

基于多类支持向量机的 3D-HEVC 深度视频帧内编码快速算法

刘晟, 彭宗举, 陈嘉丽, 陈芬, 郁梅, 蒋刚毅

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 深度视频编码中最优深度划分和模式选择过程具有非常高的计算复杂度。提出了基于多类支持向量机 (MSVM, multi-class support vector machine) 的深度视频帧内编码快速算法。该算法包括离线模型训练和快速编码 2 个部分。在离线模型训练中, 用深度视频最大编码单元 (LCU, largest coding unit) 的最优划分深度作为标签, 当前 LCU 的空域复杂度、空域相邻 LCU 的最优划分深度和彩色视频对应 LCU 的最优划分深度作为特征去构造 MSVM 模型。在编码时, 提取 LCU 的特征, 根据 MSVM 模型得到划分深度的预测值。根据该预测值提前终止编码单元递归划分和模式选择过程。实验结果表明, 提出的算法在几乎不影响虚拟视点质量的情况下, 平均节省 35.91% 的总体编码时间和 40.04% 的深度编码时间。

关键词: 深度视频编码; 3D-HEVC; 帧内编码; 最大编码单元; 多类支持向量机

中图分类号: TN919.81

文献标识码: A

Multi-class support vector machine-based fast algorithm for 3D-HEVC depth video intra coding

LIU Sheng, PENG Zong-ju, CHEN Jia-li, CHEN Fen, YU Mei, JIANG Gang-yi

(Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China)

Abstract: The recursive splitting process of largest coding unit (LCU) and the mode search process of coding unit imposed enormous computational complexity on encoder. A multi-class support vector machine-based (MSVM) fast coding unit (CU) size decision algorithm for 3D-HEVC depth video intra-coding was proposed. The algorithm included two steps: off-line training and fast CU size and mode decision. In the process of off-line training, a MSVM model was constructed, where the texture complexity of current LCU, the optimal partition depth of its spatial neighboring LCU and co-located LCU in texture video were treated as feature vectors, and the optimal partition depth of LCU was utilized as corresponding class label. In the process of fast CU size and mode decision, features of LCU were extracted before coding a LCU, then, a MSVM model was used to predict the class label. Finally, the class label that represents the largest partition depth of the current LCU was employed to terminate the CU recursive splitting process and CU mode search process. Experimental results show that the proposed algorithm saves the encoding time of 3D-HEVC by 35.91% on average, and the encoding time of depth video by 40.04% on average, with negligible rendered virtual view image degradation.

Key words: depth video coding, 3D-HEVC, intra coding, largest coding unit, multi-class support vector machine

收稿日期: 2015-10-16; 修回日期: 2016-10-08

基金项目: 国家高技术研究发展计划 (“863” 计划) 基金资助项目 (No.2015AA015901); 国家自然科学基金资助项目 (No.61620106012, No.U1301257, No.61271270); 浙江省自然科学基金资助项目 (No.LY16F010002, No.LY15F010005, No.LY17F010005); 宁波市自然科学基金资助项目 (No.2015A610127, No.2015A610124); 宁波大学科研基金(理)/学科基金资助项目(No.xkx11502)

Foundation Items: The National High-Tech R&D Program of China (863 Program) (No. 2015AA015901), The National Natural Science Foundation of China (No.61620106012, No.U1301257, No.61271270), The Natural Science Foundation of Zhejiang Province (No.LY16F010002, No.LY15F010005, No.LY17F010005), The Natural Science Foundation of Ningbo (No.2015A610127, No.2015A610124), Ningbo University Research Foundation (Science) / Discipline Project(No.xkx11502)

1 引言

三维 (3D, three-dimensional) 视频系统支持用户自由选择观看角度, 提供更加真实的立体视频感知体验, 已成为视频编码和通信领域的研究热点。与 2D 视频应用相比, 3D 视频系统需要传输更多的视频数据, 给 3D 网络带来了极大的负担。为了减少 3D 视频传输所需要的数据量, 目前广泛采用多视点彩色加深度(MVD, multiview video plus depth)的三维场景表示方式^[1,2]。MVD 信号包括多个视点的彩色视频及对应的深度视频。深度视频可以通过软件估计^[3]和深度相机采集得到, 其主要功能是辅助虚拟视点的绘制^[4], 形成虚拟视点提供给终端用户。

高效视频编码 (HEVC, high efficiency video coding) 是最新的视频编码标准^[5], 其 3D 扩展 (3D-HEVC, 3D extension of HEVC) 是最新的 3D 视频编码结构^[6]。3D-HEVC 的帧内编码不仅包含了 HEVC 的四叉树编码结构和 35 种帧内预测模式, 还增加了深度模型模式 (DMM, depth modeling mode)^[7]。DMM 的引入极大地增加了深度视频的编码复杂度。3D-HEVC 采用基于块的压缩方案, 一幅图像会被划分为许多个最大编码单元 (LCU, largest coding unit), 每一个 LCU 会采用递归的方式进行四叉树结构划分 (64×64 、 32×32 、 16×16 和 8×8), 利用率失真代价优化准则^[8]得到 LCU 最优的划分深度和编码单元 (CU, coding unit) 的最优预测模式。显然, LCU 划分和模式选择具有极高的编码复杂度。

