,\mathrm{lrc}_i\right),\ \min\left(1, b\,\mathrm{var}_i\right)\right), \tag{5}$$

где $a = 40$, $b = 0.5$, i — индекс пикселя. Итоговое значение доверия лежит в диапазоне $[0, 1]$: $\mathrm{conf}_i \in [0, 1]$. Пример построенной карты диспаратности и соответствующей ей карты доверия представлен на рис. 3.

Для вычисления областей открытия/закрытия также используется мера достоверности сопоставления LRC путем бинаризации построенных карт значений данной меры.

5. Метод одновременной оценки искажений цвета и резкости. Для оценки данных искажений используется сверточная нейронная сеть, которой на вход 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. Итоговая оценка искажений по цвету m^{c} и резкости m^{d} в стереопаре формируется на основе предсказанных карт различий следующим образом:

$$m^{\mathrm{c}} = \frac{\sum\limits_{i=1}^{n} \mathrm{conf}_i \left(\hat{c}_i^{\mathrm{Y}} + \hat{c}_i^{\mathrm{U}} + \hat{c}_i^{\mathrm{V}}\right)}{3 \sum\limits_{i=1}^{n} \mathrm{conf}_i}, \qquad m^{\mathrm{d}} = \frac{\sum\limits_{i=1}^{n} \mathrm{conf}_i\, \hat{d}_i}{\sum\limits_{i=1}^{n} \mathrm{conf}_i}, \tag{6}$$

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.

В качестве архитектуры сети для предсказания карт различий по цвету и резкости была использована сверточная нейронная сеть типа GridNet [65], представляющая собой 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

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

между вышестоящим и нижестоящим потоками. Такая структура нейронной сети обобщает архитектуру кодировщик-декодировщик, в которой 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 карт признаков в одном потоке и двух типов вертикальных блоков для уменьшения и увеличения разрешения. Каждый блок состоит из нескольких сверточных блоков с функциями активации PReLU [66]. Коэффициент дилатации каждого сверточного слоя в сети равен 1, а шаг изменяется от 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. 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) [67] — фильтр, использующийся для распространения данных предсказанных карт искажений по маске карты доверия с учетом границ исходного изображения. Общая архитектура сети представлена на рис. 4.

Для обучения нейросети был использован набор данных, состоящий из 9488 вырезанных из различных 3D-фильмов стереопар в разрешении 960×540, к которым применялись случайные искажения цвета и резкости. 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:

$$L_{\mathrm{c}}\left(\hat{c}, c\right) = \frac{\sum\limits_{i=1}^{n} \mathrm{conf}_i \left(\left(\hat{c}_i^{\mathrm{Y}} - c_i^{\mathrm{Y}}\right)^2 + \left(\hat{c}_i^{\mathrm{U}} - c_i^{\mathrm{U}}\right)^2 + \left(\hat{c}_i^{\mathrm{V}} - c_i^{\mathrm{V}}\right)^2 \right)}{3 \sum\limits_{i=1}^{n} \mathrm{conf}_i}, \tag{7}$$

$$L_{\mathrm{d}}\left(\hat{d}, d\right) = \frac{\sum\limits_{i=1}^{n} \mathrm{conf}_i \left(\hat{d}_i - d_i\right)^2}{\sum\limits_{i=1}^{n} \mathrm{conf}_i}, \tag{8}$$

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:

$$L_2\left(\Theta\right) = \lambda \sum_{i=1}^{k} \Theta_i^2, \tag{9}$$

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

$$L\left(\hat{c}, c, \hat{d}, d, \Theta\right) = L_{\mathrm{c}}\left(\hat{c}, c\right) + L_{\mathrm{d}}\left(\hat{d}, d\right) + L_2\left(\Theta\right). \tag{10}$$

Для инициализации весов сверточных слоев в начале обучения был использован метод инициализации Xavier [68]. В качестве метода оптимизации был выбран алгоритм Adam [69]. Модель обучалась 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.

6. Метод оценки геометрических искажений. Для оценки параметров геометрических искажений используется нейросетевая архитектура, аналогичная ResNet-18 [70]. Вначале входной тензор обрабатывается сверточным слоем размера 7×7 с шагом 2. Далее следуют четыре последовательных остаточных

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

блока, в каждом из которых вычисляется 64, 128, 256, 512 карт признаков соответственно. Количество подблоков в каждом блоке было выбрано равным 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 $\theta = \begin{bmatrix}\alpha & k & t\end{bmatrix}$, $\theta \in \mathbb{R}^3$, 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:

$$\left(\Delta x_i', \Delta y_i'\right) = \left(\frac{2\Delta x_i}{w}, \frac{2\Delta y_i}{h}\right), \quad i = \overline{1,n}, \quad n = h \times w. \tag{11}$$

Пространственные размеры входного тензора могут быть произвольными — перед полносвязными слоями используется глобальный слой субдискретизации с выбором среднего. Также в отличие от исходной архитектуры в предложенной модели не используется батч-нормализация [71]. Использование батч-нормализации приводило к ухудшению сходимости модели и замедлению скорости обучения.

Для обучения нейронных сетей был использован набор данных, состоящий из 15500 вырезанных из различных 3D-фильмов стереопар в разрешении 960×540, к которым применялись случайные аффинные преобразования. Чтобы модель могла успешно оценивать геометрические искажения, предлагается оптимизировать следующий функционал:

$$L\left(\theta, \theta_{\text{gt}}, I^{\text{R}}, I^{\text{R}}_{\text{gt}}, \theta_{\text{b}}\right) = L_{\text{SE}}\left(\theta, \theta_{\text{gt}}\right) + L_{\text{Grid}}\left(\theta, \theta_{\text{gt}}\right) + L_{\text{Warp}}\left(\theta, I^{\text{R}}, I^{\text{R}}_{\text{gt}}\right) + L_{\text{Siam}}\left(\theta, \theta_{\text{b}}\right), \tag{12}$$

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^{R}_{gt} — правый ракурс стереопары, содержащий и не содержащий 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 L_{SE} оптимизируемого функционала представляет собой взвешенную сумму квадратичных разниц между вычисленными и эталонными значениями геометрических искажений с эмпирически подобранными весами для каждого типа искажений:

$$L_{\text{SE}}\left(\theta, \theta_{\text{gt}}\right) = w_\alpha\ \left(\alpha - \alpha_{\text{gt}}\right)^2 + w_k\ \left(k - k_{\text{gt}}\right)^2 + w_t\ \left(t - t_{\text{gt}}\right)^2, \tag{13}$$

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

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 квадрате $[-1, 1] \times [-1, 1]$ с шагом 0.1, таким образом $H = W = 21$. При вычислении L_{Grid} вектор параметров геометрических искажений разбивается на три различных вектора: $\theta^{\text{rotate}} = \begin{bmatrix}\alpha & 0 & 0\end{bmatrix}$, $\theta^{\text{scale}} = \begin{bmatrix}0 & k & 0\end{bmatrix}$, $\theta^{\text{shift}} = \begin{bmatrix}0 & 0 & t\end{bmatrix}$. 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:

$$\begin{aligned} G^\alpha &= T\left(G, \theta^\alpha\right), & G^\alpha_{\text{gt}} &= T\left(G, \theta^\alpha_{\text{gt}}\right), \\ G^k &= T\left(G^\alpha, \theta^k\right), & G^k_{\text{gt}} &= T\left(G^\alpha_{\text{gt}}, \theta^k_{\text{gt}}\right), \\ G^t &= T\left(G^k, \theta^t\right), & G^t_{\text{gt}} &= T\left(G^k_{\text{gt}}, \theta^t_{\text{gt}}\right), \end{aligned}$$

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

$$L_{\text{Grid}}\left(\theta, \theta_{\text{gt}}\right) = w^\alpha_{\text{Grid}}\, \text{MSE}\left(G^\alpha, G^\alpha_{\text{gt}}\right) + w^k_{\text{Grid}}\, \text{MSE}\left(G^k, G^k_{\text{gt}}\right) + w^t_{\text{Grid}}\, \text{MSE}\left(G^t, G^t_{\text{gt}}\right), \tag{14}$$

где $\text{MSE}\left(G^1, G^2\right) = \dfrac{1}{HW} \sum\limits_{i=1}^{H} \sum\limits_{j=1}^{W} \left(G^1_{\{i,j\}} - G^2_{\{i,j\}}\right)^2$, $w^\alpha_{\text{Grid}} = 5000$, $w^k_{\text{Grid}} = 3000$, $w^t_{\text{Grid}} = 3000$.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

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

$$L_{\text{Warp}}\left(\theta, I^{\text{R}}, I^{\text{R}}_{\text{gt}}\right) = \text{MSE}\left(I^{\text{R}}, I^{\text{R}}_{\text{Warp}}\right). \tag{15}$$

Вторая 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, $\theta = -1 \cdot \theta_{\text{b}}$. Таким образом, четвертая компонента оптимизируемого функционала штрафует разницу между предсказанными векторами параметров для левого и правого ракурсов:

$$L_{\text{Siam}}(\theta, \theta_{\text{b}}) = L_{\text{SE}}(\theta, -\theta_{\text{b}}). \tag{16}$$

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 модели был использован метод инициализации весов He [66], а также оптимизационный метод Adam [69], для которого использовались стандартные параметры, за исключением равного 10^{-4} коэффициента скорости обучения, изменявшемуся по косинусному правилу с уменьшающейся амплитудой. Нейросетевой регрессор обучался в течение 120 эпох.

