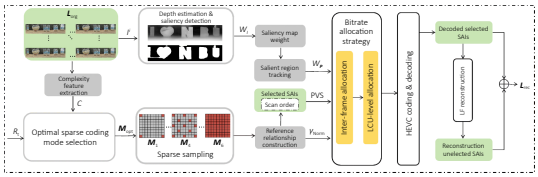
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 5 月, 第 53 卷, 第 5 期

目标码率与图像内容感知的宽视场光场图像编码



朱　磊，陈晔曜，封哲宇，金攀麒，蒋刚毅*

宁波大学信息科学与工程学院，浙江 宁波 315211

摘要：宽视场光场成像能以更大视野反映场景中光线方向与强度信息，其巨大的数据量也对编码压缩技术提出更高要求。针对宽视场光场图像数据量大、目标码率可变及内容不确定的编码挑战，提出一种目标码率与图像内容感知的编码方法。首先，构建稀疏采样模式集，基于光场图像复杂度特征与目标码率约束，通过 K 近邻算法自适应优选最佳稀疏采样模式，降低选择复杂度；其次，以关键视图帧间距离为依据设计帧级码率分配策略，提升角度维质量一致性；最后，结合场景显著性、深度与区域复杂度，提出 LCU 级内容感知码率分配策略，实现细粒度资源优化。用户端通过深度学习网络重建非关键视图，获得完整光场图像。实验结果表明，所提方法能根据目标码率与光场图像内容快速地获得优选的稀疏采样模式，显著降低稀疏采样模式选择的复杂度，提升在网络传输条件变化下的宽视场光场图像编码性能；对比现有方法，平均实现 45.284% 码率节省与 1.641 dB 峰值信噪比提升，且角度维一致性得到增强，随机访问惩罚最高仅 20.29%，表现出良好的综合编码性能。

关键词：宽视场光场图像；稀疏采样模式；码率分配策略；角度维质量一致性

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250320 | CSTR: 32245.14.oee.2026.250320 | 中图分类号：TP391 | 文献标识码：A

朱磊，陈晔曜，封哲宇，等. 目标码率与图像内容感知的宽视场光场图像编码 [J]. 光电工程，2026，**53**(5): 250320

Zhu L, Chen Y Y, Feng Z Y, et al. Target bitrate and image content-aware wide field-of-view light field image coding[J]. *Opto-Electron Eng*, 2026, **53**(5): 250320

1　引　言

光场成像能同时记录自由空间中光线的强度与方向信息[1]，是实现沉浸式体验的关键技术之一，已被应用于视差/深度估计[2-3]、超分辨率重建[4-5]、数字重聚焦[6]与显著性检测[7]等。现有光场相机因硬件限制导致成像视场窄，利用光场拼接技术可扩展视场[8-9]，形成宽视场光场图像；然而，更宽视场的光场图像包含更高的数据规模，给高效存储与传输带来了挑战。

光场图像可表示为视图阵列，由于其视图内与视图间的固有冗余而具有高度可压缩性[10]。现有基于视图预测的光场图像编码研究主要有两条技术路线：采用传统视频编码框架的方式，以及稀疏编码与重建的方式。第一条技术路线主要利用已有视频编码框架来压缩光场图像，对视图阵列进行扫描得到伪视频序列，利用帧内和帧间预测技术来去除光场图像冗余。相关扫描策略主要包括蛇形扫描[11]、锯齿扫描和螺旋扫描[12]等。Conceicao 等[13]设计了一种上下文自适应编码方案，其通过计算当前视图和其他视图间的差异自适应地排列成伪视频，再由 H.265/HEVC 进行压缩。Li 等[14]提出了一种基于伪序

收稿日期：2025-10-28；修回日期：2026-02-02；录用日期：2026-02-11；网络出版日期：2026-05-25

基金项目：国家自然科学基金项目 (62271276, 62401301)；浙江省自然科学基金项目 (LQ24F010002)

*通信作者：蒋刚毅，jianggangyi@nbu.edu.cn

版权所有©2026 中国科学院光电技术研究所

列的二维分层编码结构，并设计三种算法来优化二维编码结构。Yu 等[15] 提出一种基于增强重采样重建和角度注意力估计的光场传输方法，先对原始光场图像进行下采样得到低分辨率光场图像，再将其转换为伪视频序列进行编码，在传输到用户端后，使用残差图对低分辨率光场图像进行上采样和重建。这些方法与现有编码标准相兼容且能去除大部分数据冗余，但对所有视图进行编码可能会限制其编码效率与随机访问性能、增加复杂度。

第二条技术路线是基于稀疏编码与重建的方式，对光场图像视图阵列进行采样，分解为关键视图与非关键视图，仅对关键视图进行编码，而对非关键视图则由用户端的视图合成网络来执行重建，得到完整的光场图像。Huang 等[16] 对选取的关键视图及非关键视图的视差图进行编码，关键视图是通过对光场图像视图阵列沿着水平方向进行整列视图采样而成，在用户端利用解码后的关键视图及非关键视图的视差图通过视图绘制方式重建完整光场图像。Bakir 等[17] 从 8×8 光场图像视图阵列中固定采集 16 个视图，只对这 16 个关键视图进行 H.266/VVC 编码，而其余非关键视图则在解码端采用具有双鉴别器的生成对抗网络进行重建。Liu 等[18] 提出了一种多流密集视图重建网络，从 8×8 光场图像视图阵列中固定地采集 8 个视图作为关键视图，在用户端重建时，先基于解码的关键视图构建多视差几何结构，再设计一种多流视图重建网络来重建非关键视图形成完整光场。Jin 等[19] 设计了双码流方式进行光场图像编码，采用多种稀疏采样模式产生关键视图集合，并利用光场图像中心视图方差进行稀疏采样模式选择。上述方法通过仅对稀疏的关键视图进行编码，提高了编码效率；但对光场图像内容与码率可变情况下的优化编码还可进一步研究，包括对网络传输带宽的适应性、角度维质量一致性、显著性区域的感知编码性能等。

近年来，深度学习的快速发展为光场图像处理、压缩与重建提供有效手段。Tong 等[20] 提出了一种用于光场图像编码的端到端空间角度去相关网络，利用卷积网络对光场图像进行信息解耦以压缩相应的空间和角度信息。Feng 等[21] 提出了一种端到端的光场图像编码方法，使用解耦表示和非对称条带卷积来提高编码效率。Takhtardeshir 等[22] 提出了一个基于变分自动编码器的光场图像编码框架。基于深度学习的光场图像编码研究发展较快，但由于与现有视频编码标准不兼容，限制了其实际应用。

综上所述，在基于稀疏编码与重建的光场图像编码中，为了获得更高的编码性能，应根据网络传输过程中的目标码率变化、被编码光场图像内容特征合理地选择关键视图集合，即最优稀疏采样模式是动态变化的。显然，如何根据目标码率与光场图像内容设计网络传输适应性好、综合编码性能优的宽视场图像编码方法是该领域的一个重要挑战。为此，本文提出一种目标码率与图像内容感知的宽视场光场图像编码方法，根据网络传输的目标码率与光场图像的内容特征，设计一种宽视场光场图像稀疏采样模式自适应优选方案，在优化编码过程与提升编码效率的同时显著减少了计算复杂度。同时，在帧级与最大编码单元 (largest coding unit, LCU) 级编码过程中，设计不同的码率分配策略，提高了可变码率下宽视场光场图像编码性能。其主要贡献点如下：

1) 针对目标码率可变与光场图像内容差异，设计了一种基于光场图像内容特征与目标码率约束的稀疏采样模式自适应优选方案。该方法能够在可变码率下高效确定优选的稀疏采样模式，在保持与现有编码标准兼容的同时，降低计算复杂度、提高编码效率与随机访问性能。

2) 针对解码光场图像角度维质量一致性问题，提出了将关键视图集合中当前编码帧与其参考帧的帧间距离作为帧级码率预分配的依据，明显减小了传统编码框架中码率分配策略导致编码后视图质量不一致现象。

3) 为了提升敏感区域编码质量，设计了一种显著性区域跟踪的 LCU 级感知码率分配策略，在保证压缩非显著性区域质量的同时有效提升显著性区域的质量。

2 所提方法

为了适应网络传输带宽变化、提高感知编码性能、增强角度维质量一致性并保留良好的随机访问特性，本文提出了一种目标码率与图像内容感知的宽视场光场图像编码方法，如图 1 所示。

首先，设计了一种面向可变目标码率 R_{t} 与宽视场光场图像 $\boldsymbol{L}_{\mathrm{org}}$ 的稀疏采样模式自适应优选方案，其根据宽视场光场图像 $\boldsymbol{L}_{\mathrm{org}}$ 的复杂度特征 C，在不同 R_{t} 约束下，通过 K 近邻 (K near neighbor, KNN) 算法预测得到优选稀疏采样模式 $\boldsymbol{M}_{\mathrm{opt}}$，并对关键视图构建预测关系、采用 H.265/HEVC 视频编码技术进行压缩。其次，为提升编码后光场图像的角度维一致性，将关键视图集合中的 I 帧与各视图的帧间距离作为质量一致性调节的依据，实

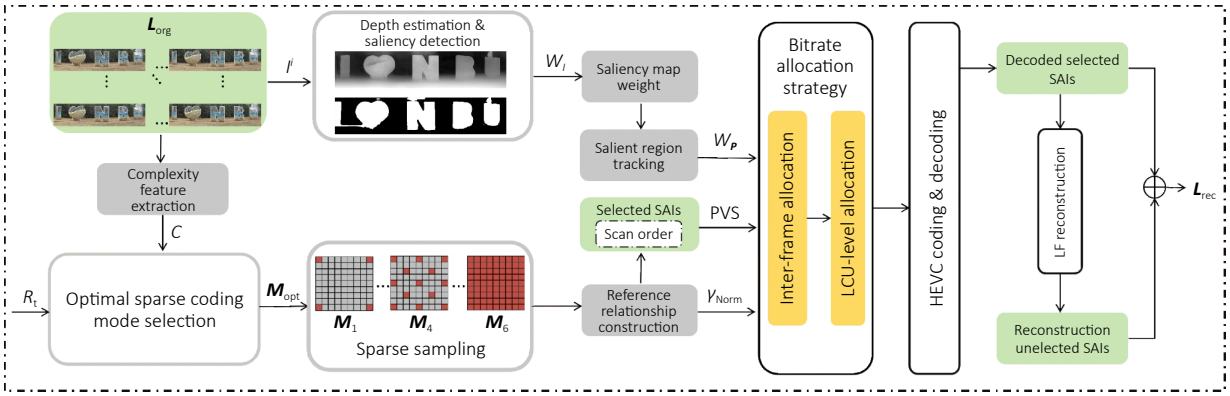


图 1 所提目标码率与图像内容感知的宽视场光场图像编码方法框图

Fig. 1 Framework of the proposed wide field-of-view (FOV) light field image coding method with target bitrate and content awareness

现帧级码率预分配方案。之后，以关键视图集合中的 I 帧为基础，通过显著性检测与深度估计生成其 LCU 级显著性权重，并结合显著性区域在角度维平移关系传播得到其余关键视图的显著性权重图。最后，将预分配的码率与 LCU 的显著性与复杂度联合，在 H.265/HEVC 平台中实现细粒度码率分配并进行关键视图编码，在用户端利用深度学习网络重建非关键视图以合成完整光场图像 $\boldsymbol{L}_{\rm rec}$。

2.1 稀疏采样模式选择

当目标码率 $R_{\rm t}$ 可变和宽视场光场图像 $\boldsymbol{L}_{\rm org}$ 内容不确定情况下，选择适配的模式与编码参数是实现编码性能高与网络传输适应性好的宽视场光场图像压缩有效手段。$\boldsymbol{L}_{\rm org}$ 可表示为视图阵列，令 $\boldsymbol{S}_{i,j}$ 是 $\boldsymbol{L}_{\rm org}$ 中第 (i, j) 个视图，则 $\boldsymbol{L}_{\rm org}=\{\boldsymbol{S}_{i,j};\ 1\leqslant i\leqslant N_I,\ 1\leqslant j\leqslant N_J\}$。在基于稀疏编码与重建的光场图像编码方式中，$\boldsymbol{L}_{\rm org}$ 可分解为关键视图集合 $\boldsymbol{L}_{\rm key}$ $(\boldsymbol{L}_{\rm key}\subseteq\boldsymbol{L}_{\rm org})$ 与非关键视图集合 $\boldsymbol{L}_{\rm non}$，即 $\boldsymbol{L}_{\rm org}=\{\boldsymbol{L}_{\rm key},\ \boldsymbol{L}_{\rm non}\}$。$\boldsymbol{L}_{\rm key}$ 作为 $\boldsymbol{L}_{\rm org}$ 的子集，可采用稀疏视图采样方式 $\varphi(\cdot)$ 而产生：

$$\boldsymbol{L}_{\rm key}=\varphi(\{\boldsymbol{S}_{i,j}\},\{m_{i,j}\})\,. \tag{1}$$

