

结合显著性的MVD视频整帧丢失错误隐藏

崔金鹏, 周 洋, 殷海兵, 黄晓峰, 陆 宇

(杭州电子科技大学通信工程学院, 浙江杭州 310018)

摘 要: 基于多视点视频加深度(Multiview Video plus Depth, MVD)格式的三维视频能给用户提供多视角的沉浸式视觉体验. 针对MVD视频流传输中彩色帧发生整帧丢失的情况, 提出了一种结合视觉显著性的三维视频整帧丢失错误隐藏算法. 先将丢失帧各区域划分为三种不同的显著等级, 然后对低、中、高显著区域, 分别采用时域直接拷贝、时域和视点间像素自适应搜索法和新型的运动矢量补偿法进行恢复. 相较于目前的对比算法, 经提出算法重建后的丢失帧的峰值信噪比(Peak Signal to Noise Ratio, PSNR)和结构相似度值(Structural Similarity Index Measure, SSIM)能分别提高0.99~2.61 dB和0.005~0.012, 并且重建图像主观视觉质量更佳.

关键词: 多视点加深度; 错误隐藏; 整帧丢失; 视觉显著性; 运动补偿; 视差补偿

中图分类号: TP391; TP943

文献标识码: A

文章编号: 0372-2112(2022)09-2146-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210393

Error Concealment for Frame Losses in MVD Video Using Saliency

CUI Jin-peng, ZHOU Yang, YIN Hai-bing, HUANG Xiao-feng, LU Yu

(School of Telecommunication Engineering, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China)

Abstract: The 3D video based on multiview video plus depth(MVD) format can provide users with immersive experiences from different viewpoints. To solve the whole color frame losses in MVD video stream transmission, we propose a novel error concealment algorithm of whole-frames loss using visual saliency. Firstly, each region of the lost frame is classified into three different saliency levels. The low, medium and high saliency regions are recovered by direct temporal copy, adaptive pixel matching search between temporally and inter-view domains, and a new motion vector compensation, respectively. Compared with the recent algorithms, the lost frame concealed by the proposed algorithm achieves significant improvements in the average PSNR and SSIM with approximate gains of 0.99~2.61 dB and 0.005~0.012, respectively. In addition, our concealed results have better visual quality than compared methods.

Key words: multiview video plus depth; error concealment; whole frame loss; visual saliency; motion compensation; disparity compensation

1 引言

目前三维(Three Dimension, 3D)视频日渐获得人们的关注与喜爱, 为此国际标准化组织推出了高性能3D视频编码标准(3D High Efficiency Video Coding, 3D-HEVC)来对3D视频数据进行高效压缩. 然而压缩后的3D数据包在无线信道下传输时易发生数据包丢失, 进而造成解码出错^[1]. 错误隐藏技术是利用码流帧间和帧内的冗余性来重建丢失帧信息, 但由于传统二维视频编码主要基于IPPP预测结构, 其相应的错误隐藏方法对于基于分层B帧预测结构(Hierarchical B Predic-

tion, HBP)的3D视频适用性较差.

根据编码中是否有深度信息, 3D视频数据格式主要分为无深度信息的多视点视频与多视点视频加深度序列两类. 针对多(或双)视点视频传输中的整帧丢失, 文献[2, 3]利用相邻视点运动矢量(Motion Vector, MV)的相似性来重建当前视点的丢失帧, 文献[4]利用MV外推技术来恢复丢失帧. 但文献[2, 3]是基于传统IPPP预测结构, 未面向HBP预测结构的多视点视频. 文献[4]中的运动矢量外推技术更适用于IPPP编码结构, 在双向分层编码的HBP结构下难取得良好的效果. 针对

收稿日期: 2021-03-25; 修回日期: 2022-02-08; 责任编辑: 覃怀银

基金项目: 浙江省自然科学基金(No. Y21F020076); 国家自然科学基金(No. 61972123); 浙江省影视媒体技术研究重点实验室开放基金(No. CM2021001)

多视点视频中部分宏块(Macro Block, MB)丢失的错误隐藏,主要方法有宏块分类修复法^[5]、数据隐藏^[6]、贝叶斯滤波法^[7].但文献[5~7]只能解决单帧图像中的部分块出错并未面向整帧丢失的问题.此外,文献[2~7]都是基于H.264/AVC(Advanced Video Coding)编码标准压缩后的视频流进行错误隐藏,未针对目前流行的H.265/HEVC压缩视频流.

多视点视频加深度的错误隐藏除了利用视点内时空相关性与视点间相关性外,还可采用深度信息来提升重建质量.其中,深度序列的时域平滑度^[8]、视点间两对应块间的深度值差异^[9]以及重建块与周边块之间的深度差^[10]等常被用于丢失块的MV和视差矢量(Disparity Vector, DV)恢复上.面向视频帧整帧丢失, Tai等^[11]采用高精度运动估计和深度图信息外推来重建丢失帧. Khattak等^[12]采用一致性模型来对中间视点整帧丢失进行错误隐藏. Lin等^[13]利用深度差对参考像素进行加权来恢复丢失帧.此外,针对深度图出错问题,近来常用方法有基于模糊推理的自适应错误隐藏法^[14]、结合轮廓曲线拟合的三步错误隐藏法^[15]、采用像素聚类的空洞隐藏法^[16]、绘制驱动的多模式隐藏法^[17].但是,文献[8~16]是针对H.264/AVC标准压缩后的视频流进行错误隐藏,且大多数码流都是基于IPPP编码结构的,尚未针对基于HBP编码结构的3D-HEVC标准展开研究.文献[14~17]都是面向深度图丢失的隐藏算法,无法有效地重建丢失的彩色图.

当3D视频在传输中发生彩色图整帧丢失时,首先由于没有相邻块作为参考信息,无法利用基于空域的邻块信息来恢复丢失块,因此面向部分宏块丢失的错误隐藏算法不能直接应用于整帧丢失中;其次,如果依靠对应的深度图信息进行重建,由于彩色图整帧丢失不能准确地提取深度图与彩色图之间的相关性,导致其恢复效果不够理想;最后由于3D-HEVC编码标准在编码端采用双向分层预测的HBP结构,解码端的错误隐藏方法也需基于HBP结构来设计.

