适用于 OMAF 流媒体工作流的 360°视频快速多码率编码

AmrithaPremkumar,
ChristianHerglotz勃兰登堡工业大学科特布斯-森
夫滕贝格分校计算机工程系，德国科特布斯

arXiv:2601.17568v1 [eess.IV] 24 Jan 2026

摘要
为 HTTP 自适应流媒体准备高质量 360° 视频，需要将每个序列编码为多个表示，这些表示涵盖不同的分辨率和量化参数（QPs）。对于 8K 等超高分辨率沉浸式内容 360° 视频，由于表示数量庞大且现代编解码器复杂度很高，这一过程的计算开销十分巨大。本文探讨快速多码率编码策略，通过在不同量化参数和分辨率间复用编码器分析信息来缩短编码时间。我们评估了两种跨分辨率信息复用流程，它们在参考编码跨分辨率传播的方式上有所不同：（i）严格的高清 → 4K → 8K级联方案，配合缩放后的分析复用；（ii）分辨率锚定方案，为每种分辨率使用各自的最高码率参考进行初始化，再以此指导依赖编码。除在标准等距圆柱投影内容上评估这两种流程外，我们还将其应用于立方体贴图投影（CMP）分块，其中每个360° 帧都被划分为独立编码的图块。CMP 带来了大量并行性，同时仍能从所提出的多码率分析复用策略中获益。使用 SJTU 8K 360° 数据集进行的实验结果表明，分层分析复用显著加快了高效视频编码（HEVC）的编码速度，同时对等距圆柱投影和 CMP 分块内容的率失真影响极小；等距圆柱投影的编码时间缩短约 33%–59%，CMP 平均缩短约 51%，Bjøntegaard编码时间增量（BDET）收益接近 −50%，实际运行时间加速比最高达 4.2×。

CCS 概念
• **信息系统** →**多媒体流媒体**。

关键词
360度视频,快速多码率编码, 高效视频编码, 立方体贴图投影, 自适应流媒体传输

ACM 参考格式：
Amritha Premkumar、Christian Herglotz。2026。面向 OMAF 流媒体工作流的 360°视频快速多码率编码。载于*Mile-High*视频会议（*MHV '26*），*2026*年*2*月*2*日至*5*日，美国科罗拉多州丹佛。 ACM，美国纽约州纽约，7页。https://doi.org/10.1145/3789239.3793270

请使用 nonacm 选项或 ACM Engage 类以启用 CC 许可。本作品采用 Creative Commons 署名 4.0 国际许可协议授权。*MHV '26，2026*年*2*月*2*日至*5*日，美国科罗拉多州丹佛 © 2026 版权所有，归所有者/作者所有。ACM
ISBN 979-8-4007-2466-4/2026/02
https://doi.org/10.1145/3789239.3793270

1 引言
沉浸式媒体是虚拟现实（VR）、增强现实（AR）、远程临场、教育和互动娱乐的核心。常见格式之一是 360° 视频，它可以捕捉完整的球形环境，让用户以三个旋转自由度探索场景 [1, 2]。要以高感知质量传输此类内容颇具挑战，因为这需要超高分辨率（4K–8K 及更高）、高帧率（60–120fps）和专用投影格式，从而给采集、投影、压缩和流媒体传输 [3–6]带来沉重的计算和带宽负担。

现代沉浸式流媒体传输依赖 HTTP 自适应流（HAS）方法，例如 MPEG-DASH 和 HLS [7, 8],，其中每个视频都被编码为一个比特率阶梯，包含多个具有不同分辨率、比特率和量化参数的表示 [9–11]。客户端会在这些表示之间切换，以保持播放不中断。对于 360° 内容，自适应通常与视口相关或基于分块的传输相结合，从而仅以高质量流式传输用户视场内的内容 [12–16]。先前的研究涵盖了视口预测、显著性、码率分配、能源效率以及基于强化学习的自适应 [6, 15, 17–23], , 但大多数方法都假定完整的多码率表示集已经准备就绪，忽略了制作底层多码率表示所需的巨大成本——这正是本研究明确关注的空白 [24], 也是先前 360° HAS研究未曾解决的问题。

这一挑战在 OMAF 工作流程中尤为突出 [25], , 该流程定义了投影元数据、视口自适应信令以及分块或立方体贴图分解等基于区域的机制 [12, 26, 27]。这些特性提高了传输效率，但需要针对不同分辨率和区域编码大量码流。在商用硬件上编码一段 8K 360° 视频的 15 个表示，耗时可能超过 30 CPU小时 [28], , 这使得针对大型 VR 内容库或直播等对延迟敏感的场景进行穷举全搜索编码变得不切实际 [29]。近期研究进一步强调，在自适应视频流媒体工作流程中，应将编码延迟和能耗视为首要优化目标 [24]。图 1 展示了这种计算成本：随着分辨率提高，编码时间急剧增加，而除非显著提高比特率，否则 WSPSNR 的增益会趋于饱和。因此，准备完整的比特率阶梯需要反复进行高成本编码，这促使人们开发快速多码率编码技术。

快速多码率和多分辨率编码利用了一项关键观察：同一内容在不同表示形式中的编码器决策高度相关。先前的工作 [4, 5, 31] 表明，CTU 划分、帧内／帧间模式选择、运动估计和变换决策在不同量化参数和分辨率下均保持结构稳定。因此，复用先前编码表示的分析结果，可以避免大量冗余搜索，大幅

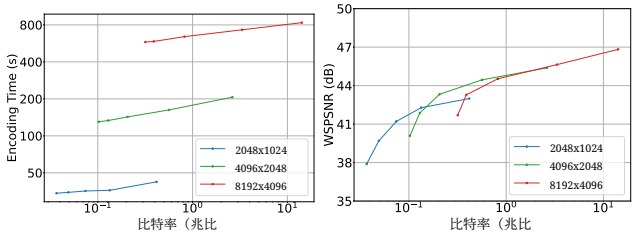
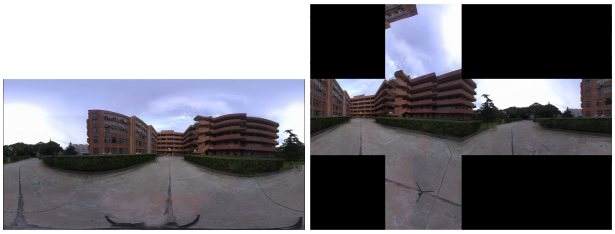


