

Анализ видеоданных беспилотного летательного аппарата на основе индекса структурного сходства

П.А. Ляхов^{1,2}, А.Р. Оразаев¹

¹ Северо-Кавказский федеральный университет, 355017, Россия, г. Ставрополь, ул. Пушкина, д. 1;

² Северо-Кавказский центр математических исследований, 355017, Россия, г. Ставрополь, ул. Пушкина, д. 1

Аннотация

В статье предложена метрика для анализа видеоданных, зарегистрированных беспилотным летательным аппаратом, использующая для оценки индекс структурного сходства. Метрика заключается в сравнении кадров на предмет сходства яркости, контраста и структуры в пикселях и последующей оценке состояния кадра видео. Был проведен сравнительный анализ предложенной и современной метрики. Исследование включало моделирование на аналоговых и цифровых видео с разной частотой кадров. Результаты показали, что разработанная метрика успешно выявляет задержки, искажения кадров и динамические изменения в сцене видео. Предлагаемая метрика может найти широкое применение в прикладных областях использования беспилотных летательных аппаратов: строительстве, сельском хозяйстве, геологии и картографии.

Ключевые слова: индекс структурного сходства, цифровая обработка изображений, оценка качества изображений, беспилотный летательный аппарат, анализ видеоданных.

Цитирование: Ляхов, П.А. Анализ видеоданных беспилотного летательного аппарата на основе индекса структурного сходства / П.А. Ляхов, А.Р. Оразаев // Компьютерная оптика. – 2025. – Т. 49, № 4. – С. 624–633. – DOI: 10.18287/2412-6179-CO-1569.

Citation: Lyakhov PA, Orazaev AR. Analysis of video data from an unmanned aerial vehicle based on a structural similarity index. Computer Optics 2025; 49(4): 624–633. DOI: 10.18287/2412-6179-CO-1569.

Введение

Регистрация фото и видео беспилотными летательными аппаратами (БПЛА) используется в различных прикладных областях, в том числе для создания высококачественных фотоснимков и трёхмерных моделей местности [1], для обновления картографических данных [2]. Это полезный инструмент для планирования строительства [3], мониторинга изменений в окружающей среде и поддержки навигации и автомобильного транспорта [4]. Также БПЛА применяется в сельском хозяйстве, например, при мониторинге созревания и уборки урожая [5].

Существуют потенциальные проблемы при передаче видео БПЛА: задержка, возникновение шума и снижение качества видео. Задержка – одна из основных проблем, особенно при использовании беспроводных технологий передачи данных. Задержка может привести к ошибке пилота или автономной системы, что критически опасно во время полета [6].

При передаче видео качество изображения может снижаться из-за сжатия данных, для экономии пропускной способности, особенно при высоком разрешении видео, если доступная пропускная способность ограничена или разделяется с другими важными данными, такими как управление полетом, особенно с использованием средств виртуальной и дополненной реальности [7]. Это может усложнить визуальное восприятие для пилота или анализирующих систем.

Задержка, возникновение шума и снижение качества видео могут быть вызваны интерференцией и

потерей сигнала. Под «задержкой» понимается случай, при котором один или несколько кадров видео отображаются дольше, чем предусмотрено частотой кадров видео. Проблемы с интерференцией встречаются в густо застроенных районах или в условиях сильного электромагнитного шума. Потеря сигнала также может произойти из-за наличия физических препятствий между БПЛА и его наземной станцией или из-за выхода за пределы дальности передачи [8].

Запись и трансляция видео в БПЛА может осуществляться в форме аналоговой или цифровой передачи данных. Основные различия между ними заключаются в формате представления данных. В случае аналоговой передачи видео и других данных информация представлена непрерывными изменениями физических параметров, таких как амплитуда и частота сигнала [9]. Например, видеосигнал может быть представлен в виде непрерывной аналоговой волны, а аналоговые датчики могут предоставлять аналоговые значения физических величин, таких как температура или давление. В цифровой передаче данные представлены дискретными значениями в двоичной битовой форме (0 и 1). Видеосигнал кодируется в виде чисел, которые представляют пиксели изображения [10]. В контексте устойчивости к шуму аналоговые сигналы более подвержены помехам и искажениям, в то время как цифровые сигналы обычно более устойчивы к помехам и могут быть восстановлены с большей точностью, поскольку использование двоичного кодирования позволяет обнаруживать и исправлять ошибки. Оба подхода имеют свои преимущества и недостатки. Выбор между аналоговой и цифровой пе-

редачей данных в БПЛА зависит от конкретной задачи, а также от условий окружающей среды и потенциальных помех [11].

В данной работе мы предлагаем новую метрику для анализа видеоданных БПЛА на основе индекса структурного сходства. Мы продемонстрируем, что разработанная метрика успешно выявляет задержки, искажения кадров и динамические изменения в сцене видео. Таким образом, целью данной работы является разработка новой метрики для оценки анализа видеоданных беспилотного летательного аппарата, которая не требует больших вычислительных затрат и может быть использована операторами БПЛА.

Наш вклад в область цифровой обработки видеоданных БПЛА описывается следующим образом.

1. Предложена метрика для анализа потокового видео, регистрируемого при помощи БПЛА, который основан на отношении характеристик индекса структурного сходства (SSIM) между кадрами видео.
2. Проведено моделирование расчета метрики для анализа потокового видео, регистрируемого при помощи БПЛА. Предложен алгоритм для применения разработанной метрики. Продемонстрировано, что предложенная метрика обнаруживает задержку, фиксирует искажения и определяет динамику в видео.
3. Выполнено сравнение предложенного и известного подходов. Сделаны выводы и даны практические рекомендации.

Основное содержание статьи организовано следующим образом. Предварительные сведения о метриках для оценки качества изображений приведены в параграфе 1. В параграфе 2 мы представляем предлагаемую метрику. В параграфе 3 продемонстрировано сравнение с известными аналогами, а также представлены экспериментальные результаты работы предложенной метрики.

