

УДК 004.02
<https://doi.org/10.32362/2500-316X-2022-10-4-7-17>



ОБЗОР

Сравнительный анализ алгоритмов сжатия четырёхмерных световых полей

Р.Г. Болбаков[@],
В.А. Мордвинов,
А.Д. Макаревич

МИРЭА – Российский технологический университет, Москва, 119454 Россия

[@] Автор для переписки, e-mail: bolbakov@mirea.ru

Резюме

Цели. Широкое распространение систем захвата световых полей обусловлено высоким качеством воспроизводимого изображения. Этот вид захвата, хоть и качественно превосходит традиционные подходы к захвату объемных изображений, генерирует огромное количество данных, необходимых для восстановления исходного заснятого четырехмерного светового поля. Цель работы – рассмотреть традиционные и расширенные до четырехмерной размерности алгоритмы сжатия изображений, провести их сравнительный анализ и определить наиболее подходящие из них.

Методы. Использованы математические методы обработки сигналов и методы статистического анализа.

Результаты. Проведены сравнение и анализ алгоритмов применительно к сжатию четырехмерных световых полей с использованием метрики PSNR. Установлено, что на выбранный критерий оценивания влияет не только размерность алгоритма сжатия, но также и расстояние базовой линии установки захвата, так как разница между изображениями увеличивается в зависимости от расстояния между оптическими центрами каждой матрицы камеры. Так для установок, состоящих из массива камер машинного зрения, находящихся на стойках и расставленных в помещении, очевидным выбором будет применение обычных методов сжатия изображений. Также, исходя из оценки произвольностей методов сжатия видео, замечено, что алгоритм XVC остается недооцененным, хотя его результаты оказываются выше остальных. Следующим по значимости можно считать алгоритм AV1. Установлено, что новейшие алгоритмы сжатия показывают более высокую производительность по отношению к своим предшественникам. Продемонстрировано, что при небольшом расстоянии между оптическими центрами запечатленных изображений применение алгоритмов сжатия видео более предпочтительно, чем применение алгоритмов сжатия изображений, так как они показывают более высокие результаты как в трехмерном, так и в четырехмерном варианте.

Выводы. Сравнение полученных результатов показывает необходимость применения на установках с длинной базовой линией (установленных на стойках камеры) алгоритмов из семейства сжатия видеозаписей (XVC, AV1). При работе с интегрированными камерами светового поля (Lytro) и установкой захвата с короткой базовой линией рекомендуется использовать алгоритмы сжатия изображений (JPEG). В общем случае рекомендуется использовать алгоритмы сжатия видео, в частности XVC, поскольку в среднем он показывает приемлемый уровень PSNR как в случае с короткой, так и с длинной базовой линией установки.

Ключевые слова: 3D-визуализации, 4D-световое поле, сжатие световых полей

• Поступила: 10.01.2022 • Доработана: 04.05.2022 • Принята к опубликованию: 20.06.2022

Для цитирования: Болбаков Р.Г., Мордвинов В.А., Макаревич А.Д. Сравнительный анализ алгоритмов сжатия четырехмерных световых полей. *Russ. Technol. J.* 2022;10(4):7–17. <https://doi.org/10.32362/2500-316X-2022-10-4-7-17>

Прозрачность финансовой деятельности: Авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Авторы заявляют об отсутствии конфликта интересов.

REVIEW ARTICLE

Comparative analysis of compression algorithms for four-dimensional light fields

Roman G. Bolbakov[@],
Vladimir A. Mordvinov,
Artem D. Makarevich

MIREA – Russian Technological University, Moscow, 119454 Russia

[@] Corresponding author, e-mail: bolbakov@mirea.ru

Abstract

Objectives. The widespread use of systems for capturing light fields is due to the high quality of the reproduced image. This type of capture, although qualitatively superior to traditional methods to capturing volumetric images, generates a huge amount of data needed to reconstruct the original captured 4D light field. The purpose of the work is to consider traditional and extended to four-dimensional image compression algorithms, to perform a comparative analysis and determine the most suitable.

Methods. Mathematical methods of signal processing and methods of statistical analysis are used.

Results. Algorithms are compared and analyzed in relation to the compression of four-dimensional light fields using the PSNR metric. The selected evaluation criterion is affected not only by the dimension of the compression algorithm, but also by the distance of the baseline of the capture setting, since the difference between images increases with the distance between the optical centers of each camera matrix. Thus, for installations consisting of an array of machine vision cameras located on racks and placed in a room, the obvious choice would be to use conventional image compression methods. Furthermore, based on the assessment of the arbitrariness of video compression methods, it should be noted that the XVC algorithm remains undervalued, although its results are higher. Algorithm AV1 can be considered the next in order of importance. It has been established that the latest compression algorithms show higher performance if compared to their predecessors. It has also been shown that with a small distance between the optical centers of the captured images, the use of video compression algorithms is preferable to the use of image compression algorithms, since they show better results in both three-dimensional and four-dimensional versions.

Conclusions. A comparison of the results obtained shows the need to use algorithms from the video compression family (XVC, AV1) on installations with a long baseline (mounted on camera stands). When working with integrated light field cameras (Lytro) and setting the capture with a short baseline, it is recommended to use image compression algorithms (JPEG). In general, video compression algorithms are recommended, in particular XVC, since on average it shows an acceptable level of PSNR in both the case of a short and long installation baseline.

Keywords: 3D visualization, 4D light field, light field compression

• Submitted: 10.01.2022 • Revised: 04.05.2022 • Accepted: 20.06.2022

For citation: Bolbakov R.G., Mordvinov V.A., Makarevich A.D. Comparative analysis of compression algorithms for four-dimensional light fields. *Russ. Technol. J.* 2022;10(4):7–17. <https://doi.org/10.32362/2500-316X-2022-10-4-7-17>

Financial disclosure: The authors have no a financial or property interest in any material or method mentioned.

The authors declare no conflicts of interest.