图1：*Kaix-uanmenye pano* [30]的率失真曲线和码率-编码时间曲线。此处，质量采用 WSP-SNR 表示。



(a) 等矩形投影　(b) 立方体贴图投影

图 2：360° 投影格式比较：等矩形投影与立方体贴图投影。

缩短编码时间，同时对率失真（RD）的影响较小 [32, 33]。虽然 x265 提供了分析复用原语 [10, 34], , 但其在360° 布局、OMAF 驱动的码率阶梯和跨分辨率流水线中的表现尚未得到表征；我们的工作首次对复用流程进行了投影感知评估。

然而，将此类复用应用于球面视频颇具挑战。等距圆柱投影（ERP）是主流的 360° 布局，在靠近两极处会引入严重的几何失真 [35, 36], , 影响空间与运动统计特征，并削弱跨分辨率一致性。旋转运动、视差和视口动态变化进一步增加了时间变化性 [26, 37]。立方体贴图投影（CMP）提供了一种更均匀的替代方案：它将球面分解为六个正方形面（图2），从而减少几何畸变，并产生更加均匀的块统计 [13, 38]。CMP 也与 OMAF 的基于区域的设计自然契合，可实现独立的并行逐面编码，并进一步发挥分析复用的优势。

本研究的主要贡献如下：

(1) 面向投影的多码率编码˚视频： 我们将多码率和多分辨率分析复用方法扩展到球面视频，以解决 ERP 投影失真和视口驱动的变化问题，并表明 CMP 可提高跨分辨率复用稳定性。
(2) 对 *8K 360˚*内容的全面评估： 利用 SJTU 8K 360° 数据集 [30],，我们量化了所提技术在符合 OMAF 标准的流水线中的编码时间缩减、率失真权衡和可扩展性，结果表明其加速比显著，同时率失真损失极小。

论文结构： 第 2 节回顾相关工作。第 3 节介绍所提出的架构，第 4 节详细说明实验设置，第 5 节展示结果，第 6 节总结全文并展望未来方向。

2 快速多码率编码

为 HAS 生成完整的码率阶梯需要大量计算，尤其是使用高效视频编码 (HEVC) 和通用视频编码 (VVC) 等现代编解码器时 [39–41]。快速多码率编码通过在同一视频的不同表示之间复用编码器决策——编码单元分区、预测模式、运动矢量和变换选择——来解决这一问题，并利用这些决策在不同量化参数和分辨率下的高度相关性 [10, 32, 42]。这种复用可以大幅缩短编码时间，同时仅造成极小的率失真损失。主要有两种范式：自顶向下 和自底向上。

2.1 自顶向下与自底向上范式

自顶向下 方法将决策从质量最高的表示（量化参数最低或分辨率最高）传递到其他表示 [32, 42]。虽然率失真性能出色，但串行处理限制了并行性，因此不适用于对时间敏感或大规模的工作流程。相反，自底向上 方法从编码速度最快的表示（分辨率最低或量化参数最高）开始 [10, 33, 43]。分析结果可以较早获得，从而实现高并发性和可扩展性。尽管其率失真损失略高于自顶向下方法，但对于需要数十种表示的 VR/360° 流水线而言，自底向上方法实用得多。

2.2 跨分辨率与跨量化参数预测

编码器分析在相邻量化参数和分辨率之间具有良好的迁移性。跨量化参数复用利用稳定的块划分模式，而跨分辨率复用则利用空间自相似性 42] [33, 。

然而，在不同分辨率之间映射决策，需要精确缩放 CTU坐标和块大小。近期的启发式方法和轻量级机器学习模型提升了稳健性，但超高分辨率编码（4K–8K）仍然具有显著的复杂度，因此有必要进一步优化。

2.3 360°视频面临的挑战

由于像素密度不均匀，球面视频中的复用更加复杂。作为主流布局的等矩形投影会造成严重的极区畸变，改变空间与运动统计特征，并降低迁移的可靠性。因此，现有多码率方法应用于360°内容时性能会下降。立方体贴图投影通过将球面分解为六个特性均匀的正方面，缓解了这些问题。立方体贴图投影改善了运动估计，稳定了块统计，并支持各面独立编码，从而释放出巨大的并行性。对于需要数十种表示的 OMAF 式处理流程，这些特性使立方体贴图投影成为实现实用、快速多码率编码的关键。

如果没有投影感知加速，为 8K 360°视频准备完整的码率阶梯仍然耗时过长，因此需要基于软件的快速多码率流水线；硬件编码器虽然吞吐量高，但通常不支持导出/导入 CTU 级分析结果，因此无法参与跨量化参数或跨分辨率复用。我们的方法通过结合 CMP 与高效的跨分辨率和跨量化参数复用弥补了这一不足，在不牺牲率失真性能的前提下实现可扩展性。

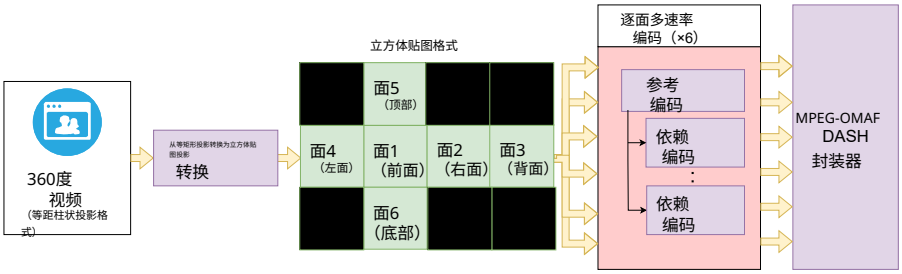


图3：所提出的快速多码率编码流程概览。可选的 ERP→CMP 转换仅应用于 CMP-CRC 和 CMP-PRA。对于 ERP-CRC 和 ERP-PRA，流程省略投影框；多码率复用逻辑（锚点 +依赖项、分析数据保存/加载）保持不变。