1. Предварительные сведения

В области задачи анализа видеоданных БПЛА существуют различные методы. В работе [12] предложен метод глубокого мультимодального обучения для оценки эстетического качества видео БПЛА. Нейронная сеть состоит из модулей пространственного внешнего вида, движения камеры БПЛА и структуры сцены. Метод выполняет суждение о том, было ли видео с БПЛА снято профессиональными фотографами или любителями, а также классифицирует тип сцены. В работе [13] предложен метод, включающий систематическое построение онтологий для поддержки методов компьютерного зрения в анализе видеосцен. В контексте анализа видео онтологическая поддержка позволяет кодировать результаты анализа видео, такие как отслеживание объектов, классификация целей и событий в форме онтологических утверждений. С помощью семантического вывода новые утверждения, выведенные на основе ра-

нее сгенерированных знаний, позволяют получить общее представление о содержании видео. Данный метод обеспечивает семантическую поддержку обнаружения объектов и описания сценариев в сочетании с методами машинного и глубокого обучения. Глубокие нейросетевые методы субъективны, интерпретировать и объяснить их решения бывает сложно, что создает проблемы с доверием к результатам. Также нейросетевые методы чувствительны к качеству входных данных, а шум, артефакты и искажения в видео могут существенно повлиять на точность анализа.

В работе [14] авторы оценивают шум сенсора на основе локально адаптивного дискретного косинусного преобразования, затем коррелируют остаток шума исследуемого кадра n с шумом сенсора и с остатком шума предыдущего кадра для выявления цифрового монтажа. В работе [15] проведен анализ эффективности системы обработки видеоконтента с подвижных камер, основанной на методах обнаружения и отслеживания движущихся объектов. Система может достичь сокращения данных до 10 000 раз без значительной потери событий интереса. Ранее работ по анализу видеоданных на предмет задержек и помех в видео, а также по отслеживанию изменения динамики в видео не проводилось. Предполагается, что данный анализ можно провести с помощью методов оценки качества изображений и видео. Оценка динамики геометрии сцены требует серьезных вычислительных ресурсов, что является серьезным практическим недостатком при использовании в мобильных устройствах, обеспечивающих полет БПЛА. Это связано с критической зависимостью от заряда аккумулятора в указанных устройствах.

В отличие от методов на основе машинного обучения и оценки динамики геометрии, метрики оценки качества изображений позволяют проводить анализ с меньшими затратами энергии. Использование метрик оценки качества изображений обеспечивает не только более быстрое оценивание качества видео, но и способствует оптимизации энергопотребления, что является критически важным для современных технологий обработки видео в условиях ограниченных ресурсов, в том числе управление полетом БПЛА.

В настоящее время известны несколько метрик оценки качества видео: полностью связанные метрики оценки качества (full reference, FR) и метрики оценки качества без эталона (no-reference, NR) [16]. FR-метрики сравнивают оцениваемое видео с эталонным (оригинальным) видео. NR-метрики оценивают качество видео без доступа к эталонному видео, обычно с использованием статистических моделей или машинного обучения для предсказания качества. NR-метрики в данной задаче не способны обнаруживать задержки и определять искажения на кадрах видео, но при этом способны определять размытие в кадре.

Индекс структурного сходства (SSIM) является FR-метрикой оценки качества изображений. Другими

популярными FR-метриками являются пиковое отношение сигнала к шуму (PSNR) [17], а также методы на основе среднеквадратичной ошибки (MSE). Метрика Visual Information Fidelity (VIF) применяется для оценки качества изображений и использует меры информационной достоверности для оценки сходства между исходным и искаженным изображениями. VIF основан на модели смеси гауссовских масштабов, которая вычисляет информационную достоверность изображений на основе статистических характеристик их вейвлет-поддиапазонов на разных уровнях [18]. Метрика Visual Multimethod Assessment Fusion (VMAF) применяется для оценки качества видео и использует объединение различных методов и метрик оценки для предсказания восприятия качества видео человеком с помощью метода машинного обучения Support Vector Regressor (SVR) [19]. VMAF широко используется в области стриминга и сжатия видео, для оценки качества видеоконтента. Он представляет собой современную метрику, которая помогает оптимизировать процессы кодирования и улучшать восприятие качества видео со стороны зрителей [20]. Недостатками метрик на основе машинного обучения являются высокая алгоритмическая сложность, а также весьма затруднительная интерпретируемость и объяснимость.

2. Предлагаемая метрика для анализа видеоданных беспилотного летательного аппарата на основе индекса структурного сходства

В задаче оценки качества изображений с использованием FR-метрик сравнивают искаженные изображения с эталонными неискаженными. Мы предлагаем выполнять оценку метрик между текущим и последующим кадрами в задаче анализа видеоданных БПЛА. Предполагается, что при достаточно высокой частоте кадров разница между кадрами видео будет невысокая. Следовательно, чем выше частота кадров, тем выше будет оценка FR-метрики. Метрика SSIM не использует методы машинного обучения, что снижает вычислительную сложность, а также снижает элемент случайности. Также метрика SSIM не требует оценки группы пользователей и в отличие от метрики PSNR учитывает особенности яркости, контрастности и структуры. Примером оценки группы пользователей является проект LIVE (в сотрудничестве с департаментом психологии Техасского университета в Остине), где был проведен масштабный эксперимент по получению оценок от испытуемых за ряд изображений, измененных различными типами искажений. Эти изображения были получены в рамках исследовательского проекта по сопоставлению и распознаванию общих форм. Каждому наблюдателю в случайном порядке показывали изображения. Наблюдателям было предложено представить свое восприятие качества по непрерывной линей-

ной шкале, которая была разделена на пять равных областей, отмеченных оценками «Очень плохо», «Плохо», «Удовлетворительно», «Хорошо» и «Отлично».

