

面向监控视频内容自适应的通用视频编码 量化参数级联

公衍超^{1,2,3}, 张森森¹, 何书婷¹, 李宝贵¹, 刘颖^{1,2,3}, 林庆帆^{1,4}

(1. 西安邮电大学通信与信息工程学院, 710121, 西安; 2. 陕西省法庭科学电子信息实验研究中心, 710121, 西安; 3. 陕西省无线通信与信息处理技术国际联合研究中心, 710121, 西安;
4. 新加坡 XsecPro 公司, 787820, 新加坡)

摘要: 最新通用视频编码(VVC)标准测试模型 VTM 推荐的 JCTVC-X0038 量化参数级联方法未考虑监控视频的内容特性, 编码效率低。针对这一问题, 提出了一种有效考虑监控视频空时域内容特性的 VVC 量化参数级联方法(SCA-QPC)。基于 VVC 混合编码框架原理, 理论分析得出时间层及监控视频的空时域内容特性是影响量化参数(QP)选择的关键因素, 设计编码实验对理论分析结果进行验证; 采用梯度和背景帧差分别评测监控视频的空域和时域特性, 将梯度和背景帧差作为反向传播神经网络的输入, 构建预测低时延结构时间层对应最优 QP 的模型; 提出适用于 VVC 标准有效提高监控视频编码效率的量化参数级联方法。与当前最先进的 JCTVC-X0038 量化参数级联方法相比, 对于所有测试序列, SCA-QPC 对应的平均 BD-rate 为 -4.26%。研究对于 VVC 编码框架中量化参数级联和码率控制模块的优化具有指导意义。

关键词: 视频编码; 通用视频编码; 量化参数级联; 监控视频; 视频内容特性

中图分类号: TN911.7 **文献标志码:** A

DOI: 10.7652/xjtuxb202312004 **文章编号:** 0253-987X(2023)12-0030-11

Content Adaptive Quantization Parameter Cascading for Surveillance Video Coding Using Versatile Video Coding

GONG Yanchao^{1,2,3}, ZHANG Sensen¹, HE Shuting¹, LI Baogui¹,
LIU Ying^{1,2,3}, LIM Keng-Pang^{1,4}

(1. School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China;
2. Electronic Information Experimental Research Center for Forensic Science of Shaanxi Province, Xi'an 710121, China;
3. International Joint Research Center for Wireless Communication and Information Processing Technology of Shaanxi Province, Xi'an 710121, China; 4. XsecPro Pte. Ltd., 787820, Singapore)

Abstract: The JCTVC-X0038 quantization parameter cascading (QPC) method recommended by the latest versatile video coding (VVC) standard test model VTM has low coding efficiency because the surveillance video content characteristics are not considered. For this problem, a VVC QPC method considering the spatio-temporal content characteristics of the surveillance videos, referred to as SCA-QPC, was proposed. First, based on the VVC hybrid coding framework principle, it was found that the temporal layer and spatio-temporal content characteristics of the surveillance videos were the key factors affecting the selection of quantization parameters (QP) by

收稿日期: 2023-05-30。 作者简介: 公衍超(1986—), 男, 副教授, 硕士生导师。 基金项目: 国家自然科学基金资助项目(62277036, 61801381); 西安邮电大学研究生创新基金资助项目(CXJJZL2022005)。

网络出版时间: 2023-08-16

网络出版地址: <https://kns.cnki.net/kcms/detail/61.1069.T.20230815.1905.002.html>

theoretical analysis. And the coding experiments were further carried out to verify the results of the theoretical analysis. Then the gradient and background frame difference were used to measure the spatio-temporal characteristics of the surveillance videos respectively and further used as the inputs to the back propagation neural network to construct a model predicting the optimal QPs corresponding to temporal layers in low delay structure. Finally, the SCA-QPC suitable for VVC standard was proposed to effectively improve the surveillance video coding efficiency. Compared with the state-of-the-art JCTVC-X0038 QPC method, the average BD-rate of SCA-QPC is -4.26% for all test sequences. This study can guide the optimization of QPC and rate control modules in the VVC coding framework.