针对上述存在的问题,本文基于3D-HEVC编码标准,提出了一种结合立体视觉显著性的彩色图整帧丢失错误隐藏方法.不同于先前算法,本文算法是首个基于3D-HEVC编码标准的多视点视频加深度格式的3D视频彩色图整帧丢失错误隐藏算法;同时,结合了3D视频的视觉显著性,根据整帧丢失图像各编码单元的显著性等级不同进行分区错误隐藏.

2 显著性图提取

首先提取丢失图像的分块显著性图,然后根据各丢失区域的视觉显著性等级不同采用不同的错误隐藏方法.图1所示为分块显著性图提取总体流程框图,先

分别求出运动显著性图、深度显著图和深度边缘显著图,然后将运动显著图和深度显著图相互融合获取运动深度显著图,最后将块化后的运动深度显著图与块化后深度边缘显著图相互融合,获取最终的分块显著等级图.

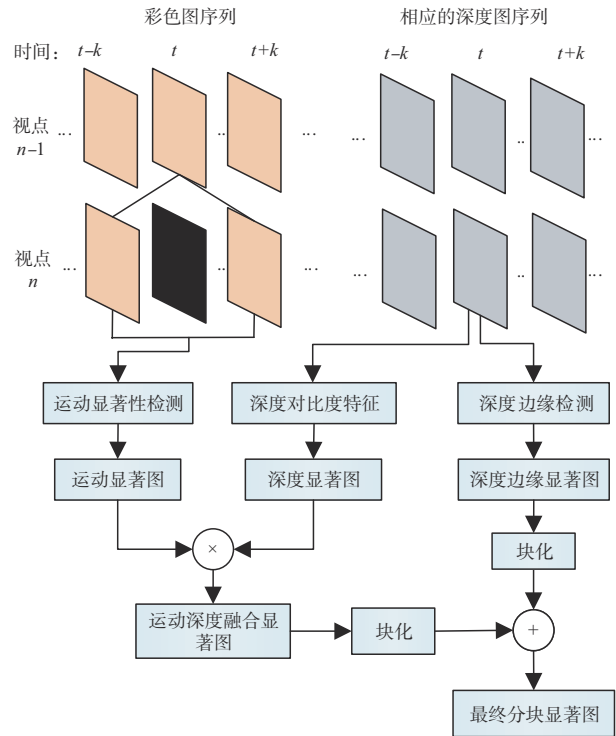


图1 分块显著性图提取流程框图

2.1 运动显著性图

运动显著性图能反映视频序列中物体的运动显著性^[18].将相邻可用视点作为参考视点,在时域上获取当前丢失帧 F_n 的前向参考帧 $F_{n,t-k}$ 和后向参考帧 $F_{n,t+k}$,并在视点域上获取当前时刻的相邻视点参考帧 $F_{n-1,t}$,求得丢失帧中像素 p 的运动显著性 $S_m(p)$ 如下:

$$S_m(p) = (|p_{n,t-k} - p_{n-1,t}| - |p_{n,t+k} - p_{n-1,t}|) \quad (1)$$

其中下标 n 表示视点位置, t 表示时域位置, $p_{n,t-k}$, $p_{n,t+k}$ 和 $p_{n-1,t}$ 分别表示在 $F_{n,t-k}$, $F_{n,t+k}$ 和 $F_{n-1,t}$ 与 p 同位置的像素值.

2.2 深度显著性图

通常来说,深度图中像素值大的前景区域对人眼产生的视觉刺激大,像素值小的背景区域产生的视觉刺激小.但深度像素值并不能直接决定物体的显著程度,有时距离相机较近的物体的显著程度不一定高,比如地面.物体的显著程度还与深度值变化快慢有关,图像中深度值变化大的区域一般是显著程度高的区域^[19].

考虑到深度图只包含一路反映拍摄景物距离远近

的灰度信息,而彩色图携带包含丰富纹理信息的Y/U/V三路通道的信息,深度图编码后的比特流要远小于彩色图的编码比特流,因此在无线传输中深度图发生数据包丢失的概率也低.当彩色图发生整帧丢失时,可假设其对应的深度图未丢失.深度图显著性求取分两步,先对深度图进行加权预处理:

$$d'_z = d_z \times \frac{|d_z - \bar{d}_z| - |d_z - \bar{d}_z|_{\min}}{|d_z - \bar{d}_z|_{\max} - |d_z - \bar{d}_z|_{\min}} \quad (2)$$

其中, z 是深度图中与 p 同位置的像素点, d_z 表示深度图中 z 点像素的深度值, \bar{d}_z 表示深度图中 z 点所在行的平均像素值, d'_z 表示处理后 z 点的深度值.

然后再结合深度图的深度值对比度来提取显著性,基于区域对比度^[20]的像素点 z 的深度显著性 $S_d(z)$ 计算公式如下:

$$S_d(z) = \sum_{r=0}^{255} C(r) \times |d'_z - r| \quad (3)$$

其中, $C(r)$ 表示预处理后深度图中像素值为 r 的像素点出现频率,即 $C(r) = m / (W \times H)$, m 表示深度值为 r 的像素点个数, W 、 H 分别为视频图像的宽度和高度.

2.3 深度边缘显著性图

物体的边缘区域是前景与背景的分界,其深度差较大,往往更受人眼的关注^[21].提取深度边缘显著性是采用Sobel梯度算子来计算深度图各像素点的边缘梯度.设输入深度图为 I_d ,将它分别与Sobel横向、纵向算子作卷积:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * I_d \quad (4)$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I_d \quad (5)$$

其中, G_x 和 G_y 分别代表水平梯度值和垂直梯度值.深度图中每个像素点 z 的梯度幅值 $G(z)$ 为

$$G(z) = |G_x(z)| + |G_y(z)| \quad (6)$$

考虑到处于背景区域的物体边缘和深度值变化平缓的边缘并不显著,需要对各像素点的边缘梯度幅值进行阈值化处理.将边缘梯度幅值大于阈值 t_1 的深度像素点所对应的像素值设为255,其余设为0,得到像素 z 处的边缘显著性值 $S_e(z)$ 为

$$S_e(z) = \begin{cases} 255, & \text{if } G(z) > t_1 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

其中,二值化阈值 t_1 是综合6种不同测试序列的平均实验结果获得,在本文实验中 t_1 值设定为120.

