

# 基于学习模型的3D-HEVC提前Merge模式终止算法

李跃<sup>1,2</sup>, 杨高波<sup>2</sup>, 丁湘陵<sup>3</sup>, 朱亚培<sup>4</sup>

(1. 南华大学计算机学院, 湖南 衡阳 421001; 2. 湖南大学信息科学与工程学院, 湖南 长沙 410082;  
3. 湖南科技大学计算机科学与工程学院, 湖南 湘潭 411201; 4. 衡阳师范学院物理与电子学院, 湖南 衡阳 421002)

**摘要:** 作为高效视频编码(HEVC)的扩展, 3D-HEVC标准有效地提高了3D视频的压缩效率, 但是也带来了很高的编码计算复杂度。为了显著地降低3D-HEVC编码复杂度, 提出了一种提前Merge模式终止算法。首先, 提取Merge模式编码后的残差信号作为特征信息; 然后, 根据当前编码帧内已经编码的编码单元(CU)的最优Merge模式残差信号建立学习模型; 最后, 提取当前CU的Merge模式的残差信号, 并且利用学习模型预测Merge模式是否为最优模式。实验结果表明, 提出的提前Merge模式终止算法分别将3D-HEVC纹理视图和深度图编码的时间降低了41.9%和24.3%, 且编码性能的降低几乎可忽略不计。相较于现有的提前Merge模式算法, 提出的提前Merge模式终止算法能进一步降低3D-HEVC的编码时间, 并且设计简单, 易于集成到3D-HEVC测试模型。

**关键词:** 3D高效视频编码; 快速Merge模式决定; 残差信号; 学习模型

中图分类号: TN919.8

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2019165

## Early Merge mode decision algorithm for 3D-HEVC based on learning model

LI Yue<sup>1,2</sup>, YANG Gaobo<sup>2</sup>, DING Xiangling<sup>3</sup>, ZHU Yapei<sup>4</sup>

1. Computer School, University of South China, Hengyang 421001, China

2. School of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China

3. School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, China

4. Faculty of Physics and Electronic Information Science, Hengyang Normal University, Hengyang 421002, China

**Abstract:** 3D-high efficiency video coding (3D-HEVC) standard is an extension of HEVC. Though 3D-HEVC effectively improves the compression efficiency of 3D video, it also brings huge computational complexity. To greatly reduce the 3D-HEVC coding complexity, an early Merge mode decision approach was proposed. The residual signal that encoded by the Merge mode was firstly extracted as feature information. A learning model was established in terms of the residual signals of the coding unit (CU) in current frame that used early Merge mode as the optimal mode. Finally, the residual signal was extracted for the Merge mode of current CU, and the learning model was used to predict whether the Merge mode was the optimal mode or not. Experimental results show that the proposed early Merge mode decision approach reduces the coding times of 3D-HEVC texture views and depth maps about 41.9% and 24.3%, respectively, and the coding performance degradation is almost negligible. Compared with existing early Merge mode decision approaches, the proposed approach further reduces the coding time, and can be easily integrated into the 3D-HEVC test model due to its design simplicity.

**Key words:** 3D-high efficiency video coding, early Merge mode decision, residual signal, learning model

收稿日期: 2019-04-11; 修回日期: 2019-06-27

通信作者: 杨高波, yanggaobo@hnu.edu.cn

基金项目: 国家重点研发计划基金资助项目(No.2018YFB1003205); 国家自然科学基金资助项目(No.61572183)

**Foundation Items:** The National Key R & D Program of China (No.2018YFB1003205), The National Natural Science Foundation of China (No.61572183)

## 1 引言

随着 3D 内容获取和显示技术的快速发展,3D 视频因其能提供更真实的视觉体验越来越受观众欢迎。3D 视频数据由多个纹理视点及其相应的深度图(纹理视点的几何信息)组成。然而,与单个纹理视频相比,3D 视频数据更多、更复杂。因此,高效压缩对于 3D 视频数据实时传输的应用非常重要<sup>[1]</sup>。

HEVC (high efficiency video coding) 由 JCT-VC (joint collaborative team on video coding) 开发并用于单个纹理视点压缩<sup>[2]</sup>, 为了进一步压缩 3D 视频数据, JCT-3V (joint collaborative team on 3D video coding) 在 HEVC 的基础上进一步推出了针对 3D 视频压缩的 3D-HEVC 压缩标准<sup>[3]</sup>。3D-HEVC 编码的预测结构如图 1 所示。图 1 中, View 轴表示编码的视点, Temporal 轴表示每个视点在时序上的编码帧, 基本视点 ( $V_0$ ) 为独立视点, 其余视点 ( $V_1$ 、 $V_2$ ) 为非独立视点,  $T_i$  ( $i=0,1,2$ ) 为纹理视点,  $D_i$  ( $i=0,1,2$ ) 为在  $T_i$  之后编码的深度图, P、B、I 表示采用了不同技术的编码帧。为了进一步提高压缩效率, 3D-HEVC 采用了一些新的编码技术。例如, 为减少视点间相关性, 非独立视点编码采用了视差补偿预测 (DCP, disparity compensated prediction) 技术, 特别对于深度图, 它由锐利边缘分开的光滑区域组成。为了保护锐利边缘质量, 进一步提高深度图编码的效率, 采用了几种额外的编码工具, 如 DMM (depth modeling mode)<sup>[4]</sup>、DIS (depth intra skip)<sup>[5]</sup>、SDC (segment-wise DC coding)<sup>[6]</sup>和 VSO (view synthesis optimization) 等。虽然这些先进的编码技术可以提高 3D-HEVC 纹理视点和深度图的压缩效率, 但同时也增加了 3D-HEVC 的编码复杂度。因此, 有必要研究快速编码方法以降低 3D-HEVC 编码器复杂度, 同时保持编码性能基本不下降。

为了使 3D-HEVC 能应用于实时需求场景或者计算资源有限的场景, 学者们提出了许多快速编码方法以降低 3D-HEVC 的编码复杂度, 主要分为 2 种类型: 帧内快速编码方法和帧间快速编码方法。

### 1) 帧内快速编码方法

已提出的帧内快速 3D-HEVC 编码方法主要分为帧内快速 CU 大小决策方法和帧内预测模式快速决策方法, 主要采用基于相关性分析和基于机器学习的方法, 进行快速 CU 大小和预测模式决策<sup>[7-8]</sup>。