3 所提出的框架

我们的框架接收输入的 360° 等矩形投影视频，并采用以下两种方式之一进行处理：(i) 直接编码；或 (ii) 将其转换为 CMP布局，以降低投影失真并释放更多并行性。随后，我们通过完整编码选定的 锚点，并复用其分析结果来加速其余的 依赖 编码，从而实现快速多码率编码。最后，将所有生成的码流打包为符合 OMAF标准的DASH演示（图 3）。

我们通过组合投影格式（ERP 或 CMP）与分析共享策略（CRC 或 PRA），评估四种实用变体：

- **ERP-CRC**： 采用级联高清→4K→8K流程的 ERP投影。
- **ERP-PRA**： 采用各分辨率锚点的 ERP投影。
- **CMP-CRC**： 采用按面应用级联流程的立方体贴图投影。
- **立方体贴图投影-PRA**： 采用逐面、逐分辨率锚点的立方体贴图投影。

等矩形投影变体在单个全帧图块上运行，而立方体贴图投影变体则按面独立复用分析，从而实现更高的并行性，并受益于立方体贴图投影更均匀的采样。

在每个分辨率层（高清、4K、8K），我们选择一种锚点表示（低质量（LQ）、中等质量（MQ）或高质量（HQ）），对其执行完整的率失真优化。其决策会被复用于加速同一分辨率下的所有依赖表示；在 CRC 中，也会用于更高分辨率。

3.1 球面到立方体贴图投影

设 360° 内容由辐射函数定义在球面 $\mathbb{S}^2$上

公式

$$\mathcal{V} : \mathbb{S}^2 \times \mathbb{N} \to \mathbb{R}^3, \qquad \mathcal{V}(\boldsymbol{\omega}, t) = (Y, Cb, Cr), \tag{1}$$

其中， $\boldsymbol{\omega}$表示观察方向，而（Y、Cb、Cr）表示 YCbCr 分量。ERP投影通过以下方式对该函数进行采样

$$E(u, v, t) = \mathcal{V}(\boldsymbol{\omega}_{\mathrm{ERP}}(u, v), t), \qquad (u, v) \in [0, W) \times [0, H). \tag{2}$$

CMP布局将 $\mathbb{S}^2$ 映射到六个正方形域 Ω_k，具体如下

$$F_k(x, y, t) = \mathcal{V}(\boldsymbol{\omega}_k(x, y), t), \qquad (x, y) \in \Omega_k, \tag{3}$$

其中， $\boldsymbol{\omega}_k(\cdot)$ 表示针对各面的逆投影。立方体贴图投影可减少几何畸变并稳定编码器统计信息，从而提高跨分辨率复用的可靠性。

3.2 码率阶梯和表示结构

直观来说。 我们为每个面编码一组精简的质量级别（CRF）。只有锚点执行完整的率失真优化；依赖项复用其分析结果，跳过大部分分区、模式和运动搜索。

公式化。 对于每个面 k，CRF阶梯为

$$Q = \{q_0, q_1, \ldots, q_{L-1}\}, \qquad q_0 < q_1 < \cdots < q_{L-1}, \tag{4}$$

其中， q_0 表示质量最高的锚点。表示 ℓ 的编码方式为

$$R_{k,\ell} = \mathcal{E}(F_k; q_\ell), \qquad \ell \in \{0, \ldots, L-1\}, \tag{5}$$

其中，E(·)表示启用分析保存/加载功能的 x265 编码。依赖项加载并复用锚点的分析结果，以加速编码。

3.3 参考编码：完整率失真优化与分析提取

每个分辨率（或面）的锚点都会执行完整的率失真优化，穷举搜索块划分、预测模式和运动矢量。每帧和每个编码单元的最优决策均会存储为：

$$d_{k,0}^{(n,i)} = \left(s_{k,0}^{(n,i)}, m_{k,0}^{(n,i)}, \mathbf{v}_{k,0}^{(n,i)}\right), \tag{6}$$

其中， s是划分模式， m是预测模式，而 $\mathbf{v}$ 是运动矢量。

3.4 通过分析复用实现依赖编码

依赖编码会复用锚点的决策作为强先验，从而缩小搜索空间并避免高成本的评估。

概述

公式。对于每个 $\ell \geq 1$:

$$\mathbf{d}_{k,\ell} = \arg\min_{\mathbf{d} \in \mathcal{S}_{k,0}} \left(D(\mathbf{d}) + \lambda(q_\ell)\, R(\mathbf{d})\right), \tag{7}$$

其中，可行决策满足：

- **Split reuse:** $s_{k,\ell}^{(n,i)} \preceq s_{k,0}^{(n,i)}$ (no deeper splits).
- **Mode reuse:** $m_{k,\ell}^{(n,i)} = m_{k,0}^{(n,i)}$ (skip mode search).
- **MV reuse:** $\mathbf{v}_{k,\ell}^{(n,i)} = \mathbf{v}_{k,0}^{(n,i)}$ (perform only local refinement).

对于跨分辨率复用（高清→4K、4K→8K），x265 会将锚点的分析缩放至目标编码单元网格。

表1：结合投影与分析共享策略的框架变体。

变体	投影	策略（流程）
ERP-CRC	等矩形投影（单个图块）	级联高清→4K→8K
ERP-PRA	等矩形投影（单个图块）	各分辨率锚点（高清、4K、8K）
CMP-CRC	立方体贴图投影（6 个面）	每面级联高清→4K→8K
CMP-PRA	立方体贴图投影（6 个面）	各分辨率锚点（每面）

3.5 分析共享方案：CRC 和 PRA

我们比较两种实用的多分辨率策略：

- **CRC：**严格的自底向上级联（高清→4K→8K）。每种分辨率只有一个锚点，其分析结果可加速所有依赖项，并为下一种分辨率提供初始值。这种方案能以最低的内存开销最大限度地实现跨分辨率复用。

- **PRA：** 每种分辨率都会独立选择一个锚点（低/中/高质量）依赖项会复用该锚点，锚点也有助于初始化下一种分辨率。PRA 提供了更大的灵活性，但每种分辨率都需要存储一个分析文件。