目前, 许多学者提出了彩色视频快速编码算法^[9~11]和深度视频快速编码算法^[12~20]。Pan 等^[9]利用当前 CU 在 Inter-2N \times 2N 模式下的运动矢量信息以及在 Merge 模式下的残差信息优化了模式选择过程, 减少了彩色视频帧间编码复杂度。Chiang 等^[10]建立了 2 个反向传播神经网络结构, 第一个神经网络具有 3 层神经元, 用于预测 CU 的最优尺寸, 第二个神经网络具有 3 层神经元, 用于参考帧选取, 减少了彩色视频帧间模式搜索复杂度。Correa 等^[11]建立基于决策树方法的模型用以预测当前 CU 和预测单元的最优尺寸, 降低了帧间彩色视频编码复杂度。Shen 等^[12]利用空域相邻 CU、时域相关 CU 和相邻视点响应块 CU 的最优分割尺寸大小及预测模式信息优化预测模式候选数目以及遍历的分割深度层数, 减少了帧间预测复杂度。

Zhang 等^[13]利用彩色视频和深度视频的编码相关性提出了模式遍历提前终止的算法, 还根据 CU 类型设置不同的运动搜索范围减少了帧间运动估计复杂度。Park^[14]对 CU 进行哈达玛变换得到哈达玛变换系数, 判定 CU 是否存在边缘, 如果 CU 不存在边缘或者只存在水平边缘和垂直边缘, 那么跳过 DMM 搜索, 减少了帧内模式复杂度。Gu 等^[15]通过统计实验发现, 如果 CU 在遍历完 35 种帧内预测模式后的最小率失真代价值小于一定的阈值, 那么跳过 DMM 模式, 降低了深度视频帧内编码复杂度。Sanchez 等^[16]提取 CU 的边缘, 如果边缘强度小于阈值则跳过 DMM 搜索, 降低了帧内模式选择复杂度。Gu 等^[17]通过统计实验提出了深度视频帧内快速编码算法, 如果 CU 的最有可能模式 (MPM, most probable modes) 包含平面模式, 则跳过 DMM 搜索。Lei 等^[18]利用彩色视频与深度视频的编码模式相关性, 提出了快速深度视频编码算法。如果彩色视频 CU 最优预测模式为 Skip 模式, 那么对应块深度视频 CU 也直接采用 Skip 模式的算法。Shen 等^[19]根据空域相邻 CU 以及对应块彩色视频 CU 的空域相邻 CU 的运动矢量信息和最优预测模式信息调整当前 CU 的运动估计范围和预测模式搜索范围, 减少了深度视频帧间编码复杂度。Lei 等^[20]针对深度视频锚节点帧和非锚节点帧, 分别利用相关性采用快速模式选择来提高深度视频帧间编码速度。由于彩色视频与深度视频不同的统计特性, 彩色视频快速编码算法^[9~11]难以直接用到深度视频中。上述深度视频快速编码算法通过利用编码统计特性^[12,13]、减少 DMM 搜索^[14~17]来提高深度编码速度, 性能可以进一步提高。文献[18~20]提出的降低深度视频编码复杂度算法是以多视点视频编码器 (JMVC, joint multi-view video coding) 为平台, 然而 JMVC 与 3D-HEVC 相比, 分割尺寸大小以及模式数目都有所改变, 这些方法不适合直接运用于 3D-HEVC。

深度视频中存在大部分缓慢变化的平坦区域和少量尖锐边缘的复杂区域, 而平坦区域 LCU 的最优划分深度一般为 0 或 1。本文结合深度视频编码的特性, 提出一种基于多类支持向量机 (MSVM, multi-class support vector machine) 的 3D-HEVC 深度视频帧内编码快速算法。本文算法包括离线模型训练和快速编码 2 个阶段。在离线模型训练阶段, 用深度视频 LCU 的最优划分深度作为标签, 用当前

LCU 的空域复杂度、空域相邻 LCU 的最优划分深度和彩色视频对应 LCU 的最优划分深度作为特征，训练得到 MSVM 模型。在深度视频编码阶段，提取 LCU 的特征，导入到模型中预测出当前 LCU 的最优编码划分深度。根据预测的划分深度提前终止 LCU 最优深度划分和模式选择过程。实验结果表明，本文算法在保证虚拟视点绘制质量的基础上，有效地降低了编码端复杂度。

