

Квантование видеопотока при FPV-управлении беспилотными системами в гибридных сетях СВЯЗИ

А. А. Ченский

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

chenskii.aa@sut.ru

А. А. Березкин

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

berezkin.aa@sut.ru

Аннотация. В настоящей статье рассматриваются три алгоритма квантования латентного представления квантованного вариационного автокодировщика VQ-f16, входящего в состав оригинального квантизированного нейросетевого кодера видеопотока при FPV-управлении беспилотными системами. К данным алгоритмам относятся: линейный, степенной и модифицированный логистический. Модификация последнего заключается в сдвиге не на среднее значение, а на минимальное. В результате была показана возможность увеличения степени сжатия по сравнению с линейным алгоритмом: при степенном на 11 % и при логистическом на 20,5 % без существенных изменений в метриках качества изображений SSIM и PSNR.

Ключевые слова: нейронные сети; квантование; беспилотная система; управление от первого лица

I. ВВЕДЕНИЕ

Цифровизация всех секторов экономика растёт, и вместе с тем всё более широко проявляется запрос пользователей на гарантированное снижение задержки сигнала в мобильных сетях связи. В отдельных случаях, таких как приложения тактильного интернета [1–3], требуется снижение задержки до менее 1 мс. К другим требованиям относится скорость соединения в 1–10 Гбит/с, что особенно важно при FPV-управлении беспилотными системами (БС).

Среди тенденций технологического и рыночного развития сектора телекоммуникаций стоит отдельно отметить появление и развитие в 2020-х годах негеостационарных многоспутниковых систем связи. В результате их развития появляются гибридные орбитально-наземные сети связи (ГОНСС).

На спутниковых радиointерфейсах данных сетей внедряются стандарты мобильной связи LTE и 5G, что в свою очередь позволяет космическим аппаратам и мобильным устройствам устанавливать связь напрямую. Стандартизация наземных, воздушных и космических мобильных сетей 5G/6G ожидается с 2028 по 2033 года.

ГОНСС обеспечит осуществление высокоскоростной передачи данных с малыми задержками, а также FPV-управление беспилотными летательными аппаратами

Научная статья подготовлена в рамках прикладных научных исследований СПбГУТ, регистрационный номер 1023031600087-9 в ЕГИСУ НИОКТР.

(БПЛА) на большом расстоянии, что указано в Стратегии развития отрасли связи Российской Федерации до 2035 года [4]. Минимизация задержки критична для эффективной работы БС в гибридных сетях, что в свою очередь требует разработки новых моделей и методов, которые бы позволили уменьшить сетевые издержки при передаче команд управления и потока видеоданных между БС и станцией внешнего пилота (СВП).

II. КВАНТИЗИРОВАННЫЙ НЕЙРОННЫЙ КОДЕР

Структурная схема квантизированного нейронного кодера (КНК), который предназначен для использования в канале передачи потока видеоданных в режиме управления от первого лица (УПЛ) от БС к СВП, представлена на рис. 1.

Функционирование данного нейросетевого кодера базируется на использовании вариационного автокодировщика VAE из состава модели стабильной диффузии (Stable Diffusion, SD) [5]. В [6] представлено описание принципа работы КНК и его блоков.

В настоящей статье проводится исследование применения векторного квантованного вариационного автокодировщика (VQ-VAE) с коэффициентом пространственного сжатия $f=16$ (VQ-f16) и алгоритмы квантования его выхода, используемые для уменьшения объёма данных. При кодировании VQ-f16 изображения размерности (3, 512, 512) получается внутренний латентный тензор размерности (1, 16, 32, 32).

VQ-VAE представляет собой разновидность VAE, сочетающего в себе сильные стороны автокодировщиков и векторного квантования для изучения осмысленных дискретных представлений данных. Работает он следующим образом: кодирует данные, поступающие на вход, в непрерывное скрытое пространство, после чего отображает их в конечный набор изученных представлений (embeddings), используя векторное квантование. В результате получается дискретное представление, которое для восстановления исходных данных можно декодировать. Его главное преимущество состоит в отделении важной информации от шума, что делает его пригодным для задач, требующих надёжного и компактного представления.