2.4 各维显著性图融合与块化处理

先将深度显著图与运动显著图相互融合得到运动深度显著性图,再将块化后的运动深度显著性图与块

化后的深度边缘显著性图相融合得到最终显著性图.

目前融合算法主要有最大值融合和乘性融合两种方法.实验发现采用最大值融合后的显著性图中像素值大的像素点比较稀疏,这会导致块化过程中阈值选取存在误差,而乘性融合后的显著图的像素点十分密集,不易受阈值选取误差的影响,因此本算法采用乘性融合法.将运动显著性图中各像素值与深度显著性图中对应同位置像素值相乘,并做二值化处理(乘积大于阈值 t_2 的像素值设为128,其余设为0),得到运动深度融合显著性图 S_{md} ,其融合后像素 z 处的显著性值为

$$S_{md}(z) = \begin{cases} 128, & \text{if } S_m(p) \times S_d(z) > t_2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

对于运动深度融合显著性图 S_{md} 中每个尺寸为 64×64 的编码单元块(CTU),如果有超过 t_3 个像素值为128的像素点,则将该CTU内全部像素点的像素值都设为128,得到块化后的显著性图 S_{mdB} .

同理,对深度边缘显著性图 S_e 进行块化处理.对于 S_e 中的各个CTU,如果有超过 t_3 个像素值为255的像素点,则该CTU的像素值都设为255,得到块化 S_{eB} .在这里,阈值 t_2 和 t_3 是在3D-HEVC标准组织推荐的6种多视点测试序列上,通过实验结果比较后将 t_2 和 t_3 值分别设为50和10.

融合 S_{mdB} 和 S_{eB} 得到最终的块化显著性图 S_B :

$$S_B(z) = \begin{cases} 255, & \text{if } S_{eB}(z) = 255 \ \& \ S_{mdB}(z) = 128 \\ 0, & \text{if } S_{mdB}(z) = 0 \\ 128, & \text{otherwise.} \end{cases} \quad (9)$$

然后将 S_B 中像素值为255的区域设为高显著等级区域,像素值为0的区域设为低显著等级区域,像素值为128的区域设为中显著等级区域.各显著图计算结果见图2.



图2 各维显著性图与块化后的显著性图

图2(f)所示为最终显著性图,图中灰色区域是中显著等级区域,该区域没有显著物体的边缘,属于人眼较为关注区域,白色区域是人眼最关注的高显著等级区域,即深度边缘显著性图与融合显著性图重合的区域,剩下的黑色区域是人眼不关注的低显著等级区域.

3 丢失帧重建

3.1 重建低显著等级区域

低显著等级区域一般由静止块或背景块构成,可采用传统的时域像素拷贝法便能达到良好的重建效果,因此本算法选取当前视点后向参考帧的同位置像素来填补处于低显著等级区域的丢失块.

3.2 重建中显著等级区域

中等显著等级区域中的丢失块多为前景块或运动块.考虑到运动块的运动特征,本算法对均匀运动块和非均匀运动块,分别采用时域搜索像素法和视点间搜索像素法进行填补.先将丢失帧中每个尺寸为 64×64 的丢失块以树形编码结构的方式划分为4个 32×32 的子块,下文仅以一个丢失子块 U 为例进行说明.

3.2.1 时域搜索像素填补

如果丢失块 U 的运动是均匀的,那么对 U 设定一个时域搜索值,经过子像素精度搜索总能在当前视点前向参考帧和后向参考帧中找到同尺寸的匹配块 U_f 和 U_b .通过计算 U_f 和 U_b 内像素的平均绝对差值的最小值来确定最优搜索值:

$$(x_{s0}, y_{s0}) = \arg \min_{(x_s, y_s)} \left\{ \frac{1}{N \times N} \sum_{(x, y) \in U} |p_{n, t-k}(x+x_s, y+y_s) - p_{n, t+k}(x-x_s, y-y_s)| \right\} \quad (10)$$

其中 (x, y) 表示丢失块内像素的坐标, x_s 和 y_s 分别为水平和垂直方向的搜索值,通过实验设定 x_s 和 y_s 的范围为 $[-16, 16]$, (x_{s0}, y_{s0}) 为最优搜索值, $N \times N$ 表示 U 的尺寸.

求得两块内像素的平均绝对差值的最小值为 D ,将 D 和运动阈值 $t4$ 进行比较.若 $D < t4$,则认为对应的块 U_f 和块 U_b 相似.其中,阈值 $t4$ 取值由实验得到,当 $t4$ 为3时获得的平均实验结果相对最优.

根据时域相关性,丢失块 U 与 U_b 和 U_f 也相似,因此丢失块 U 中的像素值可由两块像素的加权平均值填补:

$$p_{n, t}(x, y) = \frac{p_{n, t+k}(x-x_{s0}, y-y_{s0}) + p_{n, t-k}(x+x_{s0}, y+y_{s0})}{2} \quad (11)$$

3.2.2 视点间搜索像素填补

若求得的最小差值 $D \geq t4$,则搜索到的 U_b 与 U_f 不相似,说明丢失块 U 的运动是不均匀的.本算法利用视点

间相关性在视点域搜索匹配块进行隐藏.图3所示为非均匀运动示意图,对 U_f 设定一个视点间搜索值 d_f ,在参考视点R的前向参考帧中搜索匹配块 $U_{R,f}$,同理对 U_b 设定搜索值 d_b ,在后向参考帧中寻找匹配块 $U_{R,b}$;然后根据相对最优的视点间搜索值(视差值)计算出丢失块 U 的视点间匹配块 U_R .