2 复杂度分析

LCU 递归划分，如图 1 所示。CU₀ 代表 LCU，在编码过程中会被分割成 4 个大小一样的 CU₁，CU₁ 会被继续分割为 4 个大小一样的 CU₂，一直划分到最小编码单元 CU₃。要确定一个 LCU 的最优划分深度需要经过 $341(4^0+4^1+4^2+4^3+4^4=341)$ 次递归过程，这种 LCU 的递归划分引入了极大的编码复杂度。

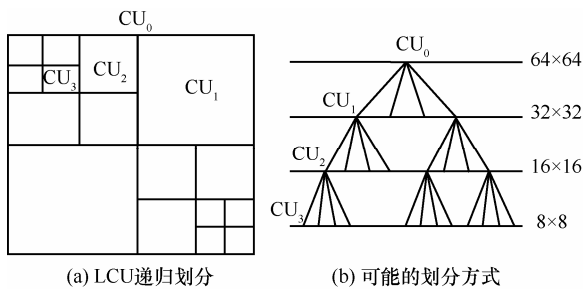


图 1 LCU 递归划分示意和可能的划分方式

图 2 所示为 Kendo 第 3 视点第 5 帧深度视频 LCU 分割深度结果，编号为 1 的 LCU 的最优划分深度为 0，编号为 2 的 LCU 最优分割深度为 3。在编码过程中，所有 LCU 都会递归分割到最大划分深度。提前终止图 2 中最优划分深度小于 3 的 LCU 的最优分割深度递归过程，可以有效地降低编码复杂度。

表 1 统计了各个序列在帧内编码配置下 LCU 最优划分深度的分布情况。LCU 最优划分深度为 0 的平均比例达到了 56.10%，最优划分深度等于 3 的

比例只有 24.99%，最优划分深度小于 3 的比例高达 75.01%。因此，在 CU 分割递归过程中，当前分割深度为最优分割深度时，提前终止递归过程，可以减少编码复杂度，而不会对率失真性能产生任何影响。

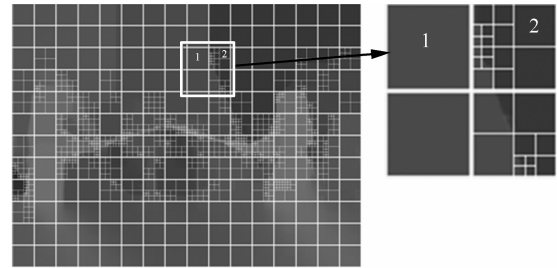


图 2 Kendo 第 3 视点第 5 帧编码结果

3 基于多类支持向量机的深度视频帧内编码快速算法

LCU 的最优划分深度包括 0、1、2 或 3 共 4 种类型，本文运用 MSVM 解决这样的分类问题，并进一步提出了深度视频帧内编码快速算法。首先把深度视频 LCU 的最优划分深度作为标签，然后把当前 LCU 的空域复杂度、空域相邻 LCU 的最优划分深度和彩色视频对应 LCU 的最优划分深度作为特征，最后通过离线训练得到 MSVM 模型。编码时，把 LCU 的特征输入到模型中得到 LCU 最优划分深度的预测值，再根据该预测值决定是否终止深度划分和模式选择过程。

3.1 MSVM 模型的建立

3.1.1 特征选取

特征选取可以减少训练时间、预测时间和提高模型预测的准确度。在 3D-HEVC 的深度视频编码中，当前 LCU 的最优划分深度与当前 LCU 的空域复杂度、彩色视频对应 LCU 的最优划分深度、空域相邻 LCU 的最优划分深度有关。此外，本文的目标是维持编码率失真性能的同时减少编码复杂度，因此，特征的提取不能有很高的计算复杂度。本文采用以下 6 个特征来预测当前 LCU 的最优划分深度值。

表 1 深度视频 LCU 深度值分布比例比较

分割深度	Balloons	Kendo	Newspaper	GhostTownFly	Shark	平均
0	48.47%	11.52%	7.35%	32.66%	56.01%	56.10%
1	57.03%	12.74%	7.82%	22.41%	9.16%	11.43%
2	38.31%	15.89%	12.68%	33.12%	5.86%	7.48%
3	80.67%	7.82%	3.69%	7.82%	28.96%	24.99%

- 1) 对应彩色视频编码块 LCU 的最优划分深度, 记作 x_{ld} ;
- 2) 左上边 LCU 最优划分深度, 记作 x_{lad} ;
- 3) 上边 LCU 最优划分深度, 记作 x_{ad} ;
- 4) 左边 LCU 最优划分深度, 记作 x_{ld} ;
- 5) 当前 LCU 每个像素与对应行像素平均值的差值平方和, 记作 x_{Xmse} , 计算方法为

$$x_{Xmse} = \sum_{i=0}^{63} \sum_{j=0}^{63} (p(i, j) - \bar{X}_i)^2 \quad (1)$$

其中, $p(i, j)$ 代表当前 LCU 的中 (i, j) 处的像素值, X_i 代表 LCU 第 i 行的像素平均值;