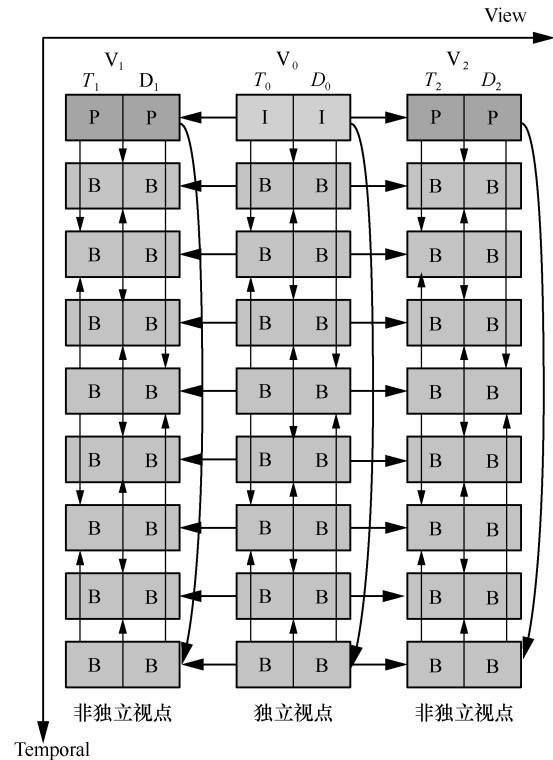


图 1 3D-HEVC 预测编码结构

### 2) 帧间快速编码方法

3D-HEVC 帧间快速方法同样可分为帧间快速 CU 大小决策方法和帧间预测模式快速决策方法, 采用的方法主要是基于相关性分析和编码信息进行快速决策, 相关性分析主要包括模式相关性、时空相关性、视点间相关性等, 编码信息主要包括率失真代价 (RD cost, rate distortion cost)、运动矢量 (MV, motion vector)、编码块标记 (CBF, coded block flag) 等<sup>[9-17]</sup>。

例如, Shen 等<sup>[10]</sup>提出利用时空相关性、视点间相关性、纹理深度相关性和层间相关性来加速 3D-HEVC 中的模式决策。类似地, Zhang 等<sup>[11]</sup>根据空间-时间和视点间相关性来建立编码复杂度模型, 然后根据复杂度模型提出 2 个低复杂度方法来降低 3D-HEVC 编码的复杂度。Tohidypour 等<sup>[12]</sup>利用运动信息和 RD cost, 提出了一种用于 3D-HEVC 非独立纹理视点的复杂度降低方法。Zhang 等<sup>[13]</sup>提出了一种基于运动同一性的快速方法, 以降低 3D-HEVC 纹理视频编码复杂度, 该方法包括提前 CU 大小决策、提前 SKIP/Merge 模式决策和自适应运动搜索范围调整。Li 等<sup>[14]</sup>通过利用 RD cost、CBF 等编码信息相关性建立混合终止模型用于降低 3D-HEVC 非独立纹理视频编码复杂度。Chen 等<sup>[15]</sup>利用视点间的纹理相关性, 提出了一种快速

3D-HEVC 帧间模式决策方法。Lei 等<sup>[16]</sup>利用灰度相似度和视点间相关性，提出了一种用于 3D-HEVC 深度图编码的快速模式决策方法。Liao 等<sup>[17]</sup>利用时空相关性及视点间相关性，提出了一种有效的帧间预测方法以降低 3D-HEVC 深度图编码复杂度。

由于在 HEVC 或 3D-HEVC 中，SKIP/ Merge 模式是一种特殊的帧间预测模式，通常在背景或运动缓慢区域被选为最优模式，并且其不需要进行复杂的运动估计 (ME, motion estimation)<sup>[18]</sup>以获得预测信息。因此，研究者提出了一些提前 SKIP/Merge 模式决策方法以降低编码复杂度。例如，Yang 等<sup>[19]</sup>提出了一种提前 SKIP 模式决策方法，该方法首先计算 Inter<sub>2N×2N</sub> 和 Merge 模式，如果当前编码 CU 满足运动矢量差 (MVD, motion vector difference) 和 Inter<sub>2N×2N</sub> 模式的残差都为零的条件，则跳过当前 CU 深度其余模式计算。其中，SKIP 模式为 Merge 模式的一种特殊情况，既不计算 ME 也不编码残差。Li 等<sup>[20]</sup>提出了一种单峰停止模型用于 HEVC 提前 SKIP 模式决策。Pan 等<sup>[21]</sup>利用全零块 (AZB, all zero block) 和 Inter<sub>2N×2N</sub> 模式的运动信息，提出了一种提前 Merge 模式决策方法。利用视点间相关性，研究者提出了一些提前 Merge 模式决策方法，分别用于 3D-HEVC 的非独立纹理视点<sup>[22-24]</sup>和非独立深度图<sup>[25]</sup>。Li 等<sup>[26]</sup>通过建立概率模型，提出了一种提前 Merge 模式决策方法以降低 3D-HEVC 非独立视点编码复杂度。

然而，现有 3D-HEVC 提前 Merge 模式决策方法仍未充分探索所有纹理视点、深度图的独立视点和非独立视点的 Merge 模式提前决策。本文提出了一种基于残差的学习模型，用于所有纹理视点和深度图提前 Merge 模式决策，以降低 3D-HEVC 的编码复杂度。本文的主要贡献包括：1) 提取纹理视点和深度图的预测残差信号作为特征，其与传统的量化后残差系数不同；2) 建立学习模型，为每个 P 帧和 B 帧选择自适应阈值；3) 针对 3D-HEVC 中的纹理视点和深度图，提出了提前 Merge 模式决策方法，该方法在保持编码性能的同时降低了编码复杂度。

## 2 问题分析

为了分析 3D-HEVC 中预测单元 (PU, prediction unit) 的预测模式的分布，采用 2 个不同运动特性的视频序列进行测试，“GT\_Fly” (1 920×1 088) 为缓慢运动视频序列，“Poznan\_Hall2” (1 920×1 088) 为快速运动视频序列。测试条件为：对于纹理视点