式中：$\varphi(\cdot)$ 表示当 $m_{i,j}=1$ 时，$\boldsymbol{L}_{\rm org}$ 在 (i, j) 位置的视图 $\boldsymbol{S}_{i,j}$ 为关键视图；反之，为非关键视图；$\boldsymbol{M}_{\rm key}=\{m_{i,j};\ 1\leqslant i\leqslant N_I, 1\leqslant j\leqslant N_J\}$称为稀疏采样模式。

令 $\boldsymbol{L}'_{\rm key}$ 是 $\boldsymbol{L}_{\rm key}$ 被编码传输到用户端并解码得到关键视图集合，即

$$\boldsymbol{L}'_{\rm key}=\phi(\boldsymbol{L}_{\rm key})\,, \tag{2}$$

式中：$\phi(\cdot)$ 是对关键视图集合的伪视频编码方法。

在用户端，利用重建算法 $\zeta(\cdot)$ 由 $\boldsymbol{L}'_{\rm key}$ 对 $\boldsymbol{L}_{\rm non}$ 重建得到 $\boldsymbol{L}'_{\rm non}$，$\boldsymbol{L}'_{\rm non}=\zeta(\boldsymbol{L}'_{\rm key})$。显然，$\boldsymbol{L}_{\rm key}$ 的编码失真也影响到 $\boldsymbol{L}_{\rm non}$ 的重建质量。用户端重建宽视场光场图像为 $\boldsymbol{L}_{\rm rec}=\{\boldsymbol{L}'_{\rm key},\ \boldsymbol{L}'_{\rm non}\}$。

$\boldsymbol{L}_{\rm org}$ 经稀疏采样编码传输后，在用户端解码并重建得到的光场图像 $\boldsymbol{L}_{\rm rec}$ 可进一步表示为

$$\boldsymbol{L}_{\rm rec}=\{\boldsymbol{L}'_{\rm key},\zeta(\boldsymbol{L}'_{\rm key})\}\,, \tag{3}$$

$$\boldsymbol{L}_{\rm rec}=\{\phi(\boldsymbol{L}_{\rm key}),\zeta(\phi(\boldsymbol{L}_{\rm key}))\}\,, \tag{4}$$

$$\boldsymbol{L}_{\rm rec}=\{\phi(\varphi(\boldsymbol{L}_{\rm org})),\zeta(\phi(\varphi(\boldsymbol{L}_{\rm org})))\}\,. \tag{5}$$

显然，用户端解码与重建的宽视场光场图像 $\boldsymbol{L}_{\rm rec}$ 的质量与 φ、ϕ、ζ 等编码参数紧密关联。为了提升 $\boldsymbol{L}_{\rm rec}$ 的质量，可根据 $\boldsymbol{L}_{\rm rec}$ 关于 $\boldsymbol{L}_{\rm org}$ 的失真 (D) 对 φ、ϕ、ζ 等编码参数进行优化选择，可表示为

$$D=\Theta_q(\boldsymbol{L}_{\rm org},\phi,\varphi,\zeta)\,, \tag{6}$$

其中：Θ_q 为失真计算函数，表示“输入视图阵列与编码参数，输出重建失真”的计算过程。

式 (6) 可为获取宽视场光场图像失真感知特性与编码参数的关联模型提供依据，以优化宽视场光场图像编码。根据视频图像编码的率失真代价函数 $J_{\rm cost}=R_{\rm t}+\lambda D$，对于 $R_{\rm t}$、$\boldsymbol{L}_{\rm org}$ 的编码优化过程可表示为

$$\arg\min_{R_{\rm t},\boldsymbol{L}_{\rm org},\varphi,\phi,\zeta}(D)\,. \tag{7}$$

当 ϕ 与 ζ 固定时，所采用的 φ (即稀疏采样方式) 在不同的 $R_{\rm t}$ 与 $\boldsymbol{L}_{\rm org}$ 下，其编码性能也可能不同。因此，可以根据 $R_{\rm t}$ 与 $\boldsymbol{L}_{\rm org}$，通过统计分析，设计出稀疏采样模式自适应优选方法。

为避免光场相机成像时的渐晕效应，仅选用其成像中心的 9×9 视图阵列，采用典型稀疏视图配置方式，图 2 为角度分辨率为 9×9 的光场图像的 6 种代表性的稀疏采样模式[19]，这里，$N_I\times N_J=9\times 9$。图中每种稀疏采样

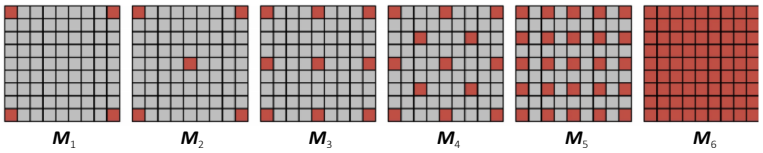


图 2 光场图像的稀疏采样模式集合 $\boldsymbol{M}$

Fig. 2 Set $\boldsymbol{M}$ of sparse sampling modes for the light field image

模式对应各自的关键视图集合 $\boldsymbol{L}_{\rm key}$，并按一定规则形成伪视频进行编码 (如式 (3) 所示)；这样，形成稀疏采样模式集合 $\boldsymbol{M}$={$\boldsymbol{M}_1$, $\boldsymbol{M}_2$, …, $\boldsymbol{M}_6$}。图中红色区域表示需要编码的关键视图，灰色区域表示无需编码、仅在用户端通过视图重建网络[23] 生成的非关键视图。模式 $\boldsymbol{M}_1$ 表示最稀疏的采样结构，其仅保留四角点视图；而模式 $\boldsymbol{M}_6$ 则表示最密集的采样结构，此时 $\boldsymbol{L}_{\rm org}$ 中全部 9×9=81 个视图均为关键视图并参与编码，没有非关键视图及其相应的重建。

对于不同的 $R_{\rm t}$ 与 $\boldsymbol{L}_{\rm org}$，所对应图 2 中的最佳稀疏采样模式可能不同。作为一个实例，表 1 为图 3(b) 中 4 幅不同宽视场光场图像在 5 个目标码率 (R_{01},R_{02},R_{03},R_{04},R_{05}) 下的最佳稀疏采样模式。这里，最佳稀疏采样模式采用了穷举搜索法 (brute force search, BFS)，以图 2 中每一种稀疏采样模式对光场图像分别进行稀疏编码与重建，再比较率失真性能，根据式 (7) 从 $\boldsymbol{M}$ 的 6 个模式中选出最佳稀疏采样模式 $\boldsymbol{M}_{\rm opt}$；显然，BFS 算法复杂度非常高。由表 1 可知，当 $R_{\rm t}$ 越小时，最佳稀疏采样模式中所包含的关键视图数量越少。给定 $R_{\rm t}$，不同光场图像 $\boldsymbol{L}_{\rm org}$ 由于其内容差异导致其所对应的最佳稀疏采样模式 $\boldsymbol{M}_{\rm opt}$ 可能不同。

显然，快速找到最佳稀疏采样模式不仅能提高编码效率还能减少编码复杂度。为了提升编码速度与编码效率，采用大量典型光场图像集合以合适的目标码率进行预编码实验、生成最佳稀疏采样模式集合，构建出相应的训练集，通过训练学习、生成不同目标码率与光场图像的最佳稀疏采样模式区间，为快速稀疏采样模式自适应优选方法的设计提供相关实验依据。训练集构建如下：

表 1 不同目标码率下的最佳稀疏采样模式 (R_1 至 R_5 表示目标码率值逐渐增大)

Table 1 Optimal sparse sampling patterns at different bitrates (R_1–R_5 denote increasing levels)

Scenes	R_1	R_2	R_3	R_4	R_5
I01	$\boldsymbol{M}_1$	$\boldsymbol{M}_2$	$\boldsymbol{M}_4$	$\boldsymbol{M}_5$	$\boldsymbol{M}_5$
I02	$\boldsymbol{M}_1$	$\boldsymbol{M}_2$	$\boldsymbol{M}_3$	$\boldsymbol{M}_5$	$\boldsymbol{M}_6$
I03	$\boldsymbol{M}_2$	$\boldsymbol{M}_2$	$\boldsymbol{M}_3$	$\boldsymbol{M}_5$	$\boldsymbol{M}_6$
I04	$\boldsymbol{M}_1$	$\boldsymbol{M}_1$	$\boldsymbol{M}_3$	$\boldsymbol{M}_5$	$\boldsymbol{M}_6$

1) 选定一组代表性光场图像{$\boldsymbol{L}_{\rm train}$ ($m_{\rm tr}$);1≤$m_{\rm tr}$≤$N_{\rm LFI}$}、一组典型目标码率值{$b_{\rm train}$ ($n_{\rm tr}$);1≤$n_{\rm tr}$≤$N_{\rm bit}$}。$m_{\rm tr}$ 是训练集光场图像的索引，$n_{\rm tr}$ 是目标码率的索引。

2) 以 $b_{\rm train}$ ($n_{\rm tr}$) 为目标码率对光场图像 $\boldsymbol{L}_{\rm train}$ ($m_{\rm tr}$) 分别采用图 2 中稀疏采样模式 $\boldsymbol{M}_1$, $\boldsymbol{M}_2$, …, $\boldsymbol{M}_6$ 进行编码压缩，并获得相应的最佳稀疏采样模式集合{$\boldsymbol{M}_{\rm train_opt}$ ($m_{\rm tr}$,$n_{\rm tr}$)}，

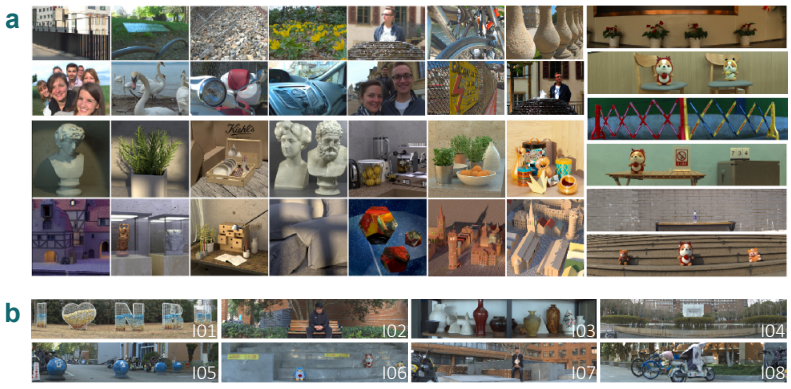


图 3 训练与测试光场图像集合的光场图像中心视图。(a) 训练集中代表性光场图像集合；(b) 测试光场图像集合

Fig. 3 Center views of the light field images from the training and testing light field image sets. (a) Training light field image set; (b) Test light field image set

$\boldsymbol{M}_{\text{train_opt}}(m_{\text{tr}},n_{\text{tr}})\in\boldsymbol{M}$。图 3 给出了相关训练集中代表性光场图像集合与测试用的光场图像集合，图中显示的是光场图像中心视图。图 3(a) 为训练集中代表性光场图像集合，包括 14 幅由 Lytro Illum 光场相机采集的窄视场光场图像 (取自于 EPFL 数据库[24])、14 幅合成窄视场光场图像 (取自于 HCI new 数据库[25])，以及 6 幅由拼接与融合算法[26] 产生的宽视场光场图像[27]。图 3(b) 为后续实验验证的测试光场图像集合，以此来分析被编码光场图像的最佳稀疏采样模式估计的准确性以及编码性能。