- 6) 当前 LCU 每个像素与对应列像素平均值的差值平方和, 记作 x_{Ymse} , 计算方法为

$$x_{Ymse} = \sum_{j=0}^{63} \sum_{i=0}^{63} (p(i, j) - \bar{Y}_j)^2 \quad (2)$$

其中, Y_j 代表 LCU 第 j 列像素的平均值。

3.1.2 离线训练

传统的支持向量机分类器一般只是分出 2 类, 而本文的 LCU 分割深度取 4 个不同的值 0、1、2 或 3, 所以需要运用 MSVM 分类器。MSVM 分类器在传统的支持向量机分类器的基础上建立 4 个超平面, 每一个超平面把对应类 LCU 与其余类 LCU 分割开。超平面函数表示为

$$f_m(\mathbf{x}) = (\mathbf{w}_m)^T \phi(\mathbf{x}) + b_m, \quad m \in \{0, 1, 2, 3\} \quad (3)$$

其中, \mathbf{w}_m 是第 m 类的预测权重矢量, $(\mathbf{w}_m)^T$ 表示 \mathbf{w}_m 的转置, $\phi(\bullet)$ 代表非线性核函数^[21], \mathbf{x} 是由 x_{ld} 、 x_{lad} 、 x_{ad} 、 x_{ld} 、 x_{Xmse} 和 x_{Ymse} 组成的矢量, b_m 表示第 m 类的偏置量。一般情况下, 在低维空间中, 几乎找不到一个超平面能把不同类别 y_i 的 LCU 线性分割开。根据 Cover 定理^[22], 采用核函数把不同类别 y_i 所对应的特征矢量 \mathbf{x}_i 映射到高维空间后, 总能找到一个最优的超平面把不同类别 y_i 的 LCU 线性分割开。然后通过最小化代价函数 $J_m(\mathbf{w}_m)$ 构造超平面函数 $f_m(\mathbf{x})$ 。

$$J_m(\mathbf{w}_m) = \frac{1}{2} (\mathbf{w}_m)^T \mathbf{w}_m = \frac{1}{2} \|\mathbf{w}_m\|^2 \quad \text{s.t.} (\mathbf{w}_m)^T \phi(\mathbf{x}_{i,m}) + b_m \geq 1 \quad (4)$$

其中, 约束条件表示所有的训练样本 LCU 都正确地分类, $\mathbf{x}_{i,m}$ 为的第 i 个训练样本 (类型为 m) 的特征矢量。但实际上有些不同 y_i 的 LCU 的 \mathbf{x}_i 却十分相似, 为了避免这些特殊的 LCU 对超平面函数产

生严重的影响, 允许这些特殊的 LCU 被错误分类, 通常需要引入松弛变量 $\xi_{i,m}$ 、惩罚因子 C 和用户设定的权重 W_m , 代价函数转化为

$$\begin{aligned} J_m(\mathbf{w}_m) &= \frac{1}{2} (\mathbf{w}_m)^T \mathbf{w}_m + C W_m \sum_{i=1}^k \xi_{i,m} \\ \text{s.t.} (\mathbf{w}_m)^T \phi(\mathbf{x}_{i,m}) + b_m &\geq 1 - \xi_{i,m}, y_i = m \\ (\mathbf{w}_m)^T \phi(\mathbf{x}_{i,m}) + b_m &\leq -1 + \xi_{i,m}, y_i \neq m \\ \xi_{i,m} &\geq 0 \end{aligned} \quad (5)$$

其中, k 表示训练样本 LCU 的个数。利用拉格朗日乘子法对式(5)构造优化函数, 对优化函数求取偏导数, 可得到超平面函数

$$f_m(\mathbf{x}) = \sum_{i=1}^k \alpha_{i,m} K(\mathbf{x}_{i,m}, \mathbf{x}) + b_m, m \in \{0, 1, 2, 3\} \quad (6)$$

其中, $K(\mathbf{x}_{i,m}, \mathbf{x}) = \phi^T(\mathbf{x}_{i,m})\phi(\mathbf{x})$, $\alpha_{i,m}$ 为第 m 类的拉格朗日乘子。编码时, 预测样本 LCU 的特征矢量代入到式(6)得到 4 个决策函数值, 最大的决策函数值为编码 LCU 的深度的预测值 $dpre$, 可以表示为

$$dpre = \arg \max_m (f_m(\mathbf{x})) \quad (7)$$

离线训练参数设置如表 2 所示。参数 W_m 的选择时考虑了样本标签分布的不均匀性。同一类别标签所占比例越小, 该类标签预测误差则尽可能越小, 也就是代价函数值应该大一些, 这样能提高模型的顽健性。本文把训练样本标签最多的一类所对应的权重设为 1, 其余类别标签所对应的权重为样本标签最多一类的样本数除以对应类训练样本数。 C 越大表示对分类总体错误越敏感, 在量化参数(QP, quantization parameter)小的时候应尽可能使总体误差小些, 减少编码率失真性能的损失。因此, QP 不同, 对应 C 也不同。通常核函数有线性函数、多项式函数、径向基函数(RBF, radial basis function)和 Sigmoid 函数。因为 RBF 核函数可以把不同类别 y_i 所对应的 \mathbf{x}_i 变换到高维空间, 使其变得线性可分^[23], 所以本文选择 RBF 核函数。