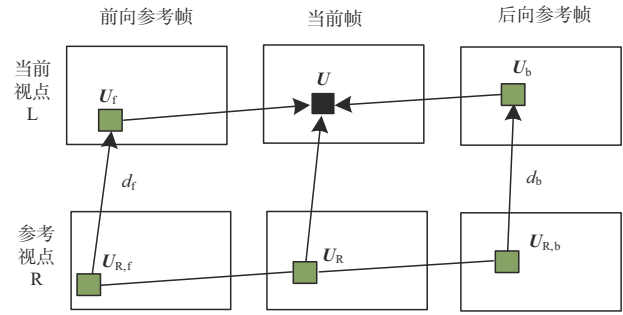


图3 非均匀运动示意图

由于平行排列的多视点相机主要是水平视差,只需在水平方向搜索,因此视点间搜索范围可适当扩大, d_f 和 d_b 的取值范围可设为 $[-64, 64]$.计算块 U_f 和块 $U_{R,f}$ 内像素值的平均绝对差值 D_f :

$$D_f = \frac{1}{N \times N} \sum_{(x, y) \in U} |p_{n, t-k}(x+x_{s0}, y+y_{s0}) - p_{n-1, t-k}(x+x_{s0}+d_f, y+y_{s0})| \quad (12)$$

求得前向视点间最优搜索值 d_{f0} 和 D_f 的最小值 D_{f0} ,即:

$$d_{f0} = \arg \min_{d_f} (D_f) \quad (13)$$

$$D_{f0} = \min_{d_f} (D_f) \quad (14)$$

同理,计算 U_b 与 $U_{R,b}$ 的像素值平均绝对差值 D_b :

$$D_b = \frac{1}{N \times N} \sum_{(x, y) \in U} |p_{n, t+k}(x-x_{s0}, y-y_{s0}) - p_{n-1, t+k}(x-x_{s0}+d_b, y-y_{s0})| \quad (15)$$

记录最优后向搜索值 d_{b0} 和最小值 D_{b0} :

$$d_{b0} = \arg \min_{d_b} (D_b) \quad (16)$$

$$D_{b0} = \min_{d_b} (D_b) \quad (17)$$

若 $D_{b0} \geq D_{f0}$,采用 d_{f0} 做视差补偿得到丢失块 U 的视点间匹配块 U_R ,并用 U_R 中的像素值填充丢失块 U :

$$p_{n, t}(x, y) = p_{n-1, t}(x+d_{f0}, y) \quad (18)$$

否则,采用 d_{b0} 作视差补偿得到匹配块 U_R ,并填充丢失块 U :

$$p_{n, t}(x, y) = p_{n-1, t}(x+d_{b0}, y) \quad (19)$$

3.3 重建高显著等级区域

高显著等级区域是人眼最关注的前景物体或者运动物体的边缘区域,若单纯采用视点间搜索像素填补,

当参考视点前后参考帧的搜索块与当前视点前后参考帧的搜索块相似度不高时,产生的误差会对最终重建帧的主观质量造成较大影响.因此本算法在视点间两块相似时,采用视点间搜索法进行重建;当视点间两块不相似时,进一步提出了一种新的MV选择方法进行运动补偿重建丢失块.

3.3.1 视点间搜索像素填补

若 D_{r0} 和 D_{b0} 都小于视点间块相似阈值 $t5$ 时,则认为视点间两匹配块相似,否则两块不相似,其中 $t5$ 的取值由实验比较获得, $t5$ 取值为2.当视点间匹配块 U_f 与 $U_{r,f}$ 相似, U_b 与 $U_{r,b}$ 相似,那么块 U 与块 U_r 也相似;则采用节3.2.2中的视点间搜索像素填补法重建 U 中的像素.

3.3.2 新型MV运动补偿

若视点间匹配块不相似,先计算丢失块 U 的视点间对应块 U_{RC} 的初始坐标:

$$x_{R,f0} = x_0 + x_{s0} + d_{f0} \quad (20)$$

$$x_{R,b0} = x_0 - x_{s0} + d_{b0} \quad (21)$$

$$x_{R,0} = \frac{x_{R,b0} + x_{R,f0}}{2} \quad (22)$$

其中 x_0 、 $x_{R,f0}$ 、 $x_{R,b0}$ 和 $x_{R,0}$ 分别为块 U 、 $U_{r,f}$ 、 $U_{r,b}$ 和 U_{RC} 左上角像素的水平坐标.

令 $(x_{R,0}, y_{R,0})$ 为块 U_{RC} 的左上角像素坐标,而 U_{RC} 必隶属于参考视点当前帧中 64×64 (CTU大小)的编码单元块 C .设块 C 左上角像素坐标为 (x_n, y_n) ,获取 C 中16个 4×4 块(HEVC最小处理单元)的MV并构建候选运动矢量集,记为 u_1, u_2, \dots, u_{16} ,如图4所示(蓝色区域表示块 U_{RC} ,灰色区域表示编码单元块 C ,标号1至16的小块为所选 4×4 块).将 32×32 大小的块 U_{RC} 划分为4个 16×16 的子块,记为块 a, b, c 和 d .虽然存在 $y_{R,0} = y_n$ 和 $y_{R,0} = y_n + 32$ 两种情况,但两种情况的处理方法相似,下面仅以 $y_{R,0} = y_n$ 为例说明.

首先,当 $x_{R,0} = x_n$ 时,根据HEVC中运动矢量的选取

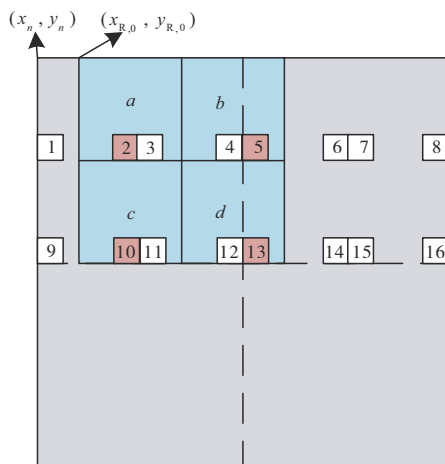


图4 U_{RC} 与 C 关系示意图

方法,块 a, b, c, d 应分别采用2、4、10、12号块的运动矢量 u_2, u_4, u_{10}, u_{12} 进行重建;当 $x_{R,0} = x_n + 32$ 时,采取6、8、14、16号块中的运动矢量 u_6, u_8, u_{14}, u_{16} 进行重建.考虑到 U_{RC} 是通过估计得到,其位置会在块 C 中水平移动,还需分四种情况进行运动补偿:

(1) $0 < x_{R,0} - x_n < 16$: 块 a, b, c, d 分别选用2、5、10、13号块的运动矢量 u_2, u_5, u_{10}, u_{13} 进行运动补偿.

(2) $16 \leq x_{R,0} - x_n < 32$: 块 a, b, c, d 分别选用4、7、12、15号块中的运动矢量 u_4, u_7, u_{12}, u_{15} 进行运动补偿.