ВВЕДЕНИЕ

Информационные данные светового поля по своей природе описывают параметры света, испускаемого из точки пространства, захваченного, например, множеством камер [1] или камерой светового поля (light-field imaging) [2], основанной на пленочном принципе [3]. Требования к хранению и передаче таких данных сводятся зачастую к увеличению объема их хранилища и повышению производительности каналов передачи, поэтому разработка и исследование методов сжатия для световых полей в последние годы становятся все более актуальными. Несмотря на то, что существует множество эффективных форматов сжатия для неподвижных и движущихся изображений, в литературе выполнено лишь небольшое исследование влияния этих методов на свойства изображений светового поля [4]. В данной работе проведена оценка влияния современных методов сжатия изображений и видео на качество изображений, получаемых на основе данных светового поля. Эти методы включают в себя последние стандарты сжатия видео, особенно AV1¹ [5]

¹ AOMedia Video 1 (AV1) – открытый стандарт сжатия видео, предназначенный для кодирования видео, передаваемого по сети Интернет. Он заменяет формат кодирования видео VP9, разработанный Google. Согласно [5], AV1 превосходит H.265/HEVC на 17%, а VP9 – на 13% в широком диапазоне битрейта/разрешений. Разрабатывается Альянсом за открытые медиа (AOMedia), созданным в 2015 г. и состоящим из компаний, занимающихся производством электроники, распространением видео по запросу и разработкой веб-браузеров (AMD, Apple, Arm, Broadcom, Intel, Nvidia, Amazon, Facebook, Google, Hulu, Netflix, Mozilla, Microsoft). Timothy B. Terriberry. Progress in the Alliance for Open Media (slides). URL: <https://people.xiph.org/~tterribe/pubs/lca2017/aom.pdf> (18 January 2017), дата обращения: 22.06.2017. [AOMedia Video 1 (AV1) is an open video compression standard for encoding video transmitted over the Internet. It replaces the VP9 video encoding format developed by Google. According to [5], AV1 outperforms H.265/HEVC by 17% and VP9 by 13% over a wide range of bitrates/resolutions. Developed by the Alliance for Open Media (AOMedia), created in 2015 and comprised of electronics, video-on-demand, and web browser companies (AMD, Apple, Arm, Broadcom, Intel, Nvidia, Amazon, Facebook, Google, Hulu, Netflix, Mozilla, Microsoft). Timothy B. Terriberry. Progress in the Alliance for Open Media (slides). URL: <https://people.xiph.org/~tterribe/pubs/lca2017/aom.pdf> (18 January 2017). Accessed June 22, 2017.]

и XVC² [6]. Чтобы в полной мере использовать потенциал распространенных методов сжатия изображений на четырехмерных изображениях светового поля, мы расширяем эти методы на трех- и четырехмерные измерения. В статье продемонстрировано, что данные четырехмерного светового поля могут быть сжаты гораздо сильнее, чем независимые неподвижные изображения при сохранении того же визуального качества воспринимаемого изображения.

Для описания трехмерной сцены из любого возможного положения наблюдателя определим функцию $P(x, y, z, \phi, \psi)$ [3], где (ϕ, ψ) – угол обзора камеры (в сферических координатах); (x, y, z) – абсолютное положение точки просмотра (рис. 1). Результатом выполнения функции P является цвет, описывающий визуальное отображение наблюдаемой точки пространства. Определение функции P может быть дополнено параметром t (время) для описания динамичной сцены.

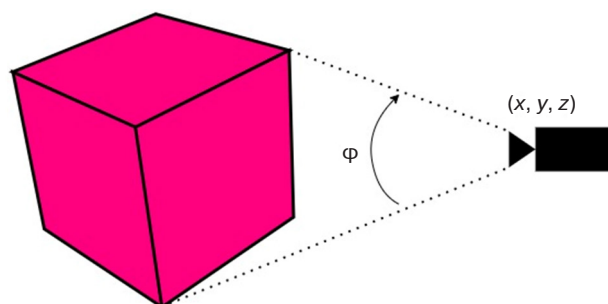


Рис. 1. Захват сцены с одной позиции просмотра.
Для простоты угол обзора указан для одной сферической координаты

Наша задача – описать сцену путем съемки либо с помощью множества камер (camera array), либо с помощью одной компактной матрицы, которой предшествуют микролинзы, например, как в Lytro. В этом случае апертура может быть изображена в виде сетки видов (камер), расположенных на двумерной плоскости. Этот случай проиллюстрирован

² XVC – формат кодирования видео с сильным акцентом на потоковые приложения с низким битрейтом. На официальном сайте (URL: <https://xvc.io/>, дата обращения 04.12.2021) утверждается, что кодек превосходит AV1, H.265/HEVC и VP9. [XVC is a video encoding format with a strong focus on low bitrate streaming applications. The official website (URL: <https://xvc.io/>. Accessed December 04, 2021) claims that the codec is superior to AV1, H.265/HEVC, and VP9.]

на рис. 2, где расстояние на линии, проходящей через точки основания между отдельными видами, описывается параметром d . Это представление часто называют 4D-световым полем (англ. light field – LF), поскольку мы имеем дело с функцией светового поля L , дискретизированной в четырех измерениях (k, l, m, n) , где (m, n) – координаты пикселей; (k, l) – индексы изображения с субапертурой.

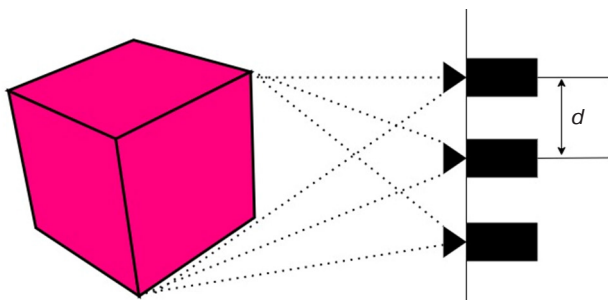


Рис. 2. Захват четырехмерного светового поля при помощи массива камер

Световые поля, получаемые одним компактным единым устройством, имеют ограничения по углу обзора. Световые поля, основанные на множестве камер, обеспечивают большие углы обзора за счет распределения матриц камер в пространстве вокруг объекта. На практике количество видов, расположенных на двумерной плоскости, колеблется от пары единиц. Учитывая высокое разрешение датчиков, неудивительно, что объем данных о световом поле достигает огромных размеров. В качестве примера рассмотрим световое поле «Treasure Chest»³ (рис. 3), взятое из Стэнфордского архива световых полей. Это поле снимается с помощью сетки камер 17×17 с разрешением изображения 1536×1280 пикселей. Размер несжатых данных превышает 1 Гб. При использовании фото-последовательностей для организации видеоряда требования к хранению и передаче будут соразмерно увеличены.

