高技术通讯 2026 年 第36卷 第7期:713 - 723

doi:10.3772/j.issn.1002 - 0470.2026.07.005

基于纹理结构与 CA-Transformer 的深度图帧内快速编码算法①

李竟泽② 贾克斌③

(北京工业大学信息科学技术学院 北京 100124)

(先进信息网络北京实验室 北京 100124)

摘 要 针对三维高效视频编码(three-dimensional high efficiency video coding,3D-HEVC)在深度图编码中复杂度过高的问题,本文提出了一种基于纹理划分结构与深度学习并行的算法实现深度图的快速编码算法。首先,提出了基于纹理图划分结构加速深度图编码的方法,提前识别并简化编码树单元(coding tree unit,CTU),得到快速划分结果。其次,针对无法使用纹理图划分结构快速划分的 CTU,设计了一种名为 CA-Transformer 的神经网络。应用卷积神经网络对局部特征的敏感性与视觉 Transformer 对全局特征的敏感性,将结构复杂的深度图图像输入到本文的 CA-Transformer 网络中,得到对深度图 CTU 划分结构的预测。使用本文算法替换原始编码标准的 CTU 划分过程,实验结果表明,本文算法在保证率失真性能与合成视点质量几乎不受影响的前提下,平均减少了 56.00% 的编码时间。

关键词 三维高效视频编码;纹理图;深度图;编码树单元划分;帧内编码

随着信息与数字技术的迅猛发展,视频已经成为了人们信息交流与生活娱乐的主要媒介之一。在此背景下,技术在不断进步,视频质量要求也在提高,特别是高清、甚至超高清视频的出现,为人们的生活带来了更加清晰的高品质视觉体验。与此同时,人们开始倾向于更沉浸式的感受,三维(three-dimension,3D)视频的出现满足了这一需求。3D 视频通过多视点视频合成,带来了立体化效果,提高了交互性和沉浸度,也让虚拟现实、增强现实等技术离我们的生活越来越近。然而随着视频质量的提高和视点数的增加,3D 视频数据量直线上升,传输和存储成为 3D 视频广泛使用的难点,如何能在有限的网络带宽内高效传输 3D 视频成为了急需解决的问题。

2012 年 7 月,视频编码专家组(video coding experts group,VCEG)和动态图像专家组(moving pictures experts group,MPEG)共同指定了多视点视频加深度(multi-view video plus depth,MVD)[1]视频格式,将深度信息引入到 3D 视频编码的制定中。随着高效视频编码(high efficiency video coding,HEVC)[2]标准在 2013 年设计完成,在 2015 年 JCT-3V 小组开发出多视点高效视频编码(multiview HEVC,MV-HEVC)[3]标准和 3D-HEVC[4]标准,其中 MV-HEVC 仅包含纹理图,而 3D-HEVC 包含表示信息的纹理图和表示距离的深度图;针对不同视点以及纹理图深度图采用的编码技术不尽相同,因此 3D-HEVC 的复杂度更高,更多工作者将重心放在降低 3D-HEVC 复杂度上。在 3D-HEVC 中,编码树单元(coding tree unit,CTU)的划分带来了极大的编码复杂度,由于深度图编码过程中加入了更新的技术,因此深度图 CTU 的划分更为复杂耗时,导致 3D-HEVC 的使用范围并不宽广,需要降低 3D-HEVC 深

① 北京市自然科学基金(4212001)资助项目。
② 男,2001 年生,硕士生;研究方向:三维视频编码技术;E-mail:2229279934@qq.com。
③ 通讯作者,E-mail:kebinj@bjut.edu.cn。
(收稿日期:2025 - 02 - 25)

度图的编码过程展开研究。

近年来,众多学者为了降低 3D-HEVC 帧内编码复杂度,提出了大量的方法与算法。大致可以分为基于传统方法的算法和基于学习的算法。

传统方法常通过利用深度图的纹理边缘等特性,提前跳过 CTU 的编码流程,降低计算复杂度。Song 等人[5]通过编码单元的复杂度提前确定编码中的预测模式选择,加速编码流程。Lee 等人[6]对预测模式的代价和概率进行计算,保留更少的可选择预测模式,降低了编码复杂度。Lin 等人[7]通过纹理与深度图的性质,将编码树单元 CTU 分为复杂、中等与均匀三类,并且实行不同的早期终止准则,加速编码过程。Huo 等人[8]提出了一种跳过率失真代价(rate-distortion cost,RD)成本检查从而减少预测模式的快速算法,显著降低了编码复杂度。

传统方法主要基于公式阈值计算设计,普适性不强,编码质量难以得到保障,因此基于学习的算法在近几年得到广泛发展。首先,基于机器学习的算法得到了应用。Hamout[9]使用自组织图和高效特征选择性跳过了深度图不必要的预测模式,降低编码时间。Li 等人[10]将 CTU 划分结构视为聚类问题,并建立了 3 种聚类模型,在编码时间大幅降低的情况下,合成视点的 BDBR(Bjφntegaard delta bit rate)几乎没有上升。Bakkouri 等人[11]利用梯度增强机(gradient boosting machine,GBM)分析图像平均局部方差(average local variance,ALV),通过自适应阈值对 CTU 快速划分,编码损失效率可以忽略不计。Wang 等人[12]通过贝叶斯决策理论加速了 3D-HEVC 中深度图编码,并成功进行应用。Saldanha 等人[13]用编码器上下文属性构建了静态决策树,从而定义编码单元是否继续分割。Zou 等人[14]利用贝叶斯决策和纹理深度视频与空间相关性分析深度图的特征,从而跳过部分预测模式以实现快速编码。

