文章编号:1001-5078(2025)01-0111-09　　· 光学技术 ·

面向增强随机访问的光场图像可伸缩编码

郭锴鸿,蒋刚毅,陈晔曜,郁　梅
(宁波大学 信息科学与工程学院,浙江 宁波 315211)

摘　要:光场成像技术给用户提供了更高质量的沉浸式视觉体验,然而,由于其捕捉了大量的空间和角度信息,因此需要有效的编码方法来压缩光场图像庞大的数据量。在光场图像编码方法中,除了要注重压缩效率外,还需考虑其他重要因素,如随机访问和视口可伸缩性等。鉴于这些需求,本文提出一种新型的光场图像可伸缩编码方法,同时兼顾编码效率以及随机访问、可伸缩性等性能。首先,将光场图像子孔径阵列进行稀疏采样,在编码端将第五个视口层丢弃。其次,根据视点分层策略,对前四个视口层中的视点构建不同的多参考帧预测关系进行编码;在解码端通过视点合成网络对第五个视口层中的视点进行合成,从而实现光场图像的压缩以及视口可伸缩性。实验结果表明,相较于 JPEG Pleno 通用测试标准中所提出的 JPEG Pleno Anchor,真实场景下的 BDBR 减少了 16.166 %;合成场景下的 BDBR 减少了 6.796 %,并且提供了更有效的随机访问性能,在码率为 0.75 bpp 时,真实场景的相对随机访问惩罚平均 18.91 %,合成场景的相对随机访问惩罚平均为 17.00 %。所提方法实现了编码效率和随机访问性能的更佳平衡。

关键词:光场图像编码;随机访问;可伸缩编码;视图合成
中图分类号:TN911.7;TN919.8　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2025.01.017

Scalable coding of light field images for enhanced random access

GUO Kai-hong, JIANG Gang-yi, CHEN Ye-yao, YU Mei
(Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China)

Abstract: Light field imaging technology provides users with a higher quality immersive visual experience. However, since it captures a vast amount of spatial and angular information, effective encoding methods are required to compress the massive data volume in light field images. In addition to focusing on compression efficiency, other important factors such as random access and viewport scalability need to be considered in light field image coding methods. Given these requirements, a novel light field image scalable coding method for light-field images is proposed, which simultaneously takes into account the coding efficiency as well as the performance of random access and scalability. Firstly, the optical field image sub-aperture array is sparsely sampled and the fifth viewport layer is discarded at the coding end. Then, based on the view hierarchical strategy, different multiple reference frame prediction relationships are conducted for encoding the viewpoints in the first four viewport layers. At the decoding end, viewpoint synthesis network is employed to synthesize viewpoints in the fifth layer, achieving compression and viewport scalability for light field images. The experimental results demonstrate that compared to the JPEG Pleno Anchor proposed in the JPEG Pleno general testing standard, the BDBR in the real scene is reduced by 16.166 %, the BDBR in the synthetic scenes is reduced by

基金项目:国家自然科学基金项目(No. 62271276)资助。
作者简介:郭锴鸿(1999 -),男,回族,硕士研究生,研究方向为光场图像编码。E-mail:1525137192@qq.com
通讯作者:蒋刚毅(1964 -),男,博士,教授,博士生导师,主要从事多媒体信号处理与通信、计算成像、视觉感知与编码、图像与视频质量评价等方面的研究工作。E-mail:jianggangyi@126.com
收稿日期:2024-05-18

6.796 %, and provides a more effective random-access performance, with a relative random-access penalty of the real scene at a bit rate of 0.75 bpp 18.91 % on average and the relative random-access penalty for synthetic scenes is 17.00 % on average. The proposed method achieves a better balance between coding efficiency and random-access performance.

Keywords: light field image encoding; random-access; scalable coding; view synthesis

1 引 言

光场成像技术通过捕捉场景中光线的强度和方向信息,获取了三维环境丰富的信息,为用户提供更加沉浸式的体验,并实现了一系列后处理功能,如视点切换、深度估计[1]、视图渲染[2]和数字重聚焦[3]等。与传统的 2D 图像相比,光场图像(Light Field Image, LFI)不仅可以呈现出图像的空间域信息,同时还包括独特的角度域信息。基于双平面光参数化模型,光场图像信号可以表示为一个 4D 函数 $L(x, y, s, t) \in \mathbb{R}^{(X\times Y)\cdot(S\times T)}$,其中,$(x, y)$ 和 (s, t) 分别代表所记录光线的空间位置和角度位置,$(X \times Y)$ 代表空间分辨率,$(S \times T)$ 代表角度分辨率。光场图像中丰富的信息为人们提供了更真实、更细致的视觉体验,但也意味着巨大的数据量,这给光场图像的存储和传输带来了极大的挑战。光场图像压缩除了要解决数据的存储和传输问题,还应满足用户端不同的功能需求,如随机访问、视口可伸缩性等。

光场图像编码方法可分为普通光场图像编码和可伸缩性的光场图像编码。其中,普通光场图像编码可分为基于变换的光场图像编码方法、基于伪视频序列的光场图像编码方法和基于预测的光场图像编码方法。基于变换的光场图像编码方法使用各种变换编码技术来解相关光场图像,然后去除相邻视图之间的冗余信息。例如,Carvalho 等人[4]提出了一种基于四维离散余弦变换(4D Discrete Cosine Transform, 4D-DCT)的编码器,称为多维光场编码器(Multidimensional Light Field Encoder, MuLE),将微透镜图像转换为全视差多视图形式,划分为 4D 块,然后通过使用 4D-DCT 转换和十六叉树对光场微透镜图像进行压缩。该方案已被用于光场图像编码的 JPEG Pleno 验证模型[5]所采用。

