Методика тестирования конфигураций нейросетевых видеокодеков

А. А. Берёзкин

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича

berezkin.aa@sut.ru

Аннотация. Нейросетевые видеокодеки обеспечивают более высокую степень сжатия кадров по сравнению с классическими. Благодаря этому свойству они позволяют эффективно передавать видеопоток при управлении беспилотными воздушными судами от первого лица. В последние годы в научной литературе была представлена система динамических конфигураций нейросетевых видеокодеков на основе требований по пропускной способности. Настоящая работа посвящена методической составляющей проверки соответствия конфигураций заявленным требованиям на примере спутникового канала связи через искусственные спутники земли на геостационарной орбите. Проведены испытания на примере конфигураций классического нейросетевого кодека. В результате показано существенное отклонение результатов работы нейросетевого видеокодека на Полученные больших конфигурациях. результаты применимы при разработке адаптивной системы сжатия кадров видеопотока.

Ключевые слова: нейронная сеть; нейросетевой кодек; видеокодек; управление от первого лица; беспилотный летательный аппарат; спутниковый канал связи

I. Введение

При передаче видеопотока в сетях связи используются видеокодеки [1]. К стандартным видеокодекам относятся: AVC/H264 [2], HEVC/H265 (high efficiency video codec) [3], VCC/H266 (versatile video codec) [4], VP8 [5], VP9 [6], AV1 [7].

Видеокодеки используются беспилотных в авиационных системах (БАС) при необходимости управления беспилотными воздушными судами (БВС) от первого лица. При управлении от первого лица в отдалённых районах сети связи с большой пропускной способностью могут быть недоступны. В этом случае использование стандартных видеокодеков затруднено. Одним из перспективных решений данной проблемы видеокодеков. является применение нейросетевых Отличительной особенностью нейросетевых видеокодеков является применение нейронных сетей для преобразования кадров видеопотока в пиксельном пространстве во внутреннее латентное пространство признаков (ЛПП).

Пусть каждый кадр содержит *size* значений типа данных uint8, каждое из которых занимает 1 байт, где

А. А. Ченский

Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича

chenskii.aa@sut.ru

$size = height \times width \times channels$,

где *height* – высота кадра (в пикселях), width – ширина кадра (в пикселях), channels - количество каналов (для RGB *channels* = 3). И пусть после кодирования значения ЛПП преобразуются в тип данных uint8. Тогда применение нейронной сети вида «вариационный автокодировщик», например, модели KL-f4 к кадру видеопотока палитры RGB (red, green, blue - красный, зелёный, синий) приводит к уменьшению количества значений в 16 раз, а модели VQ-f16 - в 96 раз. Дальнейшее сжатие ЛПП стандартными алгоритмами сжатия данных обеспечивает малый размер сжатой бинарной последовательности и, соответственно, высокую степень сжатия кадра видеопотока.

II. КЛАССИЧЕСКИЙ НЕЙРОСЕТЕВОЙ КОДЕК

Классический нейросетевой кодек (видеокодек) был предложен в исследованиях [8, 9]. Он состоит из кодера и декодера. Кодер последовательно преобразует кадр видеопотока блоками: интерполяции (методом суперсэмплинга до разрешения 512×512); кодирования (вариационным автокодировщиком); квантования (преобразования значений в тип данных uint8) и сжатия. Декодер последовательно восстанавливает кадр из ЛПП блоками: декомпрессии, деквантования (восстановления float16-значений ЛПП); декодирования (вариационным автокодировщиком) и интерполяции (бикубическим методом до исходного разрешения).

Состав каждого из блоков, включая модель вариационного автокодировщика, зависит от конфигурации классического нейросетевого кодека. Множество конфигураций классического нейросетевого кодека необходимо для использования в сетях связи с разной пропускной способностью: чем меньше размер сжатой бинарной последовательности, тем меньшего качества восстанавливаемых кадров возможно добиться. Для смены конфигураций без перезапуска нейросетевого кодека предназначена система адаптивного сжатия кадров видеопотока [10].

В [8, 9] были предложены адаптивные конфигурации для классического нейросетевого кодека (табл. 1). Каждая адаптивная конфигурация предназначена для применения при разной полосе пропускания канала связи (*bitrate*). Значения kbps указаны при $FPS = 10(\kappa/c)$, где κ/c – кадр в секунду.

Конфигурации характеризуются: номером (#) и значениями метрик SSIM (structure similarity index

Научная статья подготовлена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1024032900105-1-2.2.4;2.2.5;2.2.6 в ЕГИСУ НИОКТР.

measure – мера структурного сходства) и *PSNR* (peak sound noise ratio – пиковое отношение сигнала к шуму) [11] и требуемой полосой пропускания *bitrate*.