SSIM – это метрика качества изображения, которая в основном используется для оценки сходства между двумя изображениями. SSIM учитывает сходство воспринимаемых яркостей, контраста и структуры в пикселях изображений [21]. Под структурой понимается организация текстур и деталей на изображениях. Общий индекс представляет собой мультипликативную комбинацию трех составляющих. SSIM генерирует значение от 0 до 1, где 1 указывает на полное сходство между изображениями, а 0 – на полное несходство. Чем ближе значение к 1, тем более похожи изображения. Преимущества SSIM включают учет особенностей восприятия человеческого глаза, возможность обнаружения как глобальных, так и локальных искажений, а также устойчивость к некоторым видам шума.

Индекс SSIM для двух изображений x, y рассчитывается по формуле:

$$S = SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (1)$$

где

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3},$$

где μ_x, μ_y – локальные средние значения, σ_x, σ_y – стандартные отклонения, σ_{xy} – перекрестная ковариация для изображений x, y ; l, c, s – сходство по яркости, контрасту и структуре соответственно; α, β, γ – коэффициенты влияния l, c, s . Если $\alpha = \beta = \gamma = 1$ (по умолчанию для экспонент) и $C_3 = C_2/2$ (общепринятый выбор C_3 по умолчанию), то формула вычисления SSIM упрощается до

$$S = SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (2)$$

Если $S = 1$, то предполагается, что во время фиксации или передачи видеоданных БПЛА произошла задержка. Если значение S снижается, то можно сделать вывод о повышении динамики сцены в видео. Для обнаружения искажений в кадре видео мы предлагаем вычислять отношения SSIM между кадрами видео:

$$\delta_{i-1}^{(SSIM)} = \frac{SSIM(k_{i-1}, k_i)}{SSIM(k_{i-2}, k_{i-1})}. \quad (3)$$

Подставляя (2) в (3), получим

$$\delta_{i-1}^{(SSIM)} = \frac{(\mu_{k_{i-2}}^2 + \mu_{k_{i-1}}^2 + C_1)(\sigma_{k_{i-2}}^2 + \sigma_{k_{i-1}}^2 + C_2)(2\mu_{k_{i-1}}\mu_{k_i} + C_1)(2\sigma_{k_{i-1},k_i} + C_2)}{(\mu_{k_{i-1}}^2 + \mu_{k_i}^2 + C_1)(\sigma_{k_{i-1}}^2 + \sigma_{k_i}^2 + C_2)(2\mu_{k_{i-2}}\mu_{k_{i-1}} + C_1)(2\sigma_{k_{i-2},k_{i-1}} + C_2)} \quad (4)$$



Рис. 1. Диапазоны значений $\delta_{i-1}^{(SSIM)}$ для кадра видеоданных БПЛА

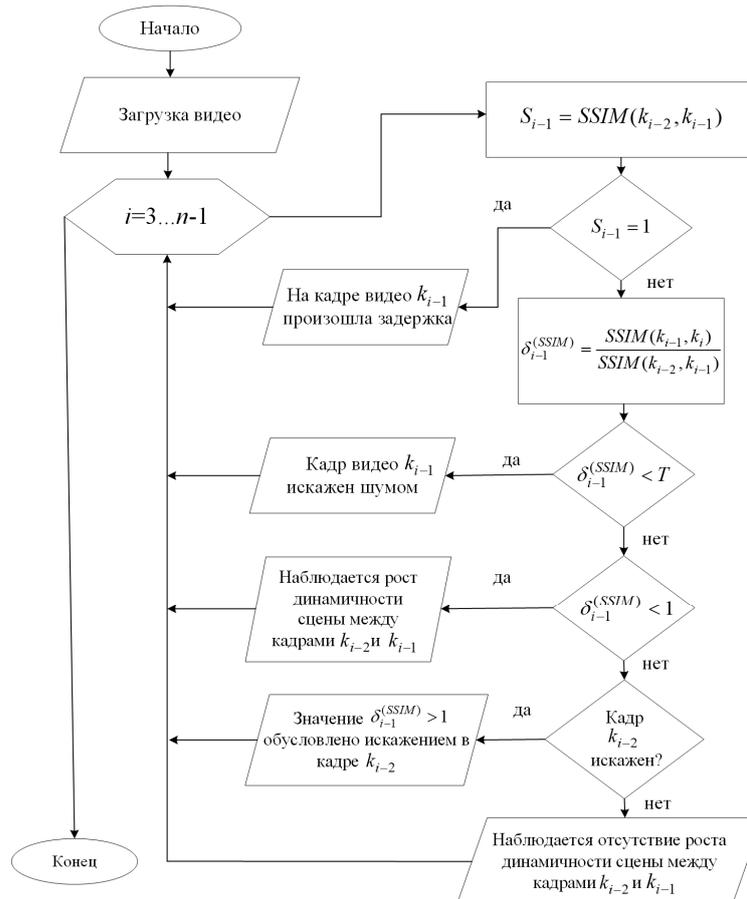


Рис. 2. Блок-схема применения разработанной метрики для анализа видеоданных БПЛА на основе индекса структурного сходства

Искаженные кадры видео БПЛА могут быть найдены с использованием порога T , задаваемого пользователем. Если $\delta_{i-1}^{(SSIM)} < T$, то можно сделать вывод о том, что кадр видео искажен шумом. Если $T < \delta_{i-1}^{(SSIM)} < 1$, то можно сделать вывод о том, что динамика в кадре увеличивается. Наглядное представление распределения значений $\delta_{i-1}^{(SSIM)}$ и последующих выводов о качестве кадра видео БПЛА изображено на рис. 1. В цифровой передаче данных качество видео выше, чем при аналоговой передаче, следовательно, значения $\delta_{i-1}^{(SSIM)}$ могут быть выше, что может потребовать более высоких значений порога T . Кроме того, на выбор величины порога T должна влиять частота кадров видео с БПЛА, так как в видео с более высокой частотой кадров будет наблюдаться менее динамичное изменение сцены.