7. Метод оценки перепутанных ракурсов. Для предсказания вероятности наличия перепутанных ракурсов в кадре была также использована архитектура нейронной сети, аналогичная ResNet-18 [70], как и в методе по оценке геометрических искажений. В модификации архитектуры используются четыре остаточных блока перед каждым увеличением размера канала признаков, а батч-нормализация [71] не используется. Помимо этого, последний слой сети предсказывает вектор из двух значений, представляющих собой 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, 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. Результатом работы предложенного алгоритма для оценки перепутанности ракурсов в сцене является число

$$\overline{p} = \frac{1}{n_A} \sum_{j \in A} p_j, \tag{17}$$

где p_j — значение нейросетевого признака j-го кадра анализируемой сцены, $j = \overline{1, n}$, n — число кадров в сцене, $n_A = |A|$ — число подходящих для анализа кадров в сцене, $A = \{k_j \mid 1 \leqslant k_j \leqslant n\}$ — множество номеров кадров сцены, подходящих 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 диспаратности и векторов движения. При этом возникающий дискомфорт при просмотре “темных” кадров с перепутанными ракурсами значительно меньше, чем при просмотре “ярких” кадров [72], что в целом позволяет не учитывать такие кадры при анализе стереофильмов.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

Также при использовании предложенной модели 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.

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:

$$L_{CE}(y,p) = -\frac{1}{N}\sum_{i=1}^{N}\big(y_i \log(p_{i1}) + (1-y_i)\log(p_{i2})\big), \tag{18}$$

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. При этом сам порядок ракурсов выбирался во время обучения случайно.

В качестве метода инициализации весов сети был использован метод Xavier [68], а для оптимизации был выбран алгоритм Adam [69]. 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.

8. Экспериментальная оценка. В данном разделе приводятся результаты сравнения используемых методов оценки стереоскопических искажений с аналогами, а также результаты применения разработанной системы для оценки качества 1000 видео в формате VR180.

8.1. Сравнение с аналогами. Для тестирования метода одновременной оценки искажений цвета и резкости была подготовлена тестовая выборка на основе набора данных Sintel [73]. 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 случайные искусственные искажения. 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 табл. 1. Используемый в системе метод превосходит по качеству другие методы как по корреляции Пирсона, так и по корреляции Спирмена.

Тестирование метода оценки геометрических искажений проводилось на наборе данных из 3700 вырезанных из 3D-фильмов стереопар, к которым применялись случайные аффинные преобразования. Сравнение проводилось как с нейросетевыми аналогами, так и с методом, ранее применявшимся при анализе полнометражных стереоскопических фильмов. Результаты представлены в табл. 2. 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. Для предложенного метода 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.

Также на тестовой выборке, состоящей из 900 сцен длиной в 30 кадров, было проведено сравнение предложенного алгоритма поиска перепутанных ракурсов с аналогами, применявшимися на практике при анализе полнометражных стереофильмов, а также с нейросетевыми методами построения карт дис-

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

Таблица 1. Результаты тестирования методов оценки различий по цвету и резкости между ракурсами стереовидео на искусственном наборе данных Sintel

Table 1. Results of testing color and sharpness mismatch estimation methods on the Sintel synthetic dataset

Method	**Pearson correlation**	**Spearman correlation**
Color distortion		
MAE	0.1254	0.1626
MAE with compensation	0.1338	0.2039
Method [33]	-0.4430	-0.4093
Method [38]	0.8136	0.8760
The proposed method	**0.9696**	**0.9602**
Sharpness distortions		
Method [23]	0.1310	0.0692
Method [24]	0.9564	0.8047
Method [22]	0.5176	0.3152
Method [37]	0.7686	0.6815
Method [29]	0.8151	0.4488
The proposed method	**0.9762**	**0.9078**

Таблица 2. Результаты тестирования метода оценки геометрических искажений. В таблице представлена абсолютная погрешность вычислений по каждому геометрическому искажению

Table 2. Results of testing the geometric distortion estimation method. The table shows the absolute error for each geometric distortion

Method	**Angle**	**Scale**	**Shift**
Zero vector	0.634	0.651	0.575
Method [49]	0.437	1.236	0.825
Method [47]	0.051	0.108	0.191
Method [38]	0.012	0.026	0.020
The proposed method	**0.0099**	**0.0002**	**9.1e-05**

Таблица 3. Результаты тестирования алгоритмов поиска перепутанных ракурсов в стереовидео

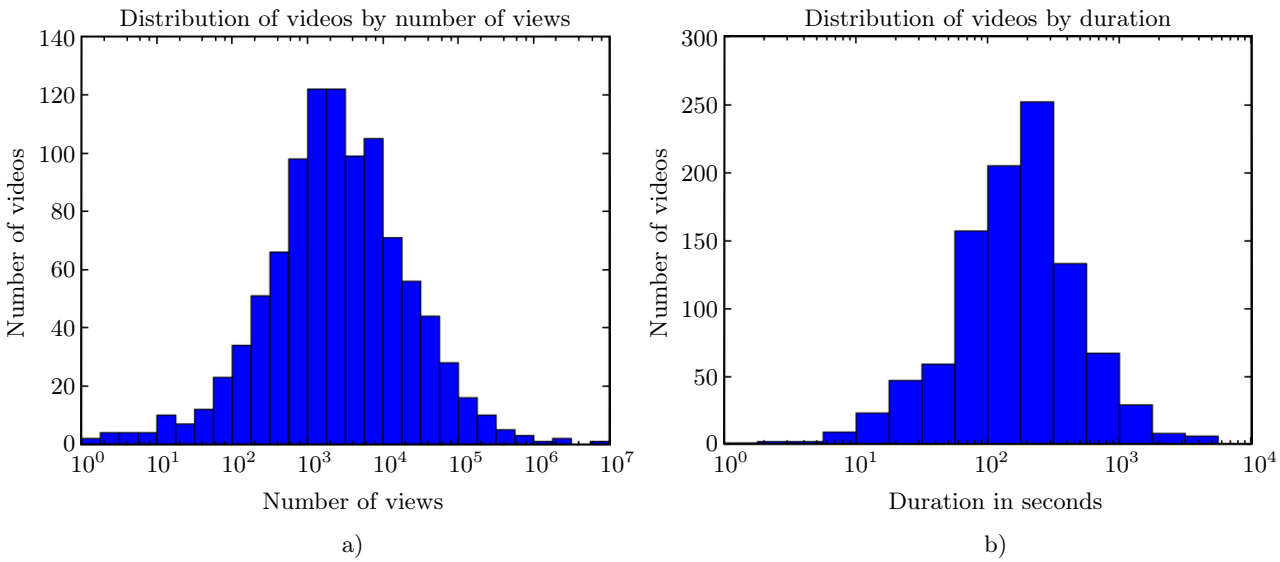
Table 3. Results of testing algorithms for channel mismatch detection in stereoscopic videos

Algorithm	AUC of ROC	Precision	F-measure
Method [60]	0.7223	0.6614	0.6683
Method [62]	0.901	0.8378	0.8409
Method [8]	0.957	0.8946	0.8928
Method [57]	0.9913	0.8394	0.8613
Method [58]	0.9899	0.8256	0.8515
The proposed method	**0.9963**	**0.9784**	**0.9789**

AUC of ROC — площадь под ROC-кривой

паратности по одному кадру. Для использования последних методов в задаче оценки порядка ракурсов вычислялась корреляция Пирсона между предсказанными картами диспаратности и картами, вычисленными с помощью блочного метода компенсации движения. 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. Полученные показатели представлены в табл. 3. По результатам тестирования видно, что предложенный метод поиска перепутанных ракурсов в стереовидео превосходит существующие аналоги по качеству классификации.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323



a) b)

Рис. 5. Общие статистики по набору данных VR180-видео: a) распределение по числу просмотров; b) распределение по длительности

Fig. 5. General statistics for the VR180 video dataset: a) distribution by number of views; b) distribution by duration

8.2. 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. В набор данных были включены только доступные для загрузки видео в стереоскопическом формате, обладающие высоким разрешением.

На рис. 5 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 мин.