和深度图，量化参数 (QP, quantization parameter) 分别设置为 25、30、35 和 40，编码树单元 (CTU, coding tree unit) 大小为 64×64。图 2 给出了纹理视点和深度图在不同 QP 下的 Merge 模式分布情况。从图 2 中可以看出，在纹理视点和深度图编码中，大部分 Merge 模式被选择为最优模式。随着 QP 的增加，对于纹理视点和深度图编码，Merge 模式的百分比也逐渐增加。同时，对于不同的运动特性视频序列，非独立视点的 Merge 模式的百分比大于独立视点的 Merge 模式的百分比。特别地，对于具有缓慢运动的“GT\_Fly”视频序列，超过 75%和 77%的 CU 分别选择 Merge 模式作为纹理视点和深度图编码中的最优模式。由于“Poznan Hall2”视频序列中包含快速运动对象，因此，在 QP=25 时，仅有 34%和 54%的独立和非独立深度图选择 Merge 模式为最优模式。但是，在较大的 QP 编码中，Merge 模式的百分比仍然相对较高。在 3D-HEVC 的模式决策过程中，Merge 模式都在其他模式之前进行计算，并且纹理和深度图视频序列中都存在大量的静止或运动缓慢区域，而这些区域大多选择 Merge 模式作为最优模式。因此，如果可以准确地提前终止 Merge 模式，则将跳过帧间和帧内的剩余模式计算过程，从而减少编码的计算复杂度。因此，可以设计合理的条件提前终止 Merge 模式决策，以减少编码时间。

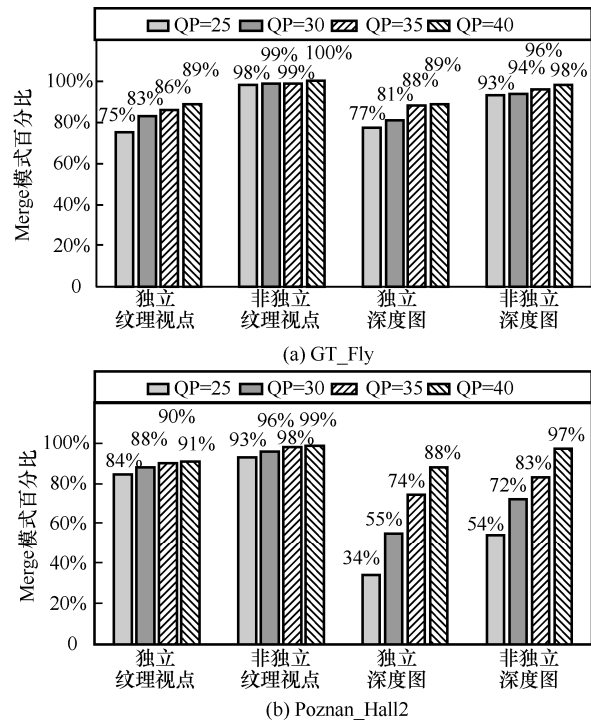


图 2 不同 QP 下纹理视点和深度图的 Merge 模式百分比

### 3 提前 Merge 模式决策方法

#### 3.1 整体框架

本文提出的方法基于学习框架,其目标是提前终止 Merge 模式决策并且保持 3D-HEVC 纹理视点和深度图的编码性能基本不变。与传统的基于相关性分析方法不同,本文提出了一种新的基于残差的学习方法提前终止 Merge 模式决策。具体地,本文将提前 Merge 模式决策的过程建模为二元分类问题。因此,本文建立了一个基于残差的学习模型以确定 Merge 模式是否为最优模式。图3为针对每个 CU 深度  $i$  的提前 Merge 模式决策方法的流程。首先,计算每个 CU 深度  $i$  的 Merge 模式。然后,提取 Merge 模式的残差信号特征值,将其与学习模型的判定参数值进行比较,以确定当前 CU 深度  $i$  的 Merge 模式是否为最优模式。学习模型的决策参数值从先前已编码的最优 Merge 模式中学习。

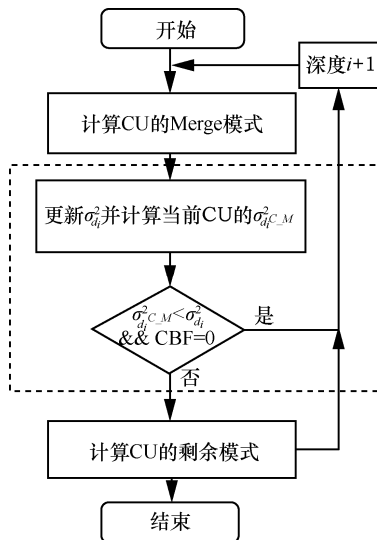


图3 提前 Merge 模式决策方法的流程

#### 3.2 特征分析和提取

为每一个 CU 深度  $i$  获得最优模式,所有预测模式  $m$  都需要通过式(1)计算其 RD cost,即

$$RD_m = \arg \min_{m \in Q} \{D_m + \lambda B_m\} \quad (1)$$

其中,  $RD_m$ 、 $D_m$ 、 $B_m$  和  $\lambda$  分别表示模式  $m$  的 RD cost、预测失真、编码比特率和拉格朗日因子。 $Q$  表示所有的帧间帧内预测模式。在式(1)中,编码比特率  $B_m$  也可以分为两部分,头部比特率和残差系数比特率。因此,式(1)可以重写为

$$RD_m = \arg \min_{m \in Q} \{D_m + \lambda(B_m^h + B_m^c)\} \quad (2)$$

其中,  $B_m^h$  和  $B_m^c$  分别表示头部比特率和残差系数比特率。从式(2)可以看出,失真  $D_m$  和残差系数  $B_m^c$  都与 PU 预测模式的残差信号有很强的正相关关系,这是因为残差信号值越大,失真  $D_m$  值越大,残差信号的编码比特率可能越大。其相关性可以表示为

$$D_m = f(D)R_S \quad (3)$$

$$B_m^c = g(B)R_S \quad (4)$$

其中,  $R_S$  表示残差信号,  $f(D)$  和  $g(B)$  分别表示失真和残差系数比特率的正相关系数。然后,结合式(2)~式(4),模式  $m$  的 RD cost 计算可以重写为

$$RD_m = \arg \min_{m \in Q} \{f(D)R_S + \lambda(B_m^h + g(B)R_S)\} = \arg \min_{m \in Q} \{R_S(f(D) + \lambda g(B)) + \lambda B_m^h\} \quad (5)$$