基于伪视频序列的光场图像编码方法将光场微透镜图像转换成子孔径阵列,按照特定的扫描顺序排列成伪视频序列,利用现有的视频编码技术实现高效压缩,有效地消除子孔径图像(Sub-Aperture Image, SAI)的帧内和帧间冗余信息。例如,Conceição 等人[6]提出了一种自适应排序方式,其基于内容相似度,有助于更好地识别相似区域,并进行相应的排序。这有助于减少重复信息,从而提高压缩效率。Li 等人[7]提出了一种四叉树结构,结合参考帧选择优化和最佳比特分配,以实现高密度相机阵列光场图像的高效压缩。

基于预测的光场图像编码方法利用光场图像的几何结构特性,通过稀疏采样得到部分子孔径图像进行编码,以此在解码端重构全部子孔径图像。例如,Çetinkaya 等人[8]分别在编码端对子孔径阵列进行空间下采样、角度下采样、空角下采样、角空下采样来去除更多的冗余信息,在解码端采用空间和角度超分辨率进行重建。Huang 等人[9]通过提取深度图,采用高效视频编码(High Efficiency Video Coding, HEVC)标准对深度图以及稀疏的子孔径图像进行压缩,解码端利用深度图来对未编码子孔径图像进行重建。

在视口可伸缩光场图像编码方法方面,Gomes 等人[10]将每四个视点为一个子孔径图像块,将光场图像由中心向外螺旋扫描成 4 个伪视频序列,再利用多视点高效视频编码(Multi-view High Efficiency Video Coding, MV-HEVC)进行压缩。Monteiro 等人[11]提出一种基于 HEVC 随机访问配置的改进方法,先将子孔径图像分为 6 层,每层都构建成为伪视频来编码,并对参考列表、分块数、参考层数进行控制。

然而,上述方法虽然有效地提升了光场图像的压缩效率以及视口可伸缩性,但是在随机访问方面还有一定可提高的空间。因此,本文在兼顾编码效率和随机访问性能的基础上,提出了一种具有多参考帧预测结构的光场图像可伸缩编码方法。通过对子孔径图像阵列的不同视点进行分层处理,将不同视点的子孔径图像划分到各自的视口层。随后,使用深度学习合成网络来处理每个视口层之间的不同子孔径图像之间的关系,最终生成合成的视图或合成的参考图像。以此在编码效率方面取得较好性

能,并同时保持对图像的随机访问能力。

2　光场图像分层编码方法

2.1　整体框图

为增强随机访问性能,本文提出了一种具有多参考帧预测结构的光场图像可伸缩编码方法,如图1所示。在编码端,首先对原始光场 $\boldsymbol{I}^{\mathrm{org}}$ 进行稀疏采样,将选中的SAI集按照所提出的视点分层策略进行分层操作,得到四个视口层 $\boldsymbol{L}_1$、$\boldsymbol{L}_2$、$\boldsymbol{L}_3$、$\boldsymbol{L}_4$。然后,构建每个视口层之间的多参考帧预测关系,通过HEVC编码器进行编码。在解码端,利用解码重建的四个视口层通过视点合成网络进行视点合成得到第五个视口层 $\boldsymbol{L}_5^{\mathrm{syn}}$,进而得到完整的解码光场 $\boldsymbol{I}^{\mathrm{rec}}$。

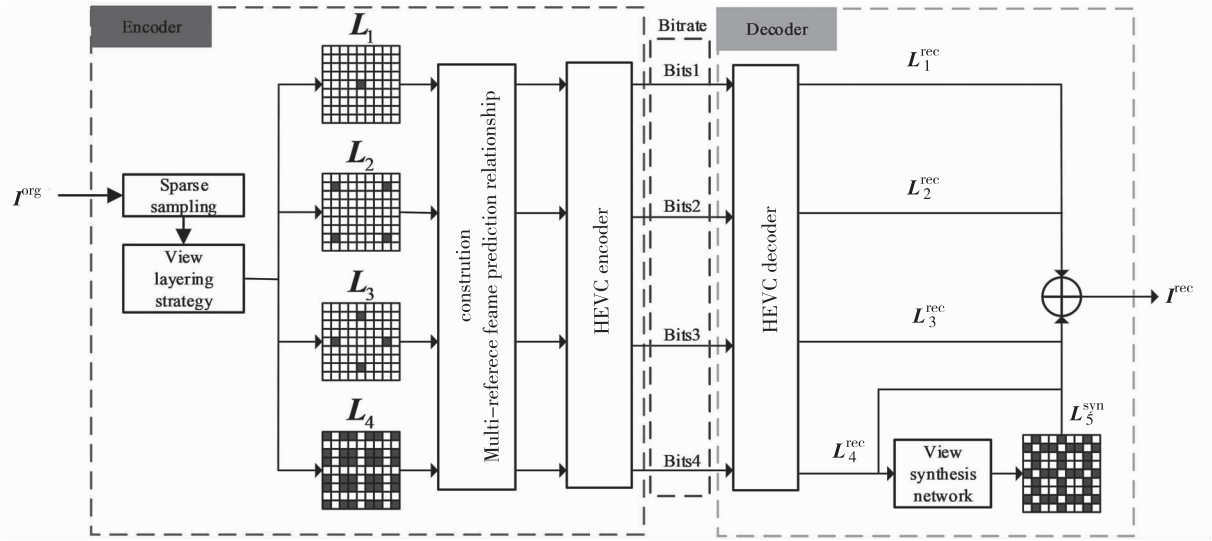


图1　面向增强随机访问的光场图像可伸缩编码总体框图

Fig. 1 Overall framework of enhanced random accessoriented light field imagescalable coding