8.3. Результаты оценки цветовых, резкостных и геометрических искажений VR180-видео. Для всех 1 000 видео в формате VR180 было проведено измерение силы цветовых искажений, искажений резкости и геометрических искажений. Результаты для цветовых искажений и вертикального сдвига представлены на рис. 6, 7 соответственно. Результаты по остальным искажениям доступны в отчете `https://videoprocessing.ai/stereo_quality/report12.html`. Результаты анализа продемонстрированы а) относительно количества просмотров на YouTube, б) даты публикации и в) длительности 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-му перцентилю, а нижняя — 66-му 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.
- 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.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

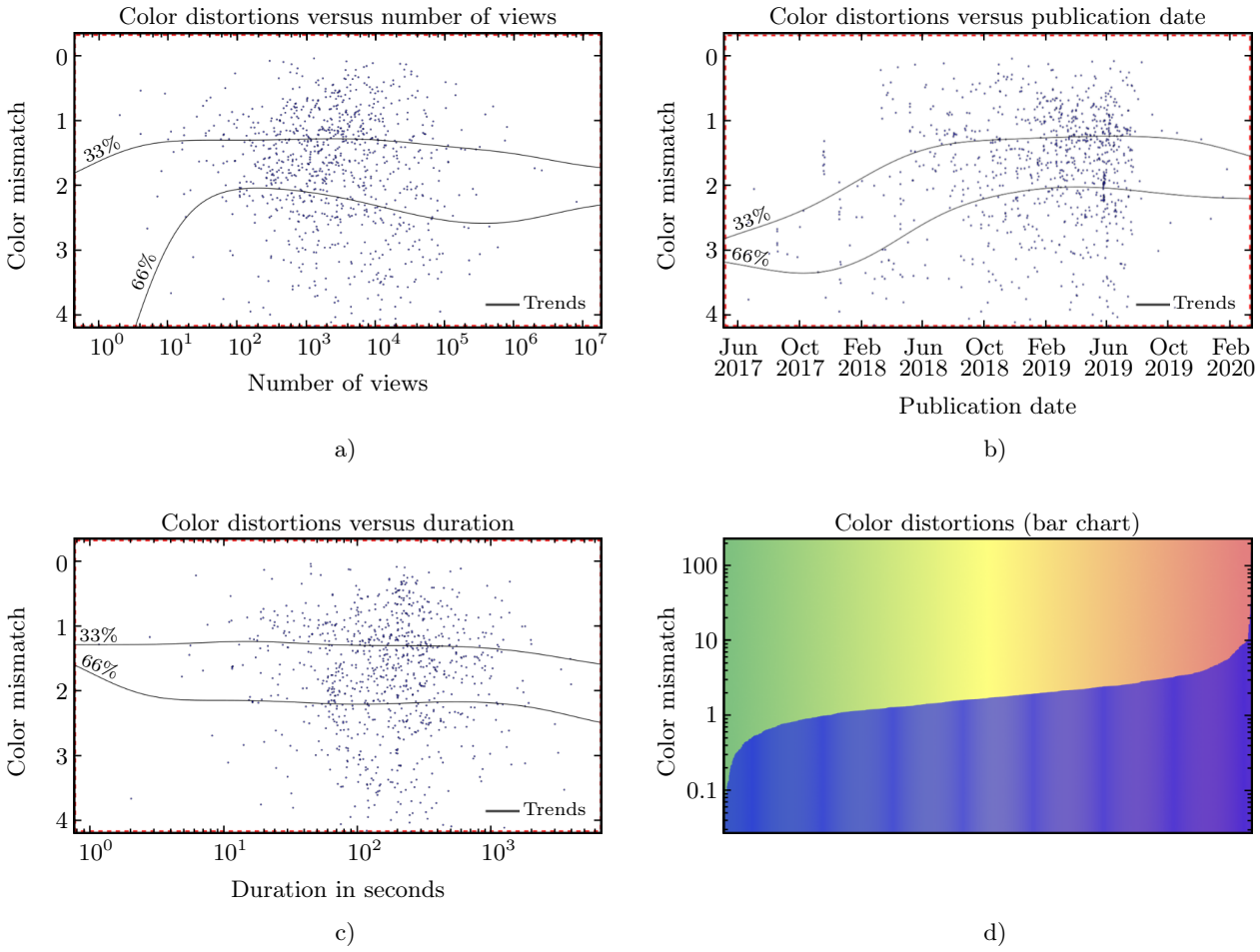


Рис. 6. Результаты анализа цветовых искажений в VR180-видео: a) по количеству просмотров на YouTube; b) по дате публикации видео; c) по длительности видео; d) по средним значениям искажений для каждого видео

Fig. 6. Results of color distortion analysis in VR180 video: a) by number of views on YouTube; b) by date of video publication; c) by video duration; d) by average distortion values for each video

При этом значительное количество проанализированных видео в формате VR180 демонстрирует наличие по крайней мере одного стереоскопического артефакта из рассмотренной группы искажений. Рис. 6 d, 7 d 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 соответствуют либо “плоским” видео с одинаковыми ракурсами, либо видео на основе компьютерной графики. Примеры найденных искажений продемонстрированы на рис. 8.

8.4. Результаты поиска перепутанных ракурсов в VR180-видео. Для поиска перепутанных ракурсов в VR180 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 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. Пример сцены с перепутанными ракурсами показан на рис. 9.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

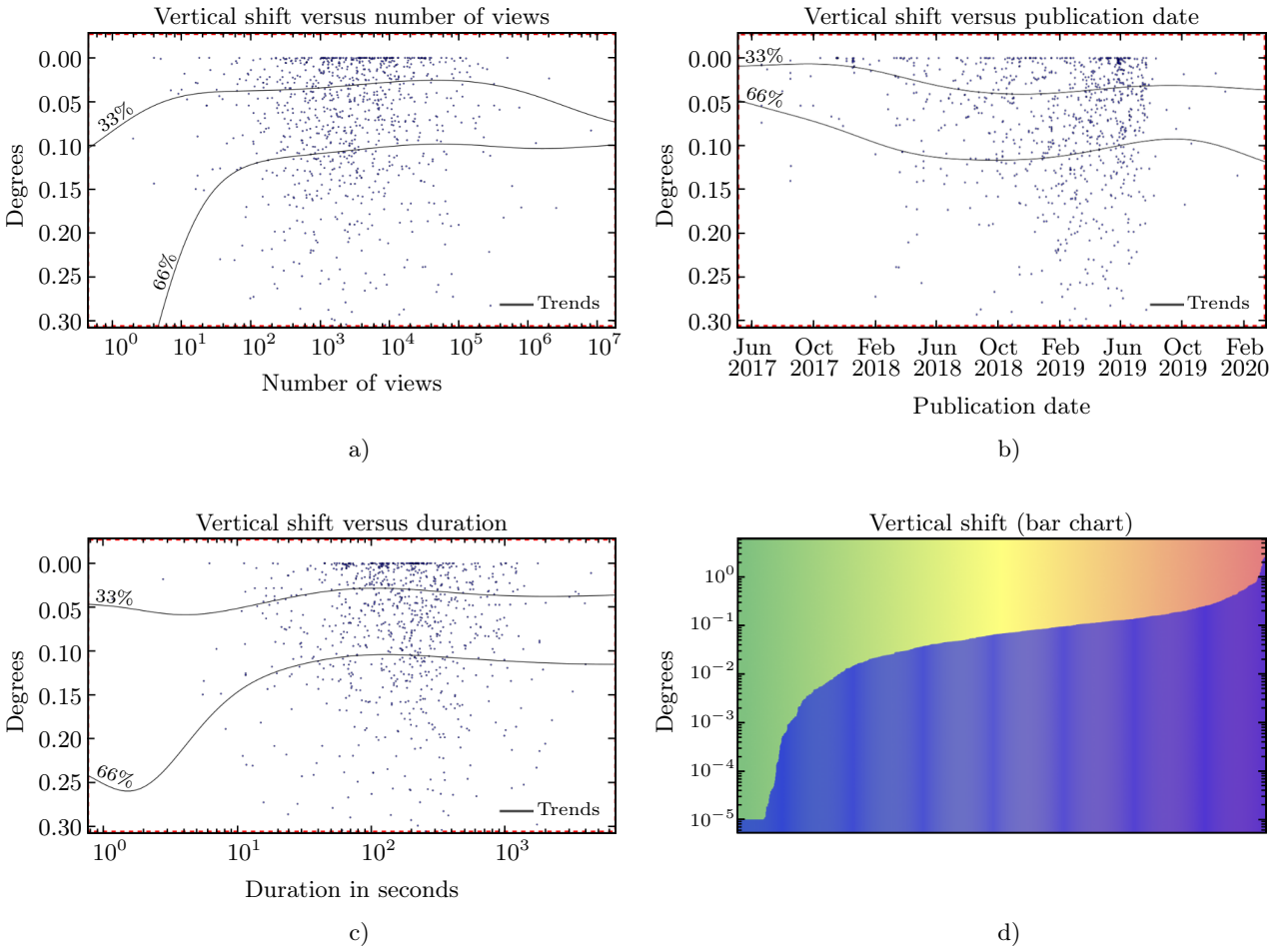


Рис. 7. Результаты анализа вертикального сдвига в VR180-видео: a) по количеству просмотров на YouTube; b) по дате публикации видео; c) по длительности видео; d) по средним значениям искажений для каждого видео

Fig. 7. Results of vertical shift analysis in VR180 video: a) by number of views on YouTube; b) by date of video publication; c) by video duration; d) by average distortion values for each video

9. Ограничения. Предложенная программная система поиска стереоскопических искажений обладает рядом следующих ограничений:

- *Отсутствие оценки степени зрительского дискомфорта.* Предложенная система в первую очередь направлена на поиск соответствующих стереоскопических искажений для их дальнейшего исправления на этапе производства стереоскопического контента. На данный момент не существует наборов данных с субъективными оценками видео в формате VR180, которые можно было бы использовать для обучения модели предсказания зрительского дискомфорта. При этом первым шагом в задаче подготовки такого набора данных является поиск видео с рассматриваемыми стереоскопическими искажениями, что и было осуществлено в рамках данной работы. Дальнейшим шагом в этом направлении будет организация масштабного субъективного сравнения для получения оценок степени дискомфорта, что является одним из направлений будущих исследований в рамках рассматриваемой темы.
- *Потенциальные ложноотрицательные срабатывания.* Предложенные методы поиска стереоскопических искажений в том или ином виде используют в качестве входных данных карты диспаратности и точность их работы напрямую зависит от точности построенной карты диспаратности. В случае отсутствия высокочастотной информации в кадре или его низкой яркости сопоставление ракурсов будет неточным, что может привести к ложноположительным и ложноотрицательным срабатываниям методов поиска искажений. Для того чтобы предложенная программная система не выдавала

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

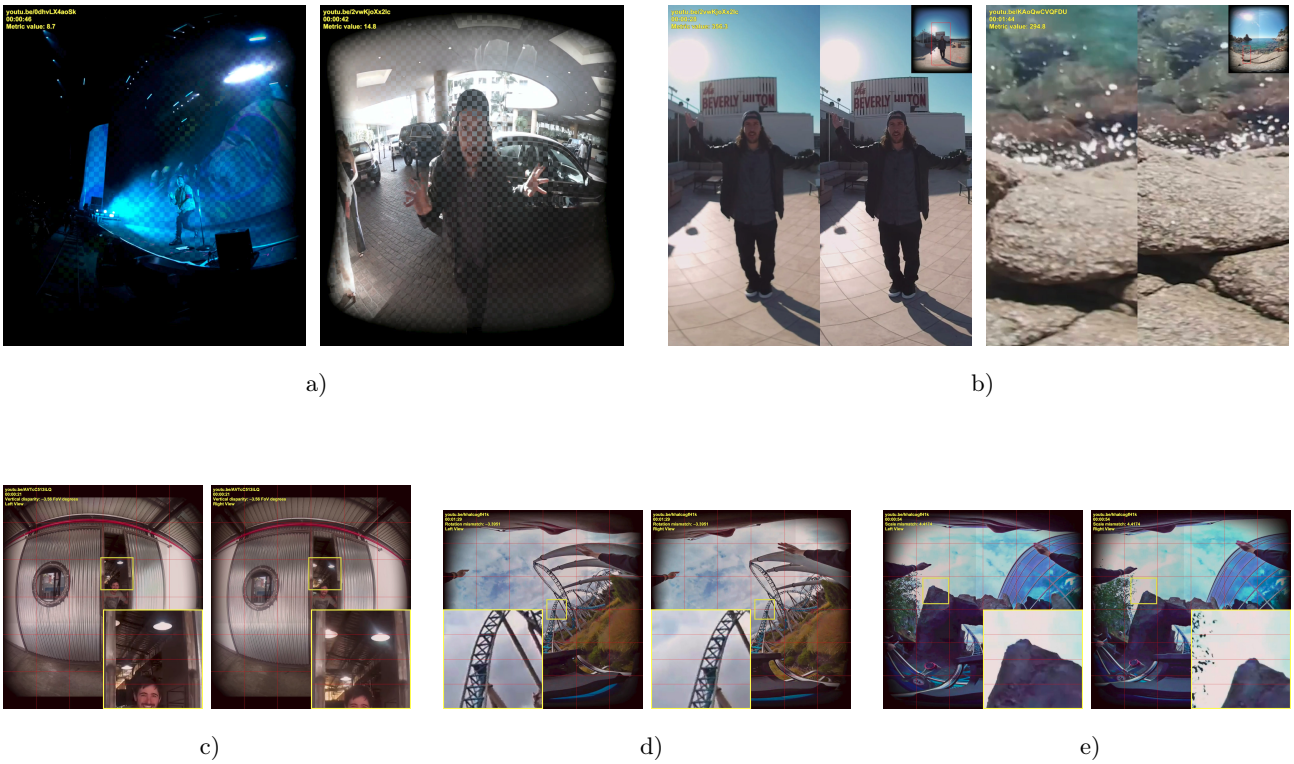


Рис. 8. Примеры найденных стереоскопических искажений в проанализированных видео в формате VR180: a) цветовые; b) резкости; c) поворот; d) масштабирование; e) вертикальный сдвиг

Fig. 8. Examples of stereoscopic distortions found in analyzed videos in VR180 format: a) color; b) sharpness; c) rotation; d) scaling; e) vertical shift

Рис. 9. Пример сцены с перепутанными ракурсами, возникшими из-за неправильного наложения элементов компьютерной графики титров

Fig. 9. An example of a scene with channel mismatch caused by incorrectly superimposed CG elements in the titles

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

много ложноположительных срабатываний в таких случаях, при построении карты диспаратности также строится карта доверия, которая далее задействована в методах поиска искажений. Данная карта позволяет не учитывать некорректно сопоставленные участки кадров, а в случае большого количества таких областей в кадре — отфильтровать результаты по среднему доверию. Такой подход решает проблему ложноположительных срабатываний, но может привести к ложноотрицательным. Однако в случае возникновения стереоскопических искажений в этих условиях они будут практически незаметны для глаза, как было показано в [1], что практически нивелирует необходимость их исправления по сравнению с более заметными искажениями.

- *Перенос домена классических 3D-видео на видео в формате VR180.* Предложенные в работе методы поиска стереоскопических искажений обучены на наборах данных, состоящих из стандартных 3D-видео в формате стереопары. Для их применения к VR180-видео используется фронтальная грань кубической проекции анализируемых видео, которые практически не отличаются от классических ракурсов. Хотя такой подход позволил успешно осуществить поиск стереоскопических искажений в VR180-видео, обучение непосредственно на VR180-видео потенциально может улучшить качество работы методов в этом домене. Однако это остается предметом дальнейшего исследования, так как потребует сбор VR180-видео без искажений (которых по результатам данного исследования оказалось немного) и адаптацию методов генерации искусственных искажений.

10. Заключение. В ходе проведенного исследования была разработана и реализована универсальная программная система, способная автоматически и с высокой точностью оценивать качество стереоскопических видео в формате VR180. Практические эксперименты подтвердили эффективность предложенных методов для обнаружения как локальных искажений (по цвету и резкости), так и глобальных (геометрических искажений и перепутанных ракурсов). Архитектура системы позволяет легко масштабировать ее под различные разрешения и форматы, а также расширять функционал за счет подключения дополнительных модулей. Автоматизированные отчеты могут помочь повысить производственную эффективность, упрощая контроль и анализ стереоконтента на всех этапах постобработки. Дальнейшее развитие системы предполагает расширение набора детектируемых искажений и разработку методов их исправления, что позволит еще более точно определять технические огрехи стереосъемки и автоматически их исправлять, что, в конечном итоге, существенно повысит комфорт при просмотре 3D- и VR180-видео.

Список литературы

1. *Antsiferova A., Vatolin D.* The influence of 3D video artifacts on discomfort of 302 viewers // 2017 Int. Conf. on 3D Immersion (IC3D). 2017. New York: IEEE Press, 1–8. doi 10.1109/IC3D.2017.8251897.
2. *Рожкова Г.И., Васильева Н.Н.* Сравнительные трудности восприятия фильмов в 2D и 3D форматах // Мир техники кино. 2010. **4**, № 2. 12–18.
3. *Рожкова Г.И., Алексеенко С.В.* Зрительный дискомфорт при восприятии стереоскопических изображений как следствие непривычного распределения нагрузки на различные механизмы зрительной системы // Мир техники кино. 2011. **5**, № 3. 12–21.
4. *Васильева Н.Н., Рожкова Г.И., Рожков С.Н.* О пользе и вреде современных технологий формирования стереокиноизображений для людей с различным состоянием зрительных функций // Мир техники кино. 2011. **5**, № 1. 7–15.
5. *Рожков С.Н., Рожкова Г.И.* Искажения пространственных образов в стереокино: иллюзии уменьшения, увеличения и уплощения объектов // Мир техники кино. 2013. **7**, № 3. 13–20.
6. *Hoffman D.M., Girshick A.R., Akeley K., Banks M.S.* Vergence-accommodation conflicts hinder visual performance and cause visual fatigue // Journal of Vision. 2008. **8**, N 3. Article Number 33. doi 10.1167/8.3.33.
7. *Khaustova D., Fournier J., Wyckens E., Le Meur O.* An objective method for 3D quality prediction using visual annoyance and acceptability level // Proc. SPIE **9391**, Stereoscopic Displays and Applications XXVI. 2015. doi 10.1117/12.2076949.
8. *Bokov A., Lavrushkin S., Erofeev M., et al.* Toward fully automatic channel-mismatch detection and discomfort prediction for S3D video // International Conference on 3D Imaging (IC3D). New York: IEEE Press, 2017. 1–7. doi 10.1109/IC3D.2016.7823462.
9. *Lavrushkin S., Lyudvichenko V., Vatolin D.* Local method of color-difference correction between stereoscopic-video views // Proc. 2018 3DTV Conf. on the True Vision — Capture, Transmission and Display of 3D Video (3DTV-CON). New York: IEEE Press, 2018. 1–4. doi 10.1109/3DTV.2018.8478453.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