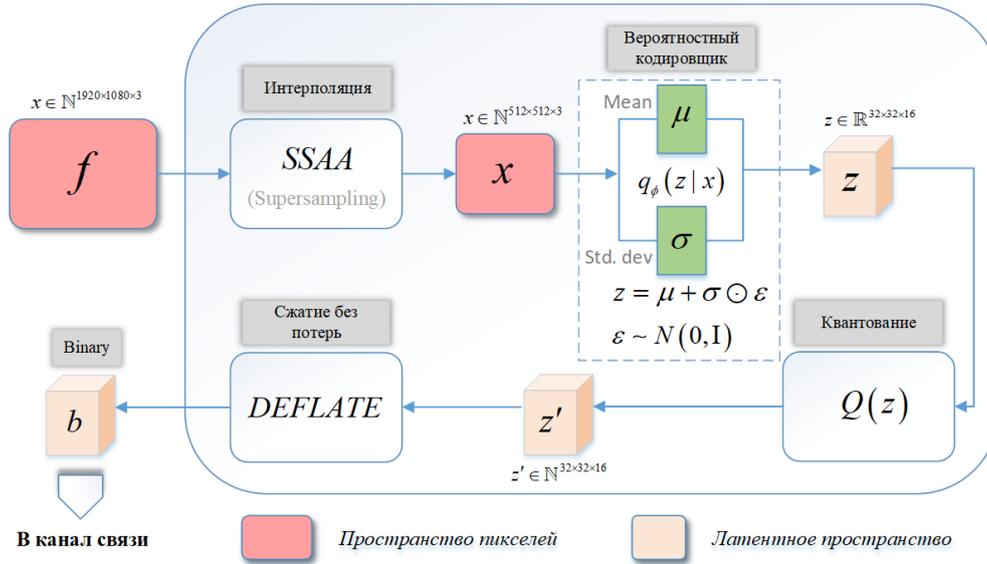


Рис. 1. Структурная схема квантизированного нейронного кодера

На вход КНК подаётся кадр видеопотока f с кафедры БС разрешения HD. Размер данного кадра при помощи блоки интерполяции уменьшается до 512×512 пикселей x . Затем выполняется кодирование полученного пиксельного пространства с использованием VQ-f16. В результате получается латентное внутреннее пространство z , которое в свою очередь квантуется ($Q(z)$) и сжимается без потерь с помощью алгоритма DEFLATE (используется в формате изображений PNG). Полученные двоичные данные передаются по каналу связи на СВП, где декомпрессируются, декодируются и отображаются оператору [7].

III. КВАНТОВАНИЕ ЛАТЕНТНОГО ПРОСТРАНСТВА

A. Принципы квантования и деквантования

Метод кодирования видеопотока, предложенный в [6], включает в себя квантование скрытого латентного пространства в качестве одного из этапов. Рассмотрим данный этап подробнее.

При кодировании кадра потока видеоданных с помощью вариационного автокодировщика VQ-f16 получается латентное пространство в виде тензора размерности (1, 16, 32, 32) типа данных float16 (чисел с плавающей точкой). Каждое значение этого типа данных находится в диапазоне $[-65504; +65504]$ с минимальными значениями $\pm 5,96E-8$ и занимает 2 байта. Распределение значений в латентном пространстве нормальное со средним, близким к нулевому, и единичной дисперсией.

Задача квантования: преобразование каждого из значений к типу данных uint8 (целые числа без знака, занимающие один байт). Каждое значение этого типа данных лежит в диапазоне $[0; 255]$ и занимает 1 байт. Таким образом, достигается сжатие в два раза. При квантовании предполагается, что вся область значений float16 разбивается на некоторые отрезки, которым ставятся в соответствие значения uint8. Перевод значений обратно в тип данных float16 проводится с помощью операции деквантования. При деквантовании предполагается, что каждому из значений uint8 ставится в соответствие наиболее подходящее значение float16, исходя из отрезков, которые получались в ходе

квантования. Отрезки могут как задаваться явно, так и получаться алгоритмически.

В данной статье рассматриваются три варианта алгоритмов квантования и деквантования, которые могут использоваться при проектировании КНК в канале передачи потока видеоданных при управлении БС от первого лица: линейный, степенной и логистический.

B. Линейное квантование

В ходе выполнения линейного квантования (ЛК) выполняется нахождение минимального и максимального значений, а также длины отрезка значений. Из них рассчитываются параметры сдвига m и масштабирования s . В этом случае преобразование значений t_i тензора T представлено в дальнейших формулах, после чего округляются до ближайших целых из диапазона типа данных uint8.

$$s = \frac{255}{\max(T) - \min(T)},$$

$$(\forall t_i \in T^*) : t_i^* = (t_i - m)s = (t_i - \min(T))s.$$

В ходе выполнения линейного деквантования предполагается, что на вход подаётся не только сам квантованный тензор T^* , но и параметры сдвига m и масштабирования s , которые используются для преобразования значений согласно следующей формуле.