机器学习算法需要手动设计特征,且特征表达能力鲁棒性较差,因此基于深度学习的算法逐渐占据了主流地位。刘畅等人[15]提出了一种基于多分支网络的深度图帧内 CTU 快速划分算法,该算法通过卷积神经网络快速预测 CTU 划分的最终结果。Sebai 等人[16]结合了 VGG19 模型加速了 CTU 的划分,确保高质量合成视点的生成。Liu 等人[17]利用 swin-Transformer,跳过冗余的 CTU 划分过程,平均编码时间压缩到一半以上,有效降低了编码复杂度。Liu 等人[18]建立了一种结合边缘复杂度和简单卷积神经网络(convolutional neural network,CNN)的 CTU 划分矫正算法,通过边缘提取网络有效将边缘信息融入到 CTU 划分中,提升了编码效率。Zhang 等人[19]提出了一种基于空间金字塔的神经网络,提取 3 种不同类型的编码单元特征,最后应用局部注意力机制加速编码。Orman 等人[20]则创造了一种名为多深度卷积神经网络(multi-deep CNN,MD-CNN 的模型,通过训练网络得到编码单元的划分结构以加速编码过程。

上述工作成果降低了 3D-HEVC 的编码复杂度且保证了一定的编码质量,但仍有可改进之处。本文有效将纹理图的编码信息与深度图的图像特征结合,提出了一种快速编码算法,跳过冗余的 CTU 率失真计算过程。实验结果表明,本文提出的算法在率失真性能与合成视点质量几乎保持不变的条件下,有效降低了 3D-HEVC 的编码复杂度。

1 3D-HEVC 标准编码原理

1.1 3D-HEVC 编码结构

3D-HEVC 编码标准采用了 MVD 视频格式:如图 1 所示,标准编码序列由 3 个不同视点的纹理图和深度图构成,共有 6 个视频序列。首先,对独立视点的纹理图进行编码,此时的编码标准是未进行修改的 HEVC 编码标准;随后对独立视点下的深度图进行编码。在对独立视点编码完成后,再对非独立视点按照先纹理图后深度图的编码顺序进行编码。按上述编码顺序直至编码到最后一帧为止。

1.2 复杂度分析

1.2.1 深度图编码技术

深度图表示的是物体距离摄像机的远近距离关系,因此 2 个距离摄像机远近不同的物体在深度图中的接触边缘是十分锋利的,此时使用常规的 HEVC 编码标准进行编码时会造成明显的编码伪影,影响后续合成虚拟视点质量。因此3D-HEVC在深

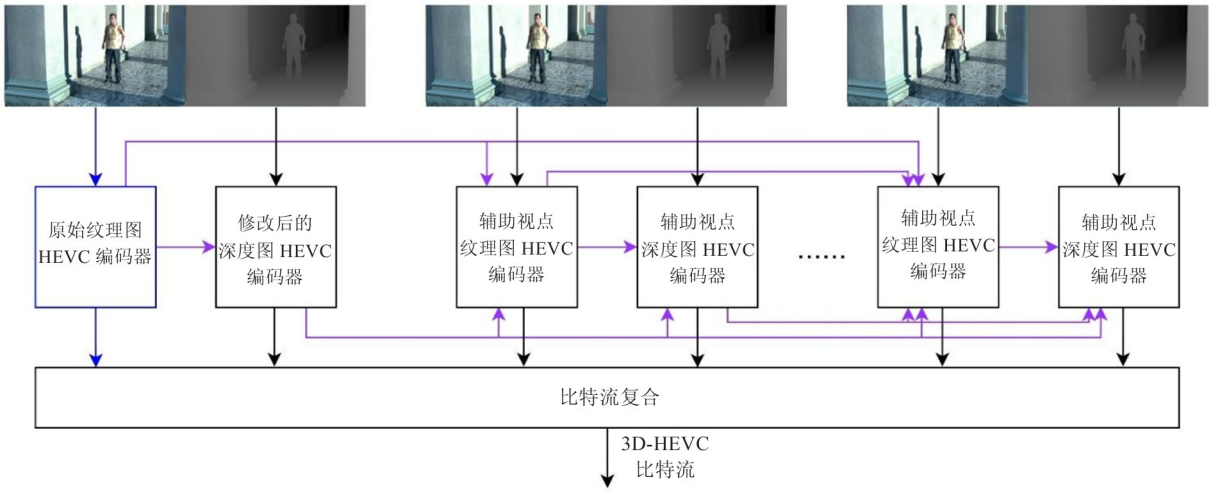


图 1　3D-HEVC 编码器结构

度图编码时加入了深度模型模式(depth modeling modes,DMM)于常规预测模式中,现在仍在使用的有楔形划分方式(Wedgelet Partition)[21]和轮廓划分方式(Contour Partition)[22]2 种。与此同时深度图编码算法还加入了分段式直流分量(direct current, DC)编码等深度图编码方法,显著增加了深度图编码复杂度。

1.2.2　CTU 划分结构

在 3D-HEVC 编码结构中,占据编码时间最久、编码复杂度最高的是 CTU 划分过程。CTU 具体的划分方式为:将待编码图像以 CTU 为单位进行编码预测,经过自上而下的划分和自下而上的比较从而得到最优的划分结构。

自上而下的划分模式:默认情况下的 CTU 大小为 64×64,1 个 64×64 大小的 CTU 被划分为 4 个 32×32 的子编码单元(coding unit, CU);每个 32×32 的 CU 被划分为 4 个 16×16 的子 CU;每个 16×16 的 CU 被划分为 4 个 8×8 的子 CU。

自下而上的比较模式:求和4 个8×8 的子 CU 与单个 16×16 的父 CU 计算各自最优的 RD-cost 并进行比较,决定是否划分;求和 4 个已确定最优划分结构的 16×16 的子 CU 与单个 32×32 的父 CU 计算各自最优的 RD 成本并进行比较,决定是否划分;求和 4 个已确定最优划分结构的 32×32 的子 CU 与 64×64 的 CTU 计算各自最优的 RD 成本并进行比较,决定是否划分。最终得到划分结构,如图 2 所示。