10. *Lavrushkin S., Vatolin D.* Channel-mismatch detection algorithm for stereoscopic video using convolutional neural network // Proc. 2018 3DTV Conf. on the True Vision —– Capture, Transmission and Display of 3D Video (3DTV-CON). New York: IEEE Press, 2018. 1–4. doi 10.1109/3DTV.2018.8478542.
11. *Lavrushkin S., Kozhemyakov K., Vatolin D.* Neural-network-based detection methods for color, sharpness, and geometry artifacts in stereoscopic and VR180 videos // Proc. 2020 Int. Conf. on 3D Immersion (IC3D), New York: IEEE Press, 2020. 1–8. doi 10.1109/IC3D51119.2020.9376385.
12. *Lavrushkin S., Molodetskikh I., Kozhemyakov K., Vatolin D.* Stereoscopic quality assessment of 1,000 VR180 videos using 8 metrics // J. Electron. Imaging. 2021. N 2, 350-1–350-7. doi 10.2352/ISSN.2470-1173.2021.2.SDA-350.
13. *Pentland A.P.* A new sense for depth of field // IEEE Trans. Pattern Anal. Mach. Intell. 1987. **9**, N 4. 523–531. doi 10.1109/TPAMI.1987.4767940.
14. *Elder J.H., Zucker S.W.* Local scale control for edge detection and blur estimation // IEEE Trans. on Pattern Anal. Mach. Intell. 1998. **20**, N 7. 699–716. doi 10.1109/34.689301.
15. *Zhuo S., Sim T.* Defocus map estimation from a single image // Pattern Recognition. 2011. **44**, N 9. 1852–1858. doi 10.1016/j.patcog.2011.03.009.
16. *Cao Y., Fang S., Wang Z.* Digital multi-focusing from a single photograph taken with an uncalibrated conventional camera // IEEE Trans. Image Process. 2013. **22**, N 9. 3703–3714. doi 10.1109/TIP.2013.2270086.
17. *Karaali A., Jung C.R.* Adaptive scale selection for multiresolution defocus blur estimation // Proc. 2014 IEEE Int. Conf. on Image Processing (ICIP). New York: IEEE Press, 2014. 4597–4601. doi 10.1109/ICIP.2014.7025932.
18. *Karaali A., Jung C.R.* Edge-based defocus blur estimation with adaptive scale selection // IEEE Trans. Image Process. 2018. **27**, N 3. 1126–1137. doi 10.1109/TIP.2017.2771563.
19. *Chakrabarti A., Zickler T., Freeman W.T.* Analyzing spatially-varying blur // Proc. 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2010. 2512–2519. doi 10.1109/CVPR.2010.5539954.
20. *Zhu X., Cohen S., Schiller S., Milanfar P.* Estimating spatially varying defocus blur from a single image // IEEE Trans. Image Process. 2013. **22**, N 12. 4879–4891. doi 10.1109/TIP.2013.2279316.
21. *D'Andrès L., Salvador J., Kochale A., Süsstrunk S.* Non-parametric blur map regression for depth of field extension // IEEE Trans. Image Process. 2016. **25**, N 4. 1660–1673. doi 10.1109/TIP.2016.2526907.
22. *Golestaneh S.A., Karam L.J.* Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes // Proc. IEEE Conf. on Computer Vision and Pattern Recognition. New York: IEEE Press, 2017. 596–605. doi 10.1109/CVPR.2017.71.
23. *Narvekar N.D., Karam L.J.* A no-reference image blur metric based on the cumulative probability of blur detection (CPBD) // IEEE Trans. Image Process. 2011. **20**, N 9. 2678–2683. doi 10.1109/TIP.2011.2131660.
24. *Kumar J., Chen F., Doermann D.* Sharpness estimation for document and scene images // Proc. 21st Int. Conf. on Pattern Recognition (ICPR), Tsukuba, Japan, 2012. 3292–3295.
25. *Zeng K., Wang Y., Mao J., et al.* A local metric for defocus blur detection based on CNN feature learning // IEEE Trans. Image Process. 2019. **28**, N 5. 2107–2115. doi 10.1109/TIP.2018.2881830.
26. *Park J., Tai Y.-W., Cho D., So Kweon I.* A unified approach of multi-scale deep and hand-crafted features for defocus estimation // https://arxiv.org/abs/1704.08992. Cited August 28, 2025.
27. *Lee J., Lee S., Cho S., Lee S.* Deep defocus map estimation using domain adaptation // Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2019. 12214–12222. doi 10.1109/CVPR.2019.01250.
28. *Tang C., Liu X., Zhu X., et al.* R^2MRF: defocus blur detection via recurrently refining multi-scale residual features // Proc. AAAI Conf. on Artificial Intelligence. 2020. **34**, N 07. 12063–12070. doi 10.1609/aaai.v34i07.6884.
29. *Cun X., Pun C.-M.* Defocus blur detection via depth distillation // European Conference on Computer Vision. 2020. 747–763. doi 10.1007/978-3-030-58601-0_44. https://arxiv.org/abs/2007.08113. Cited August 30, 2025.
30. *Karaali A., Harte N., Jung C.R.* Deep multi-scale feature learning for defocus blur estimation // IEEE Trans. Image Process. 2022. **31**. 1097–1106. doi 10.1109/TIP.2021.3139243.
31. *Li H., Qian W., Cao J., et al.* Multi-interactive enhanced for defocus blur estimation // IEEE Trans. Comput. Imaging. 2024. **10**, 640–652. doi 10.1109/TCI.2024.3354427.
32. *Zhao Z., Yang H., Liu P., et al.* Defocus blur detection via adaptive cross-level feature fusion and refinement // Vis. Comput. 2024. **40**, N 11. 8141–8153. doi 10.1007/s00371-023-03229-7.
33. *Winkler S.* Efficient measurement of stereoscopic 3D video content issues // Proc. SPIE 9016. Image Quality and System Performance XI. 2014. Page 90160Q. doi 10.1117/12.2042211.
34. *Dong Q., Zhou T., Guo Z., Xiao J.* A stereo camera distortion detecting method for 3DTV video quality assessment // Proc. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC). New York: IEEE Press, 2014. 1–4. doi 10.1109/APSIPA.2013.6694209.

35. *Devernay F., Pujades S., Vijay Ch.A.V.* Focus mismatch detection in stereoscopic content // Proc. SPIE **8288**. Stereoscopic Displays and Applications XXIII. 2012. Page 82880E. doi 10.1117/12.906209.
36. *Liu M., Müller K.* Automatic analysis of sharpness mismatch between stereoscopic views for stereo 3D videos // Proc. 2014 Int. Conf. on 3D Imaging (IC3D). New York: IEEE Press, 2015. 1–6. doi 10.1109/IC3D.2014.7032572.
37. *Vatolin D., Bokov A.* Sharpness mismatch and 6 other stereoscopic artifacts measured on 10 Chinese S3D movies // J. Electron. Imaging. **2017**, N 5. 137–144. doi 10.2352/ISSN.2470-1173.2017.5.SDA-340.
38. *Vatolin D., Bokov A., Erofeev M., Napadovsky V.* Trends in S3D-movie quality evaluated on 105 films using 10 metrics // J. Electron. Imaging. 2016. **2016**, N 5. 1–10. doi 10.2352/ISSN.2470-1173.2016.5.SDA-439.
39. *Doutre C., Pourazad M.T., Tourapis A., et al.* Correcting unsynchronized zoom in 3D video // Proc. 2010 IEEE Int. Symposium on Circuits and Systems (ISCAS). New York: IEEE Press, 2010. 3244–3247. doi 10.1109/ISCAS.2010.5537923.
40. *Lowe D.G.* Distinctive image features from scale-invariant keypoints // Int. J. Comput. Vis. 2004. **60**, N 2. 91–110. doi 10.1023/B:VISI.0000029664.99615.94.
41. *Pekkucuksen I.E., Batur A.U., Zhang B.* A real-time misalignment correction algorithm for stereoscopic 3D cameras // Proc. SPIE **8288**, Stereoscopic Displays and Applications XXIII. 2012. Page 82880J. doi 10.1117/12.906902.
42. *Voronov A., Borisov A., Vatolin D.* System for automatic detection of distorted scenes in stereo video // Proc. Sixth Int. Workshop on Video Processing and Quality Metrics (VPQM). 2012.
43. *Fischler M.A., Bolles R.C.* Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Commun. ACM. 1981. **24**, N 6. 381–395. doi 10.1145/358669.358692.
44. *Brachmann E., Krull A., Nowozin S., et al.* DSAC — differentiable RANSAC for camera localization // Proc. IEEE Conf. on Computer Vision and Pattern Recognition. 2017. New York: IEEE Press, 2017. 2492–2500. doi 10.1109/CVPR.2017.267.
45. *Brachmann E., Rother C.* Neural-guided RANSAC: learning where to sample model hypotheses // Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV). New York: IEEE Press, 2019. 4321–4330. doi 10.1109/ICCV.2019.00442.
46. *Kluger F., Rosenhahn B.* PARSAC: accelerating robust multi-model fitting with parallel sample consensus // Proc. AAAI Conf. on Artificial Intelligence. 2024. **38**, N 3. 2804–2812. doi 10.1609/aaai.v38i3.28060.
47. *Chen J., Gu Y., Luo L.* Learning to find good correspondences based on global and local attention mechanism // Proc. 2021 China Automation Congress (CAC). New York: IEEE Press, 2022. 2174–2178. doi 10.1109/CAC53003.2021.9728340.
48. *Sun W., Jiang W., Trulls E., et al.* ACNe: attentive context normalization for robust permutation-equivariant learning // Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2020. 11283–11292. doi 10.1109/CVPR42600.2020.01130.
49. *Rocco I., Arandjelović R., Sivic J.* Convolutional neural network architecture for geometric matching // IEEE Trans. Pattern Anal. Mach. Intell. 2019. **41**, N 11. 2553–2567. doi 10.1109/TPAMI.2018.2865351.
50. *Rocco I., Arandjelović R., Sivic J.* End-to-end weakly-supervised semantic alignment // Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2018. New York: IEEE Press, 2018. 6917–6925. doi 10.1109/CVPR.2018.00723.
51. *Lee J., Jung C., Kim C., Said A.* Content-based pseudoscopic view detection // Journal of Signal Processing Systems. 2012. **68**, N 2. 261–271. doi 10.1007/s11265-011-0608-8.
52. *Palou G., Salembier P.* Monocular depth ordering using T-junctions and convexity occlusion cues // IEEE Trans. Image Process. 2013. **22**, N 5. 1926–1939. doi 10.1109/TIP.2013.2240002.
53. *Lee H., Jung C., Kim C.* Depth map estimation based on geometric scene categorization // Proc. 19th Korea–Japan Joint Workshop on Frontiers of Computer Vision (FCV). New York: IEEE Press, 2013. 170–173. doi 10.1109/FCV.2013.6485482.
54. *Hoiem D., Stein A.N., Efros A.A., Hebert M.* Recovering occlusion boundaries from a single image // Proc 2007 IEEE 11th Int. Conf. on Computer Vision. New York: IEEE Press, 2007. 1–8. doi 10.1109/ICCV.2007.4408985.
55. *Fu H., Gong M., Wang C., et al.* Deep ordinal regression network for monocular depth estimation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. 2002–2011. doi 10.1109/CVPR.2018.00214. https://arxiv.org/abs/1806.02446. Cited August 29, 2025.
56. *Godard C., Mac Aodha O., Firman M., and Brostow G.* Digging into self-supervised monocular depth estimation // Proc. IEEE/CVF Int. Conf. on Computer Vision, New York: IEEE Press, 2019. 3827–3837. doi 10.1109/ICCV.2019.00393.
57. *Ranftl R., Lasinger K., Hafner D., et al.* Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer // IEEE Trans. Pattern Anal. Mach. Intell. 2022. **44**, N 3. 1623–1637. doi 10.1109/TPAMI.2020.3019967.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