(3) $32 \leq x_{R,0} - x_n < 48$: 块 a, b, c, d 分别选用6号、14号的运动矢量 u_6, u_{14} 和右相邻 C 块中1号、9号位置块的运动矢量 u_1, u_9 进行运动补偿.

(4) $48 \leq x_{R,0} - x_n < 64$: 块 a, b, c, d 分别选用8号、16号块中的运动矢量 u_8, u_{16} ,以及右相邻的 C 块中3号、11号块的运动矢量 u_3, u_{11} 进行补偿.

其中,运动补偿采用3D-HEVC中基于亚像素插值的运动补偿法,对 U_{RC} 的四个子块分别进行运动补偿后得到 U_r ,并用 U_r 中的像素值填补丢失块 U .

4 实验结果与分析

4.1 实验条件设置

为了客观评价算法性能,本文算法被嵌入至HTM16.1平台,测试数据集为Lovebird1、PoznanStreet、BookArrival、Newspaper、Kendo和Balloons这六种不同类型的MVD视频序列,其中PoznanStreet的分辨率为 1920×1088 px,其余序列为 1024×768 px.主机配置为: Intel Corei5-3407, 3.20 GHz处理器, 4.00 GB内存.实验环境设置如下:(1)各测试序列采用HBP预测结构进行编码,视频编码组(Group Of Picture, GOP)长度设置为8;(2)量化参数(Quantization Parameter, QP)设置为25、28和30;编码帧率为25 f/s,每个序列共测试了97帧;(3)实验对比采用逐帧比较的方法.在实验中,考虑到I帧是每个GOP的起始帧,也是视频编码中的关键帧,在传输过程中采用高等级不等错保护,因此假设I帧不出错,非关键帧发生整帧丢失.

基于HBP编码结构的3D-HEVC视频流整帧丢失错误隐藏算法尚未见报道,本文选用了四种典型的3D视频错误隐藏算法进行性能比较.(1) Jeon^[22]:将Jeon等提出的基于HEVC标准的整帧丢失错误隐藏算法扩展到多视点视频错误隐藏中,把参考视点的MV映射到当前视点图像块上;(2) Song^[23]:将Song等提出的基于H.264标准的多视点错误隐藏方法延伸到HEVC标准下的整帧丢失错误隐藏中,设计时域搜索、视点间搜索、多假设错误隐藏三种方法,最后根据模式选择最优准则自适应恢复;(3) Liu^[24]:一种结合3D-warping和运动矢量共享的整帧丢失错误隐藏方法,将该算法从

IPPP 编码结构移植到 HBP 编码结构; (4) Lin^[13]: 一种结合深度差异信息的彩色帧整帧丢失错误隐藏算法, 将它移植到 3D-HEVC 平台上进行比较.

4.2 客观质量比较与分析

本文采用 PSNR 和 SSIM 作为各算法重建图像客观质量的评估标准. 表 1 统计了 QP 值分别为 25、28 和 30 时, 对各测试序列使用不同的错误隐藏算法后重建 B 帧图像的平均 PSNR 值. 图 5 所示为不同算法对 PoznanStreet 第 4 视点序列重建 PSNR 值的逐帧比较折线图.

表 1 经不同错误隐藏方法后的 PSNR 值

| QP | 序列 | PSNR(dB) | | | | | |
|----|-----------|----------|-------|-------|-------|-------|--------------|
| | | 无错 | Jeon | Song | Liu | Lin | 本文 |
| 25 | Lovebird1 | 39.67 | 35.03 | 34.58 | 35.21 | 34.83 | 35.49 |
| | PoznanS | 38.95 | 31.98 | 30.61 | 32.47 | 31.57 | 34.24 |
| | BookArri | 39.89 | 30.06 | 28.54 | 30.71 | 30.07 | 32.67 |
| | Newspaper | 40.48 | 34.35 | 33.91 | 34.67 | 34.32 | 34.95 |
| | Kendo | 44.59 | 32.95 | 31.27 | 33.64 | 32.68 | 35.13 |
| | Balloons | 42.52 | 36.13 | 35.76 | 36.21 | 35.99 | 36.42 |
| 28 | Lovebird1 | 38.56 | 34.41 | 34.06 | 34.56 | 34.24 | 34.79 |
| | PoznanS | 38.17 | 31.56 | 30.22 | 32.01 | 31.12 | 33.74 |
| | BookArri | 39.14 | 29.65 | 28.18 | 30.32 | 29.68 | 32.23 |
| | Newspaper | 39.58 | 33.87 | 33.50 | 34.16 | 33.85 | 34.39 |
| | Kendo | 43.76 | 32.51 | 30.86 | 33.17 | 32.25 | 34.62 |
| | Balloons | 41.50 | 35.61 | 35.33 | 35.66 | 35.49 | 35.79 |
| 30 | Lovebird1 | 37.69 | 33.90 | 33.63 | 34.02 | 33.71 | 34.19 |
| | PoznanS | 37.46 | 31.19 | 29.91 | 31.63 | 30.74 | 33.28 |
| | BookArri | 38.52 | 29.28 | 27.86 | 29.96 | 29.34 | 31.83 |
| | Newspaper | 38.74 | 33.45 | 33.19 | 33.72 | 33.47 | 33.93 |
| | Kendo | 43.02 | 32.10 | 30.54 | 32.76 | 31.86 | 34.16 |
| | Balloons | 40.63 | 35.12 | 35.01 | 35.16 | 35.04 | 35.22 |

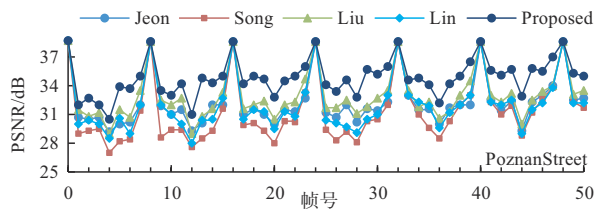


图 5 PoznanStreet 第 4 视点某 GOP 中间帧丢失逐帧折线图

从表 1 和图 5 可以看出, 本文算法对不同序列丢失帧的重建效果都有明显提升: 在 QP=25 时, 综合 6 个测试序列, 本文算法与 Jeon 算法、Song 算法、Liu 算法和 Lin 算法相比, 平均 PSNR 值能分别提高 1.41 dB、2.37 dB、0.99 dB 和 1.57 dB. 对于运动较为剧烈的多视点序列, 如 PoznanStreet, 本文方法相比以上四种算法平均 PSNR 值能分别提升 2.26 dB、3.63 dB、1.77 dB 和 2.67 dB; 对于运动不剧烈的 Balloons 和 Newspaper 序列,

本文方法相比 Jeon 算法、Song 算法、Liu 算法和 Lin 算法, 平均 PSNR 值能分别提高 0.29 和 0.60 dB、0.66 和 1.04 dB、0.21 和 0.28 dB、0.43 和 0.63 dB.