3.6 OMAF 封装与输出

编码后的等矩形投影或立方体贴图投影表示会被封装为符合 OMAF 标准的 DASH 演示文稿，其中包括投影元数据、区域信令和表示描述符，从而能够无缝集成到标准视口自适应播放器中。

ERP 会产生失真，削弱分析复用效果；而对所有码率梯级进行朴素的全搜索编码，对于 8K 360° 内容而言速度慢得难以接受。该框架结合了 CMP 的均匀性和逐面并行性、CRC/PRA 分析共享以及各分辨率锚点（低/中/高质量），实现了可扩展的多码率编码，同时将 RD 损失降至最低，并输出可用于 OMAF 的内容。

4 实验设置

4.1 x265 中的实现配置

我们的框架基于 x265 [34, 43] 的分析复用子系统构建。在每次锚点编码期间，使用 分析保存存储逐帧分析结果，之后通过 分析加载将其导入依赖编码。复用强度通过 分析保存复用级别 和 分析加载复用级别进行控制；我们采用最强设置（级别 *10*），以保留编码单元级结构、预测模式和运动信息。

跨量化参数复用（相同分辨率）。依赖编码以 加载复用级别*=10*复用编码单元划分、帧间模式和运动信息。通过帧内细化*=4*、帧间细化*=2*和运动矢量细化*=1*进行有限的细化，从而在不牺牲速度的前提下确保稳健性。

跨分辨率复用。4K 与 4K→8K 的转换，x265 使用 缩放因子*=2* 结合第 10 级复用来缩放 CTU 级决策，从而在更高分辨率下省去大部分分区、模式搜索和运动估计。

4.2 数据集

我们使用 SJTU 360° 数据集 [30],，其中包含 15 个分辨率为 8192×4096 的 ERP 序列。每段视频均长 30 秒；在我们的设置中，单个 8K 全搜索参考编码每

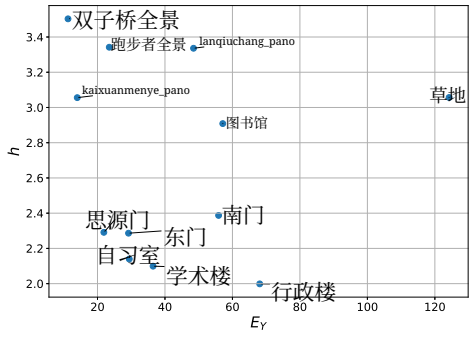


图 4：使用 VCA [44] 中的(E_Y, h)对 SJTU 数据集进行复杂度分析。取值越高，表示运动和纹理越复杂。

.

表 2：实验参数和取值。

参数	取值
$\mathcal{R}$	{ 2048x1024（高清）, 4096x2048（4K）, 8192x4096（8K）}
$\mathcal{Q}$	{ 22, 27, 32, 37, 42 }
目标编码器	x265 [中等][帧内周期=1s][4 CPU线程]
目标解码器	HM [1 CPU 线程]

个恒定速率因子都需要数分钟，这凸显了准备完整码率阶梯的成本。通过视频复杂度分析器 (VCA) [44],的 $(E_Y,\ h)$ 特征计算得出的内容复杂度分布如图 4 所示。

为进行投影感知分析并提高并行性，每个 ERP序列均使用标准的球面到立方体贴图映射转换为 CMP。这会生成六个均匀采样的面，大小为 2048×2048，分别对应{前面、后面、左面、右面、顶部、底部}。编码过程中，每个面均作为独立图块处理。

所有实验均在配备英特尔至强处理器（32个线程）和128GB内存的服务器上执行，从而支持串行运行时间评估（每个图块）和并行运行时间评估（六个图块同时进行）。表2总结了分辨率、量化参数和编解码器设置。

4.3 评价指标

所有评估均将每种所提出的方法与<style id='1'>ERP-默认（全搜索ERP编码）进行比较。

质量评价指标 我们使用 Samsung 360tool 计算每个量化参数下的峰值信噪比和加权球面峰值信噪比。[1] 图5展示了评估流程。

串行编码运行时间差异（ΔT_S）假设采用串行编码，相对运行时间变化为

$$\Delta T_{\mathrm{S}} = \frac{\sum \tau_{\mathrm{method}} - \sum \tau_{\mathrm{ref}}}{\sum \tau_{\mathrm{ref}}}. \tag{8}$$

并行编码运行时间差异（ΔT_P）假设采用并发编码，运行时间由最慢的表示决定：

$$\Delta T_{\mathrm{P}} = \frac{\max(\tau_{\mathrm{method}}) - \max(\tau_{\mathrm{ref}})}{\max(\tau_{\mathrm{ref}})}. \tag{9}$$