58. *Yin W., Zhang J., Wang O., Niklaus S., et al.* Learning to recover 3D scene shape from a single image // Proc. 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021. 204–213. doi 10.1109/CVPR46437.2021.00027.
59. *Yang L., Kang B., Huang Z., et al.* Depth anything: unleashing the power of large-scale unlabeled data // Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition. New York: IEEE Press, 2024. 10371–10381. doi 10.1109/CVPR52733.2024.00987.
60. *Knee M.* Getting machines to watch 3D for you // SMPTE Motion Imaging J. 2012. **121**, N 3. 52–58. doi 10.5594/j18162.
61. *Bouchard J., Nazzar Y., Clark J.J.* Half-occluded regions and detection of pseudoscopy // Proc. 2015 Int. Conference on 3D Vision. New York: IEEE Press, 2015. 215–223. doi 10.1109/3DV.2015.32.
62. *Shestov A., Voronov A., Vatolin D.* Detection of swapped views in stereo image // Proc. 22nd GraphiCon Int. Conf. on Computer Graphics and Vision. 2012. 23–27.
63. *Simonyan K., Grishin S., Vatolin D., Popov D.* Fast video super-resolution via classification // Proc. 2008 15th IEEE Int. Conf. on Image Processing. New York: IEEE Press, 2008. 349–352. doi 10.1109/ICIP.2008.4711763.
64. *Egnal G., Wildes R.P.* Detecting binocular half-occlusions: empirical comparisons of five approaches // IEEE Trans. Pattern Anal. Mach. Intell. 2002. **24**, N 8. 1127–1133. doi 10.1109/TPAMI.2002.1023808.
65. *Fourure D., Emonet R., Fromont E., et al.* Residual conv-deconv grid network for semantic segmentation // https://arxiv.org/abs/1707.07958. Cited August 29, 2025.
66. *He K., Zhang X., Ren S., Sun J.* Delving deep into rectifiers: surpassing human-level performance on ImageNet classification // Proc. 2015 IEEE Int. Conf. on Computer Vision (ICCV). New York: IEEE Press, 2015. 1026–1034. doi 10.1109/ICCV.2015.123.
67. *Min D., Choi S., Lu J., et al.* Fast global image smoothing based on weighted least squares // IEEE Trans. Image Process. 2014. **23**, N 12. 5638–5653. doi 10.1109/TIP.2014.2366600.
68. *Glorot X., Bengio Y.* Understanding the difficulty of training deep feedforward neural networks // Proc. Thirteenth Int. Conf. on Artificial Intelligence and Statistics. 2010. 249–256.
69. *Kingma D.P., Ba J.L.* Adam: a method for stochastic optimization // International Conference on Learning Representations (ICLR). 2015. https://arxiv.org/pdf/1412.6980. Cited August 29, 2025.
70. *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition // Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2016. 770–778. doi 10.1109/CVPR.2016.90.
71. *Ioffe S., Szegedy C.* Batch normalization: accelerating deep network training by reducing internal covariate shift // Int. Conf. on Machine Learning. 2015. 448–456. https://arxiv.org/abs/1502.03167. Cited August 29, 2025.
72. *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. 40–46.
73. *Butler D.J., Wulff J., Stanley G.B., Black M.J.* A naturalistic open source movie for optical flow evaluation // Proc. European Conference on Computer Vision. 2012. 611–625. doi 10.1007/978-3-642-33783-3_44.

Получена 17 июля 2025 г. | *Принята* 16 августа 2025 г. | *Опубликована* 20 сентября 2025 г.

Информация об авторе

Сергей Валерьевич Лаврушкин — ст. науч. сотр.; 1) Московский государственный университет имени М. В. Ломоносова, Институт перспективных исследований проблем искусственного интеллекта и интеллектуальных систем, Ломоносовский пр-кт, 27, к. 1, 119234, Москва, Российская Федерация; 2) Институт системного программирования имени В. П. Иванникова РАН, Исследовательский центр доверенного искусственного интеллекта, ул. Александра Солженицына, д. 25, 109004, Москва, Российская Федерация.

References

1. A. Antsiferova and D. Vatolin, “The Influence of 3D Video Artifacts on Discomfort of 302 Viewers,” in *Proc. 2017 Int. Conf. on 3D Immersion (IC3D), Brussels, Belgium, December 11–12, 2017* (IEEE Press, New York, 2017), pp. 1–8. doi 10.1109/IC3D.2017.8251897.
2. G. I. Rozhkova and N. N. Vasilyeva, “Comparative Perceptual Difficulties Associated with Viewing Films in 2D and 3D Format,” World of Technique of Cinema **4** (2), 12–18 (2010) [in Russian].