Keywords: video coding; versatile video coding; quantization parameter cascading; surveillance video; video content characteristics

视频监控系统被广泛应用于刑事侦查、司法取证等领域,有力维护了社会稳定与人民的人身与财产安全。在视频监控系统中,监控视频要依次经过采集、编码、传输、存储、解码、显示、分析等操作。监控视频编码的目的就是用更少的编码码率获得更高质量的重建视频,所以编码是保证整个视频监控系统有效运行的关键技术之一。当前视频监控系统中,被广泛使用的视频编码标准是由国际电信联盟电信标准化部门 (ITU-T) 和国际标准化组织 (ISO)/国际电工委员会 (IEC) 分别于 2003 年和 2013 年共同发布的 H. 264/AVC 标准^[1] 和 HEVC 标准^[2-3]。但是,随着近几年视频类型的多样化发展和视频应用需求的快速变化,H. 264/AVC 和 HEVC 标准的编码效率已经很难满足实际需求。相应地,由 ITU-T 和 ISO/IEC 联合制定的通用视频编码 (VVC) 标准^[4] 应运而生。VVC 是 2020 年发布的当前最新国际视频编码标准,其编码效率大约是 HEVC 标准的 2 倍,H. 264/AVC 标准的 4 倍。

H. 264/AVC、HEVC 和 VVC 标准都采用基于预测、变换、量化、熵编码的混合编码框架。其中,量化是将具有较大动态取值范围的变换系数进行多对一映射以减小变换系数的取值空间,在有效去除视频部分视觉冗余的同时,提高熵编码去除熵冗余的能力,有力保证了视频编码的效率。量化的核心是多对一的映射机制,而量化参数 (QP) 是控制多对一映射程度的编码参数。QP 通常取 0 及有限的正整数。QP 取值越大则对应更大取值范围的变换系数被映射为同一取值,导致视频编码的失真越大,而编码码率则越小。所以,QP 取值显著影响视频编码的效率,是量化模块中一个非常关键的编码参数。量化参数级联 (QPC) 即是研究在编码过程中如何选择最

优 QP 的一类方法。H. 264/AVC 和 HEVC 标准的官方测试模型 JM 和 HM 分别推荐文献 [5-6] 提出的 QPC 方法。但是,文献 [5-6] 提出的 QPC 方法都属于固定取值的方法,即不管编码输入因素如何变化,相邻时间层对应的 QP 差值是一样的。固定取值的方法虽然简单但是编码效率较低。为了进一步提高视频编码的效率,一些更加有效的 QPC 方法被提出^[7-14]。

基于实验统计,文献 [7] 构建了 QP 与拉格朗日因子的对数关系,并将其应用于 QP 的选择过程中。文献 [8-9] 探索了分级预测结构中不同时间层之间码率与失真的依赖关系,并分别提出了基于率失真优化的 QPC 方法。不同的是,文献 [8] 探索的是相邻时间层之间的码率与失真的依赖关系,而文献 [9] 探索的是时间 0 层与其他更高时间层之间的码率与失真的依赖关系。文献 [10-14] 分析了分级预测结构中图像间的参考依赖性对 QP 选择的影响,并分别提出了有效的 QPC 方法。不同的是,文献 [10-13] 只考虑了最近参考帧对 QP 选择的影响,而文献 [14] 则考虑了所有参考帧对 QP 选择的影响。

以上提出的 QPC 方法^[5-14] 都是面向早期的 H. 264/AVC 和 HEVC 标准。由于 VVC 标准推荐使用了大量新编码技术,所以这些 QPC 方法^[5-14] 并不适用。当前 VVC 标准的官方测试模型 VTM^[15] 推荐使用 JCTVC-X0038 提案提出的 QPC 方法^[16], 下文简称为 JCTVC-X0038 QPC 方法。JCTVC-X0038 QPC 方法忽略了视频内容特性对 QP 选择的重要影响,其编码效率还有较大的提升空间。针对以上问题,本文面向 VVC 的混合编码框架,在分析视频内容特性等因素对 QP 选择影响的基础上,