2.2　视点分层策略

光场子孔径阵列中每个视点与其八邻域的8个视点相关性最大。基于此,将光场子孔径阵列划分为若干个子区域,如图2所示。在每个子区域中,以图中第一个子区域 $\boldsymbol{M}_1$ 为例,$\boldsymbol{M}_1$ 中的中心视图与其相邻的八个视点相关性最大,如果这八个视点以中心视点作为参考图像进行编码,则可以有效地提高压缩性能。考虑到编码效率与随机访问之间的平衡,每个子区域中,只对四个角视点以及中心视点进行编码,其余四邻域视点则通过视点合成网络进行合成。

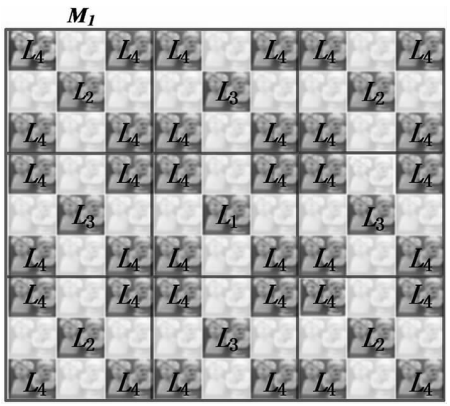


图2　分层策略示意图

Fig. 2 Diagram of layered strategy

因此,在编码端首先要对视点进行稀疏采样。对于每个子区域的中心视点,将其放在前三个视口层,第一个视口层为中心子区域的中心视点,第二个视口层为四个角子区域的中心视点,第三个视口层为其余子区域的中心视点,各个子区域中的四个角视点放在第四个视口层中。对于各子区域中不进行编码的中心视点的四邻域视点,将其放在第五个视口层并在解码端通过视点合成网络进行合成。

2.3　多参考帧预测关系构建

对于需要进行编解码的四个视口层中的视点构建多参考帧预测关系。所构建的多参考帧预测关系

如图3所示。第一个视口层 $\boldsymbol{L}_1$ 为中心视点,采用帧内编码,可以独立地进行访问。第二个视口层 $\boldsymbol{L}_2$ 为四个角子区域的中心视点,以 $\boldsymbol{L}_1$ 中的中心视点作为参考图像,彼此之间相互独立编码。第三个视口层 $\boldsymbol{L}_3$ 为剩余子区域的中心视点,其参考图像列表中有四幅参考图像,以图3中 $\boldsymbol{L}_3$ 最上方视点为例,其参考图像分别为 $\boldsymbol{L}_1$ 中的中心视点、$\boldsymbol{L}_2$ 中上方的两个中心视点以及通过视点合成网络得到的与编码图像相同位置的合成视点。第四个视口层 $\boldsymbol{L}_4$ 为每个子区域中的四个角视点,中

心子区域的四个角视点以 $\boldsymbol{L}_4$ 中的中心视点为参考图像,四个角子区域的四个角视点则以 $\boldsymbol{L}_1$ 中的中心视点以及 $\boldsymbol{L}_2$ 中对应子区域中心视点为参考图像,其余子区域参考列表中有四幅参考图像,以图 3 中 $\boldsymbol{L}_4$ 的视点为例,其参考图像分别为 $\boldsymbol{L}_1$ 中的中心视点,$\boldsymbol{L}_2$ 中下方线条所指的两个视点以及 $\boldsymbol{L}_3$ 中该子区域的中心视点。本文将视点合成网络集成在 HEVC 参考软件 HM-16.14 中。HEVC 中存在一个解码图像缓存(Decoder Picture Buffer, DPB),该缓存器通常存储先前编码的帧,并且编码器在帧间预测期间搜索参考帧。所提方法在解码完 $\boldsymbol{L}_1$ 和 $\boldsymbol{L}_2$ 后,从 DPB 中选择对应的一组图像作为输入,经过视点合成网络,将得到的合成参考图像放在参考图像列表中,进而进行帧间编码。