[1]https://github.com/Samsung/360tools

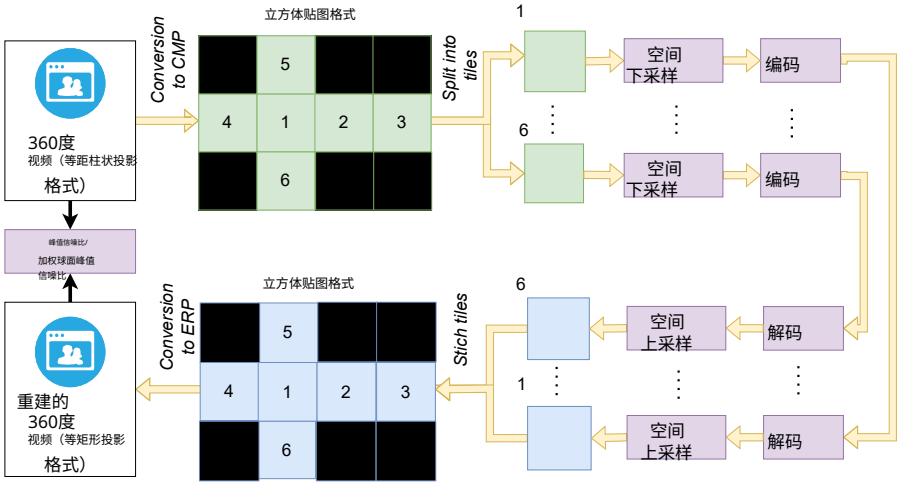


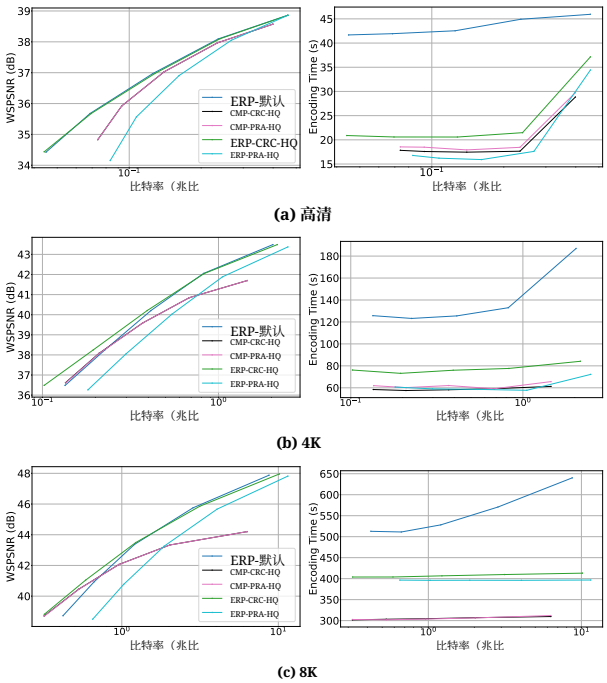
图5：质量评估框架。

Bjøntegaard差值指标BD-PSNR 和 BD-WSPSNR 用于衡量相对于 ERP-默认的平均质量差异 [45],其中正值表示提升，负值表示轻微损失。BDET [46] 用于评估相同质量下的相对编码时间节省；负值表示在相同率失真性能下编码速度更快。

5 结果与讨论

编码速度提升。 图6比较了*Academic Building*序列在高清、4K和8K分辨率下的加权球面峰值信噪比与编码时间。左侧图表显示码率–WSPSNR曲线，右侧图表显示编码时间与比特率的关系。ERP-默认在所有分辨率下的编码时间都最长，反映出穷举全搜索编码的成本。CRC和PRA都带来了显著的加速，其中PRA的耗时降幅最大。基于CMP的变体（CMP-CRC和CMP-PRA）通过逐面并行进一步放大了这些收益；在8K分辨率下优势最为显著，此时ERP-默认每种表示的耗时超过600秒。在质量方面，CRC与ERP-默认最为接近，而PRA略有偏差，体现了速度与率失真保真度之间的根本权衡。在4K分辨率下，PRA带来的改进最大；CMP变体则提供了最强的整体吞吐量，凸显了投影感知复用的优势。

表3总结了相对于ERP-默认的性能。对于ERP内容，CRC可将串行时间缩短33 %–52 %，并将并行时间缩短24 %–57 %。CMP-CRC进一步提升了这些指标，分别达到50.82 %（串行）和48.67 %（并行）。PRA实现了最大的直接降幅——平均为45 %–47 %，在4K分辨率下最高可达59 %；而CMP-PRA实现了最高的整体吞吐量，平均提升分别为51.38 %（串行）和49.88 %（并行）。这些趋势表明，即使面对球面视频更复杂的统计特性，自底向上复用依然十分有效。在生产流水线中，每个序列都需要准备数十种表示，因此这类节省尤为可贵。



(a) 高清

(b) 4K

(c) 8K

图 6：*Academic Building* 在高清、4K 和 8K 下的码率–WSPSNR 与码率–编码时间曲线。

率失真影响： CRC 的率失真行为最为稳定，ERP 的 BD-PSNR/BD-WSPSNR 差异接近于零。CMP 相较于 CRC 的损失更大（BD-PSNR 最高达 −1.07dB），而 PRA 因其采用更激进的复用粒度，偏差更为明显。不过，WSPSNR 考虑了球面采样的不均匀性，因此在 360°内容中，它比平面PSNR更能反映感知差异。因此，8K 下 CMP–CRC 的最大降幅主要体现在平面PSNR上；相应的 WSPSNR