На рис. 2 представлена блок-схема применения разработанной метрики для анализа видеоданных БПЛА на основе индекса структурного сходства. Для поступающего кадра видео вычисляется значение S_{i-1} по формуле (2). Если это значение равно 1, то делается вывод о том, что на данном кадре произошла задержка. Далее при помощи (4) можно последовательно зарегистрировать состояния, когда данный кадр искажен шумом либо возможное увеличение динамики в видео. В случае, когда предыдущий кадр видео оказался искажен, значение $\delta_{i-1}^{(SSIM)}$ для текущего кадра может оказаться больше 1, в таком случае данный кадр не будет являться искаженным. Данный алгоритм является итеративным и выполняется для каждого следующего поступающего кадра. При наличии

достаточных вычислительных ресурсов у оператора БПЛА предложенный алгоритм применения разработанной метрики может быть использован в режиме реального времени, что может быть востребовано во многих прикладных задачах беспилотной авиации. Для этого необходимо вычисление по формулам (2) и (4) за время, меньшее, чем время смены кадра, что вполне осуществимо при использовании современных контроллеров и микропроцессоров. Данное обстоятельство выгодно отличает предложенную метрику от метрик на основе машинного обучения.

Далее будет продемонстрирована работоспособность предложенной метрики, а также выполнено сравнение с другими возможными подходами к решению поставленной задачи поиска искажений и отслеживания динамики сцены в видеоданных БПЛА.

3. Экспериментальное моделирование анализа видеоданных БПЛА

Для проведения эксперимента и подтверждения работы метрики для анализа потокового видео, регистрируемого при помощи БПЛА, были использованы три видео [22]. Видео 1 зарегистрировано при помо-

щи аналогового устройства БПЛА и содержит 998 кадров с частотой 30 кадров в секунду и субъективно характеризуется нами наличием высокой динамики сцены (управление скоростным полетом с препятствиями малоразмерного БПЛА). Далее под аналоговым видео будет пониматься видео, зарегистрированное при помощи устройств аналоговой передачи данных. При записи видео оцифровывается и хранится в формате avi. Цифровое видео 2 содержит 400 кадров с частотой 30 кадров в секунду и субъективно характеризуется нами наличием высокой динамики сцены (скоростной полет с препятствиями малоразмерного БПЛА). Цифровое видео 3 содержит 1000 кадров с частотой 60 кадров в секунду и субъективно характеризуется нами наличием низкой динамики сцены (экспериментальное управление полетом авторов статьи на малоразмерном БПЛА). На кадры видео наложена информация об уровне сигнала, таймере и индикаторе записи, высоте, дальности и другая информация о БПЛА, называемая One Screen Display (OSD). В табл. 1 представлены сравнительные характеристики, а на рис. 3 приведены некоторые кадры видео, использованных в экспериментах.

Табл. 1. Сравнительные характеристики видео с БПЛА, использованных при моделировании

	Тип видео	Разрешение	Частота кадров, кадров в секунду	Количество кадров	Формат видео
Видео 1	аналоговое	720×x480	30	998	avi
Видео 2	цифровое	720×660	30	400	mp4
Видео 3	цифровое	1920×1080	60	1000	mp4



Рис. 3. Кадры видео с БПЛА, использованные при моделировании: а) видео 1; б) видео 2; в) видео 3

На рис. 4 продемонстрированы графики значений SSIM, рассчитанных по формуле (2) для каждого кадра трех видео. На рис. 4а видно, что значения SSIM между кадрами сильно изменяются и могут опускаться ниже значений 0,5, что говорит о низком качестве

видео при аналоговой передаче. На рис. 4б, в отличие от рис. 4в, значения SSIM находятся дальше от значения 1 из-за меньшей частоты кадров, а также более высокой динамики на видео. Видно, что даже при невысокой частоте кадров, высокой скорости и манев-

ренности БПЛА, значения S не опускаются ниже определенных пределов независимо от средней яркости сцены. При цифровой передаче видео значения SSIM между кадрами не опускаются ниже 0,5, даже при передаче видео, когда БПЛА движется на высокой скорости и регистрирует большое количество объектов в кадре, быстро сменяющих друг друга. На

271 кадре видео 3 происходит взлет БПЛА, что видно на рис. 4в в виде снижения значений SSIM между кадрами видео. Также на рис. 4а видно, что значение SSIM между кадрами 332 и 333 равно 1, следовательно, кадры идентичны и можно предположить, что на видео произошла задержка. На рис. 4б и в задержки обнаружены не были.

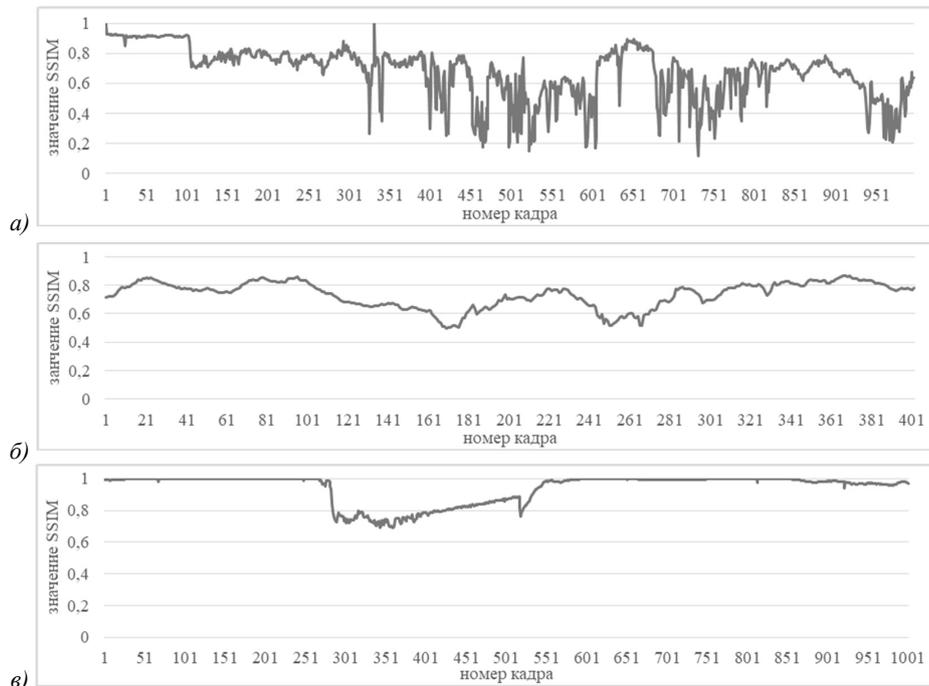


Рис. 4. Значения SSIM между кадрами видео с БПЛА: а) видео 1; б) видео 2; в) видео 3