МЕТОДЫ СЖАТИЯ 4D-СВЕТОВЫХ ПОЛЕЙ

В последнее время в литературе было предложено несколько методов сжатия 4D-световых полей [4, 5, 7–19]. С помощью некоторых из этих методов исследователи пытаются напрямую сжимать данные с датчиков, которым предшествуют микролинзы (изображения линз). С помощью других сжимают результирующее 4D-световое поле вместо обработки исходных «сырых» данных. В этой статье мы сосредоточимся только на последних.

Сравним различные современные методы сжатия, применимые к данным 4D-светового поля. Эти методы включают последние стандарты сжатия видео, особенно AV1 (утвержден в июне 2018 г.) и XVC (версия выпущена в июле 2018 г.). Чтобы провести сравнение, мы перефокусируем исходное и распакованное световое поле. Затем оценка проводится с использованием метрики PSNR (отношение пикового сигнала к шуму в исследуемой выборке, англ. peak signal-to-noise ratio) в качестве полномасштабного эталонного показателя оценки качества.

Отдельные отображения из исходного светового поля обычно не визуализируются. Поэтому нет смысла напрямую сравнивать исходное и распакованное световые поля, хотя такая методология обычно используется для оценки производительности сжатия одного вида. По этой причине будет использована методология оценки производительности сжатия для многофокусного рендеринга из [4]. Эта методология в основном заключается в оценке качества рендера представлений для нескольких контрольных точек. Визуализированные отображения получаются путем объединения пикселей из разных 4D-видов светового поля для различных фокальных плоскостей. Среднее искажение вычисляется как среднее значение PSNR для нескольких визуализированных видов в фокальной плоскости. Методика сравнения проиллюстрирована на рис. 3. Обратим внимание,

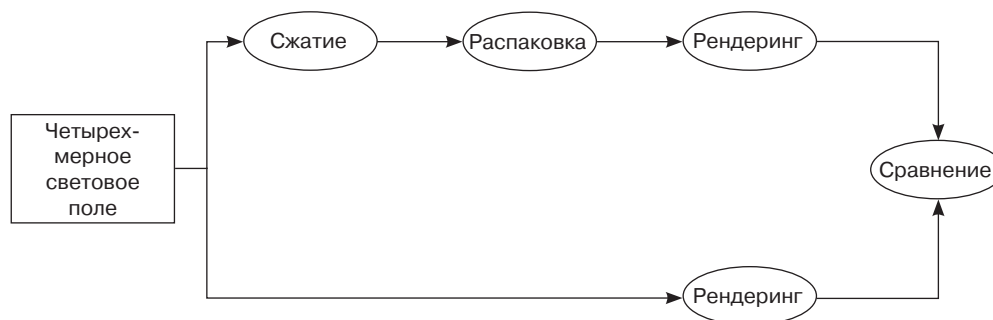


Рис. 3. Диаграмма потока данных методики сравнения алгоритмов сжатия

³ Stanford light-field dataset. URL: http://lightfield.stanford.edu/?gclid=Cj0KCQiA47GNBhDrARIsAKfZ2rD2CB3IMtzHJXPr0uXM_KJm_tElIZILviFERCFsasV9JygG55uBlaAtRTEALwwcB, дата обращения: 04.12.2021. [Stanford light-field dataset. URL: http://lightfield.stanford.edu/?gclid=Cj0KCQiA47GNBhDrARIsAKfZ2rD2CB3IMtzHJXPr0uXM_KJm_tElIZILviFERCFsasV9JygG55uBlaAtRTEALwwcB. Accessed December 04, 2021.]

что PSNR вычисляется из среднего квадрата ошибки по всем трем цветовым компонентам.

Световое поле 4D содержит двумерную сетку двумерных видов, захваченных с камер. Длина базовой линии между отдельными видами колеблется от нескольких миллиметров (микролинзы) до нескольких сантиметров (массив камер). Поэтому естественно ожидать высокого сходства видов, смежных в любом из двух направлений сетки. Это сходство открывает путь к пониманию данных 4D-светового поля как видеоряда, перемещающегося между точками обзора. Альтернативно мы можем рассматривать четырехмерное световое поле как трехмерное или непосредственно как четырехмерное тело. Описанные выше подходы также могут быть отражены в сжатии светового поля с помощью системы кодирования изображений, видео, объемных или четырехмерных изображений. Хотя возможны и другие подходы (например, использование 3D-видео).

В последние годы в сторонних работах сравнивались и оценивались характеристики сжатия различных подходов к изображениям светового поля.

В [4] авторы оценили производительность основных стандартов кодирования изображений с независимыми видами и H.265/HEVC⁴ с независимыми видами. Пометка «с независимыми видами» является обозначением того, что отдельные виды были сжаты независимо друг от друга. Подходы к кодированию видео в работе не оценивались. Как и следовало ожидать, внутренний профиль H.265/HEVC оказался наиболее эффективным методом сжатия.

В [7] авторы сравнили производительность сжатия трех стратегий с использованием кодера H.265/HEVC. В первой стратегии выполняется сжатие непосредственно на исходном планооптическом изображении. В следующей стратегии организуются виды четырехмерного светового поля в виде псевдовременной последовательности в спиральном порядке, и впоследствии происходит их сжатие. В последней стратегии сжатие происходит на подмножестве изображений, извлеченном из планооптического изображения, посредством преобразования в четырехмерное световое поле. Результаты проведенной авторами работы показывают, что кодирование четырехмерного светового поля приводит к лучшей производительности по сравнению с непосредственным кодированием изображений, полученных с камер.

⁴ H.265/HEVC (high efficiency video coding, англ.) – высокоэффективное кодирование видеоизображений; MPEG-H – стандарт сжатия видео, разработанный в качестве преемника широко используемого H.264/AVC (MPEG-4). [H.265/HEVC is a high efficiency video coding; MPEG-H is a video compression standard developed as a successor to the widely used H.264/AVC (MPEG-4).]