根据式(5)可以得出结论,模式  $m$  的 RD cost 与残差信号具有很强的正相关关系。由于最优模式决策是基于模式的 RD cost 的大小进行判断的,因此将残差信号用作评估提前 Merge 模式决策的有效特征是合理的。本文将提前 Merge 模式决策过程建模为二元分类问题,分类方法是通过特征提取和决策参数学习更新。其中预测残差  $r(i, j)$  定义为

$$r(i, j) = \text{ori}(i, j) - \text{pre}(i, j), \quad (i, j) \in (W, H) \quad (6)$$

其中,  $\text{ori}(i, j)$  和  $\text{pre}(i, j)$  表示原始亮度像素值和预测模式亮度像素值,  $W$  和  $H$  分别表示 CU 的宽和高。由于方差越小,该组中的数据越稳定;方差越大,该组中的数据越不稳定。因此,选择残差信号的方差  $\sigma^2$  作为特征,计算式为

$$\sigma^2 = \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H [r(i, j) - \frac{1}{WH} \sum_{i=1}^W \sum_{j=1}^H r(i, j)]^2 \quad (7)$$

根据式(7)可知,当前预测模式的  $\sigma^2$  值越小,原始像素值和预测像素值之间的失真越小。相反,当前预测模式的  $\sigma^2$  值越大,原始像素值和预测像素值之间的失真越大。因此,如果当前预测模式的  $\sigma^2$  很小,则可以提前终止当前模式。通常提出的方法中,采用量化后的残差信号进行快速模式选择,从式(2)可以看出,当量化残差信号时,它不能很好地反映失真  $D_m$ 。同时,量化后的大部分残差信号将等于 0。因此,本文在量化之前使用残差信号,可以更好地表示预测失真。

### 3.3 基于学习的决策参数更新

提前 Merge 模式决策被定义为一个二元分类问题，需要预先设置决策参数以预测 Merge 模式的残差信号是否为最优预测残差信号，如果 Merge 模式的预测残差信号满足提前终止模式决策参数条件，则它将提前终止。通常，可以通过基于统计学习或机器学习方法来选择决策参数  $\sigma_{d_i}^2$ 。但是，决策参数需要大量的原始样本进行训练，这将会进一步增加编码的计算复杂度，同时也增加了编码硬件设计的难度。因此，本文提出了一种简单的学习方法来实现决策参数  $\sigma_{d_i}^2$  更新，它不需要统计实验，只采用一个原始编码的最优 Merge 模式数据进行训练，并且  $\sigma_{d_i}^2$  的值为自适应更新。

图 4 给出了学习决策参数独立更新框架。为了避免不同 QP 的影响，决策参数在每个编码帧中重新进行更新，即在每个新编码帧的开始处初始化决策参数。不同 CU 深度  $i$  独立地更新决策参数  $\sigma_{d_i}^2$ ，即

$$\sigma_{d_i}^2 = \frac{1}{1 + \sum \omega_i} \sum_i \sigma_{\max}^2 \omega_i \quad (8)$$

$$\omega_i = \begin{cases} 1, & \text{深度 } i \text{ 的最优模式为 Merge} \\ 0, & \text{其他} \end{cases} \quad (9)$$

其中， $\sigma_{\max}^2$  是 CU 深度  $i$  计算完所有帧间帧内预测模式后获得的最优模式残差信号 ( $\sigma_{d_i^c}^2$ 、 $\sigma_{d_i^1}^2$ 、 $\sigma_{d_i^2}^2$ 、 $\sigma_{d_i^3}^2$  和  $\sigma_{d_i^4}^2$ ) 的最大方差值，如图 5 所示。 $\omega_i$  是权重参数，当 CU 深度  $i$  的最优模式为 Merge 模式时， $\omega_i=1$ ，否则， $\omega_i=0$ 。对于每个编码 P 帧或 B 帧，通过原始 3D-HEVC 编码产生的第一个最优 Merge

模式生成  $\sigma_{d_i}^2$  的初始值。在当前帧的编码期间，通过学习先前由本文提出的方法生成的数据，在线独立地更新  $\sigma_{d_i}^2$ 。

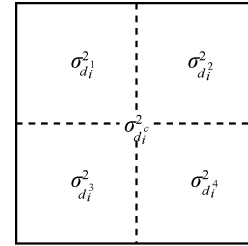


图 5 残差信号方差

### 3.4 基于学习模型的提前 Merge 模式决策

如 3.3 节所述，决策参数  $\sigma_{d_i}^2$  用于判断 Merge 模式是否为最优模式。因此，提前 Merge 模式决策条件可表示为

$$m_i \in \begin{cases} \text{Merge 模式, } \sigma_{d_i^{c-M}}^2 < \sigma_{d_i}^2 \\ \text{其他模式, 其他} \end{cases} \quad (10)$$

其中， $\sigma_{d_i}^2$  是在 CU 深度  $i$  处的先前已编码的最优 Merge 模式中的学习参数， $\sigma_{d_i^{c-M}}^2$  为当前编码 CU 深度  $i$  的 Merge 模式残差信号方差。如果当前 CU 深度  $i$  ( $\sigma_{d_i^{c-M}}^2$ ) 处的 Merge 模式的方差值小于自适应学习值 ( $\sigma_{d_i}^2$ )，则 CU 深度  $i$  处的 Merge 模式将被设置为最优模式；否则，需要计算剩余的模式选择过程。这是因为残差信号的较小方差值表示预测失真较小并且 Merge 模式可能为最优模式。