3. G. I. Rozhkova and S. V. Alekseenko, "Visual Discomfort in Conditions of Stereoscopic Image Perception as a Consequence of Unusual Distribution of Loads on Different Mechanisms of Visual System," World of Technique of Cinema **5** (3), 12–21 (2011) [in Russian].
4. N. N. Vasilyeva, G. I. Rozhkova, and S. N. Rozhkov, "On the Good and Harm from the Modern Technologies of Creating Cinematographic Stereo Images for the People with Different States of Visual Functions," World of Technique of Cinema **5** (1), 7–15 (2011) [in Russian].
5. S. N. Rozhkov and G. I. Rozhkova, "Distortions of Spatial Images in Stereo Movies: Illusions of Object Diminution, Enlargement and Flattening," World of Technique of Cinema **7** (3), 13–20 (2013) [in Russian].
6. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue," J. Vis. **8** (3), Article Number 33 (2008). doi 10.1167/8.3.33.
7. D. Khaustova, J. Fournier, E. Wyckens, and O. Le Meur, "An Objective Method for 3D Quality Prediction Using Visual Annoyance and Acceptability Level," in *Stereoscopic Displays and Applications XXVI*, Vol. 9391, 2015. doi 10.1117/12.2076949.
8. A. Bokov, S. Lavrushkin, M. Erofeev, et al., "Toward Fully Automatic Channel-Mismatch Detection and Discomfort Prediction for S3D Video," in *Proc. Int. Conf. on 3D Imaging (IC3D), Liege, Belgium, December 13–14, 2016* (IEEE Press, New York, 2017), pp. 1–7. doi 10.1109/IC3D.2016.7823462.
9. S. Lavrushkin, V. Lyudvichenko, and D. Vatolin, "Local Method of Color-Difference Correction between Stereoscopic-Video Views," in *Proc. 2018 3DTV Conf. on the True Vision — Capture, Transmission and Display of 3D Video (3DTV-CON), Helsinki, Finland, June 3–5, 2018* (IEEE Press, New York, 2018), pp. 1–4. doi 10.1109/3DTV.2018.8478453.
10. S. Lavrushkin and D. Vatolin, "Channel-Mismatch Detection Algorithm for Stereoscopic Video Using Convolutional Neural Network," in *Proc. 2018 3DTV Conf. on the True Vision — Capture, Transmission and Display of 3D Video (3DTV-CON), Helsinki, Finland, June 3–5, 2018* (IEEE Press, New York, 2018), pp. 1–4. doi 10.1109/3DTV.2018.8478542.
11. S. Lavrushkin, K. Kozhemyakov, and D. Vatolin, "Neural-Network-Based Detection Methods for Color, Sharpness, and Geometry Artifacts in Stereoscopic and VR180 Videos," in *Proc. 2020 Int. Conf. on 3D Immersion (IC3D), Brussels, Belgium, December 15, 2020* (IEEE Press, New York, 2020), pp. 1–8. doi 10.1109/IC3D51119.2020.9376385.
12. S. Lavrushkin, I. Molodetskikh, K. Kozhemyakov, and D. Vatolin, "Stereoscopic Quality Assessment of 1,000 VR180 Videos Using 8 Metrics," J. Electron. Imaging N 2, 350-1–350-7 (2021). doi 10.2352/ISSN.2470-1173.2021.2.SDA-350.
13. A. P. Pentland, "A New Sense for Depth of Field," IEEE Trans. Pattern Anal. Mach. Intell. **9** (4), 523–531 (1987). doi 10.1109/TPAMI.1987.4767940.
14. J. H. Elder and S. W. Zucker, "Local Scale Control for Edge Detection and Blur Estimation," IEEE Trans. Pattern Anal. Mach. Intell. **20** (7), 699–716 (1998). doi 10.1109/34.689301.
15. S. Zhuo and T. Sim, "Defocus Map Estimation from a Single Image," Pattern Recognit. **44** (9), 1852–1858 (2011). doi 10.1016/j.patcog.2011.03.009.
16. Y. Cao, S. Fang, and Z. Wang, "Digital Multi-Focusing from a Single Photograph Taken with an Uncalibrated Conventional Camera," IEEE Trans. Image Process. **22** (9), 3703–3714 (2013). doi 10.1109/TIP.2013.2270086.
17. A. Karaali and C. R. Jung, "Adaptive Scale Selection for Multiresolution Defocus Blur Estimation," in *Proc. 2014 IEEE Int. Conf. on Image Processing (ICIP), Paris, France, October 27–30, 2014* (IEEE Press, New York, 2014), pp. 4597–4601. doi 10.1109/ICIP.2014.7025932.
18. A. Karaali and C. R. Jung, "Edge-Based Defocus Blur Estimation with Adaptive Scale Selection," IEEE Trans. Image Process. **27** (3), 1126–1137 (2018). doi 10.1109/TIP.2017.2771563.
19. A. Chakrabarti, T. Zickler, and W. T. Freeman, "Analyzing Spatially-Varying Blur," in *Proc. 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR) San Francisco, CA, USA, June 13–18, 2010* (IEEE Press, New York, 2010), pp. 2512–2519. doi 10.1109/CVPR.2010.5539954.
20. X. Zhu, S. Cohen, S. Schiller, and P. Milanfar, "Estimating Spatially Varying Defocus Blur from a Single Image," IEEE Trans. Image Process. **22** (12), 4879–4891 (2013). doi 10.1109/TIP.2013.2279316.
21. L. D'Andrès, J. Salvador, A. Kochale, and S. Süsstrunk, "Non-Parametric Blur Map Regression for Depth of Field Extension," IEEE Trans. Image Process. **25** (4), 1660–1673 (2016). doi 10.1109/TIP.2016.2526907.
22. S. A. Golestaneh and L. J. Karam, "Spatially-Varying Blur Detection Based on Multiscale Fused and Sorted Transform Coefficients of Gradient Magnitudes," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA, July 21–26, 2017* (IEEE Press, New York, 2017), pp. 596–605. doi 10.1109/CVPR.2017.71.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

23. N. D. Narvekar and L. J. Karam, "A No-Reference Image Blur Metric Based on the Cumulative Probability of Blur Detection (CPBD)," IEEE Trans. Image Process. **20** (9), 2678–2683 (2011). doi 10.1109/TIP.2011.2131660.
24. J. Kumar, F. Chen, and D. Doermann, "Sharpness Estimation for Document and Scene Images," in *Proc. 21st Int. Conf. on Pattern Recognition (ICPR), Tsukuba, Japan, 2012*, pp. 3292–3295.
25. K. Zeng, Y. Wang, J. Mao, et al., "A Local Metric for Defocus Blur Detection Based on CNN Feature Learning," IEEE Trans. Image Process. **28** (5), 2107–2115 (2019). doi 10.1109/TIP.2018.2881830.
26. J. Park, Y.-W. Tai, D. Cho, and I. So Kweon, "A Unified Approach of Multi-Scale Deep and Hand-Crafted Features for Defocus Estimation," https://arxiv.org/abs/1704.08992. Cited August 28, 2025.
27. J. Lee, S. Lee, S. Cho, and S. Lee, "Deep Defocus Map Estimation Using Domain Adaptation," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 15–20, 2019* (IEEE Press, New York, 2019), pp. 12214–12222. doi 10.1109/CVPR.2019.01250.
28. C. Tang, X. Liu, X. Zhu, et al., "R^{2}MRF: Defocus Blur Detection via Recurrently Refining Multi-Scale Residual Features," in *Proc. AAAI Conf. on Artificial Intelligence*, Vol. 34 (07), pp. 12063–12070 (2020). doi 10.1609/aaai.v34i07.6884.
29. X. Cun and C.-M. Pun, "Defocus Blur Detection via Depth Distillation," in *Proc. European Conf. on Computer Vision (ECCV)*, pp. 747–763 (2020). doi 10.48550/arXiv.2007.08113. doi 10.1007/978-3-030-58601-0_44. https://arxiv.org/abs/2007.08113. Cited August 30, 2025.
30. A. Karaali, N. Harte, and C. R. Jung, "Deep Multi-Scale Feature Learning for Defocus Blur Estimation," IEEE Trans. Image Process. **31**, 1097–1106 (2022). doi 10.1109/TIP.2021.3139243.
31. H. Li, W. Qian, J. Cao, et al., "Multi-Interactive Enhanced for Defocus Blur Estimation," IEEE Trans. Comput. Imaging **10**, 640–652 (2024). doi 10.1109/TCI.2024.3354427.
32. Z. Zhao, H. Yang, P. Liu, et al., "Defocus Blur Detection via Adaptive Cross-Level Feature Fusion and Refinement," Vis. Comput. **40** (11), 8141–8153 (2024). doi 10.1007/s00371-023-03229-7.
33. S. Winkler, "Efficient Measurement of Stereoscopic 3D Video Content Issues," in *Proc. SPIE 9016, Image Quality and System Performance XI*, page 90160Q (2014). doi 10.1117/12.2042211.
34. Q. Dong, T. Zhou, Z. Guo, and J. Xiao, "A Stereo Camera Distortion Detecting Method for 3DTV Video Quality Assessment," in *Proc. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC), Kaohsiung, Taiwan, October 29–November 1, 2013*, (IEEE Press, New York, 2014), pp. 1–4. doi 10.1109/APSIPA.2013.6694209.
35. F. Devernay, S. Pujades, and Vijay Ch.A.V, "Focus Mismatch Detection in Stereoscopic Content," in *Proc. SPIE* **8288**. *Stereoscopic Displays and Applications XXIII*, page 82880E (2012). doi 10.1117/12.906209.
36. M. Liu and K. Müller, "Automatic Analysis of Sharpness Mismatch between Stereoscopic Views for Stereo 3D Videos," in *Proc. 2014 Int. Conf. on 3D Imaging (IC3D), Liege, Belgium, December 9–10, 2014*, (IEEE Press, New York, 2015), pp. 1–6. doi 10.1109/IC3D.2014.7032572.
37. D. Vatolin and A. Bokov, "Sharpness Mismatch and 6 Other Stereoscopic Artifacts Measured on 10 Chinese S3D Movies," J. Electron. Imaging **2017** (5), 137–144 (2017). doi 10.2352/ISSN.2470-1173.2017.5.SDA-340.
38. D. Vatolin, A. Bokov, M. Erofeev, and V. Napadovsky, "Trends in S3D-Movie Quality Evaluated on 105 Films Using 10 Metrics," J. Electron. Imaging **2016** (5), 1–10 (2016). doi 10.2352/ISSN.2470-1173.2016.5.SDA-439.
39. C. Doutre, M. T. Pourazad, A. Tourapis, et al., "Correcting Unsynchronized Zoom in 3D Video," in *Proc. 2010 IEEE Int. Symposium on Circuits and Systems (ISCAS), Paris, France, May 30–June 2, 2010* (IEEE Press, New York, 2010), pp. 3244–3247. doi 10.1109/ISCAS.2010.5537923.
40. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," Int. J. Comput. Vis. **60** (2), 91–110 (2004). doi 10.1023/B:VISI.0000029664.99615.94.
41. I. E. Pekkucuksen, A. U. Batur, and B. Zhang, "A Real-Time Misalignment Correction Algorithm for Stereoscopic 3D Cameras," in *Proc. SPIE 8288, Stereoscopic Displays and Applications XXIII*, page 82880J (2012). doi 10.1117/12.906902.
42. A. Voronov, A. Borisov, and D. Vatolin, "System for Automatic Detection of Distorted Scenes in Stereo Video," in *Proc. Sixth Int. Workshop on Video Processing and Quality Metrics (VPQM)*, 2012.
43. M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Commun. ACM **24** (6), 381–395 (1981). doi 10.1145/358669.358692.
44. E. Brachmann, A. Krull, S. Nowozin, et al., "DSAC — Differentiable RANSAC for Camera Localization," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21–26, 2017* (IEEE Press, New York, 2017), pp. 2492–2500. doi 10.1109/CVPR.2017.267.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