#	SSIM	PSNR, дБ	<i>bitrate</i> , Кбит/с
1	0,82	29,94	35,2
2	0,84	30,20	74,4
3	0,85	30,29	97,6
4	0,86	30,54	128,8
5	0,86	30,59	163,2
6	0,87	30,80	204,0
7	0,88	31,03	260,0
8	0,90	31,46	406,4
9	0,87	31,93	501,6
10	0,87	32,13	546,4
11	0,87	32,21	555,2
12	0,87	32,38	596,8
13	0,87	32,51	632,0
14	0,87	32,65	673,6

ТАБЛИЦА I. Заявляемые характеристики конфигураций

При составлении данных конфигураций в [8, 9] проводились эксперименты над 100 и 1000 кадрами достаточно однообразного и простого набора данных полёта БВС над сельской местностью. Недостатком этой методики является то, что результаты достаточно приблизительны и не отражают характеристики конфигураций в более сложных видеопотоках, таких как полёте в черте городской застройки.

Соответственно, возникает проблема проверки конфигураций заявляемым характеристикам в условиях, приближенных к реальным, которой и посвящена настоящая работа.

III. МЕТОДИКА ТЕСТИРОВАНИЯ КОНФИГУРАЦИЙ

Экспериментальная установка представляет собой стенд (рис. 1, табл. 1). На мониторе 1 отображается исходное видео интенсивностью 10 кадров в секунду. Данный видеопоток снимается камерой Evidence в разрешении HD (1280х720 пикселей) и отображается в левой части монитора 2. Затем он кодируется нейросетевым кодером заданной конфигурации, передаётся через имитатор спутникового геостационарного канала связи (используется система NetDisturb [12] с дополнительной постоянной задержкой в обоих направлениях 600 мс). После этого кадр видеопотока восстанавливается нейросетевым декодером заданной конфигурации и отображается в правой части монитора 2.



Рис. 1. Экспериментальная установка

№ п/п	Обозначение	Наименование	Функции
1	Кодер/Декодер MJPEG	GPU платформа	Кодирование/Декодирование видеопотока МЈРЕС с
			камеры Evidence
2	НС кодер/НС декодер	GPU платформа	Кодирование/декодирование кадров видеопотока
			классическим нейросетевым кодеком
3	PoE	РоЕ инжектор	Питание для камер по Ethernet
4	Спутниковый канал связи	NetDisturb	Имитация геостационарного спутникового канал с
			внесением дополнительной задержки 600 мс
5	Мониторинговый ответвитель	NetOptics TP-CU3-ZD	Ответвление трафика с нулевой задержкой для анализа
			полосы пропускания
6	Камера	Evidence Apix-22ZBox/M8	Кодирование видеопотока МЈРЕС
7	Камера	UniView	Трансляция видеопотока с метками времени
8	ПК трансляции видео	Ноутбук DELL	Трансляция тестового видеофайла
9	ПК сбора трафика	Ноутбук DELL	Вычисление реальной полосы пропускания канала связи
10	ПК обработки	ПК с видеокартой Nvidia 3060 RTX	Вычисление средней задержки кадров видеопотока с
			помощью OCR

ТАБЛИНА II.	Состав стенла
1/10/11111/111	Состявствиди

После этого с помощью камеры контроля UniView выполняется съёмка монитора 2: в левой части исходный кадр видеопотока, а в правой – восстановленный после прохождения имитации спутникового канала связи. Далее между кадрами рассчитываются метрики качества восстанавливаемых кадров PSNR и SSIM. Кроме того, на мониторинговом ответвителе производится сбор сетевого трафика. Из собранного трафика вычисляется реальная полоса пропускания канала bitrate. Данная система также позволяет рассчитывать прикладную задержку видеопотока T_B^3 при использовании модели оптического распознавания цифр на мониторе 2 в случае применения видеопотока с цифрами времени в формате «ММ:сс.ммм», где М – минута, с – секунда и м – миллисекунда и его дальнейшим распознаванием. Тем не менее, T_B^3 не рассматривается В настоящем исследовании.

Испытания проводятся для каждой из 14 конфигураций нейросетевого кодека [9].

IV. Тестирование конфигураций

В результате тестирования конфигураций по заданной методике получены характеристики конфигураций классического нейросетевого кодека в условиях работы, приближенных к реальным (рис. 2–4).

При проектировании конфигураций нейросетевого кодека в [8, 9] было выдвинуто предположение, что с увеличением номера конфигурации увеличиваются значения метрик качества восстанавливаемых кадров SSIM, PSNR (и, соответственно, качество восстанавливаемых кадров), но увеличивается и необходимая полоса пропускания.