表 2 训练参数

QP	C	W_m	核函数
34	1	1,5,6,2	RBF
39	0.8	1,5,8,2	RBF
42	0.8	1,7,11,5	RBF
45	0.2	1,9,17,10	RBF

离线训练过程如图 3 所示。本文采用 Balloons、Kendo、Newspaper、GhostTownFly 和 Shark 序列的前 25 帧进行模型离线训练。前 8 帧进行离线训练，后 17 帧进行模型测试。模型测试时，把样本的特征输入到模型中，可得到预测标签。如果预测的标签值大于真实的预测样本的最优深度值，则定义此样本为预测正确的样本，预测准确度为预测正确的样本数除以总样本数。表 3 显示了各个序列在 4 个不同的 QP 下的平均总体预测准确度。“ W_m 调节前”表示所有类采用相同的权值 1；“ W_m 调节后”为本文算法最终采用的权值。可以看出，由于训练样本的分布不均匀性，参数 W_m 在调节前训练的模型预测准确性不高，训练模型会倾向于预测出样本最多的一类；参数调节后，训练模型不会倾向于某一类预测样本，预测模型更加顽健，模型的准确性更高。



图 3 离线训练过程

表 3 参数 W_m 调节前后预测准确度对比

QP	W_m 调节前	W_m 调节后
34	78.24%	95.67%
39	76.92%	92.41%
42	75.63%	92.55%
45	76.17%	92.79%

3.2 深度视频快速编码算法

本文提出的提前终止 LCU 划分和模式选择算法流程如图 4 所示，具体描述如下。

步骤 1 如果左边和上边 LCU 不同时存在，设置预测分割深度 $dpre$ 为 3，跳到步骤 4。

步骤 2 提取当前 LCU 的特征，根据 MSVM 模型得到 $dpre$ ，进入步骤 3。

步骤 3 如果 $dpre$ 等于 0，则完成操作：遍历 35 种帧内预测模式、把 MPM 加入到率失真代价列表 L 中和跳过 DMM 模式搜索。然后，转到步骤 5。