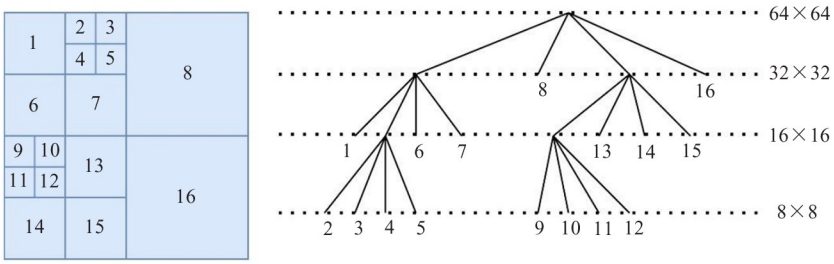


图 2　CTU 划分结构

在自下而上的比较模式中,每个 CU 都要在率失真过程中遍历帧内预测模式,并且需要计算至少 3 次(至多 8 次)的 RD 成本才可以得到最佳的帧内预测模式,对于深度图而言,由于加入了新的帧内预测技术,因此计算量相比于纹理图大量上升,非常耗时。

经过对 3D-HEVC 帧内编码流程的分析,深度图的编码时间平均占据了 80.00% 以上的编码时间,如图 3 所示。但大部分的 CTU 划分深度很浅,因此上述的 CTU 划分过程中很多计算是冗余的,如

果能直接预测 CTU 的划分结构而跳过复杂的率失真计算过程,可以实现复杂度的大幅降低。本文通过纹理图划分结构以及深度学习并行快速预测了划分结构,在编码损失处于可接受范围内大幅缩短了编码时间。

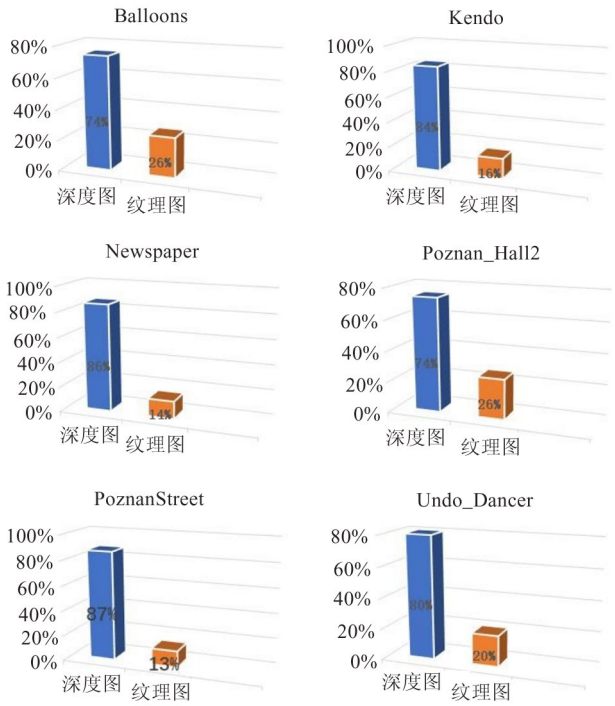


图 3　深度图与纹理图编码时间对比图

2　基于纹理结构与 CA-Transformer 算法设计

2.1　设计动机

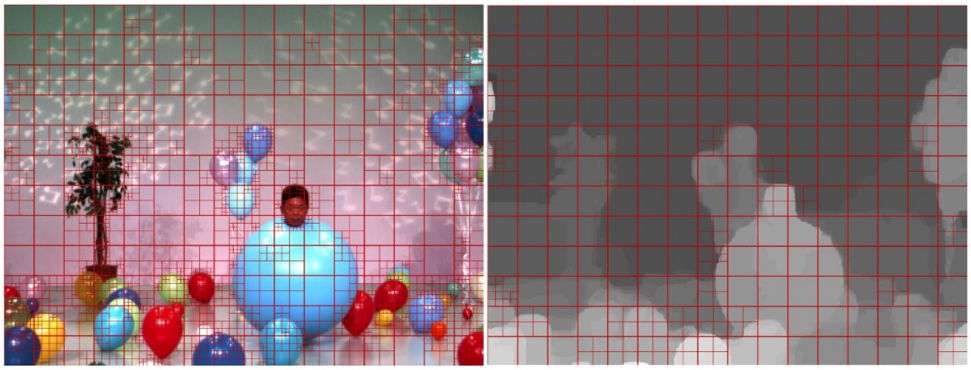
在前人的方法设计中,仅仅就深度图的特征来设计算法,但由于深度图内容略为粗糙,快速编码算法中会存在一定的精度损失,于是本文利用同视点下的纹理图信息结合深度图信息编码,在保证编码速度的前提下还可以提升部分编码质量。

同一视点下的纹理图与深度图有相同的构图结构和场景,区别是纹理图的纹理信息丰富,深度图饱含深度信息。本文方法计划结合在同一视点下的纹理图和深度图编码中的大致相同之处,通过实验寻找最佳阈值,提前确定深度图中简单 CTU 并快速结束划分。针对非简单 CTU,使用 CA-Transformer 网络进行快速预测,得到每个 CTU 的最终划分结构,实现快速编码。

2.2　基于纹理图结果的划分结构设计

在 3D-HEVC 编码标准中,同一视点下的纹理图与深度图表示相同的场景,一前一后依次进行编码,研究发现纹理图的编码简单区域与深度图的编码简单区域存在大部分重合,如图 4 所示。由于深度图存在大量的平坦区域,因此大部分 CTU 划分结构简单,而纹理图的平坦区域映射到深度图上时,平坦区域几乎没有发生太大变化,在此基础上可以考虑以纹理图的编码结果来加速深度图的编码。

编码单元 CTU 的划分结构可以简化为如图 5 所示,每个 CTU 可以通过 4×4 的架构图来显示划分结构。d_1 ~ d_{16} 表示每一个区域的划分深度,范围是 0~3。当划分深度是 0 时,则 d_1 ~ d_{16} 全部置为 0,且 CTU 不进行划分;当划分深度是 1 时,则 d_1 ~ d_{16} 全部置为 1,且 CTU 只进行一次划分;当划分深度是 2 时,会对对应的 32×32 大小的 CTU 进行划分;当深度是 3 时,只对对应的 16×16 大小的 CTU 进行划分。