在 QP=28 时, 综合 6 个测试序列, 本文算法与 Jeon 算法、Song 算法、Liu 算法和 Lin 算法相比, 平均 PSNR 值能分别提高 1.32 dB、2.23 dB、0.95 dB 和 1.49 dB. 在 QP=30 时, 综合 6 个测试序列, 本文算法与 Jeon 算法、Song 算法、Liu 算法和 Lin 算法相比, 平均 PSNR 值能分别提高 1.26 dB、2.08 dB、0.89 dB 和 1.41 dB.

同时, 重建 SSIM 值也有着较为显著的提升(如表 2 所示). 综合 6 个测试序列, 在 QP=25 时, 本文算法与 Jeon 算法、Song 算法、Liu 算法和 Lin 算法相比, 平均 SSIM 值分别能提高 0.0075、0.0122、0.0053 和 0.0080. 在 QP=28 时, 本文算法与上述四种算法相比, 平均 SSIM 值能分别提高 0.0071、0.0118、0.0052 和 0.0077. 在 QP=30 时, 本文算法与 Jeon 算法、Song 算法、Liu 算法和 Lin 算法相比, 平均 SSIM 值分别提高 0.0068、0.0112、0.0049 和 0.0073.

表 2 经不同错误隐藏方法后的 SSIM 值

| QP | 序列 | SSIM | | | | | |
|----|-----------|-------|-------|-------|-------|-------|--------------|
| | | 无错 | Jeon | Song | Liu | Lin | 本文 |
| 25 | Lovebird1 | 0.969 | 0.955 | 0.954 | 0.955 | 0.954 | 0.956 |
| | PoznanS | 0.944 | 0.904 | 0.897 | 0.907 | 0.902 | 0.918 |
| | BookArri | 0.954 | 0.893 | 0.883 | 0.897 | 0.893 | 0.909 |
| | Newspaper | 0.967 | 0.946 | 0.944 | 0.947 | 0.946 | 0.948 |
| | Kendo | 0.968 | 0.911 | 0.903 | 0.915 | 0.910 | 0.922 |
| | Balloons | 0.974 | 0.957 | 0.956 | 0.957 | 0.957 | 0.958 |
| 28 | Lovebird1 | 0.965 | 0.953 | 0.952 | 0.954 | 0.953 | 0.954 |
| | PoznanStr | 0.940 | 0.902 | 0.895 | 0.905 | 0.899 | 0.915 |
| | BookArri | 0.949 | 0.890 | 0.881 | 0.894 | 0.891 | 0.906 |
| | Newspaper | 0.963 | 0.944 | 0.943 | 0.945 | 0.944 | 0.946 |
| | Kendo | 0.964 | 0.909 | 0.901 | 0.912 | 0.908 | 0.919 |
| | Balloons | 0.971 | 0.956 | 0.955 | 0.956 | 0.955 | 0.956 |
| 30 | Lovebird1 | 0.963 | 0.952 | 0.950 | 0.952 | 0.951 | 0.953 |
| | PoznanStr | 0.938 | 0.900 | 0.893 | 0.902 | 0.898 | 0.912 |
| | BookArri | 0.945 | 0.888 | 0.879 | 0.892 | 0.889 | 0.904 |
| | Newspaper | 0.960 | 0.943 | 0.942 | 0.944 | 0.943 | 0.945 |
| | Kendo | 0.961 | 0.908 | 0.899 | 0.911 | 0.907 | 0.917 |
| | Balloons | 0.968 | 0.955 | 0.954 | 0.955 | 0.954 | 0.955 |

本文算法相比对比算法重建效果提升明显. 这是由于 Jeon 算法是将参考视点当前帧信息映射到当前视点丢失帧中进行重建, 但在 HBP 编码结构下运动物体的运动信息估计不够准确, 映射关系会存在误差. Song 算法由于在 HBP 编码结构下视频的播放顺序与解码顺序不同, 时域上前后参考帧有时相隔较远或在视点域上搜索误差过大, 导致难以搜索到匹配的像素值, 其重

建效果也不理想. Liu所提算法采用3D-warping与运动矢量共享相结合的方法,但有时通过共享获取的运动矢量并不准确,其恢复结果存在一定的失真. Lin算法是面向IPPP预测结构的,由于新标准下的HBP编码结构是采用分层双向预测的,连续帧的运动矢量并不相似,因此在HBP编码结构下不能取得良好的效果. 本文算法相比于Jeon算法,它认为运动物体及其边缘具有高视觉显著性,更加注重运动物体边缘的恢复;相比于Song算法,本算法优化了时域和视点域的像素搜索,并针对两种搜索方法都不适用的情况,提出了新的MV补偿方法;相比于Liu算法和Lin算法,本算法不使用3D-warping也获得了良好的重建质量.

在差错传播方面上,本算法的性能也优于其他三种对比算法. 图6所示为GOP中间帧丢失的差错传播图, Liu算法基于3D-warping,参考帧的误差对整个GOP的影响相对较小. 本文算法不仅利用了参考视点当前时刻的视频帧,而且用到了参考视点前后向参考帧,重建图像质量最佳,因此整个GOP受到的差错影响最小.