提出了一种有效考虑监控视频空时域内容特性的 VVC 量化参数级联方法(SCA-QPC)。

1 JCTVC-X0038 QPC 方法

在视频编码中,需要根据应用场景需求选择适合的编码结构。VTM 支持 3 种编码结构,即全帧内、随机接入、低时延(LD)^[15]。图 1 为图像组(GOP)为 8 的 LD 结构。图像序列号(POC)反映了图像在视频中的显示顺序,而编码顺序(CO)则反映了图像在视频编码中的顺序。在 LD 结构中,所有图像的 POC 和 CO 是一样的,即编码结构时延为 0。所以,LD 结构适用于视频监控、视频会议、视频通话等对视频数据传输实时性要求较高的应用场景。

在 LD 结构中,视频的第一幅图像被编码为瞬时解码刷新(IDR)帧,剩余图像则根据其 POC 组成 GOP。处在同一 GOP 的图像又根据其相对图像序列号(rPOC)被划分到不同的时间层。rPOC 反映了图像在 GOP 中的显示顺序。时间层是介于 GOP 和图像之间的一个编码层级,可以有效反映图像在 GOP 中的重要性。图 1 中由红色、粉色、蓝色、绿

色、黄色标识的图像分别处于时间 0 层、时间 1 层、时间 2 层、时间 3 层、时间 4 层。混合编码框架中的帧间编码帧,例如通常的 P 帧或 B 帧,需要利用帧间预测技术参考已经编码重建的图像以消除其包含的时域冗余。图 1 中的箭头方向表示参考图像(RP)的来源。通常,RP 的数量越多,则对应的编码率失真性能越高,但编码复杂度也会显著增大。为了有效权衡编码率失真性能和复杂度,VTM 允许支持的最大 RP 数量为 4。图 1 中的 RP1、RP2、RP3、RP4 分别表示第 1~4 个帧间参考帧的来源。通常,时间层越低的图像对整个视频编码的效率影响越大,反映在 QP 选择上就是时间层越低的图像用更小的 QP 编码,以优先保证编码质量。这也是当前大部分 QPC 方法^[5-14,16]在选择 QP 时遵循的基本规则。以上规则可表示为

$$Q_t = Q_{t-1} + \Delta_t, 1 \leq t \leq N_L \tag{1}$$

式中: Q_t 表示给第 t 个时间层图像选择的量化参数; Q_0 表示初始量化参数,通常由编码者在配置文件中初始设定; N_L 表示时间层数; Δ_t 表示第 t 个时间层图像和第 $t-1$ 个时间层图像对应的量化参数差值。