(a) 纹理图　　(b) 深度图

图 4　同视点下纹理图与深度图划分结构对比图

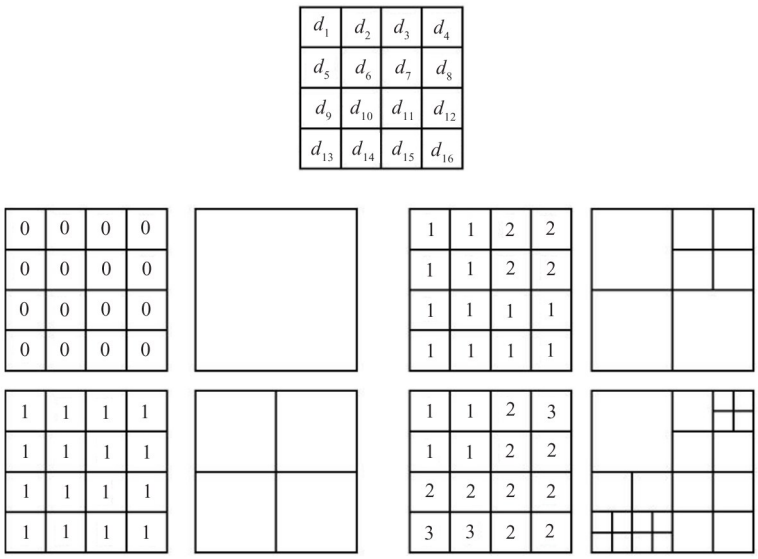


图 5　CTU 4×4 划分结构图

本文提出深度和概念,用符号 S_D 表示。S_D 统计了对应 CTU 中划分深度数据求和,用该数据来表示 CTU 的划分复杂程度。S_D 在 CTU 划分深度为 0 时也为 0; S_D 在 CTU 划分深度为 1 时为 16; S_D 在 CTU 划分深度超过 1 时,数值会从 20 起逐步增加。

$$S_D = \sum_{i=1}^{16} d_i \tag{1}$$

本文通过分析不同纹理图对应的深度图 S_D 值,得到了合适的划分阈值。我们定义深度图 CTU 的划分深度为 0 或 1 时,深度图 CTU 为简单 CTU。我们将同一深度图中不同 S_D 对应的简单 CTU 的比例与简单 CTU 预测准确率进行分析比对,得到如图 6 所示的结果。随着 S_D 数值的不断升高,简单 CTU 的划分范围也将提升,但同样会将部分复杂 CTU 划分为简单 CTU,导致一定程度的精度损失。经过对多个 S_D 数值的实验比对,我们最终选择 $S_D = 16$ 为纹理图对应深度图简单 CTU 的划分阈值,在此数值下能够保证很高的预测精确度同时也能维持较高的简单 CTU 划分数量。

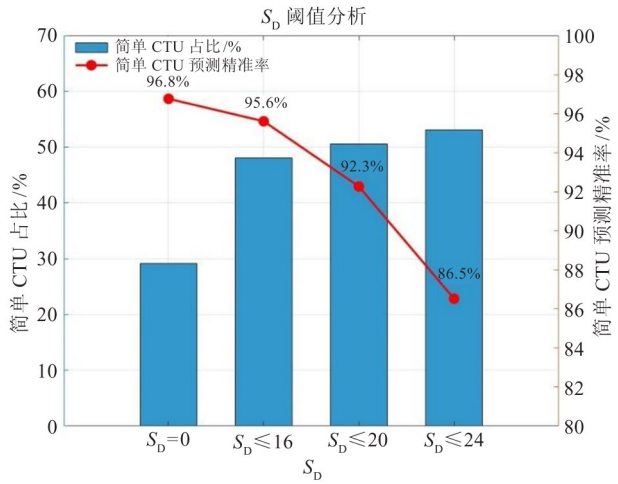


图 6　S_D 阈值图

其中, d_i 表示每个 8×8 的 CU 划分深度。本文针对序列 Balloons、Poznan _ Hall2 的纹理图和深度图各 80 帧的划分数据进行分析,当纹理图的 $S_D \leqslant 16$ 时,即纹理图 CTU 的划分深度为 0 或 1 时,深度图对应 CTU 的 S_D 数据由表 1 所示。

从表 1 分析得出,当纹理图的划分深度为 0 或 1 时,深度图的划分深度很浅,平均 90.00% 以上的深度图 CTU 划分深度为 0,5.56% 的深度图 CTU 划分深度为 1,1.47% 的深度图 CTU 仅有一块 32×32 子 CU 被划分到深度为 2 的范围。上述 3 种情况的占比平均达到了 99.12%。

当 3D-HEVC 编码时,首先对纹理图进行编码,得到的 CTU 划分结果通过计算保存下每个 CTU 的 S_D 数值,以 16 为阈值,当 $S_D \leqslant 16$ 时,将该 CTU 记录为简易 CTU,并且后续深度图编码时跳过简易 CTU 编码过程,直接将 CTU 划分结果置为 0,即不划分。

本文的这种方法达到了 90.00% 以上的准确率,且由于上述范围的深度图多处在平坦区域,仅会造成轻微少量失真,对后续合成视点的影响不大,因

此本方法可以使用。

表 1　S_D 数值分布表

序列	量化参数	$S_D=0$	$S_D=16$	$S_D=20$	$S_D\leqslant 20$
Balloons	25/34	72.99%	17.81%	5.71%	96.51%
	30/39	89.42%	8.13%	1.81%	99.37%
	35/42	94.93%	3.51%	0.96%	99.41%
	40/45	94.88%	4.35%	0.54%	99.77%
Poznan _ Hall2	25/34	89.33%	6.46%	1.77%	97.56%
	30/39	96.17%	2.70%	0.61%	99.48%
	35/42	98.64%	0.98%	0.30%	99.91%
	40/45	99.35%	0.57%	0.07%	99.98%
平均		91.96%	5.56%	1.47%	99.12%