表3：所提方法相对于 ERP-Default 编码的率失真和编码时间节省对比。

方法	参考分辨率		BD-PSNR [dB]	BD-WSPSNR [dB]	BDET (峰值信噪比) [%]	BDET (加权球面峰值信噪比) [%]	ΔT_S [%]	ΔT_P [%]
ERP-CRC	LQ	HD	−0.06 ± 0.03	−0.06 ± 0.03	−47.09 ± 4.89	−47.07 ± 4.89	−43.63 ± 5.36	−24.20 ± 5.76
		4K	0.09 ± 0.06	0.08 ± 0.06	−50.43 ± 1.68	−50.43 ± 1.66	−51.57 ± 1.20	−56.64 ± 4.24
		8K	0.02 ± 0.13	−0.02 ± 0.14	−29.63 ± 5.22	−29.67 ± 5.20	−31.04 ± 5.45	−40.84 ± 5.87
		Avg	**0.02 ± 0.07**	**0.00 ± 0.08**	**-42.38 ± 3.93**	**-42.39 ± 3.92**	**-42.08 ± 4.01**	**-40.56 ± 5.29**
	MQ	HD	−0.01 ± 0.02	−0.01 ± 0.02	−41.47 ± 4.00	−41.44 ± 3.99	−42.86 ± 4.30	−23.56 ± 3.80
		4K	0.12 ± 0.06	0.10 ± 0.06	−48.29 ± 2.09	−48.28 ± 2.10	−49.45 ± 2.64	−54.32 ± 4.81
		8K	0.03 ± 0.12	0.00 ± 0.14	−29.43 ± 5.44	−29.47 ± 5.42	−30.86 ± 5.61	−40.85 ± 5.88
		Avg	**0.05 ± 0.07**	**0.03 ± 0.07**	**-39.73 ± 3.84**	**-39.73 ± 3.84**	**-41.06 ± 4.18**	**-39.58 ± 4.83**
	HQ	HD	−0.05 ± 0.05	−0.05 ± 0.06	−40.30 ± 5.92	−40.28 ± 5.92	−33.33 ± 7.06	0.18 ± 11.62
		4K	0.13 ± 0.04	0.12 ± 0.03	−44.29 ± 2.69	−44.28 ± 2.69	−45.57 ± 2.97	−50.41 ± 7.06
		8K	0.06 ± 0.09	0.03 ± 0.10	−27.76 ± 3.93	−27.80 ± 3.90	−29.22 ± 4.46	−38.55 ± 4.85
		Avg	**0.05 ± 0.06**	**0.03 ± 0.06**	**-37.45 ± 4.18**	**-37.45 ± 4.17**	**-36.04 ± 4.83**	**-29.59 ± 7.84**
ERP-PRA	LQ	HD	−0.19 ± 0.14	−0.18 ± 0.11	−54.28 ± 4.69	−54.25 ± 4.73	−50.62 ± 4.86	−25.00 ± 5.22
		4K	−0.26 ± 0.15	−0.26 ± 0.14	−55.99 ± 2.00	−56.10 ± 1.97	−59.06 ± 0.79	−63.61 ± 3.59
		8K	−0.44 ± 0.19	−0.41 ± 0.23	−27.02 ± 3.63	−27.21 ± 3.68	−32.35 ± 5.18	−42.62 ± 5.57
		Avg	**-0.30 ± 0.16**	**-0.28 ± 0.16**	**-45.77 ± 3.44**	**-45.86 ± 3.46**	**-47.34 ± 3.61**	**-43.74 ± 4.79**
	MQ	HD	−0.16 ± 0.03	−0.15 ± 0.03	−47.29 ± 5.52	−47.25 ± 5.49	−48.62 ± 5.16	−19.46 ± 6.89
		4K	−0.21 ± 0.08	−0.22 ± 0.07	−55.97 ± 1.42	−56.03 ± 1.38	−58.78 ± 1.15	−64.31 ± 3.36
		8K	−0.45 ± 0.20	−0.48 ± 0.23	−29.49 ± 4.35	−29.67 ± 4.36	−32.39 ± 5.17	−42.71 ± 5.50
		Avg	**-0.27 ± 0.10**	**-0.29 ± 0.11**	**-44.25 ± 3.76**	**-44.31 ± 3.74**	**-46.59 ± 3.83**	**-42.16 ± 5.25**
	HQ	HD	−0.54 ± 0.08	−0.55 ± 0.10	−54.39 ± 4.14	−54.34 ± 4.14	−46.86 ± 5.66	−11.03 ± 11.02
		4K	−0.76 ± 0.14	−0.79 ± 0.14	−55.91 ± 1.65	−55.92 ± 1.65	−56.66 ± 1.67	−58.37 ± 5.78
		8K	−1.18 ± 0.37	−1.26 ± 0.42	−30.49 ± 4.51	−30.56 ± 4.48	−31.89 ± 4.69	−41.13 ± 5.13
		Avg	**-0.83 ± 0.20**	**-0.87 ± 0.22**	**-46.93 ± 3.43**	**-46.94 ± 3.42**	**-45.14 ± 4.01**	**-36.84 ± 7.31**
CMP-CRC	HQ	HD	−0.60 ± 0.31	−0.29 ± 0.43	−49.56 ± 6.03	−50.24 ± 5.72	−46.33 ± 6.54	−27.66 ± 7.25
		4K	−0.92 ± 0.08	−0.56 ± 0.19	−54.92 ± 1.84	−55.11 ± 1.98	−58.01 ± 1.81	−62.96 ± 4.87
		8K	−1.70 ± 0.88	−1.44 ± 0.77	−43.77 ± 4.23	−43.99 ± 4.17	−48.12 ± 4.23	−55.40 ± 4.51
		Avg	**-1.07 ± 0.42**	**-0.76 ± 0.46**	**-49.42 ± 4.04**	**-49.78 ± 3.96**	**-50.82 ± 4.19**	**-48.67 ± 5.54**
CMP-PRA	HQ	HD	−0.60 ± 0.31	−0.29 ± 0.43	−50.41 ± 4.24	−51.05 ± 4.00	−47.59 ± 4.46	−30.76 ± 4.22
		4K	−0.94 ± 0.07	−0.57 ± 0.18	−55.24 ± 2.79	−55.40 ± 2.90	−58.28 ± 1.90	−63.46 ± 4.20
		8K	−1.70 ± 0.89	−1.45 ± 0.77	−43.98 ± 4.14	−44.19 ± 4.09	−48.27 ± 4.17	−55.43 ± 4.54
		Avg	**-1.08 ± 0.42**	**-0.77 ± 0.46**	**-49.88 ± 3.73**	**-50.22 ± 3.66**	**-51.38 ± 3.51**	**-49.88 ± 4.32**

差异则小得多，对于典型的 VR视口而言仍在视觉可接受范围内，因为球面失真模式和以用户为中心的掩蔽降低了失真的可感知程度。

复杂度与资源考量：对于 N 分辨率，CRC 会存储 $N-1$ 次分析结果，而 PRA 会存储 N 个锚点。因此，CRC 的内存效率略高，不过对于现代服务器而言，PRA 的资源占用仍在可控范围内。这一比较凸显了实际应用中的权衡：CRC 倾向于保守复用并节省资源，而 PRA 则以少量率失真精度换取最高吞吐量。

基于*CMP*的变体及实际影响：CMP-CRC 和 CMP-PRA 遵循等矩形投影的趋势，但得益于对各面进行独立处理，可实现最高的整体加速比（最高达 4.2×）。额外的并行性，加上更加均匀的面统计特性，使基于CMP的流水线特别适用于对时效性要求较高的应用，例如实时VR流媒体，或在大型内容库中快速准备多种表示。据我们所知，这些结果首次系统地比较了符合OMAF标准的多码率编码流水线中，等矩形投影与立方体贴图投影下的分析复用行为。