光场图像的视图帧内复杂度与视图帧间差异性可应用于编码效果分析。

1) 选取中心视图的信息熵 E 与梯度能量 G 作为视图帧内复杂度特征；信息熵用于衡量图像的全局信息丰富度，梯度能量则反映图像的局部纹理变化程度。

2) 以光场图像视图阵列的中心视图分别与四个角点视图间平均绝对差值的均值 M_{mad} 来描述视图间差异性。联合视图内复杂度与视图间差异性来描述光场图像 $\boldsymbol{L}_{\text{org}}$ 的复杂度特征 $\boldsymbol{C}=(E,G,M_{\text{mad}})$。当 ϕ、ζ、$\Theta_q(\cdot)$ 等确定，采用统计分类方式 $f_\varphi()$ 在给定 R_{t} 条件下估计出对 $\boldsymbol{L}_{\text{org}}$ 压缩的最佳稀疏采样模式 $\boldsymbol{M}_{\text{opt}}$:

$$\boldsymbol{M}_{\text{opt}}=f_\varphi(R_{\text{t}},C,\{L_{\text{trian}}(m_{\text{tr}})\},\{b_{\text{trian}}(n_{\text{tr}})\},\{\boldsymbol{M}_{\text{trian_opt}}(m_{\text{tr}},n_{\text{tr}})\})\,. \tag{8}$$

在具体执行中，基于目标码率与光场图像复杂度特征的稀疏采样模式自适应优选方案分为 2 个部分，即训练部分和测试部分。训练部分主要通过训练集进行统计分析生成不同光场图像复杂度与目标码率下最佳稀疏采样模式的码率分界点集合$\{b_{\text{p}}(i);\,i=1,\ldots,5\}$。测试部分如算法 1 所示，在测试光场图像集合中验证执行；算法 1 流程如下：

1) 对于目标码率 R_{t} 和输入光场图像 $\boldsymbol{L}_{\text{org}}$，先计算 $\boldsymbol{L}_{\text{org}}$ 的复杂度特征 $\boldsymbol{C}_{\text{test}}$，利用 KNN 算法预测光场图像编码的码率分界点$\{b_{\text{p}}(i);\,i=1,\ldots,5\}$，通过计算测试光场图像与训练集中光场图像复杂度特征的欧式距离 $d_{\text{Eu}}(m_{\text{tr}})$ 来衡量它们之间的相似性，如式 (9)，选取最近的 K 个相邻样本$\{\boldsymbol{N}_{\text{test}}(k);\,1\leqslant k\leqslant K\}$，并使用高斯加权方式对这些相邻样本的码率分界点进行加权，获得测试样本的预测码率分界点 $b_{\text{p-pred}}(i)$，如式 (10)，式 (11)。

2) 将目标码率 R_{t} 映射到对应区间：根据预测得到的码率分界点，将目标码率 R_{t} 映射到其所属的区间，进而判断其落在哪个模式范围内。

3) 返回匹配的最佳匹配模式，即为该测试样本在当前目标码率 R_{t} 下优选稀疏采样模式。

$$d_{\text{Eu}}(m_{\text{tr}})=\|C_{\text{test}}-C_{\text{train}}(m_{\text{tr}})\|_2^2,m_{\text{tr}}=1,2...,N_{\text{LFI}}\,, \tag{9}$$

$$w_k=\frac{\exp(-d_{\text{Eu}}(N(k))^2/(2\sigma^2))}{\sum_{j=1}^{K}\exp(-d_{\text{Eu}}(N(j))^2/(2\sigma^2))}\,, \tag{10}$$

$$b_{\text{p-pred}}(i)=\sum_{k=1}^{K}w_k\cdot b_{\text{p}}^{N_{\text{test}}(k)}(i)\,, \tag{11}$$

式中：$d_{\text{Eu}}(N(k))$表示最近的 K 个相邻样本对应的欧式距离值；w_k 表示高斯加权系数；σ 表示 K 个距离的中位数；对于 K 值，通过固定训练集与目标码率范围，将 K 从 3 调整至 10 后统计模式预测准确率与平均 BD-PSNR，结果显示 K=7 时综合性能最优；$\{b_{\text{p}}^{N_{\text{test}}(k)}(i);\,i=1,\ldots,5\}$表示相邻样本所对应的码率分界点。

算法 1：基于目标码率与光场图像特征的稀疏采样模式自适应优选方案

输入：样本集合 (即训练光场图像合)$\{\boldsymbol{L}_{\text{train}}(m_{\text{tr}})\}$、目标码率值集合$\{b_{\text{train}}(n_{\text{tr}})\}$、对应的最佳稀疏采模式集合$\{\boldsymbol{M}_{\text{train_opt}}(m_{\text{tr}},n_{\text{tr}})\}$；训练、测试样本的特征 $\boldsymbol{C}_{\text{train}}=((E_{\text{train}},G_{\text{train}},\boldsymbol{M}_{\text{train_ad}})$、$\boldsymbol{C}_{\text{test}}=(E,G,\boldsymbol{M}_{\text{ad}})$；目标码率 R_{t}、超参数 K，样本码率分界点 $b_{\text{p}}(i)$；

输出：优选的稀疏采样模式 $\boldsymbol{M}_{\text{opt}}$

1：输入测试样本的 $\boldsymbol{C}_{\text{test}}=(E,G,\boldsymbol{M}_{\text{ad}})$ 与目标码率 R_{t}

2：计算相似性：对所有训练样本，使用式 (9) 计算与测试样本距离 $d_{\text{Eu}}(m_{\text{tr}})$

3：选取与测试样本距离 $d_{\text{Eu}}(m_{\text{tr}})$ 最小的 K 个相邻样本 $N_{\text{test}}(k)$

4：通过式 (10) 计算相邻样本 $N_{\text{test}}(k)$ 的高斯加权系数 w_k

5：利用式 (11) 预测码率分界点 $b_{\text{p-pred}}(i)$

7：**for** 每个目标码率 $R_{\text{t}}(j)$ **in** $b_{\text{p-pred}}(i)$ **do**

8： **if** $R_{\text{t}}(j)<b_{\text{p-pred}}(1)$ **then**

9： $mode(j)\leftarrow 1$

10： **else if** $R_{\text{t}}(j)\geqslant b_{\text{p-pred}}(5)$ **then**

11： $mode(j)\leftarrow 6$

12： **else**

13： 查找满足 $b_{\text{p-pred}}(i)\leqslant R_{\text{t}}(j)<b_{\text{p-pred}}(i+1)$ 的 i

14： $mode(j)\leftarrow i+1$

15： **end if**

16：**end for**

17：返回 $\boldsymbol{M}_{\text{opt}}$

2.2 光场图像内容感知驱动的码率分配策略

在给定 R_{t} 对光场图像关键视图集所形成的伪视频进行 HEVC 压缩时，其帧级码率分配策略不同于普通视频

编码的情况，它不仅仅要考虑编码效率，更需要考虑控制帧间编码质量一致性，即控制编码后关键视图间的质量一致性。在 LCU 级码率分配上，需优先保证关键视图帧显著性区域的 LCU 具有更高的编码质量，其次是合理地保障非敏感区域 LCU 的编码质量，以获得更优的编码效果。

2.2.1 面向角度维质量一致性的帧级码率预分配

当前编码帧的参考帧的质量与当前帧的帧间距离是决定编码后光场图像角度维质量一致性的关键因素之一。采用现有的帧级码率平均分配策略会导致编码后视图间质量不一致现象。为此，提出一种基于关键视图集合中当前编码帧与其参考帧的帧间距离的帧级码率预分配方案，根据光场图像的空间特性来给每一帧图像分配合理的初始码率。具体而言，首先，以光场图像中心视图作为坐标原点，建立光场视图阵列的空间坐标系。在该坐标系下，获取当前编码帧及其所有参考帧的空间索引，并据此计算当前编码帧与各参考帧之间的空间距离。之后，对这些距离取均值，以此作为衡量当前帧初始码率分配的重要依据，表示如下

$$\rho(m_k)=\frac{1}{N_k}\sum_{n_k\in c(m_k)} d(m_k,n_k), \tag{12}$$

式中：$\rho(m_k)$ 表示当前帧与所有参考帧的帧间距离的均值；N_k 表示当前帧的所有参考帧的数量；$c(m_k)$ 表示第 m_k 帧的所有参考图像索引；$d(m_k,n_k)$ 为第 m_k 帧与其所对应的参考图像 n_k 的空间距离。

然后，构造了一个调节函数 $\gamma(m_k)$，用于根据平均参考距离动态预分配码率，其中平均参考距离越远，对应的调节系数 $\beta(m_k)$ 越大，所分配的码率也越多；反之，调节系数 $\beta(m_k)$ 越小，预分配的码率也相应减少。其帧级码率预分配过程可表示为

$$\gamma(m_k)=1+\beta(m_k)\rho(m_k), \tag{13}$$

$$R_{\mathrm{PicAvg}}(m_k)=\frac{R_t}{f_{\mathrm{ps}}}, \tag{14}$$

$$R_{\mathrm{Pic}}(m_k)=R_{\mathrm{PicAvg}}(m_k)\gamma_{\mathrm{Norm}}(m_k), \tag{15}$$

式中：$R_{\mathrm{PicAvg}}(m_k)$ 表示在平均码率分配策略下各帧分配的码率数；$R_{\mathrm{Pic}}(m_k)$ 为第 m_k 帧预分配的码率；R_{t} 为目标码率；f_{ps} 为帧率。为了保证码率预分配后总码率不变，将 $\gamma(m_k)$ 归一化为 $\gamma_{\mathrm{Norm}}(m_k)$。

2.2.2 显著性区域检测与 LCU 级感知码率分配

在光场图像编码中，结合 HEVC 中的编码单元层级划分策略，尤其在其 LCU 级别进行码率控制时，可利用由被编码帧的显著性图与深度图联合提取的视觉权重信息统一指导所有被编码关键视图的码率分配。该跨视图共享权重的策略不仅能有效降低计算开销，还能在保持主观视觉质量的前提下提升压缩效率。具体地，在显著性区域即使是轻微失真也可能显著降低感知质量；而对于非显著性区域，适当的编码失真通常难以被察觉，即对视觉感知质量影响有限。因此，在 LCU 级码率分配中优先保障这些显著区域的编码质量，给予其高码率分配，以实现更优的编码效果。

为衡量显著区域，先利用深度学习网络[28-29] 对关键视图集合形成的伪视频中 I 帧估计深度图 I_{u}^{i} 和显著图 I_{l}^{i}。再将得到的深度图和显著图分割成多个与编码最大单元 LCU 同尺寸的图像，然后通过求取每个 LCU 的深度值与显著性值，并对其进行归一化后加权相加，如下所示

$$W_I(i_{\mathrm{LCU}})=\alpha_k\hat{S}(i_{\mathrm{LCU}})+(1-\alpha_k)\hat{D}(i_{\mathrm{LCU}}), \tag{16}$$

式中：$\hat{D}(i_{\mathrm{LCU}})$ 和 $\hat{S}(i_{\mathrm{LCU}})$ 分别表示归一化后的第 i_{LCU} 个 LCU 的深度值和显著性值；$i_{\mathrm{LCU}}=1,2\ldots,N_{\mathrm{LCU}}$，$N_{\mathrm{LCU}}$ 为每帧的 LCU 数量；$W_I(i_{\mathrm{LCU}})$ 为 I 帧的第 i_{LCU} 个的显著性权重；加权系数 α_k 用于确定每个 LCU 的初步码率分配权重，并通过三组方案对比验证得到最佳系数：仅仅考虑显著性 (α_k=1.0) 易导致背景复杂区域失真，等权分配 (α_k=0.5) 会降低显著性区域优先级，而 α_k=0.75 时既可保证显著性区域高编码质量，又能通过深度信息优化空间结构编码，因此采用该加权系数。

对于密集光场而言，不同视点下的子视图在空间内容上具有较高的一致性，因此直接采用中心视图 I_{c} 的显著性区域来代表所有视图似乎是可行的。然而，由于视差效应和局部像素级差异，这种做法会导致跨视图的显著边缘存在模糊不一致，并可能引起显著性结构失真。图 4 展示了中心视图与角点视图在二值化显著性区域上的差异。

图 4(a) 为中心视图，图 4(b) 的白色区域表示中心视图中的显著性区域，其边缘连续且轮廓分明；图 4(c) 的差值 (即白色像素) 集中在左上角边缘区域，这是由于左上角视图相对于中心视图存在左上方向的观测位置偏移，受视差影响，近景物体的显著性边缘在左上角视图中会向左上方平移，与中心视图的边缘位置产生错位，进而形成差值；图 4(d) 的差值主要分布在右下方，这源于右下角视图的右下方向观测偏移，近景物体显著性边缘同步向右下方平移，与中心视图边缘形成错位；图 4(e) 的差值覆盖范围最广且呈对角分布，本质是两个角点视图的观测位置完全相反，显著性区域平移方向相反，导致