2.3　基于深度学习划分方法

2.2 节中的方法,适合应用于结构简单的 CTU,因此本文对于剩余的复杂 CTU 设计了基于深度学习的快速算法,与 2.2 中的方法相结合,实现更高效的视频编码。本节使用了一种基于 CA(Coordinate Attention)[23]的 CNN 与 Dilated-Vision Transformer[24]结合的神经网络,命名为 CA-Transformer 网络。该网络通过提取图像的局部特征与全局特征,预测复杂 CTU 的划分结构,实现 3D-HEVC 的快速编码。

2.3.1　网络结构

图 7 展示了本文提出的 CA-Transformer 网络。网络的输入是待编码的 CTU,大小为 64×64,经过预处理操作后送入 CA-Transformer 网络,输出为 21 个 0 或 1 的二分类值,分别代表了 1 个 64×64 的 CTU,4 个 32×32 的 CTU,16 个 16×16 的 CTU 是否划分,如图 8 所示。将结果繁多的划分结构数据转换为二分类数据。下面介绍 CA-Transformer 各个分支的作用与功能。

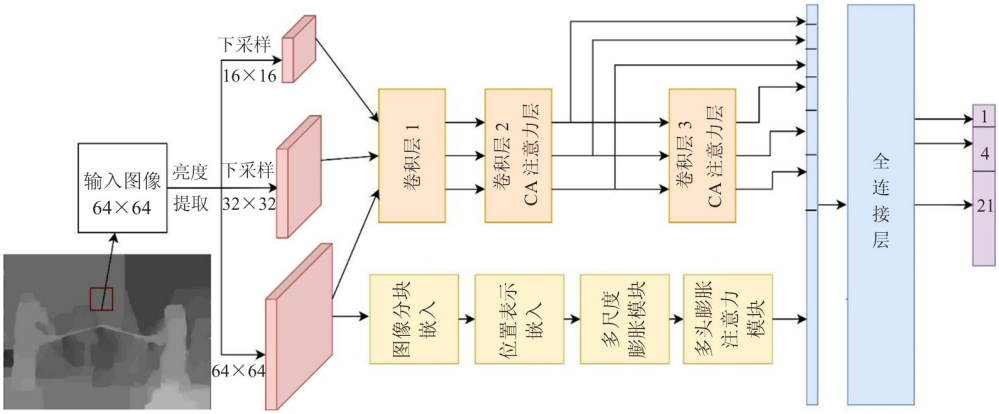


图 7　CA-Transformer 网络图

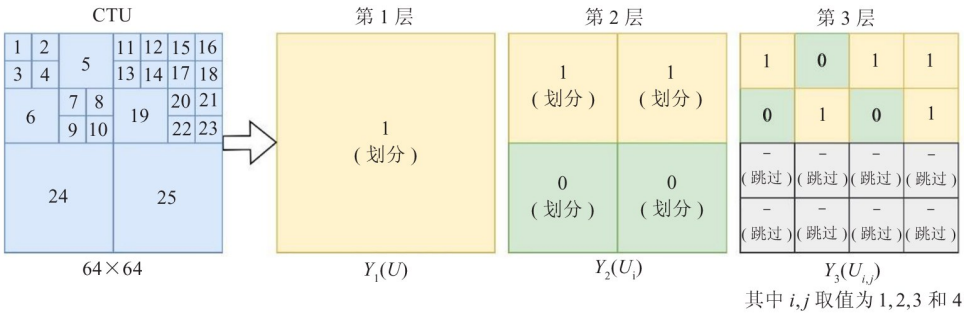


图 8　网络输出示意图

第 1 个分支由基于 CA 的 CNN 构成。首先,将输入数据通过预处理,留下亮度分量。并且将输入降采样构成 3 条支路:第 1 条支路是 16×16 为输入,第 2 条支路是 32×32 为输入,第 3 条支路是 64×64 为输入。3 条支路的输入首先经过卷积层 1,该卷积层是由 16 个 8×8 的卷积核构成,提取了输入信息的低级特征。其次,低级特征将进入以 CA 层及 24 个 4×4 卷积核为核心的卷积层 2,成功提取了输入信息的中级特征。最后,中级特征进入以 CA 层及 32 个 2×2 卷积核为核心的卷积层 3,提取了输入信息的高级特征。CA 层如图 9 所示,将输入信息按宽度维度和高度维度进行平均池化,得到不同的张量。随后拼接进行压缩卷积,通过 ReLU 激活函数得到非线性特征。将该特征重新拆分到高度支路与宽度支路中,并且通过卷积恢复通道数,再次得到归一化特征。将该自注意力特征加权到输入上并进行输出。

CA 模块通过分别处理高度和宽度方向的特征,保留了更精细的位置信息,提升了卷积神经网络的一定全局通道注意力。

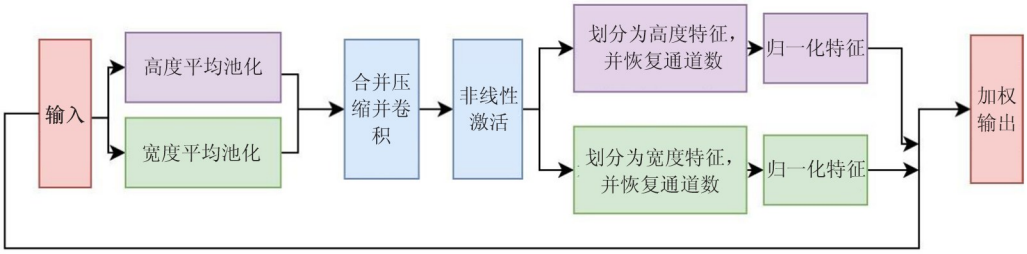


图 9　CA 模块示意图

第 2 条支路由 Dilated-Vision Transformer 所组成,用以同步提取更丰富的全局特征,捕捉各个向量之间的依赖关系。首先将输入数据提取亮度后直接送入网络,在此只使用 64×64 的亮度信息作为网络输入,可以更好地提取信息以便进行全局特征统筹。输入信息先经过预处理分块嵌入(Patch Embedding)和位置编码(Position Embedding),将二维图像转换为一维序列,且保留图像块的位置信息,使得更好地理解全局结构信息。