6 结论与未来工作

本文提出了一种适用于符合OMAF标准的流媒体工作流中立方体贴图视频的快速多码率编码框架。基于x265的分析复用能力，我们提出了两种跨分辨率策略——级联分辨率链（CRC）和逐分辨率锚定（PRA）——以及它们基于立方体贴图的对应方案CMP-CRC和CMP-PRA。基于SJTU 8K 数据集，我们的实验表明，与穷举式全搜索编码相比，所有变体都大幅缩短了编码时间；在最有利的CMP-CRC配置下，BDET最高可达 –50%，实际运行时间加速比可达 4.2×。结果凸显了明显的权衡：CRC具有更好的率失真保真度和更低的内存占用（分析仅需存储 $N-1$ 次），而PRA则以适度的率失真性能下降和略高的内存开销（N 个锚点）为代价，实现了最高编码速度。基于CMP的设计进一步利用逐面独立性实现并行编码，最高可实现4.2× 倍加速，使该框架适用于实时VR流媒体等近实时场景。重要的是，这些改进可与OMAF的基于区域的架构及DASH传输无缝集成，无需客户端修改。

未来的工作包括将所提出的框架扩展到基于VVC的流水线，探索内容感知的锚点选择与复用强度（例如，通过机器学习或在线复杂度分类），并集成视口预测或QoE驱动的码率阶梯优化。另一个有前景的方向是将所提出的编码器端加速与解码器端增强相结合，例如超分辨率或视图自适应上采样，从而支持下一代沉浸式流媒体服务。