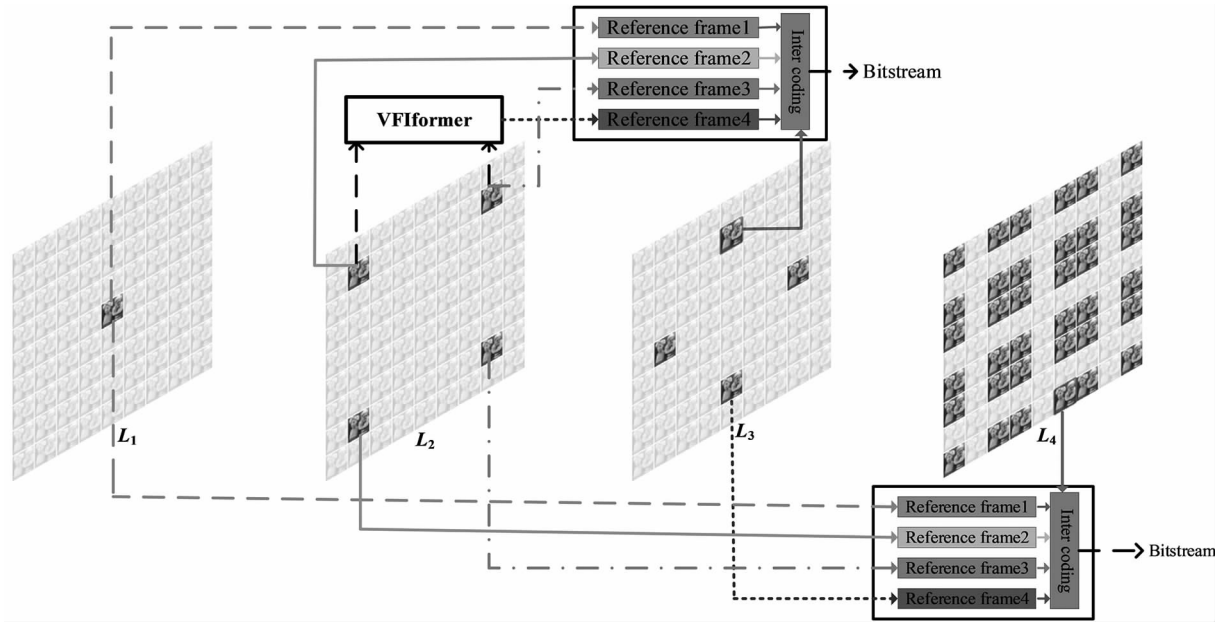


图 3 多参考帧预测关系示意图

Fig. 3 Diagram of the prediction relationship of multiple reference frames

2.4 解码重建

对于最后一个视口层 $\boldsymbol{L}_5$ 中的视图,不使用编码器对其进行编码,而是通过基于深度学习的视点合成网络进行视图合成,得到所需要的图像,如图 4 所示,将码率集中分配给前面四个视口层中的视图,可有效提升各层视图的质量。公式如下:

$$\boldsymbol{L}_5^{\mathrm{syn}} = f(\boldsymbol{L}_4^{\mathrm{rec}}) \tag{1}$$

其中,$\boldsymbol{L}_5^{\mathrm{syn}}$ 表示合成的第五个视口层视点;$\boldsymbol{L}_4^{\mathrm{rec}}$ 表示重建后的 $\boldsymbol{L}_4$,$f(\cdot)$ 表示视点合成操作。最终解码光场 $\boldsymbol{I}^{\mathrm{rec}}$ 由下式得到:

$$\boldsymbol{I}^{\mathrm{rec}} = \{\boldsymbol{L}_1^{\mathrm{rec}}, \boldsymbol{L}_2^{\mathrm{rec}}, \boldsymbol{L}_3^{\mathrm{rec}}, \boldsymbol{L}_4^{\mathrm{rec}}, \boldsymbol{L}_5^{\mathrm{syn}}\} \tag{2}$$

在光场图像编码中,基于视图合成的方法仅对光场图像视图的稀疏集进行编码,未编码的视图在解码端利用解码视图进行合成。主要方法包括光场图像超分辨率以及普通视频插帧。本文所选择的视点合成网络为视频插帧网络 VFIformer[12]。该网络利用 Transformer 对视频帧之间的长程像素相关性进行建模。此外,还使用了一种新颖的基于跨尺度窗口的注意力机制,在该机制中,跨尺度窗口相互作用,这样的设计有效地扩大了感受野并聚合了多尺度信息。

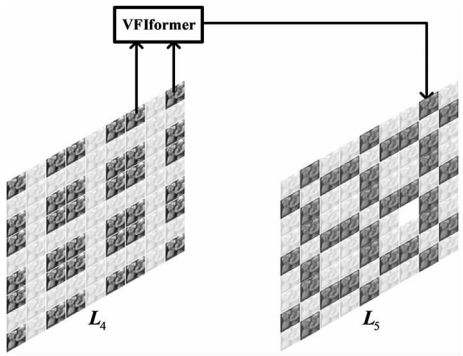


图 4 视图合成示意图

Fig. 4 Diagram of view synthesis

然而,利用视频插帧网络合成目标视图存在一个问题,即对于需要合成的同一位置的目标视图,所选择输入插帧网络的输入视图不同,得到的目标视图也不同。为了进一步说明这一点,图 5 中使用不同的输入视图组合进行插帧来获得相同位置的目标

视图,即使用两个角视点组合来作为输入视图以及距离目标视图较近的两个视点组合作为输入图像,分别来得到目标视图。图 5(b)和图 5(c)为通过不同视图组合进行插帧得到的结果及其相对于原图的放大的局部残差图,图 5(b)为两个角视点组合作为输入得到的结果,其 PSNR 为 46.6059 dB;图 5(c)为距离目标视图较近的两个视点组合作为输入得到的结果,其 PSNR 为 49.7898 dB。从图中可以看出,当选择不同的输入视图组合进行插帧时,由于光场图像的复杂性,生成的目标视图可能在细节和角度上存在差异。通常而言,距离目标视图较近的视图与目标视图的相关性更高,使用这样的输入组合进行插帧可以更好地保持光场图像的角度一致性,提高生成目标视图的质量。这种基于输入视图选择的优化策略有助于更有效地应对光场图像的复杂性,提高合成结果的整体性能。