图 4 显著性区域差值图。(a) 中心视图；(b) 中心视图的显著图；(c) 中心视图显著图与左上角视图显著图的差值；(d) 中心视图显著性图与右下角显著性值差值；(e) 左上角显著性图与右下角显著性图差值

Fig. 4 Saliency region difference maps. (a) Central view; (b) Saliency map of the central view; (c) Difference between the central-view saliency map and the top-left-view saliency map; (d) Difference between the central-view saliency map and the bottom-right-view saliency map; (e) Difference between the top-left-view saliency map and the bottom-right-view saliency map

边缘错位程度最大化。这些差异与视点偏移方向相吻合，说明光场视图阵列中的显著性区域变化可看作近似平移关系，而非大幅度的形变或失真。

基于上述分析，为了降低显著性区域计算复杂度以及其余关键帧中显著区域的模糊或失真问题，确保 LCU 级权重生成的准确性，设计了一种基于空间平移传播的显著性区域扩展方法。具体地，以中心视图 I 帧的深度图为基础，通过视差估计计算每个其余关键帧视图相对于中心视图的视差偏移量$(\Delta u,\Delta v)$，偏移量大小与其余关键帧视图的观测位置、场景物体深度正相关，确保平移方向与幅度的精准性；其次基于计算得到的视差偏移量，对中心视图的显著区域执行像素级平移操作，使得其余关键帧视图的显著性区域与自身观测到的物体边缘位置对齐，消除差值图中存在的边缘错位，解决模糊失真问题；最后将平移后的其余关键帧视图显著区域与中心视图非显著区域的 LCU 权重结合，构建其余关键帧视图的完整 LCU 级权重分布。

$$I_{\mathrm{s}}^{p}(x,y)=Shift\left(I_{\mathrm{s}}^{i}(x,y);\Delta u,\Delta v\right), \tag{17}$$

式中：$I_{\mathrm{s}}^{p}(x,y)$为其余关键帧的显著性区域；*Shift* 为像素级平移操作；$(\Delta u,\Delta v)$表示其余关键帧视图的显著区域相对于 I 帧显著区域的视差偏移量；$I_{\mathrm{s}}^{i}(x,y)$为 I 帧视图的显著性区域。

在此基础上，进一步提出一种显著性驱动的 LCU 级感知码率分配策略，旨在码率分配过程中同时兼顾视觉感知的重要性与编码区域的结构复杂度。传统基于显著性权重的码率分配方案在实际编码光场图像中存在一定局限性，即对于显著性较低但纹理复杂度较高的区域，单一显著性约束往往导致其获得的码率资源不足，进而产生明显的编码失真，影响整体视觉体验的均衡性与稳定性。

为此，采用平均绝对偏差的动态加权调节机制，以提升码率分配的精细度。具体地，对于复杂度高于全局平均水平的区域，赋予权重增益以实现显著性区域的优先级显著提升，同时对非显著但复杂的区域给予合理补偿以缓解其在压缩过程中出现的视觉失真效应。

以伪视频序列的 I 帧为例，精细的感知码率分配过程如下：对于第 i_{LCU} 个 LCU，$W_I(i_{\mathrm{LCU}})$为 I 帧的第 i_{LCU} 个 LCU 的显著性权重，其 MAD[30] 值为 $M(i_{\mathrm{LCU}})$，当前编码帧所包含的 LCU 集合为 S_{LCU}，则其全局平均复杂度定义为

$$\mu_M=\frac{1}{N_{\mathrm{LCU}}}\sum_{j_{\mathrm{LCU}}\in S_{\mathrm{LCU}}}M(j_{\mathrm{LCU}}). \tag{18}$$

当 $M(i_{\mathrm{LCU}})$ 大于 μ_M 时，该 LCU 被认为内容复杂，需要额外的码率补偿，以防过渡压缩、导致失真明显；当 $M(i_{\mathrm{LCU}})$ 小于 μ_M 时，该 LCU 被认为内容简单，不做额外码率补偿，保持原有显著权重。对于显著性区域Ω^{S}与非显著性区域分别取不同的加权系数，如下所示

$$g(i_{\mathrm{LCU}})=\begin{cases}1+a_1 & B_{\mathrm{LCU}}\in\Omega^{\mathrm{S}}\\ 1+a_0 & B_{\mathrm{LCU}}\notin\Omega^{\mathrm{S}}\end{cases}, \tag{19}$$

式中：$a_1>a_0>0$。对于显著性区域 Ω^{S} 及非显著性区域的内容简单区域则使 $g(i_{\mathrm{LCU}})=1$，即保持原有显著性权重，有效权重为

$$W^{*}(i_{\mathrm{LCU}})=W_I(i_{\mathrm{LCU}})\cdot g(i_{\mathrm{LCU}}). \tag{20}$$

最终，LCU 的感知码率分配策略表示如下：

$$R_{\mathrm{LCU}}(i_{\mathrm{LCU}})=\frac{(R_{\mathrm{Pic}}(m_k)-R_{\mathrm{h}}-R_{\mathrm{c}})\times W^{*}(i_{\mathrm{LCU}})}{W_{\mathrm{sum}}+\varepsilon}, \tag{21}$$

式中：$R_{\mathrm{LCU}}(i_{\mathrm{LCU}})$ 为第 m_k 帧所有 LCU 所分配到的码率；$R_{\mathrm{Pic}}(m_k)$ 是由式 (15) 对当前帧 m_k 的预分配码率；R_{h} 为帧头部信息所消耗的实际码率；R_{c} 为当前帧已编码 LCU 花费的实际码率；W_{sum} 为第 m_k 帧所有 LCU 的有效权重之和；ε 为一个极小值，防止除零。

对于其余关键视图帧的 LCU 级码率分配，将上述流程中式 (21) 中的$W_I(i_{\mathrm{LCU}})$替换为对应关键帧显著性权重$W_{\mathrm{P}}(i_{\mathrm{LCU}})$，并将其中的$R_{\mathrm{Pic}}(m_k)$替换为对应关键帧在帧级所分配到的码率值。

综上码率分配流程分为以下几步。步骤 1：根据式

(12) 和式 (13) 得到的调节函数 $\gamma(m_k)$ 来确定每个编码帧预分配权重；步骤 2：由式 (14) 和 (15) 计算帧级的预分配码率 $R_{\mathrm{Pic}}(m_k)$；步骤 3：将每个编码帧划分为显著区域与非显著区域，并根据式 (16) 和 (17) 得到每个编码帧 LCU 级权重；步骤 4：通过式 (18)~(20) 对每个编码帧的 LCU 权重按照区域复杂度进一步进行加权得到最终每帧的 LCU 级权重 $W^*(i_{\mathrm{LCU}})$；步骤 5：将每帧的 LCU 级权重 $W^*(i_{\mathrm{LCU}})$ 代入式 (21) 得到每个 LCU 的所分配到的码率；步骤 6：当单个 LCU 编码完成后，由编码器的 R-λ 模型进行参数更新，直到每帧的每个 LCU 编码完成。

2.3 用户端宽视场光场图像重建

用户端的宽视场光场图像解码与重建如式 (2)、式 (3) 所描述的，先对压缩的关键视图进行解码得到 $\boldsymbol{L}'_{\mathrm{key}}$，再利用重建算法 $\zeta()$，由 $\boldsymbol{L}'_{\mathrm{key}}$ 重建非关键视图 $\boldsymbol{L}'_{\mathrm{non}}$，最后得到解码与重建后的完整宽视场光场图像 $\boldsymbol{L}_{\mathrm{rec}}=\{\boldsymbol{L}'_{\mathrm{key}},\zeta(\boldsymbol{L}'_{\mathrm{key}})\}$。

在由 $\boldsymbol{L}'_{\mathrm{key}}$ 重建 $\boldsymbol{L}'_{\mathrm{non}}$ 的过程中，将采用渐进式重建过程[31]，并采用视频插帧学习网络[23] 对非关键视图进行渐进式重建，其过程如图 5 所示。图 5 为 $\boldsymbol{L}'_{\mathrm{non}}$ 的渐进式重建过程 (以 $\boldsymbol{M}_1$ 为例)，此时，$\boldsymbol{L}'_{\mathrm{key}}=\{\boldsymbol{S}'_{1,1},\boldsymbol{S}'_{9,1},\boldsymbol{S}'_{9,9},\boldsymbol{S}'_{1,9}\}$，图 5(a) 为由插帧学习网络进行首次重建的 5 个非关键视图 $\{\boldsymbol{S}'_{1,5},\boldsymbol{S}'_{5,5},\boldsymbol{S}'_{5,1},\boldsymbol{S}'_{5,9},\boldsymbol{S}'_{9,5}\}$，图 5(b) 是以左上角为例的第 2、3 次渐近重建的非关键视图，分别为 $\{\boldsymbol{S}'_{1,7},\boldsymbol{S}'_{3,7},\boldsymbol{S}'_{3,7},\boldsymbol{S}'_{3,5},\boldsymbol{S}'_{5,7}\}$ 和 $\{\boldsymbol{S}'_{1,8},\boldsymbol{S}'_{2,8},\boldsymbol{S}'_{2,9},\boldsymbol{S}'_{2,7},\boldsymbol{S}'_{3,8}\}$。利用相似的方式可以重建其他非关键视图集，得到 $\boldsymbol{L}'_{\mathrm{non}}$；最终，得到 $\boldsymbol{L}_{\mathrm{rec}}=\{\boldsymbol{L}'_{\mathrm{key}},\boldsymbol{L}'_{\mathrm{non}}\}$。

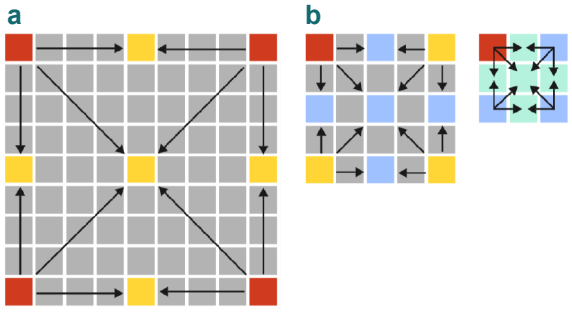


图 5 $\boldsymbol{L}'_{\mathrm{non}}$ 的渐进式重建过程 (以 $\boldsymbol{M}_1$ 为例，红色块为编码后的关键视图 $\boldsymbol{L}'_{\mathrm{key}}$)。(a) 初次重建$\{S'_{1,5}$、$S'_{5,5}$、$S'_{5,1}$、$S'_{5,9}$、$S'_{9,5}\}$; (b) 以左上角为例的第 2、3 次渐近重建

Fig. 5 Progressive reconstruction process of $\boldsymbol{L}'_{\mathrm{non}}$ (using $\boldsymbol{M}_1$ as an example, where red blocks indicate the encoded key views $\boldsymbol{L}'_{\mathrm{key}}$). (a) Initial reconstruction of $\{S'_{1,5}$、$S'_{5,5}$、$S'_{5,1}$、$S'_{5,9}$、$S'_{9,5}\}$; (b) Second and third stages of progressive reconstruction, taking the top-left region as an example

3 实验结果与分析

采用图 3(b) 所示的 8 个宽视场光场图像进行性能测试，其角度和空间分辨率为 9×9×1896×400。所提方法在 H.265/HEVC 参考软件 (HM16.20)、Pytorch (2.0.1) 和 Matlab (R2017b) 上进行测试。在实验中，所有视图均被转化为 4:4:4 YUV 格式，以五个目标码率 R_1、R_2、R_3、R_4、R_5 (即 0.0025 bpp、0.005 bpp、0.02 bpp、0.1 bpp、0.75 bpp) 进行编码。为了公平比较，所有指标在相同的硬件配置 (CPU: intel (R) Xeon (R) Silver 4210R CPU @ 2.4 Hz；GPU: NVIDIA Quadro RTX 5000) 下执行。