Авторы работы [8] сравнили производительность JPEG⁵, JPEG 2000 и SPIHT⁶ непосредственно на изображениях, полученных с матриц камер. Сравнение проводилось с использованием той же методологии, что и в данной статье. Как и следовало ожидать, JPEG 2000 демонстрирует наилучшую производительность сжатия.

В [9] авторы предложили переставить четырехмерное световое поле как представление в виде плиток большого прямоугольного изображения. Затем это изображение сжималось с помощью кодера JPEG 2000. Предложенная схема была сравнена со стандартными алгоритмами кодирования изображений, а именно, JPEG 2000 и JPEG XR. Однако неясно, как точно эти стандартные алгоритмы кодирования были применены к данным 4D-светового поля.

В [10] автор перестраивает четырехмерное световое поле в трехмерное тело. Затем трехмерный объем кодируется с использованием схемы 3D DCT на блоках $8 \times 8 \times 8$ аналогично системе кодирования JPEG.

Помимо обычных методов кодирования существует также альтернативный подход [11], который использует глубокое обучение для оценки 2D-представления из разреженных наборов 4D-представлений. Другой подход [12] предлагает собственную схему разреженного кодирования для всего четырехмерного светового поля, основанную на нескольких оптимизированных ключевых представлениях.

Метод, описанный в [13], опирается на построение суперлучей, которые ограничивают суперпиксели, образующие данный суперлуч. Это ограничение необходимо для того, чтобы суперлучи можно было использовать в качестве поддержки уменьшения угловой размерности на основе приближения матрицы низкого ранга. Затем вычисляется приближение матрицы низкого ранга для суперлучей с компенсацией несоответствия с применением разложения по сингулярным значениям (англ. singular value decomposition – SVD). Затем базовые векторы кодируются с использованием HEVC или JPEG-Pleno VM 1.1 для каждого отдельного представления.

В [14, 15] авторы предлагают иерархическую структуру кодирования для четырехмерных световых полей. Четырехмерное световое поле разбивается на несколько видов, а затем организует их в структуру кодирования в соответствии с пространственными координатами. Все представления закодированы иерархически. Схема реализована в справочном программном обеспечении H.265/HEVC.

⁵ JPEG – Joint Photographic Experts Group (англ.), по названию организации-разработчика. [JPEG – Joint Photographic Experts Group, titled after the name of the developer.]

⁶ SPIHT – set partitioning in hierarchical trees (англ.) – пространственно упорядоченные иерархические деревья. [SPIHT is a set partitioning in hierarchical trees.]

В [16] авторы предлагают схему кодирования, которая разделяет четырехмерное световое поле на несколько центральных видов и остальные смежные виды. Смежные виды вычитаются из центральных видов, и затем обе группы кодируются с помощью кодека H.265/HEVC. Авторы [17, 18] передают четырехмерное световое поле в кодек H.265/HEVC, используя режим взаимного прогнозирования для отдельных видов LF. Наконец, огромное внимание было уделено подходам сжатия на основе сверточных нейронных сетей [19, 20]. Из вышесказанного видно, что схемы кодирования JPEG 2000 и особенно H.265/HEVC довольно популярны при сжатии четырехмерных световых полей.

В настоящей статье мы сравниваем производительность основных современных методов сжатия с потерями. Эти методы можно разделить на четыре группы в зависимости от способа обработки данных 4D LF. Первая группа охватывает следующие методы кодирования изображений: JPEG и JPEG 2000. В [21] их называют методами, основанными на самоподобии. Вторая группа включает методы кодирования видео: H.265/HEVC, AV1, VP9 и XVC. В основном их называют методами, основанными на псевдопоследовательности. Третья группа расширяет методы кодирования изображений в трех измерениях. Эта группа состоит из JPEG 3D и JPEG 2000 3D. Обращаем внимание, что JPEG 3D относится к объемному изображению, а не к паре стереоскопических изображений. Четвертая группа расширяет методы кодирования изображений в четырех измерениях. Однако в этой группе существует только один метод – JPEG 4D.

Для оценки вышеуказанных методов используются следующие кодеки: OpenJPEG, x265, libaom (библиотека кодеков AV1), libvpx (SDK кодеков VP8/VP9), кодек XVC.

ИСХОДНЫЕ ДАННЫЕ

В этом разделе представлен набор данных, метод многофокусной визуализации, эксперименты, проведенные с этим набором данных, и полученные результаты.

Набор данных состоит из четырех 4D-световых полей, основанных на двух типах устройств захвата. Два световых поля были сняты с помощью пленоптической камеры Lytro Illum B01 (произв. компания Lytro, США), а два других – с помощью обычных DSLR камер.

Первое обычное световое поле камеры было снято с помощью матрицы из нескольких камер, а другое – с помощью простого моторизованной установки, оснащенного цифровой камерой Canon Rebel XT_i (произв. компания Canon, Япония). Соответствующие разрешения и соответствующие диапазоны расхождений изображений перечислены в таблице. Значения в последнем столбце – расхождение изображений – описывают

разницу в пикселях в местоположении одного и того же 3D-объекта, спроецированного на изображения, снятые камерой или вычисленные по изображению пленоптической камеры, в случае Lytro.

Как можно заметить, диапазон расхождений будет узким (от -1 до $+1$ пикселя) для светового поля с плотной выборкой (короткой базовой линией) в случае применения камеры Lytro и широким (от 40 до 90 пикселей) – для изображений, полученных массивом камер. Эти значения, очевидно, коррелируют с фокусным расстоянием камеры и расстоянием между центрами камер или пленоптических линз в случае Lytro. Для удобства центральный вид для каждого светового поля показан на рис. 4.