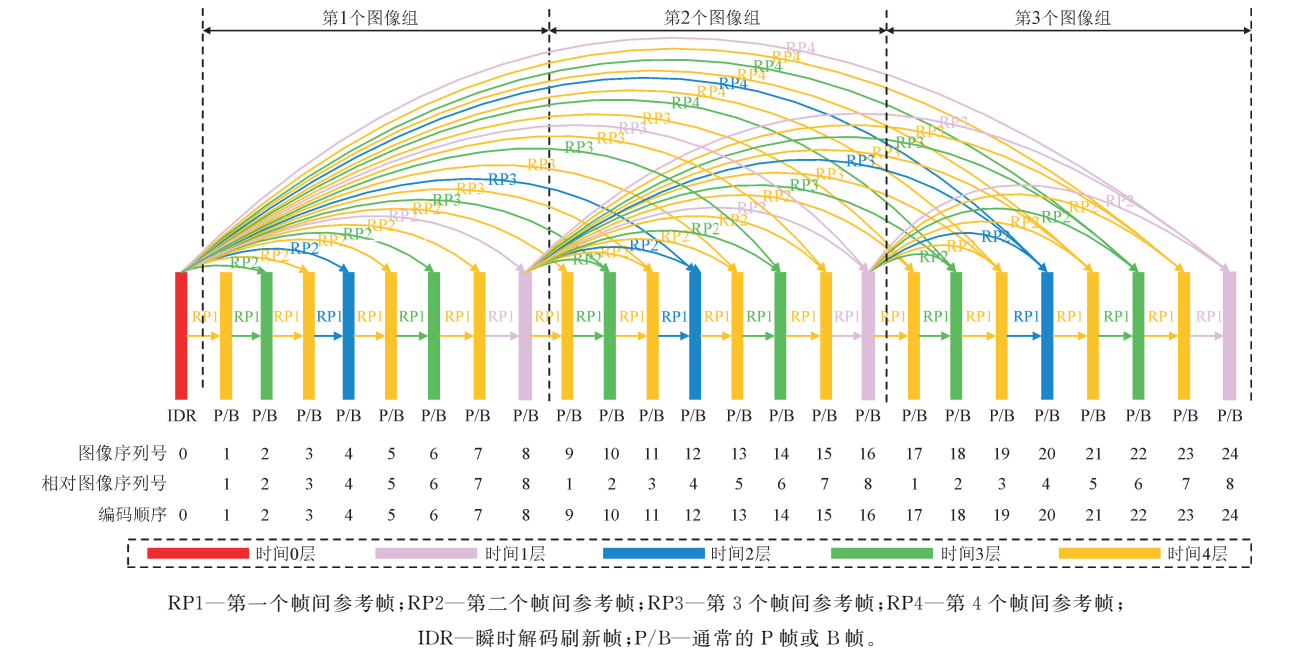


图 1 VTM 中的低时延结构
Fig. 1 Low delay structure in VTM

QPC 方法的目的是选择最优的 Δ_t , 用 $\Delta_{t,opt}$ 表示。JCTVC-X0038 QPC 方法是当前唯一面向 VVC 标准提出的 QPC 方法,其设置 QP 的方案也如式(1)所示,其中 Δ_t 被进一步表达为包括 Δ_a 、 ω 、 β 的线性模型。JCTVC-X0038 QPC 方法的 QP 设置

方案表示为

$$Q_t = Q_{t-1} + \Delta_t = Q_{t-1} + (\Delta_a + \text{Clip3}(0, 3, \omega(Q_{t-1} + \Delta_a) + \beta)) \tag{2}$$

式中: $\text{Clip3}(\gamma, \varphi, \theta)$ 表示将 θ 的取值限定在 $\gamma \sim \varphi$

范围内; ω 为缩放因子; β 为截距因子; Δ_e 为 JCTVC-X0038 QPC 方法设置的相邻时间层量化参数偏移量的差值。

Δ_e 、 ω 、 β 的取值如表 1 所示,其取值只与时间层相关,与视频内容特性等因素无关。

表 1 JCTVC-X0038 QPC 方法中的参数取值
Table 1 Parameter values in JCTVC-X0038 QPC method

时间层	Δ_e	ω	β
时间 1 层	1	0	0
时间 2 层	3	0.259	-6.5
时间 3 层	0	0.259	-6.5
时间 4 层	1	0.259	-6.5

下面,通过实验进一步确认 JCTVC-X0038 QPC 方法设定的具体 QP。选择两个典型的具有不同内容特性的监控视频 DormitoryHall 和 MeetingRoom 作为测试序列。图 2 所示为 DormitoryHall 序列和 MeetingRoom 序列的第 1 帧及第 250 帧。DormitoryHall 序列是由固定监控摄像头拍摄的宿舍走廊场景,其变化区域由绿色虚线框标注。MeetingRoom 序列是由旋转监控摄像头拍摄的会议室场景,其明显变化区域由红色虚线框标注。其中,A 区域和 B 区域分别为摄像头在旋转拍摄过程中视频消失的区域和新进入的区域。结合图 2 易得出以下结论,相比于 DormitoryHall 序列,MeetingRoom 序列包含的场景内容纹理更加复杂,且变化区域更大,序列整体运动更快。这两个测试序列的详细视频内容特性量化评价参见后文第 3.1 小节。使用的编码器为 VTM16.2^[15],其他主要编码参数设置与配置文件 encoder_lowdelay_P_vtm 中的默认设置一致,如表 2 所示。表 3 所示为 JCTVC-X0038 QPC 方法最终选择的 Δ_i 。