45. E. Brachmann and C. Rother, "Neural-Guided RANSAC: Learning Where to Sample Model Hypotheses," in *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV), Seoul, Korea (South), October 27, 2019* (IEEE Press, New York, 2019), pp. 4321–4330. doi 10.1109/ICCV.2019.00442.
46. F. Kluger and B. Rosenhahn, "PARSAC: Accelerating Robust Multi-Model Fitting with Parallel Sample Consensus," in *Proc. AAAI Conf. on Artificial Intelligence*, **38** (3), 2804–2812 (2024). doi 10.1609/aaai.v38i3.28060.
47. J. Chen, Y. Gu, and L. Luo, "Learning to Find Good Correspondences Based on Global and Local Attention Mechanism," in *Proc. 2021 China Automation Congress (CAC), Beijing, China, October 22–24, 2021* (IEEE Press, New York, 2022), pp. 2174–2178. doi 10.1109/CAC53003.2021.9728340.
48. W. Sun, W. Jiang, E. Trulls, et al., "ACNe: Attentive Context Normalization for Robust Permutation-Equivariant Learning," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, June 13–19, 2020* (IEEE Press, New York, 2020), pp. 11283–11292. doi 10.1109/CVPR42600.2020.01130.
49. I. Rocco, R. Arandjelović, and J. Sivic, "Convolutional Neural Network Architecture for Geometric Matching," IEEE Trans. Pattern Anal. Mach. Intell. **41** (11), 2553–2567 (2019). doi 10.1109/TPAMI.2018.2865351.
50. I. Rocco, R. Arandjelović, and J. Sivic, "End-to-End Weakly-Supervised Semantic Alignment," in *Proc. 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, June 18–23, 2018* (IEEE Press, New York, 2018), pp. 6917–6925. doi 10.1109/CVPR.2018.00723.
51. J. Lee, C. Jung, C. Kim, and A. Said, "Content-Based Pseudoscopic View Detection," J. Sign. Process. Syst. **68** (2), 261–271 (2012). doi 10.1007/s11265-011-0608-8.
52. G. Palou and P. Salembier, "Monocular Depth Ordering Using T-Junctions and Convexity Occlusion Cues," IEEE Trans. Image Process. **22** (5), 1926–1939 (2013). doi 10.1109/TIP.2013.2240002.
53. H. Lee, C. Jung, and C. Kim, "Depth Map Estimation Based on Geometric Scene Categorization," in *Proc. 19th Korea–Japan Joint Workshop on Frontiers of Computer Vision (FCV), Incheon, Korea (South), January 30, 2013*, (IEEE Press, New York, 2013), pp. 170–173. doi 10.1109/FCV.2013.6485482.
54. D. Hoiem, A. N. Stein, A. A. Efros, and M. Hebert, "Recovering Occlusion Boundaries from a Single Image," in *Proc 2007 IEEE 11th Int. Conf. on Computer Vision, Rio de Janeiro, Brazil, October 14–21, 2007* (IEEE Press, New York, 2007), pp. 1–8. doi 10.1109/ICCV.2007.4408985.
55. H. Fu, M. Gong, C. Wang, et al., "Deep Ordinal Regression Network for Monocular Depth Estimation," doi 10.1109/CVPR.2018.00214. https://arxiv.org/abs/1806.02446. Cited August 29, 2025.
56. C. Godard, O. Mac Aodha, M. Firman, and G. Brostow, "Digging into Self-Supervised Monocular Depth Estimation," in *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV), Seoul, Korea (South), October 27–November 2 2019* (IEEE Press, New York, 2019), pp. 3827–3837. doi 10.1109/ICCV.2019.00393.
57. R. Ranftl, K. Lasinger, D. Hafner, et al., "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-Shot Cross-Dataset Transfer," IEEE Trans. Pattern Anal. Mach. Intell. **44** (3), 1623–1637 (2022). doi 10.1109/TPAMI.2020.3019967.
58. W. Yin, J. Zhang, O. Wang, et al., "Learning to Recover 3D Scene Shape from a Single Image," in *Proc. 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, June 20–25, 2021* (IEEE Press, New York, 2021), pp. 204–213. doi 10.1109/CVPR46437.2021.00027.
59. L. Yang, B. Kang, Z. Huang, et al., "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, June 16–22, 2024* (IEEE Press, New York, 2024), pp. 10371–10381. doi 10.1109/CVPR52733.2024.00987.
60. M. Knee, "Getting Machines to Watch 3D for You," SMPTE Motion Imaging J. **121** (3), 52–58 (2012). doi 10.5594/j18162.
61. J. Bouchard, Y. Nazzar, and J. J. Clark, "Half-Occluded Regions and Detection of Pseudoscopy," in *Proc. 2015 Int. Conf. on 3D Vision (3DV), Lyon, France, October 19–22, 2015* (IEEE Press, New York, 2015), pp. 215–223. doi 10.1109/3DV.2015.32.
62. A. Shestov, A. Voronov, and D. Vatolin, "Detection of Swapped Views in Stereo Image," in *Proc. 22nd GraphiCon Int. Conf. on Computer Graphics and Vision*, pp. 23–27 (2012).
63. K. Simonyan, S. Grishin, D. Vatolin, and D. Popov, "Fast Video Super-Resolution via Classification," in *Proc. 2008 15th IEEE Int. Conf. on Image Processing (ICIP), San Diego, CA, USA, October 12–15, 2008* (IEEE Press, New York, 2008), pp. 349–352. doi 10.1109/ICIP.2008.4711763.
64. G. Egnal and R. P. Wildes, "Detecting Binocular Half-Occlusions: Empirical Comparisons of Five Approaches," IEEE Trans. Pattern Anal. Mach. Intell. **24** (8), 1127–1133 (2002). doi 10.1109/TPAMI.2002.1023808.
65. D. Fourure, R. Emonet, E. Fromont, et al., "Residual Conv-Deconv Grid Network for Semantic Segmentation," https://arxiv.org/abs/1707.07958. Cited August 29, 2025.

2025, **26** (3), 340–365. doi 10.26089/NumMet.v26r323

66. K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," in *Proc. 2015 IEEE Int. Conf. on Computer Vision (ICCV), Santiago, Chile, December 7–13, 2015* (IEEE Press, New York, 2016), pp. 1026–1034. doi 10.1109/ICCV.2015.123.
67. D. Min, S. Choi, J. Lu, et al., "Fast Global Image Smoothing Based on Weighted Least Squares," IEEE Trans. Image Process. **23** (12), 5638–5653 (2014). doi 10.1109/TIP.2014.2366600.
68. X. Glorot and Y. Bengio, "Understanding the Difficulty of Training Deep Feedforward Neural Networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 249–256 (2010).
69. D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," https://arxiv.org/pdf/1412.6980. Cited August 29, 2025.
70. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 27–30, 2016* (IEEE Press, New York, 2016), pp. 770–778. doi 10.1109/CVPR.2016.90.
71. S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," https://arxiv.org/abs/1502.03167. Cited August 29, 2025.
72. D. S. Vatolin and S. V. Lavrushkin, "Investigating and Predicting the Perceptibility Protect of Channel Mismatch in Stereoscopic Video," Vestn. Mosk. Univ., Ser. 15: Vychisl. Mat. Kibern., No. 4, 40–46 (2016) [Moscow Univ. Comput. Math. Cybern. **40** (4), 185–191 (2016)]. doi 10.3103/S0278641916040075.
73. D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A Naturalistic Open Source Movie for Optical Flow Evaluation," in *Lecture Notes in Computer Science* (Springer, Berlin, 2012), Vol. 7577, pp. 611–625. doi 10.1007/978-3-642-33783-3_44.

Received	*Accepted*	*Published*
July 17, 2025	*August 16, 2025*	*September 20, 2025*

Information about the author

Sergey V. Lavrushkin — Senior Researcher; 1) Lomonosov Moscow State University, Institute for Artificial Intelligence, Lomonosovsky Prospekt, 27, building 1, 119234, Moscow, Russia; 2) Ivannikov Institute for System Programming of RAS, Research Center for Trusted Artificial Intelligence, Alexander Solzhenitsyn ulitsa, 25, 109004, Moscow, Russia.