峰值信噪比 (Peak signal to noise ratio, PSNR) 和感知峰值信噪比 (perceptual peak signal to noise ratio, PPSNR)[32] 用于衡量解码图像的质量，计算公式如下

$$PSNR=10\log_{10}\times\frac{(2^b-1)^2\times L\times H}{\sum_{x=1}^{L}\sum_{y=1}^{H}(I(x,y)-I'(x,y))^2}, \tag{22}$$

$$PPSNR=10\log_{10}\times\frac{(2^b-1)^2\times L\times H}{\sum_{x=1}^{L}\sum_{y=1}^{H}(I(x,y)-I'(x,y))^2\delta(x,y)}, \tag{23}$$

式中：b 表示像素点的位深，此处取 8；L 和 H 分别表示图像的高度和宽度；I 和 I' 分别表示原始图像和解码图像的像素值；(x,y) 为像素的空间位置；$\delta(x,y)$ 在画面的显著性区域时为 1，在非显著性区域时为 0。若整帧是显著的，PPSNR 和 PSNR 是等效的。采用 Bjontegaard Delta-Rate (BD-rate)[33] 以及 Bjøntegaard Delta PSNR (BD-PSNR) 来衡量所提方法与现有方法的率失真性能。

为客观评估所提方法的有效性和先进性，将所提方法与现有代表性方法进行了对比，具体对比方法包括：

1) 旋转扫描法 (Rotation scanning method, RSM)[12]，其将所有宽视场光场图像螺旋扫描为伪视频，由 HEVC 进行编码压缩。

2) 线性扫描法 (Line scanning method, LSM)[12]，其将所有宽视场光场图像线性扫描为伪视频，再由 HEVC 进行编码压缩。

3) 上下文自适应编码 (Context-adaptive encoding, CAE)[13]，其将所有宽视场图像按照内容相似度自适应地扫描为伪视频，由 HEVC 进行编码压缩。

4) 穷举搜索法 (Brute forcesearch, BFS)，其联合所提框架进行编码，在六个模式中穷举搜索出编码效率最高的稀疏采样模式。

5) Jin[19] 所提的最佳稀疏采样模式选择方案联合所提框架进行编码。采用以上五种方法而非深度学习编码方法来验证所提方法先进性的核心原因在于技术路线与应用场景的本质差异：现有深度学习方法以端到端压缩性能为核心目标，网络架构与传统视频编码标准(H.265/HEVC)不兼容，难以直接融入现有光场存储与传输体系，且大多面向固定码率场景设计，缺乏对网络带宽动态变化的适应性；而所提方法以标准兼容、动态码率适配以及角度维质量一致性为核心目标，更贴合实际应用部署需求。

3.1 编码效率性能分析

为客观地评估各方法的压缩性能，表 2 和表 3 分别给出了多幅宽视场光场图像在 Y-PPSNR 和 Y-PSNR 指标下的 BD-BR 与 BD-PPSNR、BD-PSNR 对比结果。与 RSM、LSM 以及 CAE 方法相比，所提方法在 PPSNR 指标下的 BD-BR 上分别实现了 66.556%、64.670% 以及 61.469% 的码率节省，同时在显著性区域的 BD-PPSNR 指标上分别提升了 2.923 dB、2.702 dB 以及 2.527 dB，尤其在低码率下优势更加明显。与 Ref. [19] 联合所提方法框架的方法相比，所提方法在 BD-BR 上平均减少了 5.377%，并在显著性区域的 BD-PPSNR 上提升了 0.119 dB。这一性能提升主要来源于更高的稀疏采样模式匹配精度以及显著性区域合理的码率分配。尽管在与 BFS 方法的比较时，所提方法在 BD-BR 上平均增加了 0.637%，并伴随约 0.018 dB 的 BD-PPSNR 降低，但 BFS 通过穷举获得最优模式，计算成本极高，而所提方法在保持较低计算复杂度的前提下，仅以少量编码性能作为代价，换取了显著的编码效率提升和较好的随机访问能力。所提方法在 PSNR 指标下的 BD-BR 以及 BD-PSNR 均有与 PPSNR 类似的结果，这主要得益于稀疏采样带来的高匹配精度。

图 6 和图 7 展示了不同编码方法在 PPSNR 以及 PSNR 指标下的率失真曲线对比结果。可以看出，所提方法在两种指标下的五个目标码率点上均表现出较优的编码性能。相较于 RSM、LSM 以及 CAE 三种方法，所提方法在编码效率方面更具有优势。一方面，在低码率条件下，上述三种方法仍需对全部数据进行编码，这显著限制了压缩效率；另一方面，这些方法未考虑显著区域的差异化码率分配，难以将有限比特资源有效集中于

表 2 所提方法与其他宽视场光场图像编码方法在 PPSNR 指标上的编码性能对比

Table 2 Coding performance comparison of the proposed and other wide-FOV light field methods using PPSNR

Scenes	Proposed vs RSM[12]		Proposed vs LSM[12]		Proposed vs CAE[13]		Proposed vs Ref. [19]		Proposed vs BFS	
	BD-BR/%	BD-PPSNR/dB	BD-BR/%	BD-PPSNR/dB	BD-BR/%	BD-PPSNR/dB	BD-BR/%	BD-PPSNR/dB	BD-BR/%	BD-PPSNR/dB
I01	−56.257	+2.537	−54.762	+2.365	−46.198	+1.988	−1.607	+0.060	+0.814	−0.025
I02	−61.663	+2.673	−56.565	+2.215	−59.262	+2.407	−10.972	+0.265	+0.000	+0.000
I03	−78.745	+2.806	−74.183	+2.405	−72.536	+2.330	−7.717	+0.133	+0.000	+0.000
I04	−63.761	+3.006	−59.969	+2.668	−57.751	+2.511	+0.680	−0.005	+0.828	−0.026
I05	−62.943	+2.868	−65.727	+2.928	−58.891	+2.515	−4.647	+0.125	+0.000	+0.000
I06	−75.373	+3.140	−75.681	+2.996	−73.104	+2.824	−9.398	+0.121	+0.443	−0.012
I07	−68.072	+3.094	−66.476	+2.899	−63.831	+2.780	−7.847	+0.233	+0.660	−0.016
I08	−65.631	+3.264	−63.998	+3.143	−60.181	+2.863	−1.505	+0.022	+2.351	−0.066
Avg.	**−66.556**	**+2.923**	**−64.670**	**+2.702**	**−61.469**	**+2.527**	**−5.377**	**+0.119**	**+0.637**	**−0.018**

表 3 所提方法与其他宽视场光场图像编码方法在 PSNR 指标上的编码性能对比

Table 3 Coding performance comparison of the proposed and other wide-FOV light field methods using PSNR

Scenes	Proposed vs RSM[12]		Proposed vs LSM[12]		Proposed vs CAE[13]		Proposed vs Ref. [19]		Proposed vs BFS	
	BD-BR/%	BD-PSNR/dB	BD-BR/%	BD-PSNR/dB	BD-BR/%	BD-PSNR/dB	BD-BR/%	BD-PSNR/dB	BD-BR/%	BD-PSNR/dB
I01	−37.875	+1.480	−36.481	+1.435	−30.831	+1.183	−4.454	+0.194	+4.528	−0.138
I02	−42.290	+1.471	−35.726	+1.175	−32.937	+1.092	−7.345	+0.196	+0.000	−0.000
I03	−43.322	+1.070	−45.205	+1.235	−41.416	+1.009	−1.192	+0.029	+0.000	−0.000
I04	−46.524	+1.896	−38.929	+1.523	−41.524	+1.639	−0.423	+0.034	+3.558	−0.123
I05	−48.528	+1.929	−42.787	+1.636	−44.458	+1.720	−9.014	+0.262	+0.000	−0.000
I06	−55.287	+1.832	−54.942	+1.841	−51.087	+1.662	−11.887	+0.280	+1.215	−0.033
I07	−48.390	+1.855	−46.590	+1.774	−43.678	+1.622	−5.297	+0.179	+0.647	−0.018
I08	−40.058	+1.595	−38.041	+1.489	−33.608	+1.291	−2.849	+0.074	+3.418	−0.088
Avg.	**−45.284**	**+1.641**	**−42.337**	**+1.513**	**−39.942**	**+1.402**	**−5.307**	**+0.156**	**+1.671**	**−0.050**

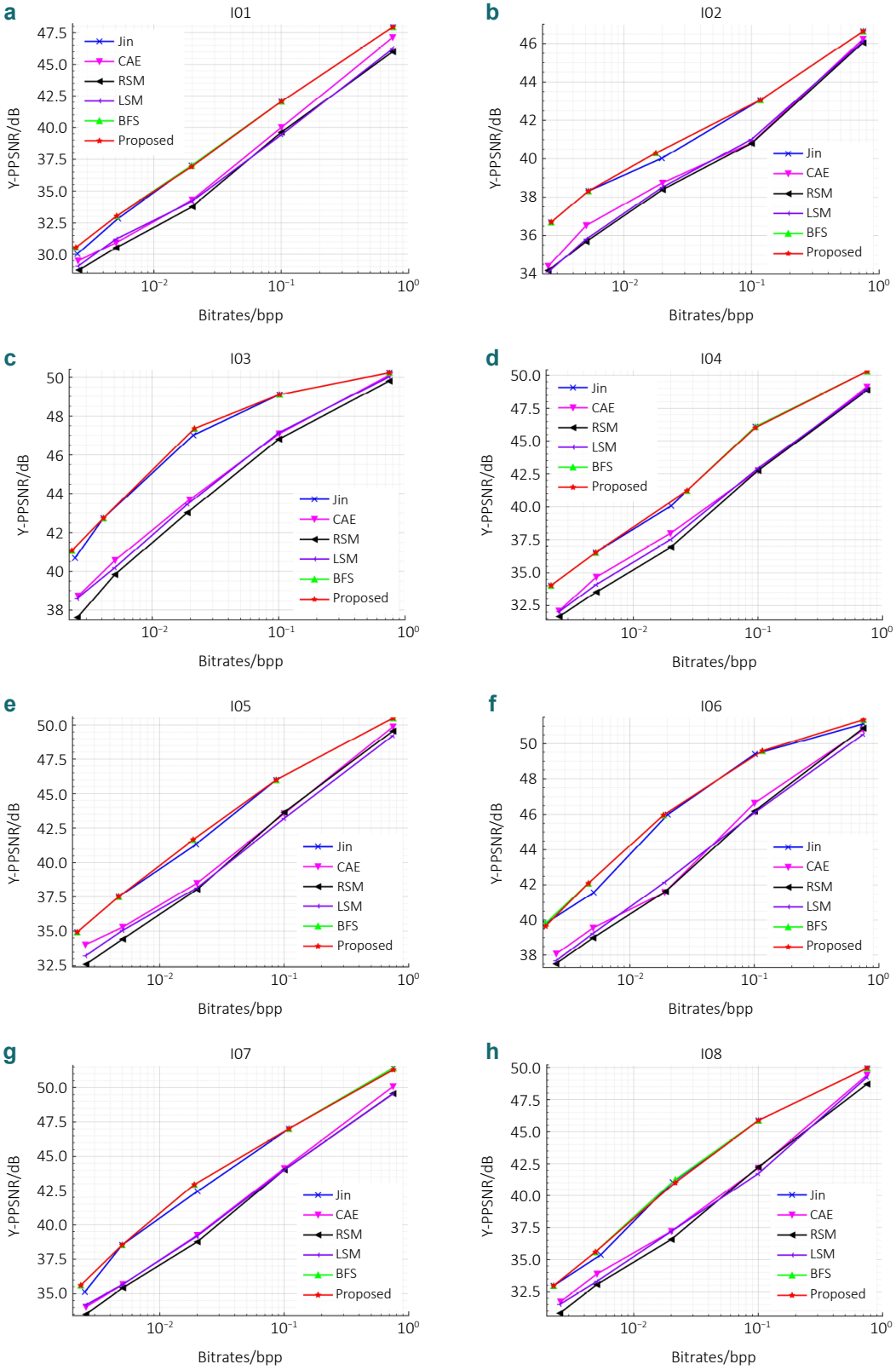


图 6 不同编码方法在 PPSNR 指标下的率失真性能对比。(a) I01；(b) I02；(c) I03；(d) I04；(e) I05；(f) I06；(g) I07；(h) I08

Fig. 6 Comparison of rate-distortion performance of different coding methods under the PPSNR metric. (a) I01；(b) I02；(c) I03；(d) I04；(e) I05；(f) I06；(g) I07；(h) I08