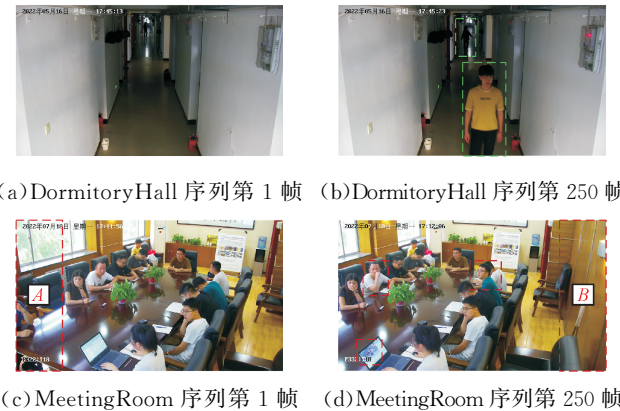


图 2 DormitoryHall 和 MeetingRoom 序列的部分帧
Fig. 2 Some frames in the DormitoryHall and MeetingRoom sequences

表 2 主要编码参数配置 Table 2 Main encoding parameters configuration	
编码参数	配置
编码结构	LD
档次	main_10
Q_0	22,27,32,37
图像组大小	8

表 3 JCTVC-X0038 QPC 方法选择的 Δ_i
Table 3 Δ_i using JCTVC-X0038 QPC method

测试序列	Q_0	Δ_1	Δ_2	Δ_3	Δ_4
DormitoryHall	22	2	3	0	1
	27	2	5	0	1
	32	2	6	0	1
	37	2	6	0	1
MeetingRoom	22	2	3	0	1
	27	2	5	0	1
	32	2	6	0	1
	37	2	6	0	1

结合式(2)、表 1 和表 3 易得出以下结论: JCTVC-X0038 QPC 方法给不同测试序列选择的 Δ_i 是一样的,即 JCTVC-X0038 QPC 方法并未考虑视频序列内容特性对 QP 选择的影响。但是,视频作为视频编码操作的输入信源,信源特性会显著影响视频编码过程中技术的参数选择与运行效率,最终会显著影响视频编码的整体效率。所以,理论上视频的内容特性会显著影响 QP 的选择,下一节也将从理论与实验两个角度证明以上推断。综上所述可知,VTM 当前推荐的 JCTVC-X0038 QPC 方法是不优化的,其编码性能还可进一步提高。

2 QP 选择影响因素分析

2.1 时间层对 QP 选择的影响

VVC 采用基于预测、变换、量化、熵编码的混合编码框架以消除原始视频中的各类冗余信息,达到数据压缩的目的。从信源的角度看,视频是一系列图像在时间维度上的集合,相邻图像内容的相关性很强,即存在很多时域冗余。VVC 采用帧间预测技术^[15,17]消除视频中的时域冗余,其基本原理表示为

$$r_{i,b} = o_{i,b} - p_{j,b} \tag{3}$$

式中: $o_{i,b}$ 表示第 i 幅图像的第 b 个编码块; $r_{i,b}$ 表示第 i 幅图像第 b 个编码块对应的残差块; $p_{j,b}$ 表示 $o_{i,b}$

在第 j 幅图像通过运动估计找到的最优匹配块,第 j 幅图像是第 i 幅图像的参考图像。相比于编码块,残差块对应的符号概率密度分布更加趋于不均匀,能量更加集中,其经过变换量化进一步处理后更利于提高熵编码技术去冗余的能力,最终提升视频编码的效率。