$$(\forall t_i \in T) : t_i = \frac{t_i^*}{s} + m.$$

C. Степенное квантование

Принципы степенного квантования (СК) и деквантования аналогичны. Отличием является вычисление параметра масштабирования s и преобразований в соответствии со следующими формулами.

$$s = \log_{(\max(T) - \min(T))} (255),$$

$$(\forall t_i^* \in T^*) : t_i^* = (t_i - m)^s = (t_i - \min(T))^s,$$

$$(\forall t_i \in T): t_i = (t_i^*)^{1/s} + m.$$

D. Логистическое квантование

Аналогичным образом реализованы и логистические квантование (ЛГК) и деквантование. Логистическое квантование основано на той идее, что для наиболее точного квантования значений требуется той области интервала значений, где их находится больше всего, сопоставить наиболее короткие отрезки. Длина же отрезков при неявном алгоритмическом способе их задания зависит от значений производных. Таким образом, исходя из этих соображений, целесообразным представляется использование логистической функции, чья производная максимальна и достигает значения 1/4 в окрестностях нуля.

Но так как целью при квантовании является максимальное сжатие данных, а не точность, проводится модификация ЛГК: в качестве параметра сдвига m берётся не среднее значение, близкое к нулю, но минимальное значение тензора T , и проводится согласно следующим формулам.

$$(\forall t_i^* \in T^*): t_i^* = \frac{1}{1 + e^{-(t_i - m)}} = \frac{1}{1 + e^{-(t_i - \min(T))}}$$

$$(\forall t_i^{**} \in T^{**}): t_i^{**} = t_i^* s = \frac{255 t_i^*}{\max(T^*) - \min(T^*)} = \frac{255 t_i^*}{\max(T^*)}$$

$$(\forall t_i \in T): t_i = -\ln\left(\frac{t_i^{**}}{s} - 1\right) + m.$$

На рис. 2 продемонстрирован принцип модификации ЛГК: распределение значений при сдвиге на среднее отображены красным, на минимальное – средним; внизу представлены графики логистической функции и её производной.

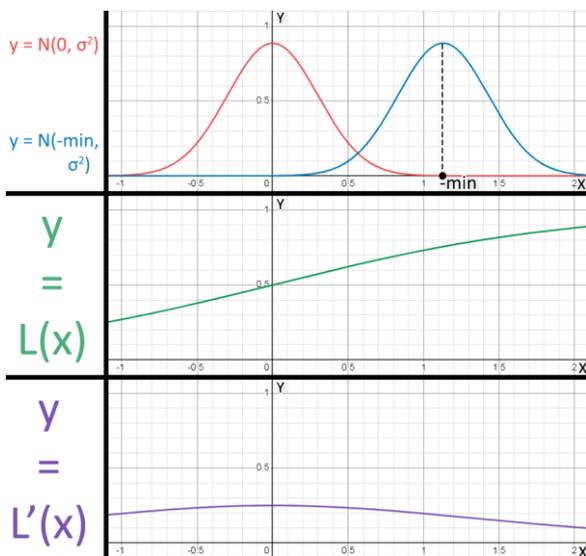


Рис. 2. Принцип модификации ЛГК

IV. ЭКСПЕРИМЕНТ

С целью проверки того, насколько целесообразно использовать те или иные алгоритмы квантования и деквантования в КНК, были проведены испытания над кадрами видео с полёта БПЛА. Примеры данных кадров представлены на рис. 3.

Испытания заключались в подаче кадров на КНК и последующем их восстановлении с помощью квантизированного нейронного декодера (КНД) [7]. Таким образом выполнялась имитация передачи кадров потока видеоданных от БС на СВП. Помехи в канале связи были приняты незначительными и не учитывались.



Рис. 3. Образцы использованных кадров

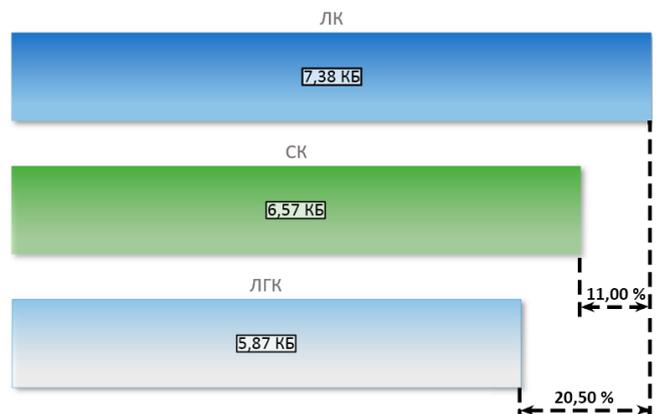


Рис. 4. Сравнение алгоритмов квантования по степени сжатия

В процессе испытаний выполнялись расчёты следующих показателей: степень сжатия изображения; метрики PSNR (пиковое отношение сигнала к шуму) и SSIM (индекс структурного сходства). Под первым понимается размер наиболее сжатого представления данных (после применения алгоритма DEFLATED), передающегося по каналу связи, в килобайтах. Последние используются для оценки ухудшения изображения после его передачи по каналу связи и восстановления.