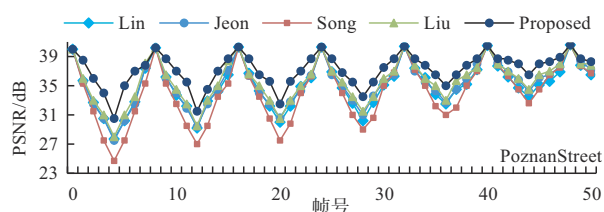


图6 PoznanStreet第4视点GOP中间帧丢失差错传播图

4.3 主观质量比较与分析

图7所示PoznanStreet序列第4视点第11帧整帧丢失时通过不同错误隐藏算法得到的重建图及其部分区域的细节放大图. Jeon算法重建结果及其区域放大如图7(b)和7(g)所示,对于时域跨度较大的运动物体,时域参考帧中视点间映射关系不一定是当前帧的视点映射关系,某些运动区域会被映射入误匹配信息,导致部分重建区域严重块失真. Song算法是通过前后参考帧的匹配像素SAD值搜索重建像素,但时域搜索的准确性受到丢失帧与参考帧的参考距离约束,跨度越大搜索所得的像素值越不准确,导致重建效果不理想,如图7(c)和7(h)所示. Liu算法对车尾后的背景区域恢复效果较差,这是由于车尾部分是3D-warping产生的空洞区域,采用MV共享法进行重建, MV获取不准确导致车尾后的背景区域被车尾区域填补. Lin算法是基于IPPP结构实现的,移植到HBP结构中后重建的丢失块运动矢量并不准确,导致有些运动区域重建模糊. 本文算法车尾和轮胎的边缘区采用视点像素搜索以及基于视点域的MV补偿,重建主观质量优于其他四种算法,同时图像其它区域也取得了较佳的重建效果,如图7(f)和7

(k)所示.

图8表示当Newspaper序列发生整帧丢失时采用不同算法的重建图. 该帧运动状态较为静止,主要比较人手与杯子形成的区域. 本文算法重建效果优于其他算法,如图8(b)~(f)所示, Jeon算法中手与杯子采用了相邻块的运动矢量从而造成了手指的部分重叠, Song算法中杯子附近人衣处的像素采用了杯子像素与衣服像素的加权值, Liu算法由于遮挡关系在warping过程中会产生细微的伪影,其手与杯子部分比较模糊, Lin算法由于其运动矢量估计并不准确,其移动的手和杯子部分比较模糊.

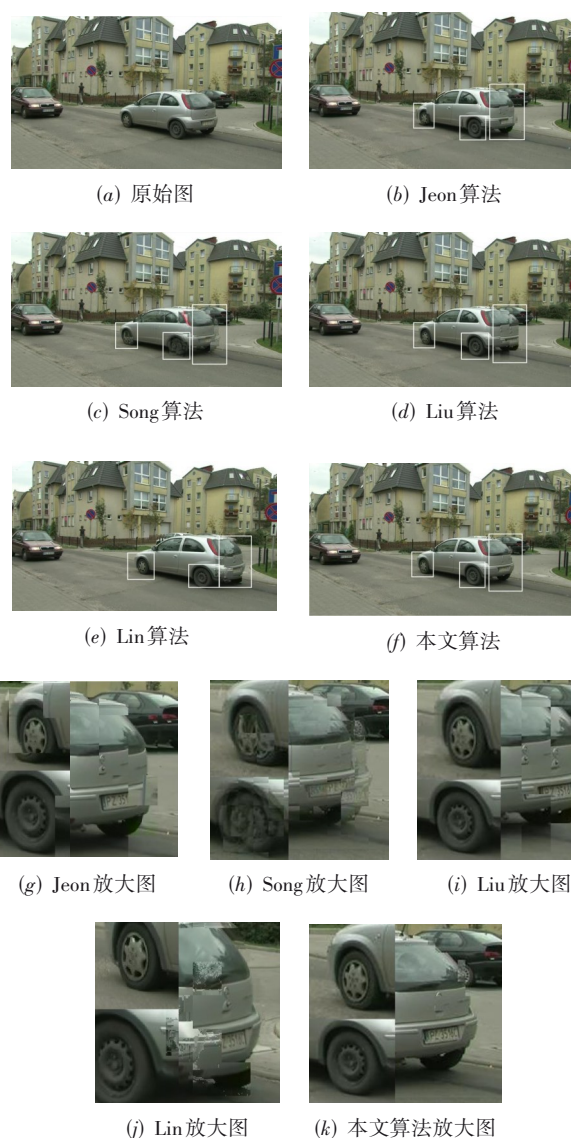


图7 PoznanStreet视点4第11帧各算法重建结果及其放大图

4.4 算法复杂度

本文通过各测试序列在不同算法下的平均解码时间来评价各算法的复杂度,计算方法如下:



图8 Newspaper第4视点第33帧各算法重建效果图

$$T = \frac{1}{N} \times \sum_{i=1}^N T_i \quad (23)$$

其中 N 为实验一共加入的丢失帧数目, T_i 表示解码第 i 个丢失帧所需要的时间. 测试结果如表 3 所示, Liu 算法和 Lin 算法都使用了 3D-warping, 绘制时间比普通解码时间长; Jeon 算法对每个图像块都采取视点映射方法填补, 因此相比 Song 算法, Jeon 算法的解码时间稍长; 本文算法与 Song 的解码时间相近, 但是综合考虑重建图像的主观质量与客观质量, 本文算法的性能是最优的.

表3 各算法重建单帧平均所需的解码时间

| 序列 | Jeon | Song | Liu | Lin | 本文 |
|--------------|-------|-------|-------|-------|-------|
| Lovebird1 | 5.403 | 5.271 | 5.731 | 5.815 | 5.384 |
| PoznanStreet | 6.141 | 5.978 | 7.632 | 7.751 | 6.041 |
| BookArrival | 5.612 | 5.403 | 5.738 | 5.887 | 5.628 |
| Newspaper | 5.584 | 5.398 | 5.742 | 5.845 | 5.559 |
| Kendo | 5.286 | 5.139 | 5.756 | 5.922 | 5.304 |
| Balloons | 5.247 | 5.054 | 5.703 | 5.764 | 5.226 |

5 结束语

针对 MVD 视频流在传输中发生彩色图整帧丢失的问题, 提出了一种基于 3D-HEVC 编码标准的视频整帧丢失错误隐藏算法. 该算法在隐藏过程中结合了人类视觉的感知特性, 将丢失帧各区域划分为不同的显著

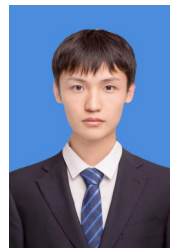
等级, 进而提出相适应的隐藏方法进行各区域重建, 实验结果表明本文算法取得了良好的性能, 可应用于基于 3D-HEVC 编码标准的 MVD 视频流错误隐藏中.