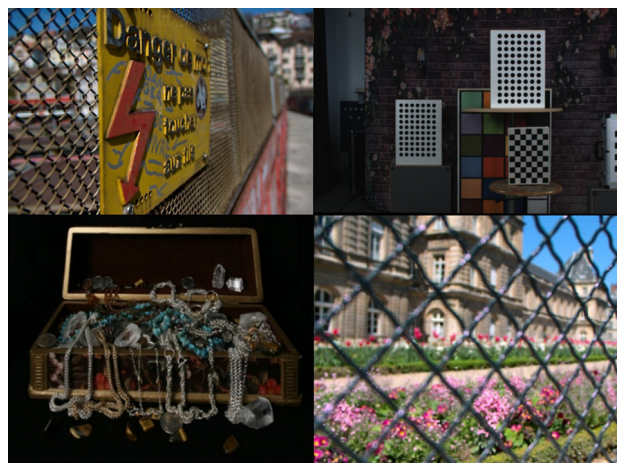


Рис. 4. Набор данных, использованных для сравнения. Слева направо: Danger de mort, Chessboard, Treasure Chest, Palais du Luxembourg

На рис. 4 первое и последнее световое поле снимаются с помощью пленоптической камеры; Chessboard снимается с помощью массива камер; Treasure Chest снимается с помощью моторизованной установки с зафиксированной на ней камерой.

Цифровая перефокусировка изображений в виртуальной фокальной плоскости достигается с помощью алгоритма суммирования сдвигов (shift-sum algorithm) [22]. Этот алгоритм сдвигает изображения (виды) с субапертурой в соответствии с оптическим центром камеры относительно базовой линии камеры относительно опорного кадра и накапливает соответствующие значения пикселей. Перефокусированное изображение будет средним значением преобразованных изображений. Вычисление значения пикселя в точке (m, n) перефокусированного изображения E_d задается уравнением:

$$E_d(m, n) = \frac{1}{N} \sum_{k, l} L(k, l, m + \alpha k, n + \alpha l), \quad (1)$$

где N – количество суммированных изображений; α – расстояние синтетической плоскости от основного объектива; k и l – индексы субапертурного

Таблица. Набор данных, использованных для сравнения

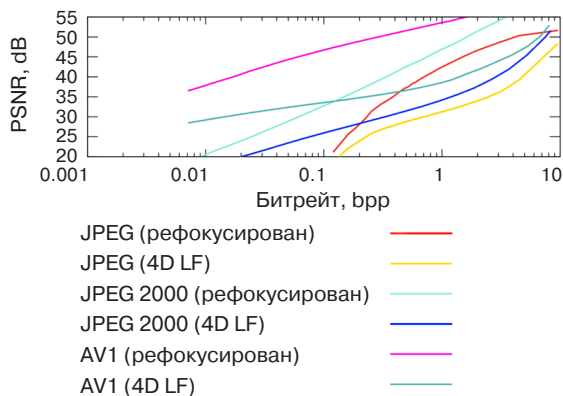
Описание	Источник	Разрешение	Расхождение в пикселях
Danger de mort	EPLF датасет	$15 \times 15 \times 625 \times 434$	От -1 до 1
Chessboard	Саарландский университет	$8 \times 8 \times 1920 \times 1080$	От 40 до 90
Treasure Chest	Стэндфордская лаборатория компьютерной графики	$17 \times 17 \times 1536 \times 1280$	От -1 до 7
Palais du Luxemburg	EPFL датасет	$15 \times 15 \times 625 \times 434$	От -1 до 1

изображения представления светового поля; ak и al – параметры сдвига по отношению к системе отсчета. Также выполнена линейная интерполяция в двух последних 4D-измерениях, чтобы преобразовать функцию выборочного светового поля в непрерывную.

ЭКСПЕРИМЕНТ 0

Предварительно стоит уточнить, действительно ли необходимо оценивать качество изображений, отображаемых для нескольких фокусных точек, а не на исходных данных (т.е. напрямую сравнивать исходный и распакованный набор изображений). Быстрый эксперимент показывает, что между этими подходами существует большая разница (рис. 5). Она составляет около 10 децибел в PSNR в зависимости от битрейта и метода сжатия. Это можно объяснить тем фактом, что любой пиксель в отображаемом виде представляет собой сумму пикселей из четырехмерного светового поля, так что эта сумма в совокупности уменьшает количество артефактов сжатия. Другими словами, мы можем позволить себе сжимать 4D-световые поля гораздо сильнее, чем независимые изображения, сохраняя при этом то же визуальное качество отображаемого изображения.

На рис. 5 показана разница в оценке качества с использованием 4D-светового поля напрямую по сравнению с использованием изображений, визуализированных в виртуальных фокальных плоскостях. Иллюстрация приведена на световом поле Danger de mort.

**Рис. 5.** Эксперимент 0. Разница в оценке качества

ЭКСПЕРИМЕНТ 1

Как видно из обзора литературы, большинство современных подходов к сжатию световых полей обрабатывают либо 2D-данные, либо их последовательность (сжатие видео). Сжатие четырехмерных изображений световых полей все еще остается относительно неисследованной областью. Поскольку четырехмерное световое поле представляет собой последовательности 2D-изображений (точек обзора), методы сжатия 2D могут использоваться для независимого кодирования точек обзора. Однако такие методы не позволяют использовать корреляции пикселей во всех четырех измерениях. Аналогичные рассуждения можно использовать и для 3D-методов. В нашем эксперименте 0 мы были заинтересованы в изучении эффектов сжатия световых полей в трех и четырех измерениях. Чтобы справедливо оценить производительность сжатия, для случаев 2D, 3D и 4D необходимо использовать идентичный метод сжатия. Таким образом, в работе используется пользовательская реализация метода сжатия JPEG с возможностью обработки данных 2D, 3D или 4D. Кроме того, известно о существовании стандарта JPEG 2000 с возможностью сжатия 2D- и 3D-данных одним и тем же способом. К сожалению, JPEG 2000 не работает с 4D-изображениями. Поскольку сходство соседних пикселей в трехмерном и четырехмерном изображении сильно зависит от базовой линии камеры, можно ожидать разных результатов. Результат этого эксперимента показан на рис. 6. На каждом графике горизонтальная ось показывает битрейт (бит на пиксель), а вертикальная ось – среднее значение PSNR для нескольких визуализированных точек обзора в фокальной плоскости.

На световых полях с небольшой базовой линией (Danger de mort и Palais du Luxemburg) оба метода 3D-сжатия явно превосходят свои 2D-аналоги по всему диапазону битрейта. Аналогично, метод 4D JPEG явно превосходит свой 3D-аналог.