通过预处理的信息首先进入多尺度空洞注意力(multi-scale dilated attention,MSDA),即多尺度膨胀模块,首先进入 DwConv 深度卷积增强局部特征;随后通过滑动窗口膨胀注意力机制部分滑动窗口空洞注意力(sliding window dilated attention,SWDA),使用不同的膨胀率为不同头分配感受野,捕获多尺度上下文信息;最后通过 MLP 非线性变换层得到输出特征图。使用 MSDA 通过滑动窗口降低了计算复杂度,也弥补了 Transformer 局部特征提取弱的缺点。

经过 MSDA 提取的特征,进入多头自注意力(multi-head self-attention,MHSA),即多头膨胀注意力模块,MHSA 的核心机制在于多头注意力,本文方法将输入信息拆成 8 个头,每个头独立计算,从不同的特征空间捕获关系,最后将输出特征拼接。输出特征通过残差连接稳定输出效果,经过 MLP 加强非线性特征后,得到最终的输出特征。MHSA 的优点在于可以捕获输入序列中任意两元素之间的关系,模型表达能力强。

由于 CNN 的局部特征提取能力强,而 Transformer 的全局特征提取能力强,本文所用网络结合了 Coordinate Attention-CNN 与 Dilated-Vision Transformer,对于输入 CTU 的局部关系和全局关系都进行了信息提取,将输出经过全连接层后,得到 21 个划分结构数据。

由于最终的输出是以 0 或 1 作为划分标志,因此本文选取常用于二分类任务的交叉熵函数作为损失函数。将上述网络嵌入到 3D-HEVC 标准官方编码平台中,实现快速深度图帧内编码。

2.4　基于纹理结构与 CA-Transformer 算法流程

图 10 展示了本文的基于纹理图划分结构与深度学习并行的深度图帧内编码单元快速划分算法。开始视频编码后,读取输入序列。对于同一视点而

言,首先读取到纹理图序列,纹理图序列进入未修改的 3D-HEVC 编码器,实现传统编码,并且保存每个 CTU 的划分结果及深度和 S_D 数据。随后编码器读取同一视点下的深度图序列,首先以 16 为阈值,将 $S_D \leqslant 16$ 的 CTU 置为简单 CTU,将简单 CTU 的划分范围置为 0;复杂 CTU 进入 CA-Transformer 网络预测划分结构。逐帧逐视点逐 CTU 进行上述快速编码流程,直至编码至最后一帧的最后一个视点深度图的最后一个 CTU 为止。

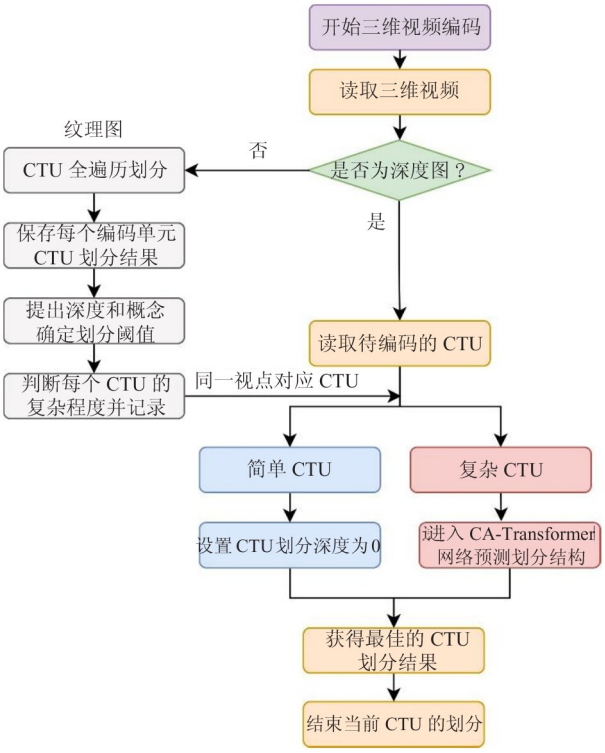


图 10　本文方法流程图

3　实验结果与分析

3.1　实验环境及配置

为了对实验结果有一致的衡量标准,本文实验在 Intel(R)Core(TM)i7-9750HCPU@2.60 GHz 的计算机上进行实验:所选取编码平台为 HTM 16.0,用 Visual Studio 2015 实现编译;深度学习划分编码算法使用 Tensorflow 1.13.1 框架,Python 语言进行编译,并且结合 HTM 16.0 进行使用。

3.2　数据集构建

数据集由 6 个官方标准测试序列的 3 个视点的深度图视频组成。其中训练集由 Newspaper 与 Poznan _ Hall2 的 0 ~ 200 帧组成,验证集由 Balloons 与 Poznan _ Street 的最后 10 帧构成;测试集则由 Kendo 和 Undo _ Dancer 的最后 30 帧构成。

3.3　评价指标计算

3.3.1　编码复杂度计算

$$\Delta T = \frac{T_{\text{ori}} - T_{\text{proposed}}}{T_{\text{ori}}} \times 100\% \tag{3}$$

式中:T_{ori} 为原始编码平台 HTM-16.0 的编码时间,T_{proposed} 为本文所提方法的编码时间。本文方法与 HTM-16.0 的时间比较在表 2 中显示。

3.3.2　合成视点失真计算

图 11 描述了计算合成视图的峰值信噪比(peak signal to noise ratio,PSNR)的计算过程。合成视图的 PSNR 是 6 个虚拟视图的平均 PSNR。

$$\text{PSNR}_{\text{synth}} = \frac{\sum \text{PSNR}_n}{6}, \quad n = 0.25, 0.50, 0.75, 1.25, 1.50, 1.75 \tag{4}$$