参考文献

- [1] WANG T, LI F, QIAO X, et al. Low-complexity error resilient HEVC video coding: a deep learning approach[J]. IEEE Transactions on Image Processing, 2021, 30(2): 1245-1260.
- [2] XIANG X, CHENG Y. A frame loss error concealment scheme for multi-view video coding[C]//Proceedings of 2014 International Conference on Instrumentation, Measurement, Computer, Communication and Control. Harbin, China: IEEE Computer Society, 2014: 857-861.
- [3] KUAN Y K, LI G L, CHEN M J, et al. Error concealment algorithm using inter-view correlation for multi-view video [J]. EURASIP Journal on Image and Video Processing, 2014, 38(1): 1-11.
- [4] ZHOU Y, XIANG W, WANG G. Frame loss concealment for multiview video transmission over wireless multimedia sensor networks[J]. IEEE Sensors Journal, 2015, 15(3): 1892-1901.
- [5] YU H, RAN M, LI P, et al. A new hybrid error concealment for multi-view video[C]//Proceedings of 2015 IEEE China Summit and International Conference on Signal and Information Processing. Chengdu, China: IEEE, 2015: 176-180.
- [6] AHAMED Z G, SHARMA V K, JAIN A. Image error concealment using robust data hiding technique[C]//Proceedings of 2020 International Conference on Innovative Mechanisms for Industry Applications. Bangalore, India: IEEE, 2020: 470-474.
- [7] EL-SHAFI W, EL-RABAIE S, EL-HALAWANY M M, et al. Recursive Bayesian filtering-based error concealment scheme for 3D video communication over severely lossy wireless channels[J]. Circuits, Systems, and Signal Processing, 2018, 37(11): 4810-4841.
- [8] YANG M, LAN X, ZHENG N, et al. Depth-assisted temporal error concealment for Intra frame slices in 3-D video [J]. IEEE Transactions on Broadcasting, 2014, 60(2): 385-393.
- [9] EL-SHAFI W, EL-RABAIE S, EL-HALAWANY M, et al. Encoder-independent decoder-dependent depth-assisted error concealment algorithm for wireless 3D video communication[J]. Multimedia Tools and Applications, 2018, 77(11): 13145-13172.

- [10] HUANG P C, LIN J R, LI G L, et al. Improved depth-assisted error concealment algorithm for 3D video transmission[J]. IEEE Transactions on Multimedia, 2017, 19(11): 2625-2632.
- [11] TAI S C, WANG C C, HONG C S, et al. An efficient full frame algorithm for object-based error concealment in 3D depth-based video[J]. Multimedia Tools and Applications, 2016, 75(16): 9927-9947.
- [12] KHATTAK S, MAUGEY T, HAMZAOUY R, et al. Temporal and inter-view consistent error concealment technique for multiview plus depth video[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(5): 829-840.
- [13] LIN T L, WANG C J, DING T L, et al. Recovery of lost color and depth frames in multiview videos[J]. IEEE Transactions on Image Processing, 2018, 27(11): 5449-5463.
- [14] LEE P J, KUO K T. An adaptive error concealment method for depth map in 3D video coding[J]. International Journal of Fuzzy Systems, 2016, 18(1): 62-71.
- [15] ASSUNCAO P A, MARCELINO S, SOARES S, et al. Spatial error concealment for intra-coded depth maps in multiview video-plus-depth[J]. Multimedia Tools and Applications, 2017, 76(12): 13835-13858.
- [16] LEE G W, HAN J K. Hole concealment for depth image using pixel classification in multiview system[C]//Proceedings of 2021 IEEE International Conference on Consumer Electronics. Las Vegas, USA: IEEE Computer Society, 2021: 1-5.
- [17] ZHANG X, ZHAO Y, LIN C, et al. Warping-driven mode selection for depth error concealment[C]//Proceedings of 2014 IEEE Global Conference on Signal and Information Processing. Atlanta, USA: IEEE, 2014: 302-306.
- [18] 丁颖,刘延伟,等. 虚拟现实全景图像显著性检测研究进展综述[J]. 电子学报, 2019, 47(7): 1575-1583.
DING Ying, LIU Yan-wei, et al. An overview of research progress on saliency detection of panoramic VR images [J]. Acta Electronica Sinica, 2019, 47(7): 1575-1583. (in Chinese)
- [19] CONG R, LEI J, FU H, et al. Going from RGB to RGBD saliency: a depth-guided transformation model[J]. IEEE Transactions on Cybernetics, 2020, 50(8): 3627-3639.
- [20] CHENG M, MITRA N J, HUANG X, et al. Global contrast based salient region detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 37(3): 409-416.
- [21] WANG J, DASILVA M P, LECALLET P, et al. Computational model of stereoscopic 3D visual saliency[J]. IEEE Transactions on Image Processing, 2013, 22(6): 2151-2165.
- [22] JEON B, HUU T, CANH T N. A frame loss concealment for spatial scalable HEVC using base layer motion[C]//Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. Jeju, Korea: IEEE, 2019: 1-4.
- [23] SONG K, CHUNG T, OH Y, et al. Error concealment of multi-view video sequences using inter-view and intra-view correlations[J]. Journal of Visual Communication and Image Representation, 2009, 20(4): 281-292.
- [24] LIU X, PENG Q, FAN X. Frame loss concealment for multi-view video plus depth[C]//Proceedings of the 15th IEEE International Symposium on Consumer Electronics. Singapore: IEEE, 2011: 208-211.

作者简介



崔金鹏 男, 1996年9月出生于山东省潍坊市. 现为杭州电子科技大学通信工程学院硕士生. 主要研究方向为3D视频编解码与差错控制.
E-mail: reader@163.com



周洋(通讯作者) 男, 1979年1月出生于浙江省台州市. 现为杭州电子科技大学通信工程学院副教授, 硕士生导师, 主要研究方向为3D视频编解码、视频图像处理.
E-mail: zhouyang@hdu.edu.cn

殷海兵 男, 1974年6月出生于湖北黄梅. 教授, 博士生导师. 主要研究方向为智能信息处理.

E-mail: yhb@hdu.edu.cn

黄晓峰 男, 1988年1月出生于江西婺源. 主要研究方向为视频芯片架构.

E-mail: xhuang@hdu.edu.cn

陆宇 男, 1977年5月出生于广西博白. 主要研究方向视频编解码.

E-mail: y_lu@hdu.edu.cn