Это неудивительно, потому что пиксели в одном и том же пространственном положении на соседних видах сильно коррелированы. Однако ситуация меняется с увеличением базовой линии. При этом (Treasure Chest и Chessboard) смежные виды становятся все

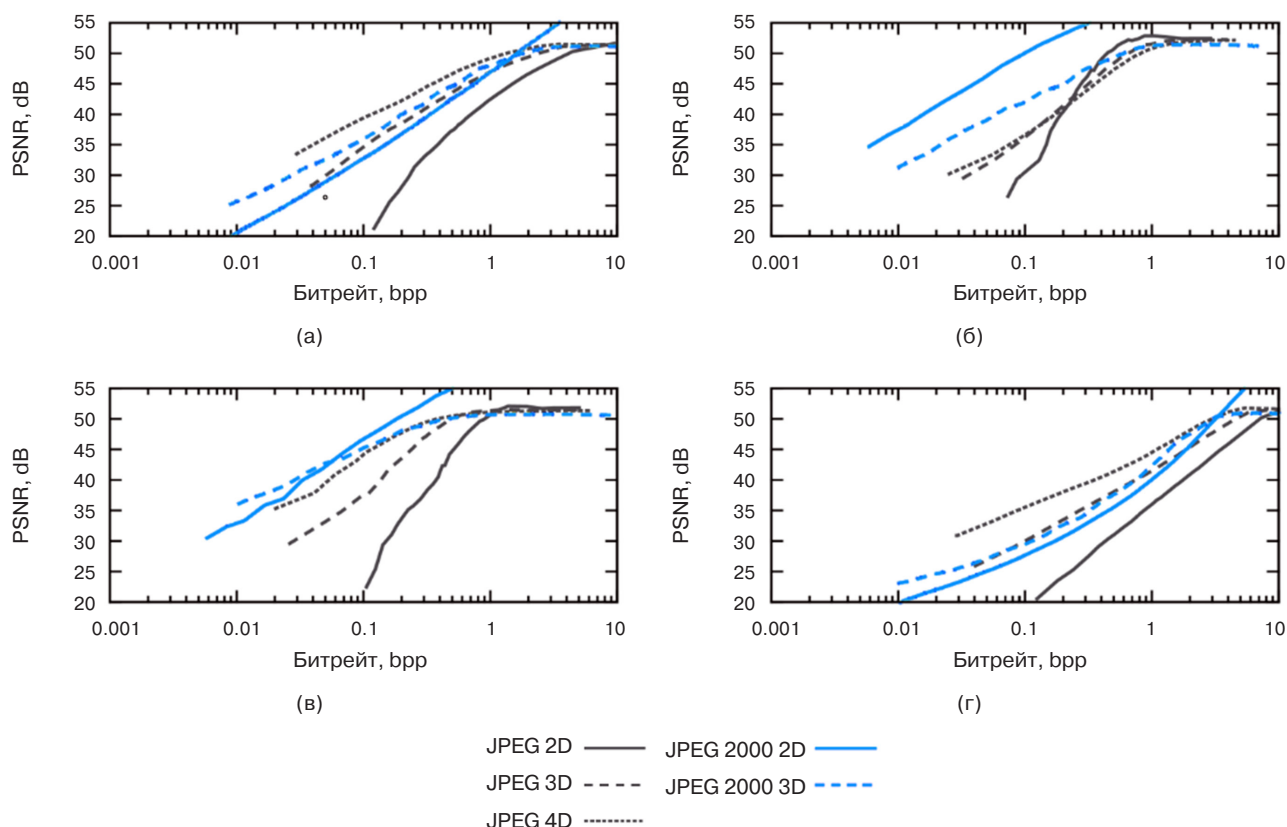


Рис. 6. Эксперимент 1. Сравнение производительности методов сжатия изображений:
(a) Danger de mort, (б) Chessboard, (в) Treasure Chest, (г) Palais du Luxemburg

менее и менее похожими, что приводит к более высоким амплитудам базовых коэффициентов преобразования. Следовательно, ситуация меняется в пользу методов сжатия с меньшими размерами.

Учитывая метод JPEG, Treasure Chest является особым случаем, поскольку он содержит большое количество черных пикселей. Оказывается, что более эффективно сжимать эти сплошные области сразу с помощью одного 4D-блока, чем с помощью нескольких 3D-блоков. Аналогично, более эффективным будет использование одного 3D-блока, чем нескольких 2D-блоков.

ЭКСПЕРИМЕНТ 2

Второе, на что следует обратить внимание в предыдущем разделе – это использование стандартов сжатия видео. Возникает вопрос: 4D-световые поля лучше сжимать в виде последовательности 2D-кадров или в виде многомерного тела? Поэтому мы измерили производительность всех вышеупомянутых стандартов сжатия видео. Результаты можно увидеть на рис. 7. На этот раз для краткости показаны результаты только для двух световых полей.

Интересно, что кодек XAVC действительно показал лучшую производительность сжатия, чем HEVC и AV1.

Чтобы ответить на вопрос: «Каков наилучший метод сжатия для данных световых полей?», было проведено дополнительное сравнение результатов с наиболее эффективными методами из эксперимента 1. Общее сравнение показано на рис. 8. Интересно, что методы сжатия видео работают лучше, чем все методы сжатия изображений, даже лучше, чем их расширения 3D и 4D.

ЗАКЛЮЧЕНИЕ

Целью работы было проведение оценки существующих методов, подходящих для сжатия 4D-световых полей с потерями. Поскольку световое поле в исходном варианте представляет из себя набор изображений, запечатленных массивом камер, методы сжатия изображений станут первыми, из чего мы будем выбирать в задачах сжатия. Эксперимент показывает, что методы, обрабатывающие 4D-световые поля непосредственно в четырех или трех измерениях, позволяют достичь более высоких результатов сжатия, чем классические реализации сжатия двумерных изображений.