Отношение $\delta^{(SSIM)}$ между кадрами видео, представленное в формуле (4), позволяет с большей точностью обнаружить кадры видео, которые сильно отличаются от предыдущих и последующих, предположительно за счет искажений на изображениях или потери цвета. Так как в данной статье рассматривается видеопоток, передаваемый с БПЛА, склеек в видео не предполагается. Первый кадр при склейке при помощи разработанной метрики будет определен как искаженный. На рис. 5 представлены значения $\delta^{(SSIM)}$.

Для оценки качества разработанной метрики была проверена возможность анализировать видеопоток с БПЛА при помощи метрики VMAF. Метрика VMAF была представлена Netflix в 2017 году и продемонстрировала хорошие результаты в задаче оценки качества изображений. Из недостатков VMAF можно отметить факт, что метрика основана на машинном обучении, причем база данных, на которых обучалась нейронная сеть, неизвестна, также вычисление данной метрики является сложной и энергозатратной операцией.

На рис. 6 продемонстрированы графики значений VMAF для каждого кадра видео из табл. 1. Поведение значений VMAF на графике 6а схоже со значениями разработанной метрики на рис. 4а. При этом метрика VMAF более чувствительна и сильно реагирует на

изменения между кадрами, что сказывается на увеличенной амплитуде графиков рис. 6 по сравнению с рис. 4. Слишком большая чувствительность является недостатком для задачи анализа видео БПЛА, так как метрика VMAF реагирует на изменения информации OSD и определяет кадр как искаженный.

Для наглядности работы метрики для анализа видеоданных беспилотного летательного аппарата на основе индекса структурного сходства в аспекте фиксирования искаженных кадров в видео 3 искусственным путем, гауссовым шумом и полярным импульсным шумом были искажены кадры под номерами 100, 200, 300, 400, 500, 600, 700, 800, 900. Кадры видео были зашумлены с нарастающей интенсивностью от 0,01 до 0,09 с шагом 0,01 плотности импульсного шума на изображениях и от 0,05 до 0,45 с шагом 0,05 плотности гауссова шума. Потерянная информация для таких кадров может быть существенна, но при этом при частоте 60 кадров в секунду во время просмотра видео пользователем практически незаметна. На рис. 8 продемонстрированы графики значений $\delta^{(SSIM)}$ и VMAF для каждого кадра видео 3 с искусственно искаженными импульсным шумом кадрами. На рис. 7а и б различимо видны изменения в поведении графика последовательности кадров видео в случае каждого искусственно зашумленного изображения. На рис. 7а видно,

что после каждого искаженного кадра видео значения $\delta^{(SSIM)}$ превышают единицу. На рис. 8 продемонстрированы графики значений $\delta^{(SSIM)}$ и VMAF для каждого кадра видео 3 с искусственно искаженными гауссовым

шумом кадрами. На рис. 8а и б также различимо видны изменения в поведении графика последовательности кадров видео в случае каждого искусственно зашумленного изображения.

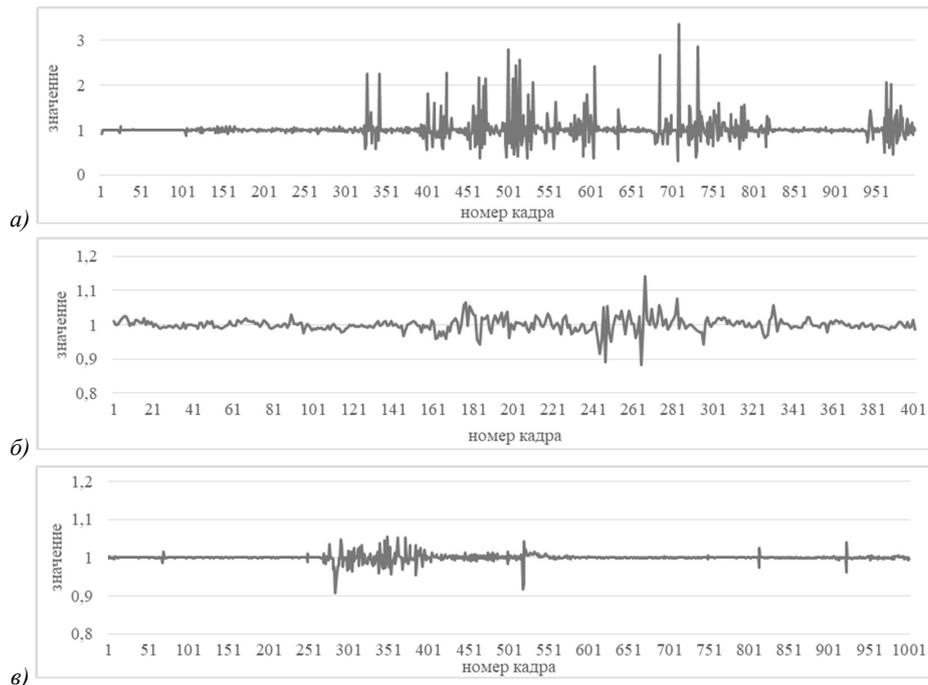


Рис. 5. Значения $\delta_{i-1}^{(SSIM)}$ между кадрами видео с БПЛА: а) видео 1; б) видео 2; в) видео 3