(a)不同的插帧方式示意图　　(b)角视点输入组合得到的结果及残差　　(c)邻视点输入组合得到的结果及残差

图 5　不同输入组合得到的插帧结果

Fig. 5 Interpolation results obtained with different input combinations

3　实验结果与分析

3.1　实验配置

将所提方法在 EPFL Light-field[13] 和 HCI[14] 数据库上进行了性能测试。实验中选用了 4 个真实场景和 4 个合成场景。真实场景采用中心 9×9 的 SAI 阵列,为了适应编码器对编码块大小的需求,将空间分辨率裁剪为 432×624。合成场景保持原始分辨率,即角度分辨率为 9×9,空间分辨率为 512×512。每个颜色分量的位深为 10 位。根据 ITU-R 建议 BT.709-6[15],在不进行下采样的情况下将视图转换为 YCbCr 格式。HCI 光场图像也从 8 比特转换为 10 比特和4:4:4 YCbCr 格式。这样设计旨在验证方法在真实和合成光场中的适用性和性能。图 6 和图 7 显示了选用场景的中心 SAI 缩略图。

3.2　编码性能

实验中通过计算 Bjontegaard Delta bitrate(BD-BR)[16] 以及 Bjontegaard Delta peak signal-to-noise rate(BD-PSNR)来衡量编码性能。负的 BD-BR 值意味着在相同质量下,所提算法相比于基准算法可以节省码率,反之则意味着要消耗更多的码率;正的 BD-PSNR 意味着,在相同码率下,所提算法相比于基准算法可以提升质量,反之则意味着降低质量。码率用每像素比特(bits per pixel,bpp)来衡量,计算如下:

(a)Bikes　　(b)Danger de Mort　　(c)Stone Pillars Outside　　(d)Fountain & Vincent 2

图 6　真实光场图像中心 SAI

Fig. 6 The central SAI of testing light field images with real scene

(a) Bicycle (b) Greek (c) Kitchen (d) Table

图 7 合成光场图像中心 SAI

Fig. 7 The central SAI of testing light field images with synthetic scene

$$bpp = \frac{R_{\mathrm{LF}}}{x \times y \times s \times t} \tag{3}$$

其中，R_{LF}表示码流的大小，(x,y)和(s,t)分别表示光场的空间分辨率和角度分辨率，本文取 4 个码率，分别为 $R_1 = 0.005$ bpp、$R_2 = 0.02$ bpp、$R_3 = 0.1$ bpp、$R_4 = 0.75$ bpp。另外，每个场景的质量用所有 SAI 的平均峰值信噪比 $PSNR_{\mathrm{avg}}$来表示，计算如下：

$$PSNR_{\mathrm{avg}} = \frac{\sum^{N} PSNR_{\mathrm{code}} + \sum^{M} PSNR_{\mathrm{syn}}}{N_{\mathrm{total}}} \tag{4}$$

其中，$PSNR_{\mathrm{code}}$代表需要被编码 SAI 的质量；$PSNR_{\mathrm{syn}}$表示合成位置 SAI 的质量；N 为被编码 SAI 的数量，实验中 N 取 45；M 表示合成视点的数量，实验中 M 取 36；N_{total}表示所有 SAI 的总数量，实验中 N_{total}取 81。这里采用 YCbCr 颜色空间中 Y 通道的 $PSNR_{\mathrm{avg}}$来衡量得到结果的质量。选择了三种对比方法来评估所提出方法的性能，分别为 LSM[17]，MuLE[18]，JPEG Pleno Anchor[19]，其中 LSM 将所有 SAI 按照线性顺序扫描为伪视频序列，再由 HEVC 进行编码；MuLE 为 JPEG Pleno 中的变换模式；JPEG Pleno Anchor 为 JPEG Pleno 通用测试条件中所提出的对比方法，其将所有 SAI 按照蛇形顺序扫描为伪视频序列，再由 HEVC 进行编码。

图 8 给出了 LSM、MuLE、JPEG Pleno Anchor 和所提方法的率失真（Rate-Distortion，RD）曲线，Y-PSNR表示仅在 Y 分量上求取 $PSNR_{\mathrm{avg}}$。表 1 给出了所提方法以其他三种对比方法为基准时的编码性能比较。相比于 LSM、JPEG Pleno Anchor 方法，所提方法在真实场景下分别取得 15.340 % 和 16.166 % 的平均码率节省，在合成场景下分别平均节省了 6.983 % 和 6.796 % 的码率。而且所提方法重建光场图像的质量也有所提升，特别是在低码率下。在考虑了随机访问性能的前提下，所提方法的编码性能优于这两种对比方法。相比于 MuLE 方法，所提方法在真实场景和合成场景中分别取得 10.111 % 和 52.442 % 的平均码率节省。