Однако стоит отметить, что на выбранный критерий оценивания, а именно PSNR, влияет не только размерность алгоритма сжатия, но также и расстояние базовой линии, так как разница между изображениями увеличивается в зависимости от расстояния между

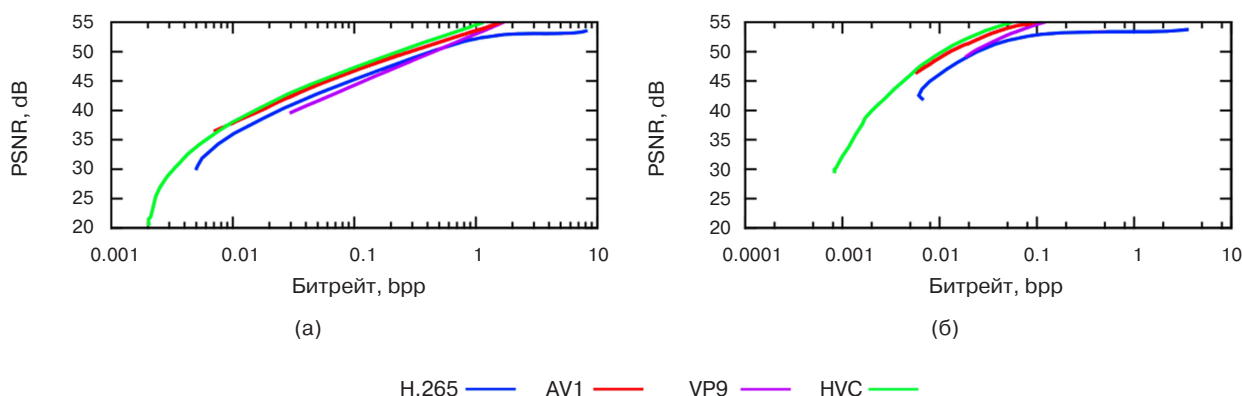


Рис. 7. Эксперимент 2. Сравнение производительности методов сжатия видео.
Использованные поля: (а) Danger de mort, (б) Chessboard

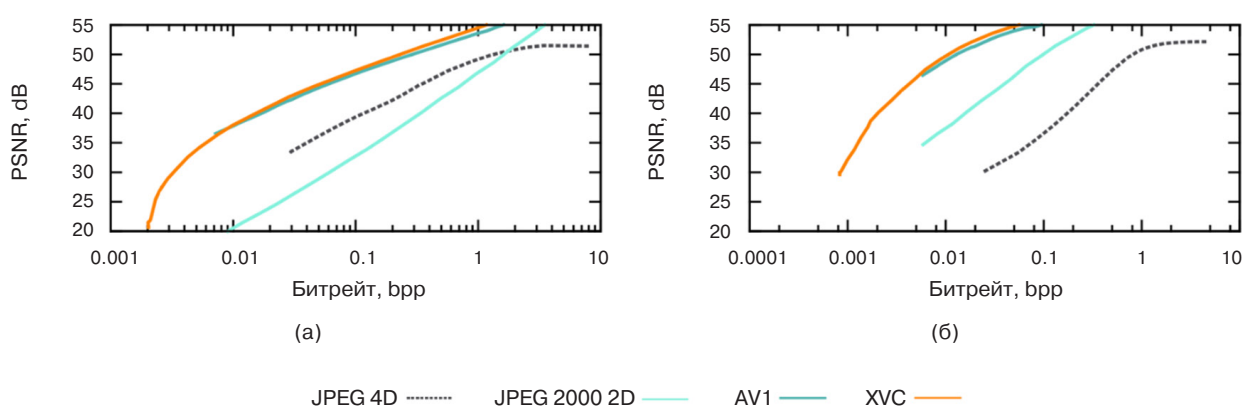


Рис. 8. Сравнение производительности методов сжатия видео по отношению к алгоритмам сжатия изображений. Использованные поля: (а) Danger de mort, (б) Chessboard

оптическими центрами каждой матрицы камеры. Так для установок, состоящих из массива камер машинного зрения, находящихся на стойках и расставленных в помещении, очевидным выбором будет применение обычных методов сжатия изображений.

Кроме того, исходя из оценки производительности методов сжатия видео, можем заметить, что алгоритм XVC остается недооцененным, хотя его результаты оказываются выше остальных. Следующим по значимости можно считать алгоритм AV1. Это подтверждает тот факт, что новейшие алгоритмы сжатия показывают более высокую производительность по отношению к своим предшественникам.

Также при небольшом расстоянии между оптическими центрами запечатленных изображений применение алгоритмов сжатия видео более предпочтительно, чем применение алгоритмов сжатия изображений, поскольку они показывают более высокие результаты как в трехмерном, так и в четырехмерном варианте.

Вклад авторов

Р.Г. Болбаков – идея, разработка дизайна исследования, консультация по вопросам проведения всех этапов исследования.

В.А. Мордвинов – анализ научной работы, критический пересмотр с внесением ценного интеллектуального содержания.

А.Д. Макаревич – анализ литературы, написание статьи, планирование исследования, проведение всех этапов исследования, формализация списка литературы.

Authors' contributions

R.G. Bolbakov – idea, development of research design, consultation on the problems of carrying out all stages of the study.

V.A. Mordvinov – analysis of scientific work, revision with the contribution of valuable intellectual content.

A.D. Makarevich – literature analysis, article writing, research planning, carrying out all stages of the study, formalization of the list of references.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Broxton M., Flynn J., Overbeck R., Erickson D., Hedman P., Matthew Duvall M., Dourgarian J., Busch J., Whalen M., Debevec P. Immersive light field video with a layered mesh representation. *ACM Trans. Graph.* 2020;39(4):86. <https://doi.org/10.1145/3386569.3392485>
2. Georgiev T., et al. Lytro camera technology: theory, algorithms, performance analysis. In: *Proc. SPIE. Multimedia Content and Mobile Devices*. 2013;8667:86671J. <https://doi.org/10.1117/12.2013581>