步骤 4 当前分割深度 $dcur$ 从 0 到 $dpre$ 依次完成操作：遍历 35 种帧内预测模式、把 MPM 加入到 L 中和遍历 DMM 模式。

步骤 5 选取 L 里的最优预测模式和 LCU 最优划分深度。

步骤 6 当前 LCU 递归划分结束。

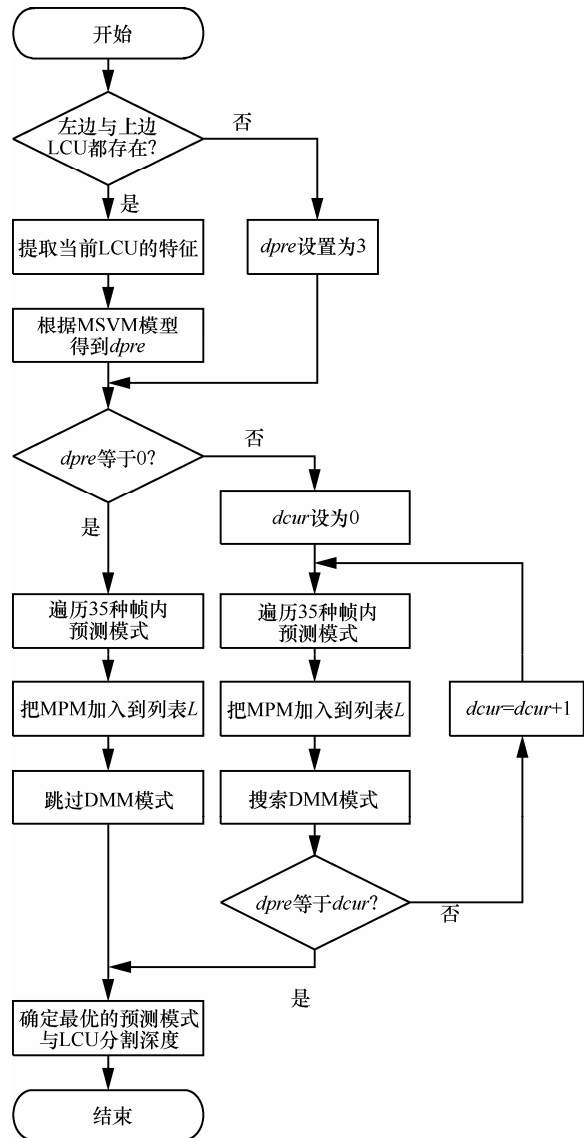


图 4 基于 MSVM 的深度视频帧内快速算法流程

4 实验结果与分析

本文算法在 3D-HEVC 参考软件 HTM-10.0 上实现。实验参数配置采用了标准测试环境^[24]，使用 All-Intra 配置文件，实验测试序列和左右视点选择以及绘制虚拟视点选择如表 4 所示。

表 4 测试环境

序列	分辨率	左视点	右视点	虚拟视点
Balloons	1 024×768	1	3	2
Kendo	1 024×768	3	5	4
Newspaper	1 024×768	4	6	5
GhostTownFly	1 920×1 088	1	9	5
Shark	1 920×1 088	1	9	5

表 5 列出了原始测试平台分别与 Park 的算法^[14]、Gu 的算法^[15]和本文算法在不同 QP (彩色视频 QP 值分别为 25、30、35 和 40, 对应的深度视频的 QP 为 34、39、42 和 45)^[24]下的编码时间对比。 DT_{Gu} 、 DT_{Park} 和 DT_{Pro} 分别表示 Gu 的算法、Park 的算法和本文算法编码总体时间(彩色视频编码时间和深度视频编码时间)与原始测试平台编码总体时间的节省时间百分比。 Dt_{Gu} 、 Dt_{Park} 和 Dt_{Pro} 分别表示 Gu 的算法、Park 的算法和本文算法深度视频编码时间与原始测试平台编码深度视频编码时间节省百分比。

从表 5 中可以看出, 本文算法与原始测试平台相比, 总体编码时间节省了 35.91%, 深度视频节省了 40.04%。本文算法编码时间节省百分比随着 QP 的增大而增多, 主要是因为 QP 越大, 满足本文条件的 LCU 越多。Gu 的算法与原始测试平台相比, 总体时间节省了 9.55%, 深度视频节省了 10.54%。Park 的算法与原始测试平台相比, 总体时间节省了 18.47%, 深度视频时间节省了 20.70%。由于不同 QP 下满足特定条件的 LCU 百分比几乎是一样的,

Gu 与 Park 的算法编码时间节省在不同的 QP 下几乎保持不变。在编码节省时间方面, 本文算法优于 Gu 与 Park 的算法。Gu 的算法主要是针对深度视频复杂区域的 CU 执行 DMM 搜索, 平坦区域的 CU 跳过 DMM 搜索, LCU 还是递归地划分到最大可允许划分深度。Park 的算法是计算 CU 的是否含有水平和垂直边缘, 加速 CU 的模式决策过程, LCU 仍然需要递归的分割到最大可允许划分深度。本文算法提前终止 CU 递归划分, 减少了 CU 划分深度遍历过程和相应深度级 CU 的模式遍历过程。

表 6 为各算法对应虚拟视点绘制质量对比。虚拟视点图像是利用 JCT-3V 提供的标准绘制算法绘制得到。负号表示虚拟视点质量下降。 $DPSNR_{Gu}$ 、 $DPSNR_{Park}$ 和 $DPSNR_{Pro}$ 分别表示 Gu 的算法、Park 的算法和本文算法与原始平台算法编码的虚拟视点 $PSNR$ 差异。Gu 的算法对应虚拟视点的 $PSNR$ 值与标准编码算法对应虚拟视点的 $PSNR$ 值相比几乎没有发生变化。Park 的算法对应虚拟视点的 $PSNR$ 值最大上升 0.01 dB, 而本文算法对应虚拟视点的