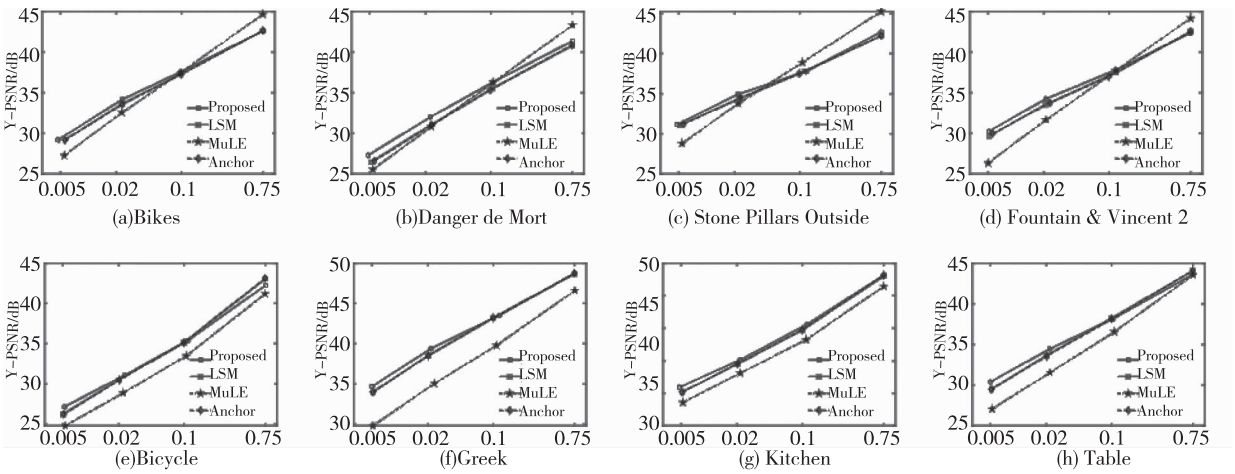


图 8 测试光场图像的率失真曲线

Fig. 8 Rate distortion curves of testing light field images

表 1　编码性能比较

Tab. 1 Coding performance comparison

LF Images	Proposed vs. LSM		Proposed vs. MuLE		Proposed vs. Anchor	
	BDBR/%	BDPSNR/dB	BDBR/%	BDPSNR/dB	BDBR/%	BDPSNR/dB
Bikes	-12.508	+0.331	-16.943	+0.443	-13.660	+0.369
Danger de Mort	-25.214	+0.815	-12.297	+0.343	-25.942	+0.844
Stone Pillars Outside	-7.940	+0.243	+15.313	-0.317	-8.103	+0.243
Fountain &Vincent 2	-15.697	+0.405	-26.518	+1.032	-16.959	+0.427
Average	**-15.340**	**+0.450**	**-10.111**	**+0.375**	**-16.166**	**+0.471**
Bicycle	-1.974	+0.034	-43.939	+1.844	-1.694	+0.018
Greek	-9.104	+0.275	-72.234	+3.839	-8.250	+0.251
Kitchen	-12.797	+0.450	-46.954	+2.147	-11.635	+0.404
Table	-4.055	+0.137	-46.642	+2.015	-5.603	+0.180
Average	**-6.983**	**+0.224**	**-52.442**	**+2.461**	**-6.796**	**+0.213**

3.3　随机访问

JPEG Pleno 在其光场图像编码通用测试条件标准[19]下定义关于随机访问的一种度量。在标准中，随机访问惩罚(Random Access Penalty，RAP)被定义为访问感兴趣区域(Region of Interest，RoI)所需的编码比特数与编码比特数总数之比，如下所示：

$$RAP = \frac{encoded\ bits\ required\ to\ access\ RoI}{encoded\ bits\ to\ decode\ the\ full\ LF} \quad (5)$$

本文中 RoI 为某一单个视图，标准中指出如果 RoI 是单个视图，则应给出访问任一视点所需的最高随机访问惩罚，即访问一个视点所需解码的最大比特数与解码整个光场所需比特数之比。因此，本文所涉及的随机访问惩罚以单个视图的随机访问为准。为了避免高随机访问代价，所提方法以比特率增加为代价独立地对每一个视口层中的视图进行编码。例如，仅使用来自 $\boldsymbol{L}_1$ 的中心子孔径图像作为帧间编码的参考图像，独立地对 $\boldsymbol{L}_2$ 中的四幅视图进行编码。为了不过于降低编码效率，$\boldsymbol{L}_3$ 中的视图使用合成视图来作为参考图像，这导致随机访问代价有所提高，如果用户对随机访问的要求较高，第三层可不用合成参考图像进行编码，但这会导致编码性能的下降。每个测试码率下访问一个视点所需解码的最大比特率如表 2 所示，随机访问惩罚如图 9 所示，该图中深灰色代表解码整个光场图像所需的总比特率，浅灰色代表访问任一视点所需最大比特率。所提方法在真实场景下随机访问惩罚最高为20.34 %，最低为 17.99 %。在合成场景下随机访问惩罚最高为 18.21 %，最低为 15.56 %。可以看出，与 LSM、JPEG Pleno Anchor 方法相比，随机访问性能得到了明显的提升。

表 2　不同码率下每个场景访问任一视图所需最大比特率

Tab. 2 Maximum bit rate required for accessing any view in each scene at different bit rates

Scene	R_1	R_2	R_3	R_4
Bikes	0.0025	0.0098	0.0279	0.1428
Danger	0.0030	0.0103	0.0344	0.1543
Stone	0.0030	0.0099	0.0279	0.1354
Vincent	0.0030	0.0107	0.0326	0.1373
Avg	**0.0029**	**0.0102**	**0.0307**	**0.1425**
Bicycle	0.0026	0.0093	0.0291	0.1376
Greek	0.0021	0.0068	0.0245	0.1169
Kitchen	0.0024	0.0082	0.0296	0.1329
Table	0.0026	0.0088	0.0246	0.1254
Avg	**0.0024**	**0.0083**	**0.0270**	**0.1282**