参考文献

[1] A. Yaqoob、T. Bi和G.-M. Muntean，“自适应360°视频流媒体综述：解决方案、挑战与机遇”，第22卷，第4期，第2801–2838页。[2] M. Zink、R. Sitaraman和K. Nahrstedt，“可扩展360°视频流传输：挑战、解决方案与机遇”，第107卷，第4期，第639–650页。[3] K.-C. Huang、P.-Y. Chien、C.-A. Chien、H.-C. Chang和J.-I. Guo，“360度全景视频系统设计”，载于《2014年国际VLSI设计、自动化与测试研讨会技术论文集》，第1–4页。[4] Y. He、X. Xiu、P. Hanhart、Y. Ye、F. Duanmu和Y. Wang，“使用混合立方体贴图投影的内容自适应360度视频编码”，载于《2018年图像编码研讨会（PCS）》，第313–317页，ISSN：2472-7822。[5] J. Carreira、S. M. M. de Faria、L. M. N. Tavora、A. Navarro和P. A. Assuncao，“使用自适应分辨率切换的360°视频通用视频编码”，载于《2020年IEEE国际图像处理会议（ICIP）》，第3398–3402页，ISSN：2381-8549。[6] X. Chen和G. Cao，“多核移动设备上的节能型360度视频流媒体”，载于《IEEE INFOCOM 2023——IEEE计算机通信会议》，第1–10页，ISSN：2641-9874。[7] I. Sodagar，“互联网多媒体流媒体的MPEG-DASH标准”，载于《IEEE MultiMedia》，第18卷，第4期，2011年，第62–67页。[8] R. Pantos和W. May，“HTTP实时流媒体”，RFC 8216，2017年。 [在线]。可访问：https://rfc-editor.org/rfc/rfc8216.txt [9] A. Bentaleb、B. Taani、A. C. Begen、C. Timmerer和R. Zimmermann，“基于HTTP的流媒体码率自适应方案综述”，《IEEE通信综述与教程》，第21卷，第1期，第562–585页，2019年。[10] V. V. Menon、H. Amirpour、M. Ghanbari和C. Timmerer，“EMES：面向高效视频编码自适应码率流媒体的高效多重编码方案”，《ACM多媒体计算、通信与应用汇刊》，第19卷，第3s期，2023年3月。[11] V. V. Menon、P. T. Rajendran、C. Feldmann、K. Schoeffmann、M. Ghanbari和C. Timmerer，“面向自适应实时流媒体的感知恰可察觉差异的两遍式逐标题编码方案”，《IEEE电路与系统视频技术汇刊》，第34卷，第2期，第1281–1294页，2024年2月。[12] D. Monakhov、I. D. Curcio和S. Mate，“关于视口依赖型流媒体中的数据浪费”，载于《2019年IEEE第21届多媒体信号处理国际研讨会（MMSP）》，第1–6页，ISSN：2473-3628。[13] Y. Shinohara、S. Shirasaki、Y. Wu、K. Kanai和J. Katto，“采用基于聚类的视口预测的基于瓦片的360度DASH流媒体性能评估”，载于《2019年IEEE国际消费电子会议——中国台湾（ICCE-TW）》，第1–2页。[14] A. Yaqoob和G.-M. Muntean，“面向360°视频的视场预测辅助视口自适应传输方案”，第67卷，第3期，第746–760页。[15] S.-Z. Qian、Y. Zhang和T. Lin，“FBRA360：面向360°视频流媒体的基于模糊逻辑的码率自适应方案”，载于《2023年IEEE国际多媒体与博览会研讨会（ICMEW）》，第122–127页。[16] M. Tang和V. W. Wong，“视口自适应360度视频流媒体的在线码率选择”，第21卷，第7期，第2506–2517页。[17] W. Feng、S. Wang和Y. Dai，“自适应360度流媒体：利用多窗口和随机视口预测进行优化”，第24卷，第7期，第5903–5915页。[18] Q. Yang、W. Gao、C. Li、H. Wang、W. Dai、J. Zou、H. Xiong和P. Frossard，“360Spred：基于三维可分离图卷积网络的360度视频显著性预测”，第34卷，第10期，第9979–9996页。[19] I. U. Khan、K. Han和J. W. Lee，“TransUser's：用于360°视频用户体验生成的基于Transformer的显著目标检测”，载于《2024年IEEE人工智能与扩展现实和虚拟现实国际会议（AIxVR）》，第256–260页。[20] M. Orduna、P. Pérez、J. Gutiérrez和N. García，“长时长360度视频中的内容沉浸式主观质量评估”，载于《2023年第15届多媒体体验质量国际会议（QoMEX）》，第73–78页。[21] J. Tu、C. Chen、Z. Yang、M. Li、Q. Xu和X. Guan，“基于感知敏感度的 360^\circ瓦片视频流媒体在工业监控中的应用”，第19卷，第9期，第9777–9789页。[22] Z. Jiang、X. Zhang、Y. Xu、Z. Ma、J. Sun和Y. Zhang，“面向360度视频流媒体的基于强化学习的码率自适应”，第67卷，第2期，第409–423页。[23] A. Katsenou、V. V. Menon、G. Laurinaviciute、B. Bross和D. Marpe，“高效自适应VVC流媒体的多目标帕累托前沿优化”， 2026 [ ] 2601 10607 。在线。可访问：https://arxiv.org/abs/ [24] 。V. V. Menon、A. Premkumar、P. T. Rajendran、A. Wieckowski、B. Bross、C. Timmerer和D. Marpe，“采用延迟感知动态分辨率编码的节能型自适应视频流媒体”，载于《第3届Mile-High Video会议论文集》会议，丛书 MHV '24，2024年，第21–27页。[25] M. M. Hannuksela 和 Y.-K. Wang，《全向媒体格式（OMAF）概述》，IEEE论文集，第109卷，第9期，第1590–1606页，2021年。[26] C. Cortés、P. Pérez、J. Gutiérrez 和 N. García，《视频时延对视口自适应沉浸式流媒体中质量、临场感和眩晕的影响》，载于《2020年第十二届多媒体体验质量国际会议（QoMEX）》，第1–4页。[27] M. Yang、W. Zou、J. Song 和 F. Yang，《提升视口自适应360度视频流媒体的体验质量：感知分析与实现》，第28卷，第4期，第64–73页。[28] J. Kufa、L. Polak、M. Simka 和 A. Stech，《全向8K视频的软件与硬件编码：性能研究》，载于《2023年第33届Radioelektronika国际会议（Radioelektronika）》，第1–5页。[29] S. Singh 和 J. Chakareski，《用于沉浸式第一人称视角无人机导航的空中360度视频传输》，载于《2023 IEEE多媒体国际研讨会（ISM）》，第139–146页。[30] X. Liu、Y. Huang、L. Song、R. Xie 和 X. Yang，《上海交通大学超高清360度沉浸式视频序列数据集》，载于《2017年虚拟现实与可视化国际会议（ICVRV）》，第400–401页，ISSN：2375-141X。[31] D. Wang、S. Du、Y. Sun、S. Xia、F. Dufaux、H. Guo、G. Wang 和 C. Zhu，《面向VVC的360度视频快速编码单元划分算法》，载于《2025 IEEE多媒体与博览会国际会议（ICME）》，第1–6页，ISSN：1945-788X。[32] D. Schroeder、P. Rehm 和 E. Steinbach，《用于多码率高效视频编码的块结构复用》，载于《2015年IEEE国际图像处理会议（ICIP）》，2015年，第3972–3976页。[33] V. V. Menon、A. Wieckowski、Y. Liu、B. Bross 和 D. Marpe，《面向自适应VVC流媒体的加速多码率编码块划分策略》， 2025 [ ] 2510 14645 。在线。网址：https://arxiv.org/abs/ [34] 。A. Mathesawaran、P. K. Karadugattu、P. Ramachandran、A. Giladi、D. Grois、P. Venkatesan 和 A. Balk，《用于低复杂度多码率HEVC编码的开源框架》，载于《数字图像处理应用 XLIII》，A. G. Tescher 和 T. Ebrahimi 编，第11510卷，国际光学与光子学学会。SPIE，2020年，第115101Y页。[35] B. Ray、J. Jung 和 M.-C. Larabi，《面向等矩形投影360度视频内容的低复杂度视频编码器》，载于《2018 IEEE声学、语音与信号处理国际会议（ICASSP）》，2018年，第1723–1727页。[36] Y. He、Y. Ye、P. Hanhart 和 X. Xiu，《360度视频编码中运动补偿预测的几何填充》，载于《2017年数据压缩会议（DCC）》，第443–443页，ISSN：2375-0359。[37] D. V. Nguyen、H. T. T. Tran 和 T. C. Thang，《时延对360度视频通信的影响》，载于《2017年TRON研讨会（TRONSHOW）》，第1–6页。[38] K. K. Sreedhar、A. Aminlou、M. M. Hannuksela 和 M. Gabbouj，《面向虚拟现实应用的360度视频视口自适应编码与流媒体传输》，载于《2016 IEEE多媒体国际研讨会（ISM）》，第583–586页。[39] G. J. Sullivan、J.-R. Ohm、W.-J. Han 和 T. Wiegand，《高效视频编码（HEVC）标准概述》，载于《IEEE电路与系统视频技术汇刊》，第22卷，第12期。IEEE，2012年12月，第1649–1668页。[40] B. Bross、Y.-K. Wang、Y. Ye、S. Liu、J. Chen、G. J. Sullivan 和 J.-R. Ohm，《通用视频编码（VVC）标准及其应用概述》，第31卷，10，2021年，3736 3764 no. pp. – 。[41] I. Siqueira、G. Correa 和 M. Grellert，《HEVC与VVC视频编码器的率失真和复杂度比较》，载于《IEEE第11届拉丁美洲电路与系统研讨会（LASCAS）》，2020年。[42] D. Schroeder、A. Ilangovan 和 E. Steinbach，《用于多种分辨率自适应HTTP流媒体的HEVC多码率编码》，载于《2015 IEEE第17届多媒体信号处理国际研讨会（MMSP）》，2015年，第1–6页。[43] K. Goswami、B. Hariharan、P. Ramachandran、A. Giladi、D. Grois、K. Sampath、A. Matheswaran、A. K. Mishra 和 K. Pikus，《面向自适应码率流媒体的自适应多分辨率编码》，载于《2018年第25届IEEE国际图像处理会议（ICIP）》，2018年，第1008–1012页。[44] V. V. Menon、C. Feldmann、K. Schoeffmann、M. Ghanbari 和 C. Timmerer，《用于自适应视频流媒体高效编码的绿色视频复杂度分析》，载于《第一届绿色多媒体系统国际研讨会论文集》，2023年，第16–18页。[45] HSTP-VID-WPOM，《使用客观指标评估视频编码效率实验的工作实践》，国际电信联盟，2020年。 [在线]。网址：http://handle.itu.int/11.1002/pub/8160e8da-en [46] C. Herglotz、A. Heindel 和 A. Kaup，《视频编码的解码能耗率失真优化》，载于《IEEE电路与系统视频技术汇刊》，第29卷，第1期，2019年1月，第171–182页。