由于量化后的 Merge 模式的残差信号通常等于 0，为了进一步提高提前 Merge 模式决策的准确性，

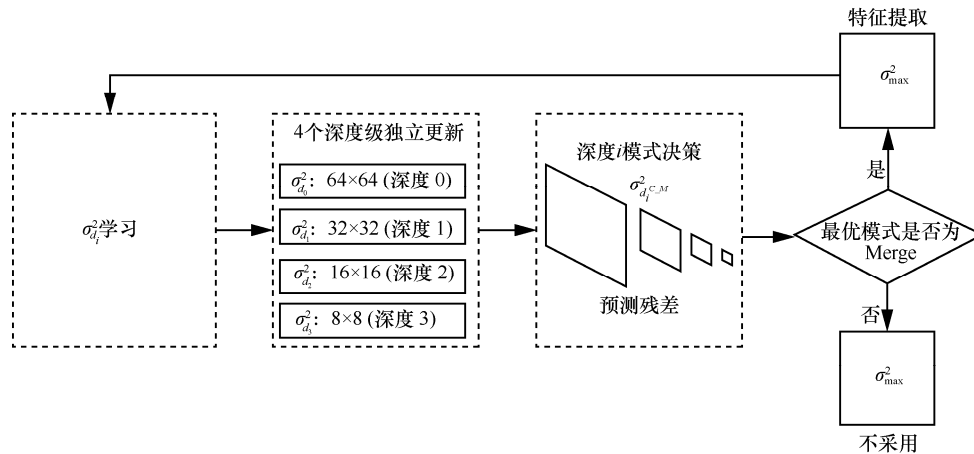


图 4 学习决策参数独立更新框架

将编码块标记 (CBF) 与式(10)进行组合, 作为 3D-HEVC 中纹理视点和深度图编码的提前 Merge 模式决策条件。根据文献[16], 如果模式的 CBF=0, 则表示当前模式很有可能为最优模式。因此, 纹理视点和深度图编码的提前 Merge 模式决策条件可表示为

$$m_i \in \begin{cases} \text{Merge 模式, } \sigma_{d_i^{c_M}}^2 < \sigma_{d_i}^2 \ \&\& \text{CBF} = 0 \\ \text{其他模式, 其他} \end{cases} \quad (11)$$

## 4 实验结果及分析

### 4.1 测试条件

为了验证本文提出的提前 Merge 模式决策算法对 3D-HEVC 中纹理视点和深度图编码性能的影响, 将提出的算法集成到 3D-HEVC 参考软件 HTM16.2 上实现, 快速编码决策方法<sup>[22]</sup>和 DIS 模式始终开启。采用通用测试条件进行评估<sup>[27]</sup>, 其中包括 JCT-3V 推荐的 8 个测试视频序列, 用于评估所提出算法的效率, 视频序列包括 2 个分辨率, 分

别是 1 024×768 (Balloons、Kendo、Newspaper) 和 1 920×1 088 (GT\_Fly、Poznan\_Hall2、Poznan\_Street、Undo\_Dancer、Shark)。每个测试视频序列包括 3 个纹理视点及其对应的深度图, 测试序列的细节如表 1 所示。编码时间节约计算为

$$TS_{\Omega} = \frac{ET_{ori}^{\Omega} - ET_{pro}^{\Omega}}{ET_{ori}^{\Omega}} \times 100\% \quad (12)$$

其中,  $ET_{ori}^{\Omega}$  表示原始平台 HTM16.2 的编码时间,  $ET_{pro}^{\Omega}$  表示本文提出的快速算法的编码时间,  $\Omega$  表示包括纹理视点编码时间、深度图编码时间以及总编码时间 (纹理和深度)。

### 4.2 实验结果

表 2~表 4 给出了本文提出的用于 3D-HEVC 纹理视点和深度图的提前 Merge 模式决策方法的编码性能, 其通过纹理视点和深度图的编码时间减少和 Bjontegaard (BD-Rate) 来测量<sup>[28]</sup>。表 2~表 4 中, video<sub>0</sub>、video<sub>1</sub> 和 video<sub>2</sub> 分别表示单个纹理视点的 BD-Rate 编码结果, video PSNR/ video

表 1 测试视频序列

视频序列	分辨率	帧率/(frame·s <sup>-1</sup> )	独立视点&非独立视点
Balloons	1 024×768	30	V <sub>3</sub> &V <sub>1</sub> , V <sub>5</sub>
Kendo	1 024×768	30	V <sub>3</sub> &V <sub>1</sub> , V <sub>5</sub>
Newspaper	1 024×768	30	V <sub>4</sub> &V <sub>2</sub> , V <sub>6</sub>
GT_Fly	1 920×1 088	25	V <sub>5</sub> &V <sub>1</sub> , V <sub>9</sub>
Poznan_Hall2	1 920×1 088	25	V <sub>6</sub> & V <sub>5</sub> , V <sub>7</sub>
Poznan_Street	1 920×1 088	25	V <sub>4</sub> & V <sub>3</sub> , V <sub>5</sub>
Undo_Dancer	1 920×1 088	25	V <sub>5</sub> &V <sub>1</sub> , V <sub>9</sub>
Shark	1 920×1 088	30	V <sub>5</sub> &V <sub>1</sub> , V <sub>9</sub>

表 2 本文提出方法的编码性能

视频序列	video <sub>0</sub>	video <sub>1</sub>	video <sub>2</sub>	V/V	V/T	S/T	TS <sub>texture</sub>	TS <sub>depth</sub>	TS <sub>total</sub>
Balloons	0.5%	0.8%	0.8%	0.6%	0.6%	0.7%	39.3%	21.3%	32.1%
Kendo	0.6%	0.9%	0.5%	0.7%	0.6%	0.7%	37.1%	18.2%	29.2%
Newspaper	0.5%	0.5%	0.3%	0.5%	0.3%	0.7%	44.1%	21.3%	33.9%
GT_Fly	1.8%	1.9%	2.1%	1.9%	1.8%	1.8%	48.7%	26.9%	39.7%
Poznan_Hall2	0.6%	0.8%	0.9%	1.0%	0.9%	0.8%	50.1%	23.1%	38.6%
Poznan_Street	0.2%	0.7%	0.3%	0.6%	0.7%	0.4%	47.5%	26.5%	38.1%
Undo_Dancer	1.0%	0.7%	0.6%	1.1%	0.9%	1.2%	37.3%	30.6%	34.6%
Shark	0.2%	-0.1%	-0.1%	0.2%	-0.1%	0.4%	31.2%	26.4%	28.9%
平均值	0.7%	0.8%	0.7%	0.8%	0.7%	0.8%	41.9%	24.3%	34.4%