3. Adelson E.H., Bergen J.R. The plenoptic function and the elements of early vision. In: Landy M., Movshon J.A. (Eds.). *Computational Models of Visual Processing*. London: The MIT Press Cambridge; 1991. P. 3–20. <https://doi.org/10.7551/mitpress/2002.003.0004>
4. Alves G., Pereira F., da Silva E.A.B. Light field imaging coding: Performance assessment methodology and standards benchmarking. In: *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE; 2016. C. 1–6. <https://doi.org/10.1109/ICMEW.2016.7574774>
5. Zbrovskiy A., Feldmann C., Timmerer C. Multi-codec DASH dataset. In: *Proceedings of the 9th ACM Multimedia Systems Conference*. 2018. P. 438–443. <https://doi.org/10.1145/3204949.3208140>
6. Samuelsson J. The XVC video code” a revolutionary software-defined video compression format. *SMPTE Motion Imaging Journal*. 2019;128(10):1–8. <https://doi.org/10.5594/JMI.2019.2937737>
7. Viola I., Řeřábek M., Ebrahimi T. Comparison and evaluation of light field image coding approaches. *IEEE Journal of Selected Topics in Signal Processing*. 2017;11(7):1092–1106. <https://doi.org/10.1109/JSTSP.2017.2740167>
8. Dlabaja D. 4D-DCT based light field image compression. *Image*. 2019;28(1):128. Available from URL: <http://excel.fit.vutbr.cz/submissions/2019/021/21.pdf>
9. Brites C., Ascenso J., Pereira F. Lenslet light field image coding: Classifying, reviewing and evaluating. *IEEE Transactions on Circuits and Systems for Video Technology*. 2020;31(1):339–354. <https://doi.org/10.1109/TCSVT.2020.2976784>
10. Aggoun A. A 3D DCT compression algorithm for omnidirectional integral images. In: *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. IEEE; 2006:2:9142643. <https://doi.org/10.1109/ICASSP.2006.1660393>
11. Liu D., et al. View synthesis-based light field image compression using a generative adversarial network. *Inf. Sci.* 2021;545:118–131. <https://doi.org/10.1016/j.ins.2020.07.073>
12. Pratapa S., Manocha D. RLFC: random access light field compression using key views and bounded integer sequence encoding. In: *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. 2019. P. 1–10. <https://doi.org/10.1145/3306131.3317018>
13. Dib E., et al. Super-ray based low rank approximation for light field compression. In: *2019 Data Compression Conference (DCC)*. IEEE; 2019. P. 369–378. <https://doi.org/10.1109/DCC.2019.00045>
14. Joshitha R., Sharma M. A hierarchical coding scheme for glasses-free 3d displays based on scalable hybrid layered representation of real-world light fields. In: *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE; 2021. P. 3491–3498. <https://doi.org/10.1109/SMC52423.2021.9658584>
15. Ravishankar J., Sharma M. A novel hierarchical light field coding scheme based on hybrid stacked multiplicative layers and Fourier disparity layers for glasses-free 3D displays. arXiv preprint arXiv:2108.12399. 2021. <https://doi.org/10.48550/arXiv.2108.12399>
16. Tu W., et al. Efficient content adaptive plenoptic video coding. *IEEE Access*. 2020;8:5797–5804. <https://doi.org/10.1109/ACCESS.2020.2964056>
17. Chao Y.H., et al. Pre-demosaic graph-based light field image compression. arXiv preprint arXiv:2102.07883. 2021. Available from URL: <https://arxiv.org/pdf/2102.07883.pdf>
18. Mathew A.G., Mathew A.A. Intra coding of plenoptic images in HEVC. In: *Emerging Technologies for Sustainability*. CRC Press; 2020. P. 459–465. Available from URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/9780429353628-58/intra-coding-plenoptic-images-hevc-ashlin-george-mathew-anu-abraham-mathew>
19. Bakir N., et al. RDO-based light field image coding using convolutional neural networks and linear approximation. In: *2019 Data Compression Conference (DCC)*. IEEE; 2019. P. 554. <https://doi.org/10.1109/DCC.2019.00066>
20. Zhong T., et al. Light field image compression using depth-based CNN in intra prediction. In: *2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE; 2019. P. 8564–8567. <https://doi.org/10.1109/ICASSP.2019.8682820>
21. Yuan X., Haimi-Cohen R. Image compression based on compressive sensing: End-to-end comparison with JPEG. *IEEE Transactions on Multimedia*. 2020;22(11):2889–2904. <https://doi.org/10.1109/TMM.2020.2967646>
22. Li Y., et al. Coding of focused plenoptic contents by displacement intra prediction. *IEEE Transactions on Circuits and Systems for Video Technology*. 2015;26(7):1308–1319. <https://doi.org/10.1109/TCSVT.2015.2450333>

Об авторах

Болбаков Роман Геннадьевич, к.т.н., доцент, заведующий кафедрой инструментального и прикладного программного обеспечения Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: bolbakov@mirea.ru. Scopus Author ID 57202836952, SPIN-код РИНЦ 4210-2560, <http://orcid.org/0000-0002-4922-7260>

Мордвинов Владимир Александрович, к.т.н., профессор кафедры инструментального и прикладного программного обеспечения Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: mordvinov@mirea.ru. SPIN-код РИНЦ 9390-1540, <http://orcid.org/0000-0003-3622-8448>

Макаревич Артём Денисович, аспирант кафедры инструментального и прикладного программного обеспечения Института информационных технологий ФГБОУ ВО «МИРЭА – Российский технологический университет» (119454, Россия, Москва, пр-т Вернадского, д. 78). E-mail: artemmakarevich1997@gmail.com. <http://orcid.org/0000-0002-2211-1241>

About the authors

Roman G. Bolbakov, Cand. Sci. (Eng.), Associate Professor, Head of the Department of Instrumental and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: bolbakov@mirea.ru. Scopus Author ID 57202836952, RSCI SPIN-code 4210-2560, <http://orcid.org/0000-0002-4922-7260>

Vladimir A. Mordvinov, Cand. Sci. (Eng.), Professor, Department of Instrumental and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: mordvinov@mirea.ru. RSCI SPIN-code 9390-1540, <http://orcid.org/0000-0003-3622-8448>

Artem D. Makarevich, Postgraduate Student, Department of Instrumental and Applied Software, Institute of Information Technologies, MIREA – Russian Technological University (78, Vernadskogo pr., Moscow, 119454 Russia). E-mail: artemmakarevich1997@gmail.com. <http://orcid.org/0000-0002-2211-1241>