Результаты проведённого эксперимента приведены в табл. I. Было показано существенное увеличение степени

сжатия: в среднем до 20,5 % при использовании модифицированного ЛГК, проиллюстрированное на рис. 4. При этом, как показано на рис. 5, отклонения в метриках качества совершенно несущественны.

Для сравнения также приведены исходные показатели при невключении в КНК и КНД блока квантования $Q(z)$ латентного пространства z , которые в сравнении с показателями, полученными при использовании данного блока с различными алгоритмами квантования, свидетельствуют о целесообразности его использования.

ТАБЛИЦА I. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТА

Квантование и деквантование	Степень сжатия, Кб	PSNR, дБ	SSIM
Отсутствует	14,71	30,93	0,75
ЛК	7,38	30,88	0,75
СК	6,57	30,54	0,75
Модифицированный ЛГК	5,87	30,90	0,75

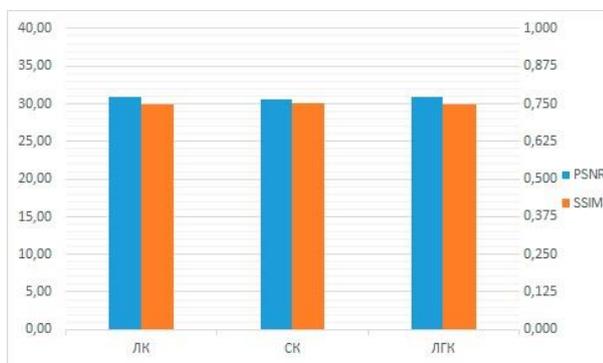


Рис. 5. Сравнение алгоритмов квантования по параметрам качества

V. ЗАКЛЮЧЕНИЕ

В настоящей работе было показано, что степень сжатия изображения без потерь может варьироваться в зависимости от используемого алгоритма в блоке квантования. Был представлен модифицированный ЛГК, который позволяет достигнуть более высоких показателей сжатия (на 11 % по сравнению со степенным и 20,5 % – линейным) с помощью алгоритма сжатия без потерь DEFLATED.

Согласно результатам проведённого эксперимента, использование блока квантования в КНК признаётся целесообразным: при отсутствии существенного влияния на качество изображения (в случае использования алгоритмов сжатия без потерь) удаётся достигнуть увеличения степени сжатия более чем в два раза в зависимости от используемого алгоритма.

СПИСОК ЛИТЕРАТУРЫ

- [1] Владимиров С.С., Кучерявый А.Е., Механизм компенсации задержек для приложений тактильного интернета // Электросвязь. 2018. № 3. С. 62-67.
- [2] Абделлах А.Р., Махмуд Д.И., Парамонов А.И., Кучерявый А.Е. Прогнозирование задержки в сетях интернета вещей и тактильного интернета с использованием машинного обучения // Электросвязь. 2021. № 1. С. 23-27.
- [3] Кучерявый А.Е., Бородин А.С., Мутханна А.С.А. и др. Искусственный интеллект в сетях связи // Актуальные проблемы инфотелекоммуникаций в науке и образовании: Международная научно-техническая и научно-методическая конференция. СПб.: СПбГУТ, 2021. С. 8-18.
- [4] Стратегия развития отрасли связи Российской Федерации на период до 2035 года // [Электронный ресурс]. URL: <http://government.ru/news/50304/> (дата обращения: 22.01.2024).
- [5] Stable Diffusion Online // [Электронный ресурс]. URL: <https://stablediffusionweb.com/> (дата обращения: 22.01.2024).
- [6] Березкин А.А., Вивчарь Р.М., Слепнев А.В. и др. Метод сжатия видеопотока при управлении беспилотными системами в гибридных орбитально-наземных сетях связи // Электросвязь. 2023. № 10. С. 48-56.
- [7] Березкин А.А., Вивчарь Р.М., Киричек Р.В., Захаров А.А. Метод декомпрессии FPV-видеопотока от беспилотных систем на основе латентной диффузионной нейросетевой модели // Электросвязь. 2024. №1. С. 42-53.