🗖 Оцениваемая SSIM 🛛 🗖 Реальная SSIM

Рис. 2. Сравнение оцениваемых и реальных SSIM



Рис. 3. Сравнение оцениваемых и реальных PSNR



Рис. 4. Сравнение оцениваемых и реальных bitrate

Тем не менее, испытания, проведённые в условиях, приближенных к реальным, показали:

- значения метрики SSIM начинают колебаться уже с 4 конфигурации, явный рост наблюдается только между 1 и 4 конфигурациями;
- значения метрики PSNR показывает неустойчивый рост на конфигурациях 1-10, а затем наблюдается спад и колебания;
- требуемая полоса пропускания канала bitrate возрастает на конфигурациях 1-8 и отдельно на 9-14, между конфигурациями 8 и 9 наблюдается значительный спад;
- оцениваемые метрики качества восстанавливаемых кадров SSIM и PSNR завышены по сравнению с реальными;
- для конфигураций 1-8 оцениваемая полоса пропускания *bitrate* существенно занижена по сравнению с реальной, а для конфигураций 9-14 – несущественно завышена.

V. ЗАКЛЮЧЕНИЕ

работе В данной представлена методика тестирования конфигураций нейросетевых кодеков заявляемым им параметрам. На основании данной методики испытаны конфигурации классического нейросетевого кодека. Результате работы могут быть составлении конфигураций применены при нейросетевых кодеков для адаптивных систем сжатия кадров видеопотока.

Испытания показали, что составленная система конфигураций не в полной мере отражает принцип их составления: увеличение качества восстанавливаемых кадров при увеличении полосы пропускания. Необходима коррекция конфигураций классического нейросетевого кодека в части разработки такого поднабор конфигураций, в котором метрики качества восстанавливаемых кадров и требуемая полоса пропускания канала строго возрастают.

Отклонения в результатах могут быть обусловлены использованием простого и однообразного видео полёта БВС над сельской местностью при составлении конфигураций, а также малым количеством кадров для экспериментов в исходном исследовании. Другая причина отклонения в результатах заключается в использовании при испытаниях дополнительного блока подавления артефактов с алгоритмом «Перенос распределения» [13].

Список литературы

- Herglotz C., Och H., Meyer A., Ramasubbu G., Eichermüller L., Kränzler M., Brand F., Fischer K., Nguyen D.T., Regensky A., Kaup A. The Bjøntegaard Bible Why Your Way of Comparing Video Codecs May Be Wrong // IEEE Transactions on Image Processing. 2024.
- [2] Advanced Video Coding for Generic Audio-Visual Services. ITU-T H.264 и ISO/IEC 14496-10. 2003
- [3] High Efficiency Video Coding. ITU-T 23008-2. 2013.
- [4] Versatile Video Coding. ITU-T H.266 и ISO/IEC 23090-3. 2020.
- [5] Bankoski J., Koleszar J., Quillio L., Salonen J., Wilkins P., Xu Y. VP8 data format and decoding guide. RFC 6386. 2011.
- [6] Grange A., De Rivaz P., Hunt J. VP9 bitstream & decoding process specification // WebM Project. 2016.
- [7] De Rivaz P., Haughton J. AV1 bitstream & decoding process specification // The Alliance for Open Media. 2018.
- [8] Берёзкин А.А., Ченский А.А., Киричек Р.В., Захаров А.А. Исследование конфигураций нейросетевых кодеков для адаптивной системы сжатия кадров FPV-видеопотока при управлении беспилотными системами. Часть І. Методика // Электросвязь. Санкт-Петербург, 2024. № 9. С. 42-51.
- [9] Берёзкин А.А., Ченский А.А., Киричек Р.В., Захаров А.А. Исследование конфигураций нейросетевых кодеков для адаптивной системы сжатия кадров FPV-видеопотока при управлении беспилотными системами. Часть II. Эксперимент // Электросвязь. Санкт-Петербург, 2024. № 10. С. 59-69.
- [10] Березкин А.А., Паршин А.А., Лазарев А.А. Адаптивный контроль интенсивности видеопотока при передаче FPV-трафика беспилотных систем // 79-я научно-техническая конференция СПб НТО РЭС им. А.С. Попова, посвященная Дню радио: сборник докладов региональной конференции, Санкт-Петербург, 22-26 апреля 2024 г. 2024. С. 158-161.
- [11] Hore A., Ziou D. Image quality metrics: PSNR vs. SSIM // 2010 20th international conference on pattern recognition. IEEE, 2010. C. 2366-2369.
- [12] NetDisturb // ZTI Communications. URL: https://www.zticommunications.com/netdisturb/ (дата обращения: 18.01.2025).
- [13] Берёзкин А.А., Ченский А.А., Киричек Р.В. Нивелирование артефактов кадров видеопотока при FPV-управлении беспилотными системами // Инфкоммуникационные технологии. 2024. Т. 22. № 1(85). С. 7-17.