3.4　视口可伸缩性

光场图像的视口可伸缩性是通过将光场子孔径阵列视图分到不同的层来实现的。这种可伸缩性增强了与显示器、网络条件等外在因素的兼容性。对于 2D 显示器，可能只需要中心子孔径图像，而立体显示器则可能需要多个子孔径图像。基于伪视频序列的方法使所有视图相互依赖，以充分利用子孔径图像之间的冗余来提高压缩效率。但是，为了访问任意视图，基于伪视频的方法需要将所有子孔径图像

都进行编码、解码和显示,从而导致一定的带宽损耗和解码延迟。考虑光场图像的视口可伸缩性可以有效地缓解这一情况。所提出方法将所有视图分为五层,每层都有一组光场图像视图。每层视图数量如图 10(a)所示。在码率 R_4 下为每个视口层分配的码率如图 10(b)所示。为提升编码性能以及合成视图的质量,码率集中在前四个视口层中,第五个视口层通过视点合成网络获取。实验结果表明,客户端可根据自己的需求渐进式请求多个适合其需要的层,也可以通过随机访问快速获取所需视点。例如,要访问 Greek 中第 1 层到第 3 层的所有视图,客户端只需要下载 0.1706 bpp,而要访问所有层,则需要 0.7511 bpp。

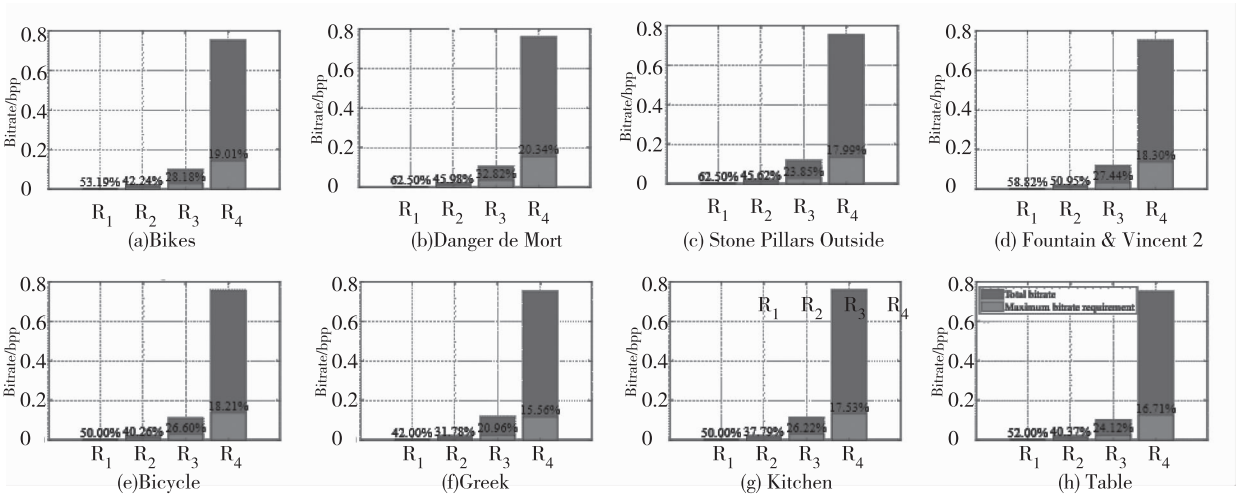
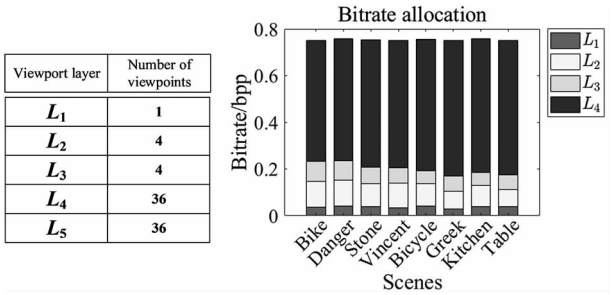


图 9 不同码率下每个场景的随机访问惩罚

Fig. 9 Random access penalty for each scene at different bitrates



Viewport layer	Number of viewpoints
L_1	1
L_2	4
L_3	4
L_4	36
L_5	36

(a)每个视口层视点数量 (b)R_4 下每层的码率分配

图 10 码率分配

Fig. 10 Bitrate allocation

3.5 质量一致性

当用户观看不同视图时,不同视图之间质量波动会显著影响用户的体验质量。表 3 给出了所提方法的解码光场质量一致性结果,表中数字为每个解码子孔径阵列 PSNR 值的方差。由于所提方法使用了插帧网络,因此需要考虑插帧网络对光场图像一致性的影响。所提方法通过设置量化参数以及每个视口层的量化步长来确保需要编码图像的质量一致性,由于所采用的插帧方式是通过相邻的视点来插帧生成新的视点,所以生成的视点质量波动不会很大;而在高码率的情况下,编码图像的质量较高,由于插帧网络达到上限,插帧得到的视图质量相对编码图像来说相对较低,然而高码率下每个视点的平均质量都比较高,在主观上没有很明显的差别,因此可以忽略高码率下的编码视图和合成视图的质量差异,如果用户对质量一致性的需求较高,则可考虑分配一定的码率来编码残差以保证质量一致性。