式中:$\text{PSNR}_{\text{synth}}$ 表示合成视图的 PSNR,PSNR_n 表示 6 个虚拟视图的 PSNR。

3.4　客观数据分析

为了验证本文方法的性能,对于不在训练集内的 4 个视频序列 Balloons、Kendo、Poznan _ Street 和 Undo _ Dancer 的前 30 帧进行了编码,得出的实验结果如表 2 所示。其中,T_Δ 表示本文算法可降低的编码时间比例;video PSNR/video bitrate 表示已编码纹理图在纹理图编码比特率上的 BDBR;video PSNR/total bitrate 表示已编码纹理图在总编码比特率上的 BDBR;synth PSNR/total bitrate 表示合成后的虚拟视点在总编码比特率上的 BDBR。

本文算法整体的编码计算时间上平均节省了 56.12% 的编码时间。同时根据表中数据,纹理图平均 BDBR 的损失为 0.70%,合成视点平均 BDBR 损失为 2.80%。说明本算法仅仅带来了微弱的失真,对视频质量的影响微乎其微。

表 3 中给出了本文算法相较于文献[5]、文献[10]、文献[18]和文献[20]的时间复杂度降低数值与纹理图 BDBR 损失比较。经过实验可知本文的时间复杂度降低最大,且纹理图 BDBR 损失相对较小,进一步证明了本文提出的深度图帧内编码单元快速划分算法的有效性。

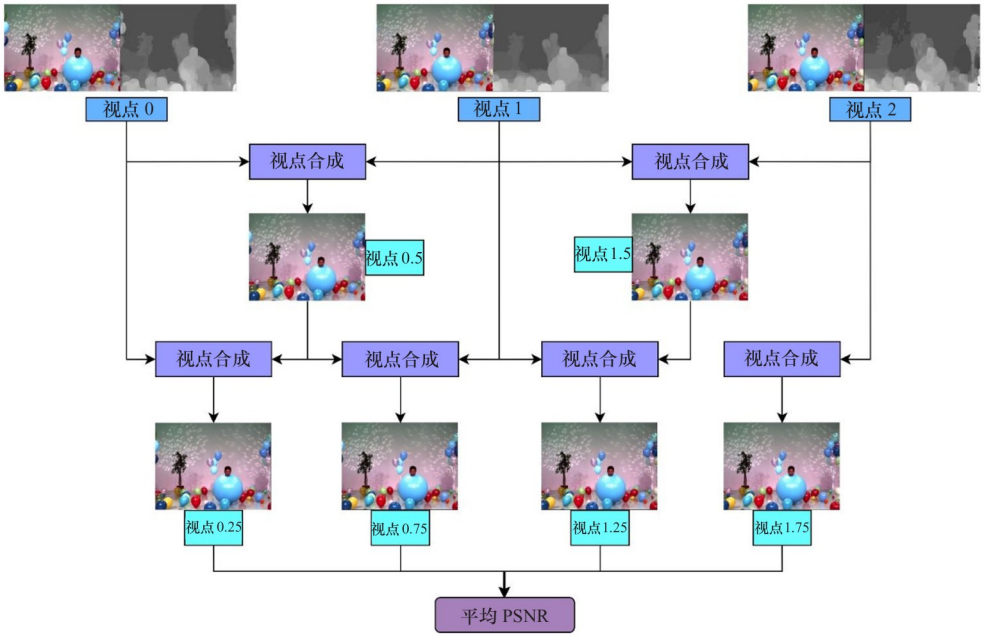


图 11　合成视点示意图

表 2　实验数据结果　%

视频序列	T_Δ	video 0/1/2	video PSNR/ video bit rate	video PSNR/ total bit rate	synth PSNR/ total bit rate
Balloons	54.17	0.00	0.00	0.80	2.70
Kendo	56.19	0.00	0.00	1.40	3.90
Poznan _ Street	58.36	0.00	0.00	0.50	1.90
Undo _ Dancer	55.75	0.00	0.00	0.20	2.70
1 024 × 768	55.18	0.00	0.00	1.10	3.30
1 920 × 1088	57.06	0.00	0.00	0.40	2.30
平均	56.12	0.00	0.00	0.70	2.80

表 3　对比实验数据结果　%

视频序列	文献[5]		文献[10]		文献[18]		文献[20]		本文算法	
	ΔT	BDBR	ΔT	BDBR	ΔT	BDBR	ΔT	BDBR	ΔT	BDBR
Balloons	37.9	0.4	45.7	0.9	49.7	1.1	42.5	2.2	54.2	0.8
Kendo	36.1	1.2	44.3	2.2	50.3	1.8	47.6	3.6	56.2	1.4
Poznan _ Street	35.7	0.2	45.8	0.6	51.4	0.9	45.3	1.8	58.4	0.5
Undo _ Dancer	35.3	0.4	44.1	0.7	49.9	0.3	47.9	1.3	55.8	0.2
平均	36.3	0.6	45.0	1.1	50.3	1.0	45.8	2.2	56.1	0.7

3.5　主观性能评价

本文算法所产生的失真与损失相对较小,如图 12所示,主观观察编码后的合成视点对比图,很难发现图像的失真情况,证明本文造成的合成视点损失在可接受范围内。

4　结　论

本文提出了一种基于纹理图划分结构与深度学习并行的深度图帧内编码单元快速划分算法,在保证视频编码质量的前提下有效降低了3D-HEVC帧

(a) 原始编码器编码　　(b) 所提出编码器编码

图 12　Kendo 测试序列在 qp25 下视点 1.5 处对比图

内编码的复杂度。

本文通过分析 3D-HEVC 视频编码标准的高复杂度以及以往快速编码方法的不足之处,有效将纹理图的已编码结果应用起来,结合深度学习方法预测复杂深度图 CTU 划分结构。使用上述方法嵌入到 3D-HEVC 中,跳过冗余的率失真计算过程,并保证视频编码质量几乎不变,降低了 56.12% 的编码时间。

由于本文的模型较为复杂,因此预测划分结构时会相比于轻量化网络消耗更多时间。未来研究会在保证编码质量的前提下继续轻量化本文网络框架。

参考文献

[1] Merkle P, Smolic A, Muller K, et al. Multi-view video plus depth representation and coding[C]//IEEE International Conference on Image Processing. San Antonio, USA: IEEE, 2007:201 – 204.