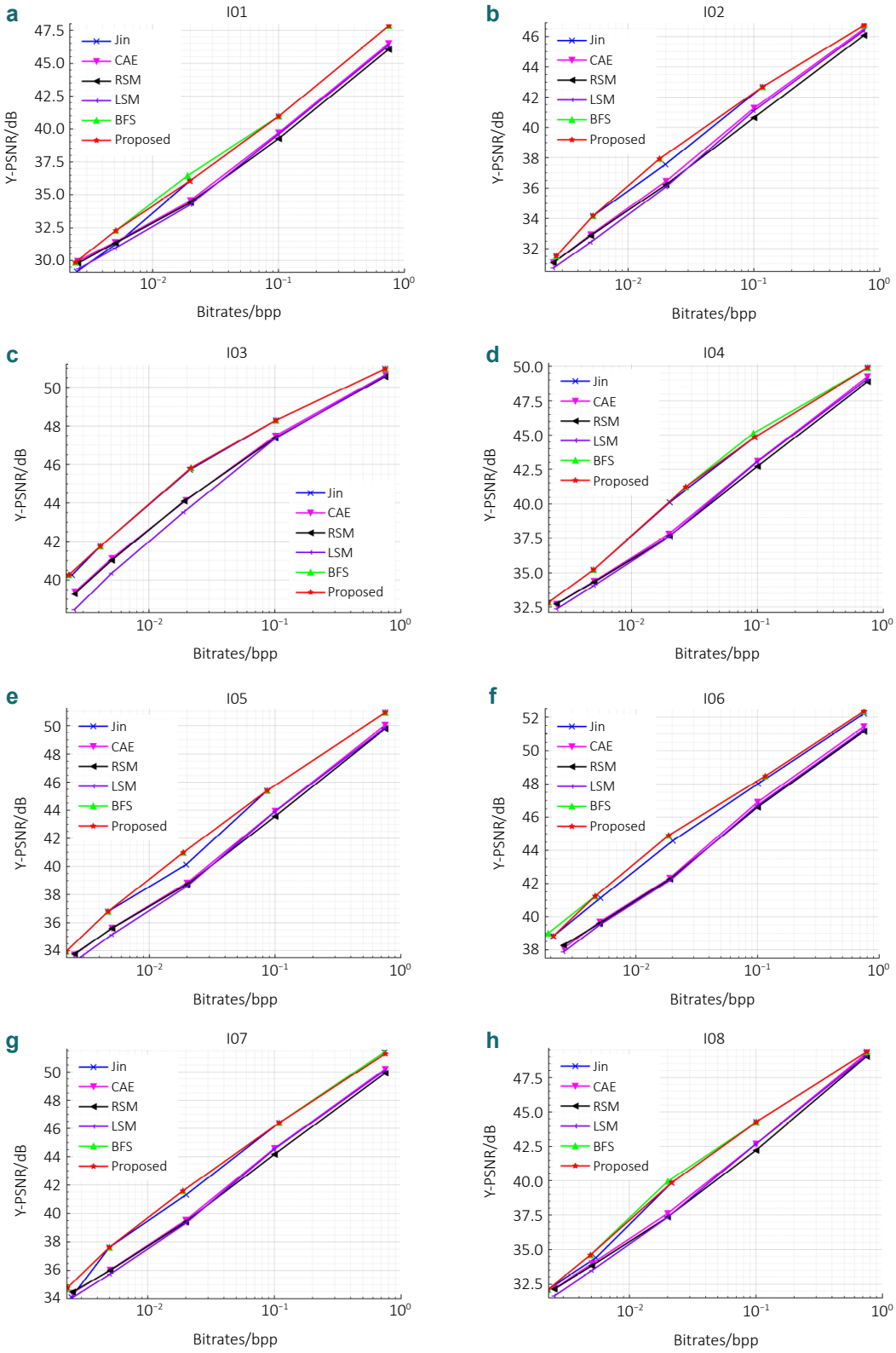


图 7 不同编码方法在 PSNR 指标下的率失真性能对比。(a) I01；(b) I02；(c) I03；(d) I04；(e) I05；(f) I06；(g) I07；(h) I08

Fig. 7 Comparison of rate-distortion performance of different coding methods under the PSNR metric. (a) I01； (b) I02； (c) I03； (d) I04； (e) I05； (f) I06； (g) I07； (h) I08

图像关键区域。进一步可以观察到，在低码率区间内，所提方法的性能优势更为明显。这主要是由于当目标码率高度受限时，若采用固定或偏密集的稀疏采样模式，有限的比特资源需要分配到过多关键视图中，容易导致关键视图质量下降，并在非关键视图重建过程中产生误差累积。相比之下，所提方法在低码率条件下倾向于选择关键视图数量较少、结构更稳定的稀疏采样模式，从而在有限码率预算内优先保障关键视图的基础质量，并为后续视图预测与重建提供更可靠的参考。

与 Jin 联合所提方法框架的方法相比，所提方法从编码角度出发，综合考虑复杂度特征以辅助选择最佳稀疏采样模式，从而在稀疏采样模式集内更有效地匹配最佳模式。随着模式选择准确率的提升，整体编码性能也随之提高。同时，结合显著性权重引导的 LCU 级码率分配策略，使得更多码率资源集中分配到感知敏感区域，进一步提升感知质量。虽然 BFS 方法可通过穷举方式确保稀疏采样模式最优、达到略高的编码性能，但这一过程伴随着极高的计算开销。相比之下，所提方法在保持优良压缩性能的同时，具备更高的计算效率。

图 8 展示了解码光场图像的中心视图的视觉比较结果。其中，红色矩形框标注的是显著性区域，蓝色矩形框标注的是非显著但内容复杂区域，GT 表示原始图像。显然，所提方法考虑了显著性区域的帧间与 LCU 级码率分配策略，能更好地保留显著性区域细节。例如，在光场图像 I02 中，所提方法在图 8(a) 中获得的 PSNR 值相较于其他三种方法均有提升，残差也相应减少。对于非显著但内容复杂度较高的区域，所提方法并未进行过度压缩，而是通过加权策略合理分配码率，从而保证了较好的视觉质量，如图 8(b) 所示。尽管编码所消耗的码率略高于其他三种方法，但在主观视觉效果和客观指标提升上是明显的。此外，在所提方法中，部分视图在用户端是通过重建网络合成得到的。因此所设计的码率分配策略不仅作用于编码的视图，同时也影响后续合成的视图，从而进一步提升整体编码效果。

3.2 随机访问性能分析

为评估编码方案的随机访问性能，采用 JPEG Pleno[34] 中的“随机访问惩罚 (random access penalty, RAP)”指标，它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

$$RAP = \frac{Bit_{\mathrm{RoI}}}{Bit_{\mathrm{LF}}}, \tag{24}$$

式中：对于光场图像编码，RoI 是指单一视图，所以计算 RAP 时应给出访问一个视点所需的最大访问惩罚，即访问一个视点所需解码的最大比特数 Bit_{RoI} 与解码整个光场所需比特数 Bit_{LF} 之比。所提方法对于不同的稀疏采样模式构建了不同的参考关系，同时减少层与层之间的视点依赖关系并适当增加码率来避免高随机访问代价。

如图 9 所示，以光场图像 I03 为例，该图展示了在测试码率下的随机访问惩罚图，横坐标以五个码率 R_1、R_2、R_3、R_4、R_5 表示，并以 bpp 衡量。蓝色区域表示解码整个光场图像所需的总码率，黄色区域表示随机接入惩罚度量且黄色区域包含于蓝色区域。所提方法在五个码率下的随机接入惩罚随着码率的升高而降低。

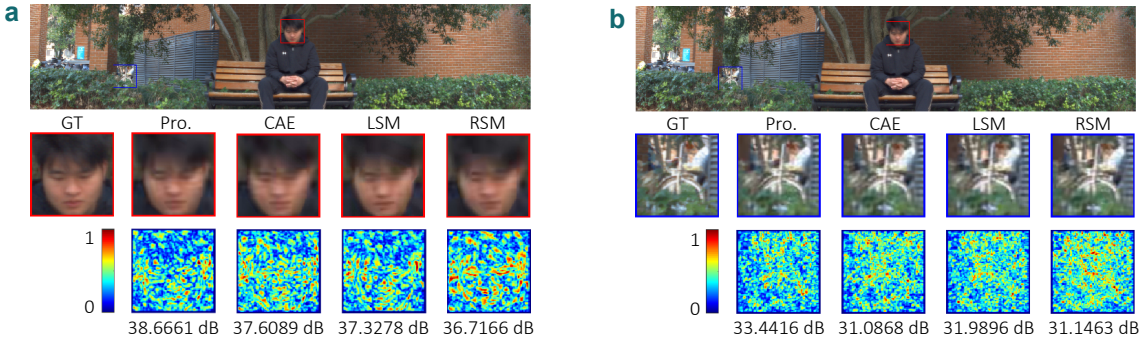


图 8 光场图像场景 I02 的 *I* 帧解码主观感知对比。所提方法的码率分别为 0.00527 bpp，CAE[13]、LSM[12] 以及 RSM[12] 方法码率为 0.00506 bpp。(a) 显著性区域视觉对比结果；(b) 非显著但内容复杂区域视觉对比结果

Fig. 8 Subjective comparison of *I*-frame decoding for light field scene I02. The proposed method uses a bitrate of 0.00527 bpp, while CAE [13], LSM[12], and RSM[12] use a bitrate of 0.00506 bpp. (a) Visual comparison results of salient regions; (b) Visual comparison results of non-salient but content-complex regions

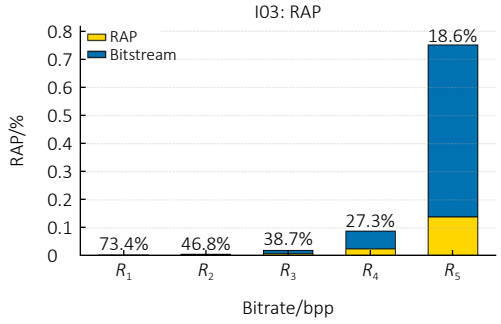


图 9 不同码率下随机访问惩罚

Fig. 9 Random access penalty at different bitrates

表 4 给出了所提方法在测试码率下访问一个视点所需解码的最大比特率。表 5 给出了所提方法在不同场景下的 RAP 值，可以看出所提方法在全码率范围内的平均 RAP 值为 19.03%，最大 RAP 值仅 20.29%，且随着码率升高，RAP 稳步下降，说明在网络带宽充足时，随机访问性能更优；对比现有方法 RSM、LSM 以及 CAE，这些方法需要解码全部视图才能访问到单个视图。

3.3 角度维质量一致性分析

用户端在接收解码的关键视图时，过大的质量波动会引起视觉体验的不适。因此，解码光场图像中的各视图应保持较为接近的质量。解码光场图像角度维质量一致调节的加入使得所提方法在整体上有着更好光场角度维质量一致性。图 10 给出了光场图像 I01 在 R_1 和 R_4 使用角度一致调节和不使用角度一致调节的视点质量波动情况。可以发现，在不使用角度一致调节的情况下，解码光场图像角度维质量方差分别达到 0.308 和 0.434，而使用角度一致调节后，方差下降到 0.166 和 0.160。这主要是由于在帧间码率的合理分配使得稀疏编码的宽视场光场图像质量一致性较好。尽管重建算法会带来额外的失真，但在视图间质量一致性较好的情况下且使用合理的渐进式重建方法会有效减少角度维质量波动。

表 6 给出了解码光场图像使用质量一致调节 (Use) 和未使用质量一致调节 (Noneuse) 的质量方差值，其中，使用质量一致调节视图间的质量波动明显降低，但随着码率以及图像编码复杂度的增加，重建失真的影响逐渐

表 4 不同码率下每个光场图像访问任一视图所需最多码率

Table 4 Maximum bitrate required to access any view of each light field image at different bitrates

Scenes	R_1	R_2	R_3	R_4	R_5
I01	0.0021	0.0027	0.0076	0.0327	0.1425
I02	0.0022	0.0025	0.0085	0.0284	0.1526
I03	0.0016	0.0022	0.0072	0.0236	0.1395
I04	0.0021	0.0024	0.0081	0.0364	0.1483
I05	0.0021	0.0036	0.0102	0.0311	0.1454
I06	0.0020	0.0029	0.0077	0.0319	0.1512
I07	0.0020	0.0026	0.0098	0.0377	0.1467
I08	0.0021	0.0025	0.0082	0.0345	0.1155
Avg.	**0.0020**	**0.0026**	**0.0084**	**0.0328**	**0.1427**