表 3 不同码率下的质量一致性(方差)

Tab. 3 Quality consistency(variance)at different bitrates

Scene	R_1	R_2	R_3	R_4
Bike	0.022	0.049	0.160	2.984
Danger	0.022	0.048	0.069	0.466
Stone	0.016	0.041	0.105	0.477
Vincent	0.037	0.075	0.160	3.882
Bicycle	0.027	0.063	0.080	6.450
Greek	0.027	0.016	0.065	7.000
Kitchen	0.041	0.042	0.079	7.012
Table	0.042	0.213	0.771	9.042

4 结 论

本文提出了一种面向增强随机访问的光场图像可伸缩编码方法。为了满足客户端不同应用的需求,所提方法对光场图像数据之间的预测结构做出优化,以在编码效率与随机访问性能之间寻

找平衡点。前四个视口层通过 HEVC 编码器进行编码,并且通过合成参考视图来提升一定的编码性能;而将码率集中在前四层,有利于提高合成视图以及整体解码视图的质量。在实验阶段,选择了当前经典的方法分别在真实场景和合成场景进行对比实验。实验结果表明,在编码性能上,所提方法在真实场景和合成场景上都优于对比方法,真实场景下的平均 BDBR 减少了 16.166 %,合成场景下的平均 BDBR 减少了 6.796 %;在随机访问性能方面,本文所提方法相对于 LSM 以及 Anchor 也得到了明显的提升。未来的研究重点将考虑如何充分利用解码图像来合成视点,以更有效地利用解码图像的空间域和角度域信息,从而进一步提升光场图像的整体质量。

参考文献:

[1] Wang T, Sheng H, Chen R, et al. Light field depth estimation: a comprehensive survey from principles to future [J]. High-Confidence Computing, 2024, 4(1): 100187.

[2] Attal B, Huang J B, Zollhöfer M, et al. Learning neural light fields with ray-space embedding [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 19819 – 19829.

[3] Jayaweera S S, Edussooriya C U S, Wijenayake C, et al. Multi-volumetric refocusing of light fields [J]. IEEE Signal Processing Letters, 2020, 28: 31 – 35.

[4] Carvalho M B D, Pereira M P, Alves G, et al. A 4D DCT-based lenslet light field codec [C]//2018 25th IEEE International Conference on Image Processing (ICIP), 2018: 435 – 439.

[5] Perra C, Astola P, Da Silva E A B, et al. Performance analysis of JPEG pleno light field coding [C]//Applications of Digital Image Processing XLII. SPIE, 2019, 11137: 402 – 413.

[6] Conceição R, Porto M, Zatt B, et al. LF-CAE: context-adaptive encoding for lenslet light fields using HEVC [C]//2018 25th IEEE International Conference on Image Processing (ICIP), 2018: 3174 – 3178.

[7] Li L, Li Z, Li B, et al. Quadtree-based coding framework for high-density camera array-based light field image [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(8): 2694 – 2708.

[8] Çetinkaya E, Amirpour H, Timmerer C. LFC-SASR: light field coding using spatial and angular super-resolution [C]//2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, 2022: 1 – 6.

[9] Huang X, An P, Chen Y, et al. Low bitrate light field compression with geometry and content consistency [J]. IEEE Transactions on Multimedia, 2020, 24: 152 – 165.

[10] Gomes P, da Silva Cruz L A. Pseudo-sequence light field image scalable encoding with improved random access [C]//2019 8th European Workshop on Visual Information Processing (EUVIP). IEEE, 2019: 16 – 21.

[11] Monteiro R J S, Rodrigues N M M, Faria S M M, et al. Light field image coding with flexible viewpoint scalability and random access [J]. Signal Processing: Image Communication, 2021, 94: 116202.

[12] Lu L, Wu R, Lin H, et al. Video frame interpolation with transformer [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 3532 – 3542.

[13] Rerabek M, Yuan L, Authier L A, et al [iso/iec jtc 1/sc 29/wg1 contribution] epfl light-field image dataset [J]. Physics Engineering, 2015: 60862229.

[14] Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields [C]//Computer Vision-ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20 – 24, 2016, Revised Selected Papers, Part III 13. Springer International Publishing, 2017: 19 – 34.

[15] BT I T U R. Parameter values for the hdtv standards for production and international programme exchange [M]. International Telecommunication Union, Recommendation, May, 2002.

[16] Barman N, Martini M G, Reznik Y. Revisiting bjontegaard delta bitrate (BD-BR) computation for codec compression efficiency comparison [C]//Proceedings of the 1st Mile-High Video Conference, 2022: 113 – 114.

[17] Dai F, Zhang J, Ma Y, et al. Lenselet image compression scheme based on subaperture images streaming [C]//2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2015: 4733 – 4737.

[18] Alves G D O, De Carvalho M B, Pagliari C L, et al. The JPEG pleno light field coding standard 4D-transform mode: how to design an efficient 4D-native codec [J]. IEEE Access, 2020, 8: 170807 – 170829.

[19] Pereira F, Pagliari C, da Silva E A B, et al. JPEG pleno light field coding common test conditions v3.3 [J]. Doc. ISO/IEC JTC, 2021, 1.