**表 3** PanTB 方法的编码性能

视频序列	video <sub>0</sub>	video <sub>1</sub>	video <sub>2</sub>	V/V	V/T	S/T	TS <sub>texture</sub>	TS <sub>depth</sub>	TS <sub>total</sub>
alloons	0.1%	1.1%	1.2%	0.6%	0.5%	0.6%	20.8%	11.3%	17.0%
Kendo	0.2%	1.9%	1.8%	0.9%	0.8%	0.7%	19.6%	8.9%	15.1%
Newspaper	0.1%	2.9%	0.8%	0.9%	0.6%	1.0%	30.0%	10.2%	21.2%
GT_Fly	0.7%	-1.0%	-0.8%	0.8%	0.6%	0.4%	20.6%	12.0%	23.0%
Poznan_Hall2	0.1%	1.4%	3.1%	0.9%	0.8%	0.8%	26.9%	11.4%	20.2%
Poznan_Street	0.1%	0.4%	0.5%	0.4%	0.3%	0.4%	33.0%	13.2%	24.2%
Undo_Dancer	0.4%	-0.8%	-0.9%	0.5%	0.3%	0.2%	26.2%	16.8%	22.5%
Shark	0.3%	-1.7%	-1.5%	0.2%	0.0%	0.2%	28.8%	11.9%	21.5%
平均值	0.3%	0.5%	0.5%	0.7%	0.5%	0.5%	25.7%	12.0%	20.6%

**表 4** CBF+ESD 方法的编码性能

视频序列	video <sub>0</sub>	video <sub>1</sub>	video <sub>2</sub>	V/V	V/T	S/T	TS <sub>texture</sub>	TS <sub>depth</sub>	TS <sub>total</sub>
Balloons	0.5%	1.0%	0.9%	0.7%	0.7%	0.5%	36.3%	6.0%	24.5%
Kendo	0.5%	1.2%	0.9%	0.8%	0.8%	0.5%	35.2%	3.7%	22.4%
Newspaper	0.3%	0.4%	0.6%	0.4%	0.4%	0.3%	37.9%	5.1%	23.6%
GT_Fly	1.5%	2.6%	2.8%	2.0%	1.9%	1.6%	37.0%	6.1%	24.6%
Poznan_Hall2	0.6%	0.8%	0.6%	0.7%	0.7%	0.6%	33.1%	4.1%	20.9%
Poznan_Street	0.2%	0.9%	1.3%	0.6%	0.6%	0.4%	34.2%	5.7%	21.8%
Undo_Dancer	0.7%	4.1%	4.7%	2.0%	1.9%	1.3%	32.4%	9.1%	23.5%
Shark	0.7%	1.6%	1.8%	1.1%	1.0%	0.8%	31.2%	5.6%	20.3%
平均值	0.6%	1.6%	1.7%	1.0%	1.0%	0.8%	34.6%	5.7%	22.7%

bitrate (V/V) 表示所有编码纹理视点的 BD-Rate 编码结果, video PSNR/total bitrate (V/T) 表示编码纹理视点和深度图的比特率之和的 BD-Rate 编码结果, synth PSNR/total bitrate (S/T) 表示合成视点的 BD-Rate 编码结果, TS<sub>texture</sub>、TS<sub>depth</sub> 和 TS<sub>total</sub> 分别表示纹理视点编码时间节约、深度图编码时间节约和总编码 (纹理视点和深度图) 时间节约。

表 2 给出了本文提出方法的编码性能。从表 2 可以看出, 本文提出的方法可以有效地节省纹理视点和深度图的编码时间, 同时保持所有测试纹理视点和深度图的相似的编码效率。对于纹理视点, 本文提出方法的编码时间节约为 31.2%~50.1%, 平均为 41.9%。同时, video<sub>0</sub>、video<sub>1</sub> 和

video<sub>2</sub> 的 BD-Rate 平均增加了 0.7%、0.8% 和 0.7%。对于深度图编码, 编码时间节约为 8.2%~30.6%, 平均减少 24.3%。同时, 总编码时间节约为 28.9%~39.7%, 平均为 34.4%。V/V、V/T 和 S/T 的 BD-Rate 平均增加了为 0.8%、0.7% 和 0.8%。以上实验结果表明, 本文提出的基于学习模型的提前 Merge 模式决策方法可以有效地提前终止 3D-HEVC 中纹理视点和深度图的 Merge 模式决策。

为了进一步评估本文提出方法的性能, 将其与提前 Merge 模式决策方法 PanTB<sup>[20]</sup>、组合快速方法 CBF+ESD<sup>[9,19]</sup> 用于编码性能的客观比较。表 3 给出了 PanTB 方法与原始测试平台 HTM16.2 相比的性能。从表 3 可以看出, PanTB 方法的纹理视点、深

度图和总编码时间平均节约了 25.7%、12%和 20.6%， $video_0$ 、 $video_1$  和  $video_2$  的 BD-Rate 平均增加了 0.3%、0.5%和 0.5%，V/V、V/T 和 S/T 的 BD-Rate 平均增加了 0.7%、0.5%和 0.5%。与 PanTB 算法相比，本文提出的方法可以进一步节约纹理视点、深度图和总编码 16.2%、12.3%和 13.8%的编码时间，同时保持相似的 BD-Rate 增加。

表 4 给出了 CBF+ESD<sup>[9,19]</sup>方法与原始测试平台 HTM16.2 相比的性能。从表 4 可以看出，CBF+ESD 方法对纹理视点、深度图和总编码的编码复杂度平均降低了 34.6%、5.7%和 22.7%， $video_0$ 、 $video_1$  和  $video_2$  的 BD-Rate 平均增加了 0.6%、1.6%和 1.7%，V/V、V/T 和 S/T 的 BD-Rate 平均增加了 1.0%、1.0%和 0.8%。与 CBF+ESD 方法相比，本文提出的方法可以进一步节约纹理视点、深度图和总编码 7.3%、18.6%和 11.7%的编码时间，同时获得了更好的编码性能。