[2] Sullivan G J, Ohm J, Han W, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012,22(12):1649 – 1668.

[3] Hannuksela M M, Yan Y, Huang X, et al. Overview of the multiview high efficiency video coding (MV-HEVC) standard[C]//2015 IEEE International Conference on Image Processing. Quebec City, Canada: IEEE, 2015: 2154 – 2158.

[4] Tech G, Chen Y, Müller K, et al. Overview of the multiview and 3D extensions of high efficiency video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016,26(1):35 – 49.

[5] Song W J, Dai P, Zhang Q. Content-adaptive mode decision for low complexity 3D-HEVC[J]. Multimedia Tools and Applications, 2023,82(17):26435 – 26450.

[6] Lee J Y, Park S H. Fast depth intra mode decision using intra prediction cost and probability in 3D-HEVC[J]. Multimedia Tools and Applications, 2024, 83(34): 80411 – 80424.

[7] Lin J R, Chen M J, Yeh C H, et al. Vision-oriented algorithm for fast decision in 3D video coding[J]. IET Image Processing, 2022,16:2263 – 2281.

[8] Huo J Y, Zhou X, Yuan H, et al. Fast rate-distortion optimization for depth maps in 3D video coding [J]. IEEE Transactions on Broadcasting, 2023,69(1):21 – 32.

[9] Hamout H, Hammani A, Elyousfi A. Fast 3D-HEVC intra-prediction for depth map based on a self-organizing map and efficient features[J]. Signal Image and Video Processing, 2024,18(3):2289 – 2296.

[10] Li Y, Yang G B, Qu A P, et al. Tunable early CU size decision for depth map intra coding in 3D-HEVC using unsupervised learning [J]. Digital Signal Processing, 2022,123:103448.

[11] Bakkouri S, Elyousfi A. An adaptive CU size decision algorithm based on gradient boosting machines for 3D-HEVC inter-coding[J]. Multimedia Tools and Applications, 2023,82:32539 – 32557.

[12] Wang X L. Application of 3D-HEVC fast coding by Internet of Things data in intelligent decision[J]. Journal of Supercomputing, 2022,78(5):7489 – 7508.

[13] Saldanha M, Sanchez G, Marcon C, et al. Fast 3D-HEVC depth map encoding using machine learning[J]. IEEE Transactions on Circuits and Systems for Video

Technology, 2020,30(3):850 – 861.

[14] Zou D, Dai P, Zhang Q. Fast depth map coding based on Bayesian decision theorem for 3D-HEVC[J]. IEEE Access, 2022,10:51120 – 51127.

[15] 刘畅, 贾克斌, 刘鹏宇. 基于多分支网络的深度图帧内编码单元快速划分算法[J]. 电子与信息学报, 2022,44(12):4357 – 4366.

[16] Sebai D, Sehli M, Ghorbel F. End-to-end variable-rate learning-based depth compression guided by deep correlation features[J]. Journal of Signal Processing Systems, 2024,96:81 – 97.

[17] Liu S, Cui S, Li T, et al. Fast CU partition algorithm based on swin-transformer for depth intra coding in 3D-HEVC[J]. Multimedia Tools and Applications, 2024, 83(42):90315 – 90329.

[18] Liu C, Jia K B, Liu P Y. Fast depth intra coding based on depth edge classification network in 3D-HEVC[J]. IEEE Transactions on Broadcasting,2022,68(1):97 – 109.

[19] Zhang H, Yao W, Huang H, et al. Adaptive coding unit size convolutional neural network for fast 3D-HEVC depth map intracoding[J]. Journal of Electronic Imaging, 2021,30(4):041405.

[20] Omran N, Maraoui A, Werda I, et al. Fast partition algorithm in depth map intra coding unit based on multi-deep convolution neural network[J]. Journal of Real-Time Image Processing, 2024,21(1):23.

[21] Lei J,Sun Z,Gu Z,et al. Simplified search algorithm for explicit wedgelet signalization mode in 3D-HEVC[C]//2017 IEEE International Conference on Multimedia and Expo (ICME). Hong Kong, China: IEEE,2017:805 – 810.

[22] Ücker M, Afonso V, Audibert L, et al. Low-power and high-throughput architecture for 3D-HEVC depth modeling mode 4[C]//2018 31st Symposium on Integrated Circuits and Systems Design. Bento Goncalve, Brazil: SBCCI, 2018:1 – 6.

[23] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE, 2021:13708 – 13717.

[24] Jiao J, Tang Y M, Lin K Y, et al. DilateFormer: multi-scale dilated Transformer for visual recognition [J]. IEEE Transactions on Multimedia,2023,25:8906 – 8919.

A fast intra-frame coding algorithm for depth maps based on texture structure and CA-Transformer

Li Jingze, Jia Kebin

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124)

(Beijing Laboratory of Advanced Information Networks, Beijing 100124)

Abstract

To address the high complexity issue in depth map coding of three-dimensional high efficiency video coding(3D-HEVC), this paper proposes a fast depth map coding algorithm that integrates texture partitioning structure and deep learning. First, a texture map partitioning structure based method is introduced to accelerate depth map coding, which identifies and simplifies coding tree unit (CTU) in advance to achieve rapid partitioning results. Second, a neural network named CA-Transformer is designed. Leveraging the sensitivity of convolutional neural networks to local features and the global feature modeling capability of vision Transformers, the proposed CA-Transformer network processes structurally complex depth map images to predict CTU partitioning structures. By replacing the original CTU partitioning process in the standard encoder with the proposed algorithm, experimental results demonstrate that the algorithm reduces encoding time by an average of 56.00% while maintaining nearly identical rate-distortion performance and synthesized view quality.

Key words: three-dimensional high efficiency video coding, texture map, depth map, coding tree unit partitioning, intra encoding