表 5 深度视频编码时间及总体编码时间减少比例比较

序列	QP	$DT_{Gu}/\%$	$Dt_{Gu}/\%$	$DT_{Park}/\%$	$Dt_{Park}/\%$	$DT_{Pro}/\%$	$Dt_{Pro}/\%$
Balloons	34	-9.66	-10.99	-20.52	-23.29	-22.57	-25.83
	39	-9.79	-10.87	-18.99	-21.77	-23.06	-26.00
	42	-9.84	-10.92	-21.88	-24.35	-31.17	-34.70
	45	-9.93	-10.87	-22.40	-24.65	-43.39	-47.73
Kendo	34	-8.98	-10.25	-24.01	-27.40	-21.14	-24.25
	39	-9.22	-10.19	-24.28	-27.36	-22.68	-25.56
	42	-10.42	-11.52	-25.56	-28.53	-35.50	-39.51
	45	-11.51	-12.87	-26.40	-29.07	-50.69	-55.72
Newspaper	34	-4.94	-5.60	-12.92	-13.92	-11.17	-12.84
	39	-6.52	-7.14	-13.34	-14.91	-11.77	-13.21
	42	-8.16	-8.90	-14.62	-16.12	-25.86	-28.47
	45	-8.29	-9.16	-14.77	-16.23	-41.98	-45.77
GhostTownFly	34	-4.34	-4.86	-12.75	-14.83	-46.58	-53.67
	39	-6.56	-7.21	-13.44	-15.20	-49.77	-55.81
	42	-16.73	-18.28	-23.22	-25.76	-62.97	-69.53
	45	-19.24	-20.89	-24.08	-26.39	-69.90	-76.45
Shark	34	-4.14	-4.60	-13.25	-15.29	-23.43	-27.07
	39	-4.55	-5.00	-13.59	-15.44	-28.34	-32.06
	42	-12.62	-13.74	-14.30	-15.94	-41.27	-45.98
	45	-15.58	-16.93	-15.17	-17.50	-55.03	-60.65
平均		-9.55	-10.54	-18.47	-20.70	-35.91	-40.04

PSNR 值最大降低 0.04 dB, 平均下降 0.01 dB, 几乎可以忽略不计。

表 6 绘制虚拟视点质量对比

序列	QP	DPSNR _{Gu} /dB	DPSNR _{Park} /dB	DPSNR _{Pro} /dB
Balloons	34	0.00	0.00	0.00
	39	0.00	0.00	0.00
	42	0.00	0.00	0.02
	45	0.00	0.01	-0.02
Kendo	34	0.00	0.00	0.00
	39	0.00	0.00	0.00
	42	0.00	0.00	-0.02
	45	0.00	0.00	-0.03
Newspaper	34	0.00	0.00	0.00
	39	0.00	0.00	0.00
	42	0.00	0.00	0.00
	45	0.00	0.00	-0.01
GhostTownFly	34	0.00	0.00	-0.02
	39	0.00	0.00	0.00
	42	0.00	0.00	-0.01
	45	0.00	0.00	-0.04
Shark	34	0.00	0.00	-0.01
	39	0.00	0.01	0.00
	42	0.00	0.00	-0.02
	45	0.00	0.00	-0.04
平均		0.00	0.00	-0.01

本文采用 BD-rate 进一步评价本文深度视频帧内编码快速算法的性能^[25], 结果如表 7 所示。本文算法相对于原始测试平台标准算法 BD-rate 变化范围是-0.21%~0.45%, 平均上升 0.11%。Gu 的算法相对于原始测试平台标准算法 BD-rate 变化在-0.08%和 0.21%之间, 平均上升 0.05%。Park 的算法相对于原始测试平台标准算法 BD-rate 变化范围为-0.19%~0.17%, 平均上升 0.01%。

表 7 编码 BD-rate 变化情况比较

序列	BDBR _{Gu} /%	BDBR _{Park} /%	BDBR _{Pro} /%
Balloons	0.05	0.17	-0.21
Kendo	0.00	-0.19	-0.14
Newspaper	0.09	0.13	0.08
GhostTownFly	0.21	-0.06	0.45
Shark	-0.08	0.01	0.36
平均	0.05	0.01	0.11

综上所述, 本文算法在编码时间节省方面优于 Gu 与 Park 的算法, 虚拟视点 PSNR 以及编码性能 BD-rate 与 Gu 和 Park 的算法相当。

5 结束语

本文分析了深度视频帧内编码 LCU 划分深度分布情况, 提出了一种基于机器学习的 3D-HEVC 深度帧内编码快速算法。根据深度视频自身的特点, 在帧内编码过程中, 有很多 LCU 的划分深度小于最大可允许划分深度。根据这种特性, 提出基于 MSVM 的提前终止 CU 分割和模式选择的算法, 从而简化了深度视频帧内编码的复杂度。实验结果表明, 本文提出的算法在绘制中间视点的质量不变的情况下, 平均节省了 40.04%的深度视频编码时间, 总体的编码时间平均节省 35.91%。

参考文献:

- [1] MÜLLER K, MERKLE P, WIEGAND T. 3-D video representation using depth maps[J]. Proceedings of the IEEE, 2011, 99(4): 643-656.
- [2] FEHN C. Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV[C]//Electronic Imaging 2004. International Society for Optics and Photonics, 2004: 93-104.
- [3] TANIMOTO M, FUJII T, SUZUKI K. View synthesis algorithm in view synthesis reference software 2.0 (VRS2.0) [S]. ISO/IEC JTC1/SC29/WG11 M, 2009, 16090.
- [4] OHM J R. Overview of 3D video coding standardization[C]//Proc of Three Dimensional Systems and Applications. 2013:1-4.
- [5] SULLIVAN G J, OHM J R, HAN W J, 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.
- [6] MULLER K R, SCHWARZ H, MARPE D, et al. 3D high-efficiency video coding for multi-view video and depth data[J]. IEEE Transactions on Image Processing, 2013, 22(9): 3366-3378.
- [7] MERKLE P, BARTNIK C, MÜLLER K, et al. 3D video: depth coding based on inter-component prediction of block partitions[C]//Picture Coding Symposium (PCS), 2012. IEEE, 2012: 149-152.
- [8] 沈燕飞, 李锦涛, 朱珍民, 等. 高效视频编码[J]. 计算机学报, 2013, 36(11): 2341-2355.
SHEN Y F, LI J T, ZHU Z M, et al. High efficiency video coding[J]. Chinese Journal of Computers, 2013, 36(11): 2341-2355.
- [9] PAN Z, KWONG S, SUN M T, et al. Early MERGE mode decision based on motion estimation and hierarchical depth correlation for HEVC[J]. IEEE Transactions on Broadcasting, 2014, 60(2): 405-412.
- [10] CHIANG J C, CHEN W C, LIU L M, et al. A fast H. 264/AVC-based stereo video encoding algorithm based on hierarchical two-stage neural classification[J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(2): 309-320.
- [11] CORREA G, ASSUNCAO P, AGOSTINI L V, et al. Fast HEVC encoding decisions using data mining[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(4): 660-673.
- [12] SHEN L, AN P, ZHANG Z, et al. A 3D-HEVC fast mode decision algorithm for real-time applications[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2015, 11(3): 34.

- [13] ZHANG Q, CHEN M, HUANG X, et al. Low-complexity depth map compression in HEVC-based 3D video coding[J]. EURASIP Journal on Image and Video Processing, 2015, 2015(1): 1-14.
- [14] PARK C S. Edge-based intra mode selection for depth-map coding in 3D-HEVC[J]. IEEE Transactions on Image Processing, 2015, 24(1): 155-162.
- [15] GU Z, ZHENG J, LING N, et al. Fast bi-partition mode selection for 3D HEVC depth intra coding[C]// 2014 IEEE International Conference on Multimedia and Expo (ICME). 2014: 1-6.
- [16] SANCHEZ G, SALDANHA M, BALOTA G, et al. DMM fast: a complexity reduction scheme for three-dimensional high-efficiency video coding intra frame depth map coding[J]. Journal of Electronic Imaging, 2015, 24(2): 023011.
- [17] GU Z, ZHENG J, LING N, et al. Fast depth modeling mode selection for 3D HEVC depth intra coding[C]// 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). 2013: 1-4.
- [18] LEI J, LI S, ZHU C, et al. Depth coding based on depth-texture motion and structure similarities[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(2): 275-286.
- [19] SHEN L, AN P, LIU Z, et al. Low complexity depth coding assisted by coding information from color video[J]. IEEE Transactions on Broadcasting, 2014, 60(1): 128-133.
- [20] LEI J, SUN J, PAN Z, et al. Fast mode decision using inter-view and inter-component correlations for multiview depth video coding[J]. IEEE Transactions on Industrial Informatics, 2015, 11(4): 978-986.
- [21] SCHÖLKOPF B, BURGESS C J C. Advances in kernel methods: support vector learning[M]. MIT Press, 1999.
- [22] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.
ZHANG X G. Introduction to statistical learning theory and support vector machines[J]. Acta Automatica Sinica, 2000, 26(1): 32-42.
- [23] SHEN X, YU L. CU splitting early termination based on weighted SVM[J]. EURASIP Journal on Image and Video Processing, 2013, 2013(1): 1-11.
- [24] MÜLLER K, VETRO A. Common test conditions of 3DV core experiments[C]//JCT3V meeting, JCT3VG1100. 2014.
- [25] BJONTEGARD B G. Calculation of average PSNR differences between RD-curves[R]. Austin: Video Coding Experts Group, 2001.

作者简介:



刘晟 (1991-), 男, 安徽来安人, 宁波大学硕士生, 主要研究方向为视频信号处理与编码。



彭宗举 (1973-), 男, 四川南部人, 博士, 宁波大学教授、博士生导师, 主要研究方向为多媒体信号处理与编码等。



陈嘉丽 (1993-), 女, 湖南邵阳人, 宁波大学硕士生, 主要研究方向为虚拟视点绘制与质量评价。



陈芬 (1973-), 女, 四川邻水人, 博士, 宁波大学副教授, 主要研究方向为视频图像处理。



郁梅 (1968-), 女, 江苏无锡人, 博士, 宁波大学教授、博士生导师, 主要研究方向为图像视频信息过程编码压缩与通信等。



蒋刚毅 (1964-), 男, 浙江绍兴人, 博士, 宁波大学教授、博士生导师, 主要研究方向为立体/多视点视频信号处理、3DTV/自由视点视频系统编码与传输等。