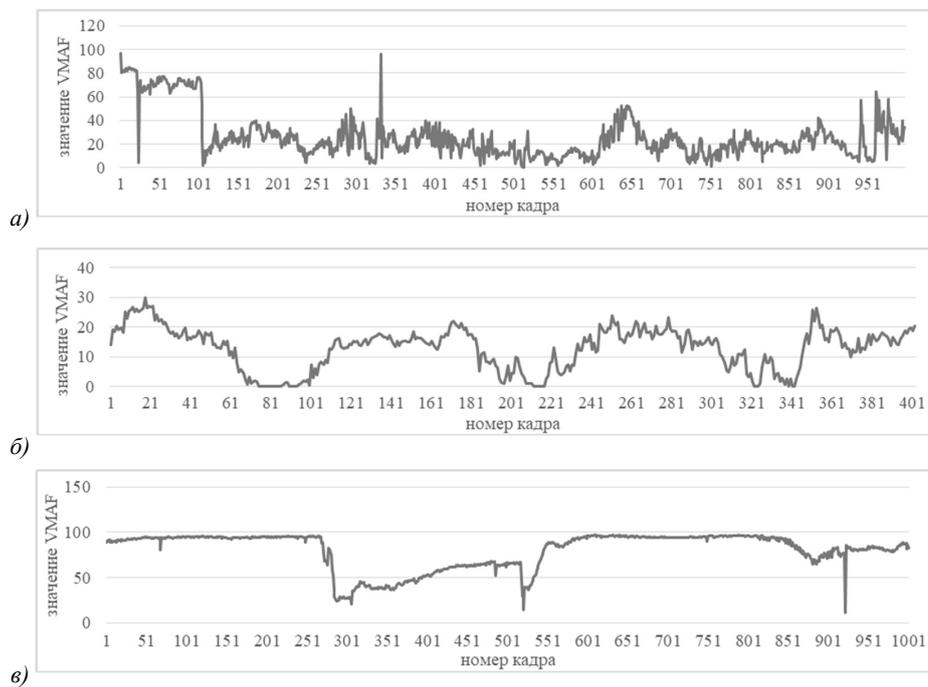


Рис. 6. Значения VMAF между кадрами видео с БПЛА: а) видео 1; б) видео 2; в) видео 3

Предлагаемая метрика при условии $\delta^{(SSIM)} < T$ позволяет принять решение о том, что кадр видео искажен. Выбор порога обусловлен классификацией кадров видео на искаженные и полезные. Анализ искусственно искаженного видео продемонстрировал оптимальность выбора порога $T_{\delta^{(SSIM)}} = 0,75$, что

видно из рис. 7 и 8. На этих рисунках видно, что выбор порога больше, чем 0,75, приведет к росту ложных срабатываний, при которых неискаженные кадры будут определяться как искаженные. Если выбрать порог меньше, чем 0,75, то кадры с низкой интенсивностью шума будут определяться как

неискаженные. Порог $T_{\delta(SSIM)}=0,75$ выявляет наличие импульсного шума с интенсивностью 0,03 и

выше и для гауссова шума с интенсивностью 0,05 и выше.

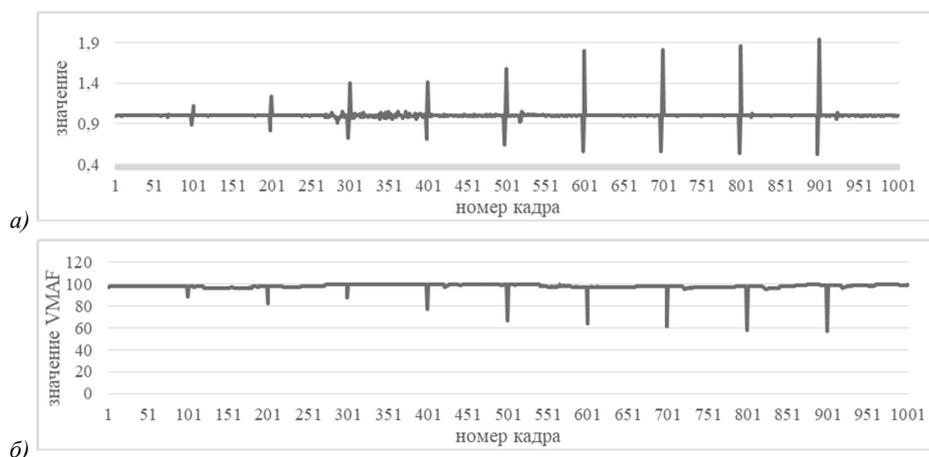


Рис. 7. Значения различных метрик для каждого кадра видео 3 с искусственно искаженными кадрами под номерами 100, 200, ..., 900 нарастающим импульсным шумом интенсивностью от 0,01 до 0,09 с шагом 0,01: а) метрика $\delta^{(SSIM)}$, б) метрика VMAF

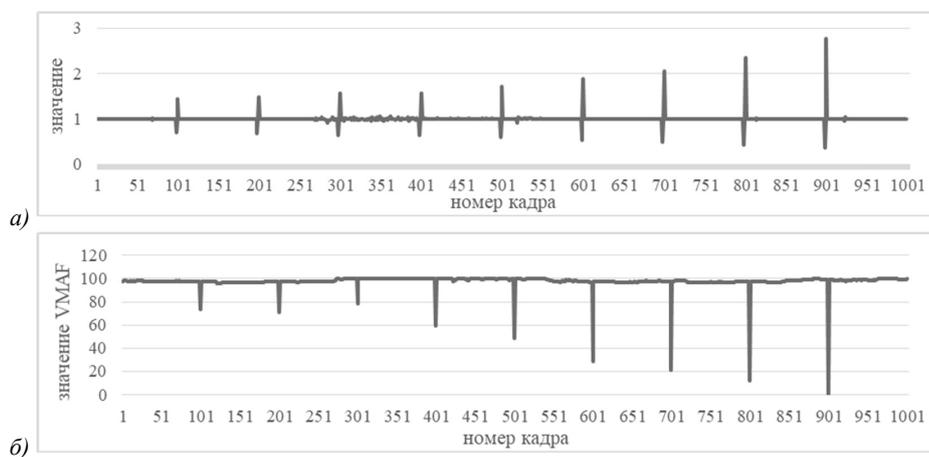


Рис. 8. Значения различных метрик для каждого кадра видео 3 с искусственно искаженными кадрами под номерами 100, 200, ..., 900 нарастающим гауссовым шумом интенсивностью от 0,05 до 0,45 с шагом 0,05: а) метрика $\delta^{(SSIM)}$, б) метрика VMAF