为了更直观地体现独立和非独立视点的编码时间节约，图 6 给出了 PanTB、CBF+ESD 和本文提出方法之间的编码时间节约对比。从图 6 中可以看出，本文提出的基于学习残差模型的提前 Merge 模式决策方法节约了更多的独立视点和非独立视点的编码时间。对于独立纹理视点编码，这些方法减少了几乎相同的编码时间。然而，对于独立和非独立的深度图编码，与 PanTB 和 CBF+ESD 方法相比，本文提出的方法可以为每个测试视频序列节约更多编码时间。特别是对于非独立纹理视点编码，PanTB 方法增加了 Poznan\_Hall2 的编码时间，原因在于 PanTB 方法中首先计算 CU 深度 0 处的 Inter<sub>2N×2N</sub> 模式，然而，集成在 3D-HEVC 中的快速算法首先计算 Merge 模式并提前终止，这可能导致 Merge 模式应该提前终止而没有提前终止，而 PanTB 方法中需要进一步计算 Inter<sub>2N×2N</sub> 模式，从而增加编码时间。

为了验证本文提出的方法单独用于 3D-HEVC 深度图的编码性能影响，表 5 给出了在深度图编码下不同快速方法的编码性能对比，其中文献[16]方法（测试平台为 HTM13.0）和文献[17]方法（测试平台为 HTM15.1）都可用于降低 3D-HEVC 深度图的编码复杂度。从表 5 中可以看出，本文提出的方法能降低较多的编码时间，得到很好的合成视点编码质量。相比文献[16]方法，本文提出的方法在相似的时间节约下，还降低了 2.1%的编码码率。相比

于文献[17]方法，本文提出的方法可以进一步节约 20.8%的深度编码时间，同时降低了 0.2%的编码码率。综上所述，将本文提出的方法用于所有 3D-HEVC 纹理视点和深度图编码以及只用于 3D-HEVC 深度图编码，相比于其他先进的快速 3D-HEVC 编码方法，都能得到更好的编码性能。

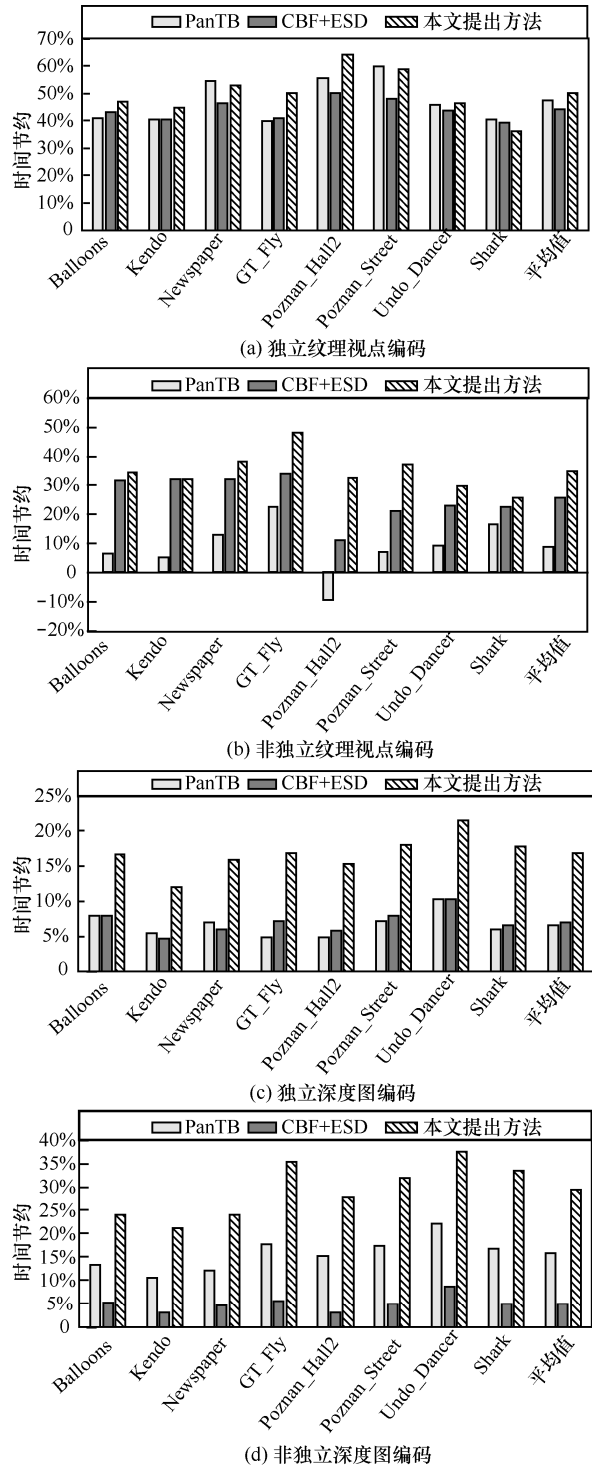


图 6 编码时间节约对比

表5 在深度图编码下不同快速方法的编码性能对比

视频序列	本文提出的方法		文献[16]方法		文献[17]方法	
	S/T	TS <sub>Depth</sub>	S/T	TS <sub>Depth</sub>	S/T	TS <sub>Depth</sub>
Balloons	0.1%	33.6%	2.4%	41.5%	0.8%	16.3%
Kendo	0.1%	28.4%	4.0%	35.5%	0.3%	16.2%
Newspaper	0.3%	32.4%	2.0%	45.0%	0.7%	15.2%
GT_Fly	0.1%	39.4%	1.6%	42.1%	—	—
Poznan_Hall2	0.2%	40.0%	0.9%	44.8%	-0.6%	17.3%
Poznan_Street	0.1%	40.0%	2.1%	47.2%	0.9%	17.0%
Undo_Dancer	0.4%	44.8%	2.6%	48.5%	—	—
Shark	0.2%	39.0%	3.0%	31.6%	0%	16.5%
平均值	0.2%	37.2%	2.3%	42.0%	0.4%	16.4%

## 5 结束语

本文提出了一种提前 Merge 模式终止方法,用于 3D-HEVC 纹理视点和深度图,通过使用基于残差的学习模型来降低 3D-HEVC 编码器的计算复杂度。首先,基于每个最优 Merge 模式中的原始亮度像素和预测亮度像素来提取残差信号。其次,基于学习残差的提前 Merge 模式决策方法采用所提出的残差信号特征来提前预测最优模式是否为 Merge 模式。实验结果表明,该方法可以有效地减少编码时间,在复杂度降低方面优于最先进的算法。此外,基于学习的方法还有几个独特的优点。首先,与传统的基于统计的方法相比,本文提出的方法可以在纹理视点和深度图编码中实现。其次,学习模型简单易实现,适用于 3D-HEVC 和 HEVC 的提前 Merge 模式决策。