表 5 所提方法在不同场景下的 RAP 值

Table 5 The RAP values of the proposed method in different scenes

Scenes	R_1/%	R_2/%	R_3/%	R_4/%	R_5/%
I01	84.30	54.62	38.53	32.71	19.05
I02	88.50	50.11	42.50	28.40	20.29
I03	73.40	46.80	38.70	27.30	18.60
I04	84.70	48.70	40.50	36.40	19.77
I05	84.10	72.50	51.25	31.13	19.38
I06	80.60	58.18	38.51	31.97	20.16
I07	80.22	52.57	49.31	37.70	19.56
I08	84.00	50.41	41.01	34.58	15.54
Avg.	80.11	52.50	42.22	32.81	19.03

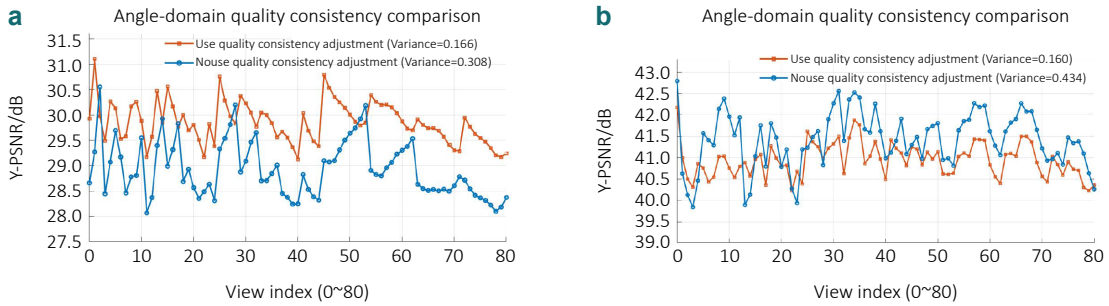


图 10 角度维质量一致对比。(a) 光场图像 I01，开启质量一致调节码率为 0.00244 bpp，未开启码率为 0.00270 bpp；(b) 光场图像 I01，开启质量一致调节码率为 0.10042 bpp，未开启码率为 0.10180 bpp

Fig. 10 Comparison of angular quality consistency. (a) I01, bitrate with quality consistency enabled; 0.00244 bpp, without quality consistency; 0.00270 bpp; (b) I01, bitrate with quality consistency enabled; 0.10042 bpp, without quality consistency: 0.10180 bpp

增加。但使用质量一致调节后整体的质量方差不会超过 1。对于 R_5，由于解码的宽视场光场中的视图的视觉质量较好，即使存在质量的波动，人眼对其也并不敏感。

3.4 稀疏采样模式选择分析

为验证所提稀疏采样模式选择方法的准确性，本文将其与基于 BFS 的穷举方法以及 Ref. [19] 提出的最佳稀疏采样模式选择方法进行对比。表 7 给出了不同目标码率下三种方法的稀疏采样模式选择结果。可以看出，Jin 方法在中低码率条件下存在一定偏差，主要由于其仅依赖单一复杂度特征，难以全面表征光场图像的编码特性。BFS 方法通过穷举搜索获得最优结果，但计算复杂度较高。相比之下，所提方法在各码率条件下均能较准确地匹配最优或近似最优模式，在显著降低计算开销的同时保持了稳定的编码性能。

3.5 稀疏采样模式选择的时间复杂度分析

本节对本文所提出的稀疏采样模式选择的复杂度以及 BFS 法搜索最佳模式的复杂度进行了对比实验。测试时间为计算单个场景 (以 I03 为例) 的最佳稀疏采样模式的运行时间，包括重建过程。表 8 给出了算法 1 (Algorithm 1) 稀疏采样模式自适应优选方案与 BFS 方法在稀疏采样模式选择阶段的实际运行时间的对比以及所提方法和 BFS 方法在每个码率点下选择最佳稀疏采样模式并且进行编码及重建的总运行时间对比。

可知 Algorithm 1 模式选择耗时仅为 BFS 方法的 0.11%~0.3%，所提方法的完整编码及重建时间仅为 BFS 的 29.08%。这是由于 BFS 方法需在各目标码率下遍历所有候选模式并重复执行编码与重建。所提方法只需要对编码的光场图像的复杂度特征进行提取并采用算法 1 中的模式选择就可以得出最优或者次最优的优选模式，然后直接进行编码及重建。

3.6 跨库实验

为验证所提方法在模式选择任务中的跨数据集泛化能力，本小节开展了基于 HCI old[35] 数据集的跨数据集

表 6 解码光场图像的质量方差值

Table 6 Quality variance of the decoded light field images

Scenes	R_1		R_2		R_3		R_4	
	Nonuse	Use	Nonuse	Use	Nonuse	Use	Nonuse	Use
I01	0.308	0.166	0.315	0.188	0.434	0.160	0.434	0.160
I02	0.412	0.212	0.388	0.251	0.631	0.151	0.534	0.297
I03	0.311	0.114	0.214	0.127	0.314	0.108	0.802	0.244
I04	0.295	0.182	0.319	0.106	0.721	0.201	1.122	0.552
I05	0.451	0.241	0.596	0.101	0.225	0.053	0.709	0.257
I06	0.332	0.123	0.411	0.097	0.224	0.065	0.922	0.351
I07	0.360	0.201	0.099	0.058	0.536	0.091	0.812	0.211
I08	0.191	0.099	0.223	0.094	0.342	0.108	0.851	0.425

验证实验，并以稀疏采样模式选择准确率作为评价指标，与 BFS、Ref. [19] 可知。由表 9 可知，所提方法 (Pro.) 在 HCI old[35] 数据集上的平均模式选择准确率为 80.0%，相较原测试集的 87.5% 仅下降 7.5%。表 10 给出了所提方法在 HCI old 数据集上与 BFS 法以及 Ref. [19] 方法编码性能对比，可以看出相比于 BFS 法，所提方法在编码性能上仅存在轻微差异，表明在不同数据来源和场景条件下所提方法仍能保持稳定的性能，具有良好的跨数据集泛化能力。

3.7 消融实验

本节针对所提方法中的三处改进进行了消融实验，包括 KNN 稀疏采样模式自适应优选算法、帧级码率分配、LCU 级码率分配，分别以 KNN、FR 以及 LCU 表示。Baseline 表示不引入任何改进模块的对比试验。表 11 展示了在测试集上消融实验的具体结果，其中，最佳性能用粗体表示。当只使用 KNN 时，模型可依据光场图像复杂度特征适配最优稀疏采样模式，相比于 Baseline 在 PSNR 以及 PPSNR 上均有提升；加入 FR 后，

表 7　稀疏采样模式选择分析

Table 7　Analysis of sparse sampling mode selection

Scenes	R_1			R_2			R_3			R_4			R_5		
	Pro.	Jin	BFS	Pro.	Jin	BFS	Pro.	Jin	BFS	Pro.	Jin	BFS	Pro.	Jin	BFS
I01	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$
I02	M_1	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$
I03	$\boldsymbol{M_2}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_3}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$
I04	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_3}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$
I05	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_4}$	$\boldsymbol{M_4}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$
I06	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_2}$	$\boldsymbol{M_3}$	$\boldsymbol{M_5}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_6}$	$\boldsymbol{M_5}$
I07	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_2}$	$\boldsymbol{M_4}$	$\boldsymbol{M_2}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$
I08	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_1}$	$\boldsymbol{M_2}$	$\boldsymbol{M_1}$	$\boldsymbol{M_3}$	$\boldsymbol{M_3}$	$\boldsymbol{M_4}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_5}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$	$\boldsymbol{M_6}$

表 8　所提方法与 BFS 方法在实际运行时间上的对比/s

Table 8　Comparison of computational time between the proposed method and BFS method/s

Method	R_1	R_2	R_3	R_4	R_5
BFS	6834.9	7902.6	9998.2	15494.7	19645.4
Algorithm 1	22.7	22.7	22.7	22.7	22.7
Proposed	425.6	526.1	996.8	2953.78	12511.9

表 9　所提方法在 HCI old 数据集上测试的结果

Table 9　The test results of the proposed method on the HCI old dataset

	Method	Accuracy/%
HCI old	Jin	66.7
	BFS	100.0
	Proposed	80.0

表 10　所提方法在 HCI old 数据集上与其他方法在 PSNR 指标上的编码性能对比

Table 10　Comparison of coding performance in terms of PSNR between the proposed method and other methods on the HCI old dataset

	Method	BD-BR/%	BD-PSNR/dB
Proposed vs	BFS	+1.604	−0.070
	Jin	−4.638	+0.131

表 11　不同模块的消融实验

Table 11　Ablation experiments of different modules

Method	KNN	FR	LCU	Evaluation		
				PSNR/dB↑	PPSNR/dB↑	Var↓
Baseline				36.441	37.994	1.122
No.1	√			37.640	38.221	1.647
No.2	√	√		37.740	38.412	**0.138**
No.3	√	√	√	**37.825**	**39.297**	0.151

算法按照视图权重分配码率，在角度维质量一致性上提升较高；最后叠加 LCU 后，模型对帧内 LCU 级码率再分配，显著区域 PPSNR 有所提升，此时整体表现最佳。

4 结 论

本文提出了一种面向目标码率与内容特性的宽视场光场图像编码方法。该方法从最佳稀疏采样模式选择、关键视图编码策略以及细粒度码率分配三个层面协同优化整体编码流程。首先，通过分析光场图像内容复杂度并结合机器学习方法实现稀疏采样模式的自适应优选，有效平衡了压缩效率与编码开销；随后在关键视图间引入帧级码率预分配策略，以帧间距离为依据提升角度维质量一致性；进一步结合显著性检测与深度估计生成区域权重，并通过视角传播机制将其扩展至所有关键视图，从而实现更具内容感知特性的最大编码单元级码率控制。最终，利用视频编码标准完成关键视图压缩，并在接收端通过深度学习网络重建非关键视图，生成完整光场图像。实验结果表明，所提方法在保持良好随机访问性能的同时，实现了显著的码率节省与视觉质量提升，并在角度维一致性方面表现较好，体现出在实际应用中的高效性与可扩展性。未来工作中，将聚焦于标准兼容型深度学习的编码框架，探索两者融合的技术路径。

作者贡献

朱磊：方法提出、论文撰写与修改、实验设计和数据整理；陈晔曜：论文审核与修改；封哲宇：实验数据校对与结果分析；金攀麒：实验数据收集与分析；蒋刚毅：论文选题指导、论文审核与修改。