Рис. 9. Примеры низкокачественных кадров видео 1, обнаруженных разработанной метрикой оценки качества видео: а) кадр 421; б) кадр 502; в) кадр 752

На рис. 9 представлены некоторые изображения, которые разработанная метрика оценила как низкокачественные. Для разработанной метрики был выбран порог $T_{\delta(SSIM)}=0,75$. С выбором данных порогов разработанной метрикой было обнаружено 56 изображений. Все изображения, которые были определены метриками как искаженные, можно изучить в источнике [22].

Анализ экспериментальных результатов, представленных в данном параграфе, позволяет сделать вывод о

достаточно высоком качестве работы представленной метрики при относительно невысокой вычислительной сложности по сравнению с методами и метриками на основе машинного обучения. Применение разработанной метрики в интерфейсах операторов БПЛА позволит повысить качество работы за счет улучшения обратной связи с транспортным средством, а также потенциально снизить вероятность возникновения аварийных ситуаций. Для этого необходимо внедрение разработанной

метрики в системы управления БПЛА с целью дополнительного контроля состояния воздушного судна и качества сигнала.

Заключение

В работе предложена метрика для анализа видеоданных БПЛА на основе SSIM. Экспериментальное моделирование разработки на различных видео с БПЛА подтвердило высокое качество предложенной метрики. Результаты моделирования продемонстрировали, что разработанная метрика способна обнаруживать задержки в видео, искаженные кадры, а также отслеживать изменение динамичности сцены. Проведено сравнение с современной FR-метрикой оценки качества видео VMAF.

На примере метрики VMAF отмечено, что недостатками известных метрик и методов на основе машинного обучения являются высокая алгоритмическая сложность и энергопотребление, что особенно критично при использовании БПЛА.

Перспективными направлениями дальнейших исследований являются разработка методов и алгоритмов шумоподавления на кадрах видео с БПЛА в режиме реального времени, анализ типов шумовых воздействий на цифровую передачу видеоданных БПЛА и методов борьбы с ними, совершенствование методов применения БПЛА в прикладных областях: строительстве, сельском хозяйстве, геологии и картографии.

Благодарности

Исследование проведено при поддержке Российского научного фонда (проект № 23-71-10013).

References

[1] Jiménez-Jiménez SI, Ojeda-Bustamante W, Marcial-Pablo MDJ, Enciso J. Digital terrain models generated with low-cost UAV photogrammetry: Methodology and accuracy. *ISPRS Int J Geo Inf* 2021; 10(5): 285. DOI: 10.3390/ijgi10050285.

[2] Reinoso JF, Gonçalves JE, Pereira C, Bleninger T. Cartography for civil engineering projects: photogrammetry supported by unmanned aerial vehicles. *Iran J Sci Technol Trans Civ Eng* 2018; 42: 91-96. DOI: 10.1007/s40996-017-0076-x.

[3] Freimuth H, König M. Planning and executing construction inspections with unmanned aerial vehicles. *Autom Constr* 2018; 96: 540-553. DOI: 10.1016/j.autcon.2018.10.016.

[4] Kataev MY, Kartashov EY, Avdeenko VD. Method for detecting road defects using images obtained from unmanned aerial vehicles. *Computer Optics* 2023; 47(3): 464-473. DOI: 10.18287/2412-6179-CO-1209.

[5] Zubarev YuN, Fomin DS, Chashchin AN, Zabolotnova MV. Use of uncleaned aircraft in agriculture [In Russian]. *Bulletin of the Perm Federal Research Center* 2019; 2: 47-51. DOI: 10.7242/2658-705X/2019.2.5.

[6] Ageev AM, Bondarev VG, Protsenko VV. Justification of the choice of radiation sources for a computer vision system in the problem of automatic landing of unmanned aerial vehicles. *Computer Optics* 2022; 46(2): 239-245. DOI: 10.18287/2412-6179-CO-875.

[7] Huang D, Cui M, Zhang G, Chu X, Lin F. Bandwidth, power and trajectory optimization for UAV base station networks with backhaul and user QoS constraints. *IEEE Access* 2020; 8: 67625-67634. DOI: 10.1109/ACCESS.2020.2986075.

[8] Li B, Fei Z, Zhang Y. UAV communications for 5G and beyond: Recent advances and future trends. *IEEE Internet of Things J* 2018; 6(2): 2241-2263. DOI: 10.1109/JIOT.2018.2887086.

[9] Khawaja W, Guvenc I, Matolak DW, Fiebig U-C, Schneckenburger N. A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles. *IEEE Commun Surv Tutor* 2019; 21(3): 2361-2391. DOI: 10.1109/COMST.2019.2915069.

[10] Saponaro M, Capolupo A, Tarantino E, Fratino U. Comparative analysis of different UAV-based photogrammetric processes to improve product accuracies. In Book: Misra S, Gervasi O, Murgante B, Stankova E, Korkhov V, Torre C, Rocha AMAC, Taniar D, Apduhan BO, Tarantino E, eds. *Computational science and its applications – ICCSA 2019. 19th International Conference, Saint Petersburg, Russia, July 1–4, 2019, Proceedings, Part IV*. Cham, Switzerland: Springer Nature Switzerland AG; 2019: 225-238. DOI: 10.1007/978-3-030-24305-0_18.

[11] Shakhatreh H, Sawalmeh AH, Al-Fuqaha A, et al. Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges. *IEEE Access* 2019; 7: 48572-48634. DOI: 10.1109/ACCESS.2019.2909530.