## 参考文献:

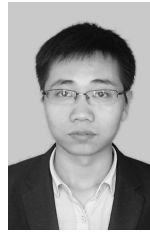
- [1] 陈亮, 顾翔, 王进. 多速率多节点 IEEE 802.11 网络的视频传输策略[J]. 通信学报, 2017, 38(3): 34-44.  
CHEN L, GU X, WANG J. Video transmission strategy in multi-rate and multi-node IEEE 802.11 networks[J]. Journal on Communications, 2017, 38(3): 34-44.
- [2] SULLIVAN G J, OHM J, HAN W, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.
- [3] MULLER K, 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.
- [4] MERKLE P, MLLER K, ZHAO X, et al. CE6.H results on simplified wedgelet search for DMM Modes 1 and 3[C]//JCT3V Meeting,

JCT3V-B0039. 2012: 1-4.

- [5] LEE J Y, PARK M W, KIM C. 3D-CE1: depth intra skip (DIS) mode[C]//JCT3V meeting, JCT3V-K0033. 2005: 1-5.
- [6] LIU H, CHEN Y. Generic segment-wise DC for 3D-HEVC depth intra coding[C]// 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014: 3219-3222.
- [7] SHEN L, LI K, FENG G, et al. Efficient intra mode selection for depth-map coding utilizing spatiotemporal, inter-component and inter-view correlations in 3D-HEVC[J]. IEEE Transactions on Image Processing, 2018, 27(9): 4195-4206.
- [8] 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, doi: 10.1109/TCSVT.2019.2898122, 2019.
- [9] GWEON R H, LEE Y L. Early termination of CU encoding to reduce HEVC complexity[C]//JCTVC Meeting, JCTVC-F045. 2011: 1-4.
- [10] 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, 2015, 11(3): 1-23.
- [11] ZHANG Q, HUANG K, WANG X, et al. Efficient multiview video plus depth coding for 3D-HEVC based on complexity classification of the treeblock[J]. Journal of Real-Time Image Processing, doi: 10.1007/S11554-017-0692-5, 2017.
- [12] TOHIDYPOUR H R, POURAZAD M T, NASIOPOULOS P. Online-learning based complexity reduction scheme for 3D-HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(10): 1870-1883.
- [13] ZHANG Q, WEI S, SU R. Low-complexity texture video coding based on motion homogeneity for 3D-HEVC[J]. Scientific Programming, 2019: 1-13.
- [14] LI Y, YANG G, ZHU Y, et al. Hybrid stopping model-based fast PU and CU decision for 3D-HEVC texture coding[J]. Journal of Real-Time Image Processing, doi: 10.1007/s11554-019-00876-9, 2019.
- [15] CHEN J, WANG B, LIAO J, et al. Fast 3D-HEVC inter mode decision algorithm based on the texture correlation of viewpoints[J]. Multimedia Tools and Applications, doi: 10.1007/s11042-018-6832-5, 2018.
- [16] LEI J, DUAN J, WU F, et al. Fast mode decision based on grayscale similarity and inter-view correlation for depth map coding in 3D-HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(3): 706-718.
- [17] LIAO Y, CHEN M, YE H C, et al. Efficient inter-prediction depth coding algorithm based on depth map segmentation for 3D-HEVC[J]. Multimedia Tools and Applications, 2019, 78(8): 10181-10205.
- [18] PAN Z, LEI J, ZHANG Y, et al. Adaptive fractional-pixel motion estimation skipped algorithm for efficient HEVC motion estimation[J]. ACM Transactions on Multimedia Computing Communications and Applications, 2018, 14(1):1-19.

- [19] YANG J, KIM J, WON K, et al. Early SKIP detection for HEVC[C]// JCTVC Meeting, JCTVC-G543. 2011: 1-6.
- [20] LI Y, YANG G, ZHU Y, et al. Unimodal stopping model based early SKIP mode decision for high efficiency video coding[J]. IEEE Transactions Multimedia, 2017, 19(7): 1431-1441.
- [21] 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.
- [22] ZHANG N, ZHAO D, CHEN Y, et al. Fast encoder decision for texture coding in 3D-HEVC[J]. Signal Processing: Image Communication, 2014, 29(9): 951-961.
- [23] ZHANG Q, WU Q, WANG X, et al. Early SKIP mode decision for three dimensional high efficiency video coding using spatial and inter-view correlations[J]. Journal of Electronic Imaging, 2014, 23(5): 1-8.
- [24] SONG Y, JIA K. Early merge mode decision for texture coding in 3D-HEVC[J]. Journal of Visual Communication and Image Representation, 2015, 23: 60-68.
- [25] CHEN H, FU C H, ZHANG Y, et al. Early merge mode decision for depth maps in 3D-HEVC[C]//22nd International Conference on Digital Signal Processing (DSP). 2017: 1-5.
- [26] LI Y, YANG G, ZHU Y, et al. Probability model-based early merge mode decision for dependent views coding in 3D-HEVC[J]. ACM Transactions on Multimedia Computing Communications and Applications, 2018, 14(4): 1-15.
- [27] MULLER K, VETRO A. Common test conditions of 3DV core experiments[C]//JCT3V Meeting, JCT3V-G1100. 2014: 1-7.
- [28] BJONTEGAARD G. Calculation of average PSNR differences between RD curves[R]. VCEG-M33, Austin: Video Coding Experts Group, 2001.

## [作者简介]



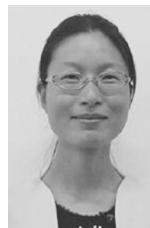
李跃(1988- )，男，湖南衡阳人，博士，南华大学讲师，主要研究方向为 HEVC、3D-HEVC、VVC 编码优化等。



杨高波(1974- )，男，湖南岳阳人，博士，湖南大学教授、博士生导师，主要研究方向为图像/视频信号处理、多媒体通信、数字媒体内容安全等。



丁湘陵(1981- )，男，湖南株洲人，博士，湖南科技大学副教授，主要研究方向为视频信息安全、图像处理等。



朱亚培(1988- )，女，河南驻马店人，衡阳师范学院讲师，主要研究方向为图像/视频压缩等。