利益冲突

本文所有作者声明无利益冲突

参考文献

1. Chen Y Y, Jiang G Y, Jin C C, et al. Multi-Attention learning and exposure guidance toward ghost-free high dynamic range light field imaging[J]. *IEEE Trans Visual Comput Graph*, 2025, **31** (9): 5304−5320.
2. Luo S C, Zhang X D, Wan L, et al. Light field depth estimation method combining image feature transfer[J]. *Comput Eng*, 2023, **49** (4): 206−216.
   罗少聪, 张旭东, 万乐, 等. 结合图像特征迁移的光场深度估计方法[J]. 计算机工程, 2023, **49** (4): 206−216.
3. Zhou W H, Lin L L, Hong Y J, et al. Beyond photometric consistency: geometry-based occlusion-aware unsupervised light field disparity estimation[J]. *IEEE Trans Neural Netw Learn Syst*, 2024, **35** (11): 15660−15674.
4. Huang L, Lv T Q, Wu Y C, et al. Two-way guided updating network for light field image super-resolution[J]. *Opto-Electron Eng*, 2024, **51** (12): 240222.
   黄莉, 吕天琪, 武迎春, 等. 基于双路引导更新的光场图像超分网络[J]. 光电工程, 2024, **51** (12): 240222.
5. Lv T Q, Wu Y C, Zhao X L. Light field image super-resolution network based on angular difference enhancement[J]. *Opto-Electron Eng*, 2023, **50** (2): 220185.
   吕天琪, 武迎春, 赵贤凌. 角度差异强化的光场图像超分网络[J]. 光电工程, 2023, **50** (2): 220185.
6. Samarakoon T, Abeywardena K, Edussooriya C U S. Arbitrary volumetric refocusing of dense and sparse light fields[Z]. arXiv: 2502.19238, 2025. https://doi.org/10.48550/arXiv.2502.19238.
7. Zheng X, Li Z Q, Liu D Y, et al. Spatial attention-guided light field salient object detection network with implicit neural representation[J]. *IEEE Trans Circuits Syst Video Technol*, 2024, **34** (12): 12437−12449.
8. Zhong C X, Chen Y Y, Zhu Z J, et al. Unsupervised large field of view light field content generation based on depth perception[J]. *Laser Optoelectron Prog*, 2025, **62** (12): 1211001.
   钟长祥, 陈晔曜, 朱仲杰, 等. 基于深度感知的无监督大视场光场内容生成[J]. 激光与光电子学进展, 2025, **62** (12): 1211001.
9. Cui Y L, Jiang G Y, Yu M, et al. Stitched wide field of view light field image quality assessment: benchmark database and objective metric[J]. *IEEE Trans Multimed*, 2024, **26**: 5092−5107.
10. Mukati M U, Zhang X, Wu X L, et al. Low-complexity ℓ_∞-compression of light field images with a deep-decompression stage[J]. *J Vis Commun Image Represent*, 2024, **99**: 104072.
11. Perra C, Assuncao P. High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement[C]//*2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Seattle, 2016: 1–4. https://doi.org/10.1109/ICMEW.2016.7574671.
12. Dai F, Zhang J, Ma Y K, et al. Lenselet image compression scheme based on subaperture images streaming[C]//*2015 IEEE International Conference on Image Processing*, Quebec City, 2015: 4733–4737. https://doi.org/10.1109/ICIP.2015.7351705.
13. 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)*, Athens, 2018: 3174–3178. https://doi.org/10.1109/ICIP.2018.8451345.
14. Li L, Li Z, Li, B, et al. Pseudo-sequence-based 2-D Hierarchical coding structure for light-field image compression[J]. *IEEE J Sel Top Signal Process*, 2017, **11** (7): 1107−1119.
15. Yu Z Y, Qu Q, Zheng W, et al. Efficient light field transmission via enhanced resampling reconstruction and angular attention estimation[J]. *ACM Trans Sensor Netw*, 2025.
16. Huang X P, An P, Chen Y L, et al. Low bitrate light field compression with geometry and content consistency[J]. *IEEE Trans Multimed*, 2022, **24**: 152−165.
17. Bakir N, Hamidouche W, Fezza S A, et al. Light field image coding using VVC standard and view synthesis based on dual discriminator GAN[J]. *IEEE Trans Multimed*, 2021, **23**: 2972−2985.
18. Liu D Y, Huang Y, Fang Y M, et al. Multi-stream dense view reconstruction network for light field image compression[J]. *IEEE Trans Multimed*, 2023, **25**: 4400−4414.
19. 金攀琪. 基于稀疏模式选择的光场图像编码方法: 中国,

2024112825009[P]. 2025-01-14

20. Tong K D, Jin X, Wang C, et al. SADN: Learned light field image compression with spatial-angular decorrelation[C]//*ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, 2022: 1870–1874. https://doi.org/10.1109/ICASSP43922.2022.9747377.
21. Feng S Y, Zhang Y, Zhu L W, et al. LFIC-DRASC: Deep light field image compression using disentangled representation and asymmetrical strip convolution[J]. *IEEE Trans Broadcast*, 2025, **71** (3): 889−902.
22. Takhtardeshir S, Olsson R, Guillemot C, et al. Efficient and fast light field compression via VAE-based spatial and angular disentanglement[J]. *IEEE Access*, 2025, **13**: 18594−18607.
23. Zhang G Z, Zhu Y H, Wang H N, et al. Extracting motion and appearance via inter-frame attention for efficient video frame interpolation[C]//*2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 2023: 5682–5692. https://doi.org/10.1109/CVPR52729.2023.00550.
24. M. Řeřábek and T. Ebrahimi, New Light Field Image Dataset[C]//8th International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 2016. https://www.epfl.ch/labs/mmspg/downloads/epfl-light-field-image-dataset/
25. Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[C]//*Proceedings of 13th Asian Conference on Computer Vision Computer Vision – ACCV 2016*, Taipei, 2017: 19–34. https://doi.org/10.1007/978-3-319-54187-7_2.
26. Li J, Wang Z M, Lai S M, et al. Parallax-tolerant image stitching based on robust elastic warping[J]. *IEEE Trans Multimed*, 2018, **20** (7): 1672−1687.
27. Zhou R, Jiang G Y, Chen Y, et al. Benchmark database and objective metric for wide-field-of-view light field image quality measurement[J]. *IEEE Trans Instrum Meas*, 2026, **75**: 1−15
28. Miangoleh S M H, Dille S, Mai L, et al. Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging[C]//*2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 2021: 9680–9689. https://doi.org/10.1109/CVPR46437.2021.00956.
29. Wang F Y, Pan J S, Xu S K, et al. Learning discriminative cross-modality features for RGB-D saliency detection[J]. *IEEE Trans Image Process*, 2022, **31**: 1285−1297.
30. Liu Z, Dong M J, Guan X H, et al. Fast ISP coding mode optimization algorithm based on CU texture complexity for VVC[J]. *EURASIP J Image Video Process*, 2021, **2021** (1): 24.
31. Amirpour H, Guillemot C, Ghanbari M, et al. Advanced scalability for light field image coding[J]. *IEEE Trans Image Process*, 2022, **31**: 7435−7448.
32. Majid M, Owais M, Anwar S M. Visual saliency based redundancy allocation in HEVC compatible multiple description video coding[J]. *Multimed Tools Appl*, 2018, **77** (16): 20955−20977.
33. Bjontegaard G. Calculation of average PSNR differences between RD-curves[S]. ITU SG16 Doc. VCEG-M33, 2001.
34. Pereira F, Pagliari C, da Silva E A B, et al. JPEG Pleno light field coding common test conditions v3.3: ISO/IEC 21794–1[S]. 2021.
35. Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for densely sampled 4D light fields[C]//*Proceedings of Vision, Modeling, and Visualization (VMV)*, 2013: 225–226. https://doi.org/10.2312/PE.VMV.VMV13.225-226.

作者简介



朱磊 (2000-)，男，硕士研究生，主要从事光场图像压缩等方面的研究。

E-mail：zhulei2025@126.com



封哲宇 (2000-)，男，硕士研究生，主要从事光场图像压缩等方面的研究。

E-mail：15068368500@163.com



陈晔曜 (1994-)，男，博士，讲师，2023 年于宁波大学获得博士学位，主要从事高动态范围图像/视频处理、光场图像处理及深度学习等方面的研究工作。

E-mail：chenyeyao@nbu.edu.cn



【通信作者】蒋刚毅 (1964-)，男，博士，教授，博士生导师，2000 年于韩国 Ajou 大学获得博士学位，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@nbu.edu.cn

Target bitrate and image content-aware wide field-of-view light field image coding

Zhu Lei, Chen Yeyao, Feng Zheyu, Jin Panqi, Jiang Gangyi*

Faculty of Information Science and Engineering, Ningbo, Zhejiang 315211, China

Abstract

Objective A sparse sampling mode set including key views and non-key views is constructed via the core strategy of sparse coding and reconstruction, providing fundamental support for the coding of wide-FOV light field images with variable bitrates. The specific implementation scheme is as follows: Firstly, the multi-dimensional complexity features of light field images are extracted, and an adaptive selection scheme for sparse sampling modes is designed combined with target bitrate constraints. Leveraging the K-nearest neighbor (KNN) algorithm, rapid matching of the optimal mode is enabled, allowing identification of the best sparse sampling strategy without traversing the entire mode set. On this basis, predictive relationships among key views are established, providing a more robust basis for subsequent coding and compression processes. Secondly, to address uneven view quality caused by traditional bitrate allocation methods, a frame-level bitrate pre-allocation strategy is designed according to the inter-frame spatial distance between the current encoded frame and its reference frames within the key view set. Dynamic adjustment of bitrate allocation weights effectively improves the angular quality consistency of the encoded views. Then, based on the I-frames in the key view set, the saliency detection and depth estimation techniques are integrated to generate largest coding unit (LCU)-level saliency weights. Utilizing the angular correlation of light field images, these weights are propagated to all key views along the angular dimension. Furthermore, saliency weights are combined with image regional complexity to develop a fine-grained LCU-level content-aware bitrate allocation strategy, achieving more accurate bitrate allocation. Finally, at the client side, key views are decoded according to the transmitted optimal sparse sampling mode, and non-key views are reconstructed using the decoded key views with a deep learning-based view synthesis network, ultimately completing the reconstruction of the wide-FOV light field image.

Methods In this paper, the core strategy of sparse coding and reconstruction is adopted to construct a sparse sampling mode set that includes key views and non-key views, thereby providing fundamental support for the coding of wide-FOV light field images with variable bitrates. The specific implementation scheme is as follows: Firstly, the multi-dimensional complexity features of light field images are extracted and an adaptive selection scheme is designed for sparse sampling modes in combination with target bitrate constraints. Leveraging the K-nearest neighbor (KNN) algorithm, this scheme enables the rapid matching of the optimal mode, which allows the identification of the best sparse sampling strategy without traversing the entire mode set. On this basis, the predictive relationships among key views are established, providing a more robust basis for subsequent coding and compression processes. Secondly, aiming to address the issue of uneven view quality caused by traditional bitrate allocation methods, a frame-level bitrate pre-allocation strategy is designed according to the inter-frame spatial distance between the current encoded frame and its reference frames within the key view set. By dynamically adjusting the bitrate allocation weights, this strategy effectively improves the angular quality consistency of the encoded views. Then, based on the I-frames in the key view set, the saliency detection and depth estimation techniques are integrated to generate LCU (largest coding unit)-level saliency weights. Utilizing the angular correlation of light field images, these weights are propagated to all key views along the angular dimension. Furthermore, the saliency weights are combined with the regional complexity of images to develop a fine-grained LCU-level content-aware bitrate allocation strategy, achieving more accurate bitrate allocation. Finally, at the client side, the key views are decoded according to the transmitted optimal sparse sampling mode, and the non-key views are reconstructed using the decoded key views with a deep learning-based view synthesis network, thus ultimately completing the reconstruction of the wide-FOV light field image.

Results and Discussions The proposed adaptive sparse sampling mode selection scheme, leveraging multi-dimensional feature fusion and deep learning prediction mechanisms, significantly reduces the computational complexity of mode selection and avoids exorbitant computational costs incurred by the traditional brute force search method. It rapidly matches diverse target bitrates and image content features to select the optimal or sub-optimal sparse coding mode. The designed two-level bitrate allocation strategy, combining frame-level pre-allocation and LCU-level perceptual allocation, effectively mitigates uneven angular view quality while maintaining compatibility with existing video coding standards, further enhancing the angular quality consistency of wide-FOV light field images. Experimental results

Foundation item: National Natural Science Foundation of China Project (62271276, 62401301); Natural Science Foundation of Zhejiang Province Project (LQ24F010002)

*Correspondence: Jiang G Y, E-mail: jianggangyi@nbu.edu.cn

show that, compared with existing representative methods, the method achieves an average bitrate saving of 45.284% and a peak signal-to-noise ratio (PSNR) improvement of 1.641 dB, with the maximum random access penalty of light field images limited to only 20.29%. These results indicate that the method not only improves coding efficiency but also ensures favorable random access performance, satisfying the demand for rapid view retrieval in practical applications.

Conclusions The proposed wide-FOV light field image coding method based on target bitrate and image content awareness effectively adapts to the coding requirements of variable bitrates and uncertain image content. It achieves satisfactory results in reducing sampling mode selection complexity and improving network transmission adaptability. Meanwhile, it fulfills multiple objectives including bitrate saving, image quality enhancement, angular quality consistency improvement, and random access capability optimization, demonstrating superior overall coding performance and favorable practical application value.

Keywords: wide field-of-view (FOV) light field images; sparse sampling mode; bitrate allocation strategy; angular quality consistency

Zhu L, Chen Y Y, Feng Z Y, et al. Target bitrate and image content-aware wide field-of-view light field image coding[J]. *Opto-Electron Eng*, 2026, **53**(5): 250320; DOI: 10.12086/oee.2026.250320

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文