[12] Kuang Q, Jin X, Zhao Q, Zhou B. Deep multimodality learning for UAV video aesthetic quality assessment. *IEEE Trans Multimedia* 2019; 22(10): 2623-2634. DOI: 10.1109/TMM.2019.2960656.

[13] Cavaliere D, Loia V, Senatore S. Towards an ontology design pattern for UAV video content analysis. *IEEE Access* 2019; 7: 105342-105353. DOI: 10.1109/ACCESS.2019.2932442.

[14] Fayyaz MA, Anjum A, Ziauddin S, Khan A, Sarfaraz A. An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues. *Multim Tools Appl* 2020; 79: 5767-5788. DOI: 10.1007/s11042-019-08236-2.

[15] Lin CC, Pankanti S, Ashour G, Porat D, Smith JR. Moving camera analytics: Emerging scenarios, challenges, and applications. *IBM J Res Dev* 2015; 59(2/3): 5:1-5:10. DOI: 10.1147/JRD.2015.239451.

[16] Wang Z, Bovik AC. Reduced- and no-reference image quality assessment. *IEEE Signal Process Mag* 2011; 28(6): 29-40. DOI: 10.1109/MSP.2011.942471.

[17] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. *2010 20th Int Conf on Pattern Recognition* 2010: 2366-2369. DOI: 10.1109/ICPR.2010.579.

[18] Sheikh HR, Bovik AC. Image information and visual quality. *IEEE Trans Image Process* 2006; 15(2): 430-444. DOI: 10.1109/TIP.2005.859378.

[19] Awad M, Khanna R. Support vector regression. In Book: Awad M, Khanna R. *Efficient learning machines. Theories, concepts, and applications for engineers and system designers*. Apress Media LLC; 2015: 67-80. DOI: 10.1007/978-1-4302-5990-9_4.

[20] Li Z, Aaron A, Katsavounidis I, Moorthy A, Manohara M. Toward a practical perceptual video quality metric. *Netflix TechBlog*. 2016. Source: <https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652>.

[21] Sara U, Akter M, Uddin MS. Image quality assessment through FSIM, SSIM, MSE and PSNR – a comparative study. *J Comput Commun* 2019; 7(3): 8-18. DOI: 10.4236/jcc.2019.73002.

[22] Bitbucket. CO_IQA. 2025. Source: <https://bitbucket.org/anzor_orazaev/co_iqa/src/main>.

Сведения об авторах

Ляхов Павел Алексеевич, 1988 года рождения, в 2009 году окончил Ставропольский государственный университет по специальности «Математика». Кандидат физико-математических наук, доцент, заведующий кафедрой математического моделирования факультета математики и компьютерных наук имени профессора Н.И. Червякова ФГАОУ ВО «Северо-Кавказский федеральный университет». Заведующий отделом модулярных вычислений и искусственного интеллекта регионального научно-образовательного математического центра «Северо-Кавказский центр математических исследований». Область научных интересов: цифровая обработка сигналов и изображений, искусственный интеллект, нейронные сети, модулярная арифметика, параллельные вычисления, высокопроизводительные вычисления, цифровые устройства, аппаратные ускорители.

E-mail: ljahov@mail.ru

Оразаев Анзор Русланович, 1995 года рождения, в 2023 году окончил аспирантуру по специальности «Информатика и вычислительная техника» в Северо-Кавказском федеральном университете, кандидат технических наук. Младший научный сотрудник департамента науки ФГАОУ ВО «Северо-Кавказский федеральный университет». Область научных интересов: цифровая обработка изображений, искусственный интеллект, адаптивная фильтрация сигналов. E-mail: arazaev@ncfu.ru

ГРНТИ: 28.23.15

Поступила в редакцию 28 мая 2024 г. Окончательный вариант – 08 ноября 2024 г.

Analysis of video data from an unmanned aerial vehicle based on a structural similarity index

P.A. Lyakhov^{1,2}, A.R. Oraziyev¹

¹North-Caucasus Federal University, 355017, Stavropol, Russia, Pushkin street, 1;

²North-Caucasus Center for Mathematical Research, 355017, Stavropol, Russia, Pushkin street, 1

Abstract

The article proposes a metric for the analysis of video data recorded by an unmanned aerial vehicle that uses a structural similarity index for evaluation. The metric consists in comparing frames in terms of brightness, contrast and pixel structure and a subsequent assessment of the video frame state. A comparative analysis of the proposed and currently employed metrics was carried out. The research included simulations on analog and digital video data at different frame rates. The results showed that the developed metric successfully detects delays, frame distortions and dynamic changes in a video scene. The proposed metric can find a wide range of applications of unmanned aerial vehicles in applied areas: construction, agriculture, geology and cartography.

Keywords: structural similarity index measure, digital image processing, image quality assessment, unmanned aerial vehicle, video data analysis.

Citation: Lyakhov PA, Oraziyev AR. Analysis of video data from an unmanned aerial vehicle based on a structural similarity index. *Computer Optics* 2025; 49(4): 624-633. DOI: 10.18287/2412-6179-CO-1569.

Acknowledgements: The work was financially supported by the Russian Science Foundation under project # 23-71-10013.

Author's information

Pavel Alekseyevich Lyakhov (b. 1988) graduated from Stavropol State University, specialty "Mathematics" in 2009. PhD in Physical and Mathematical Sciences. Head of the Department of Mathematical Modeling, North-Caucasus Federal University. Head of the Department of Modular Computing and Artificial Intelligence, regional scientific and educational mathematical center «North-Caucasus Center for Mathematical Research». Research interests are digital signal and image processing, artificial intelligence, neural networks, modular arithmetic, parallel computing, high-performance computing, digital circuits and hardware accelerators. E-mail: ljahov@mail.ru

Anzor Ruslanovich Oraziyev (b. 1995) PhD in Technical Sciences, junior researcher, Department of Science North-Caucasus Federal University. Research interests are digital image processing, artificial intelligence and adaptive filtering of signals. E-mail: aoraziyev@ncfu.ru

Received May 28, 2024. The final version – November 08, 2024.