由式(3)易看出,第 j 幅图像的质量会显著影响第 i 幅图像最优匹配块的选择效率,进而影响残差块中残差值的概率密度分布特征和熵编码的去冗余能力,最终影响第 i 幅图像的编码效率。所以,给第 j 幅图像选择的 QP 不仅影响本图像的编码效率,也显著影响以第 j 幅图像为参考图像的其他图像编码效率。另外,为了有效权衡编码复杂度和率失真性能,VTM 支持的参考图像数量和来源是固定的,其中参考图像数量最大为 4^[15],LD 结构中参考图像的来源参考图 1 中的箭头方向。由图 1 易看出,低时间层图像,例如时间 0 层与时间 1 层图像,会被更多的图像作为参考图像。综上可以得到第 1 个推论:相比于其他更高时间层图像,时间 0 层和时间 1 层图像对整个视频编码效率的影响更大。

下面从实验角度验证上述第 1 个推论的准确性。对于实验选择的测试序列,编码器、编码结构、档次、 Q_0 等都与第 1 节描述的实验设置一致。为了避免不同编码方法得到的码率相差过大,设置时间 0 层图像的实际编码 QP 为 $Q_0 - \Delta_1$, Q_i 的取值按照式(1)设置,其中 Δ_1 分别取 1、2、3、4, Δ_2 分别取 3、4、5、6, Δ_3 分别取 0、1、2, Δ_4 分别取 0、1、2。 Δ_1 、 Δ_2 、 Δ_3 、 Δ_4 取不同的值就对应着不同的 QPC 方法。与表 3 中的数据对比易看出,本实验中设置的 Δ_i 包含了 JCTVC-X0038 QPC 方法得到的所有取值,且在此基础上进行了适当扩展。

率失真性能是衡量 QPC 方法的关键指标,且通常使用 BD-rate^[9,14,17-20]作为测度,用符号 ΔR 表示。 ΔR 反映了待评价方法在与基准方法在获得同等感知质量的重建视频前提下,可以节省多少码率。首先,将 JCTVC-X0038 QPC 方法作为基准方法,计算不同 Δ_1 、 Δ_2 、 Δ_3 、 Δ_4 取值组合对应的 QPC 方法的 ΔR 。然后,进一步选择 Δ_2 、 Δ_3 、 Δ_4 取相同值而 Δ_1 取不同值的 QPC 方法,计算这些方法对应的 ΔR 随着 Δ_1 取值变化时的绝对值变化量。这一数据可以有效反映时间 0 层和时间 1 层图像的 QP 对整体视频编码率失真性能的影响程度。按照同样的原理,分别计算 Δ_2 、 Δ_3 、 Δ_4 取值变化时其对应的 ΔR 绝对值变化量。具体数值如表 4 所示。

表 4 ΔR 的绝对值变化量

Table 4 Absolute values variation of ΔR

测试序列	ΔR 的绝对值变化量			
	Δ_1 变化时	Δ_2 变化时	Δ_3 变化时	Δ_4 变化时
DormitoryHall	6.14	2.31	1.47	1.72
MeetingRoom	1.57	1.53	0.51	0.64

由表 4 数据易看出,对于不同内容特性的测试序列, Δ_1 对应的 ΔR 绝对值变化量都是最大的。所以,相比于其他时间层,时间 0 层和时间 1 层对于整个视频编码率失真性能的影响是最大的。另外,相比于 Δ_3 、 Δ_4 对应的 ΔR 绝对值变化量, Δ_2 对应的 ΔR 绝对值变化量也明显更大。 Δ_2 反映的是时间 1 层和时间 2 层 QP 之间的差值,所以 Δ_2 对应的 ΔR 绝对值变化量也部分反映了时间 1 层图像对于整个视频编码率失真性能的影响。综上可知,表 4 数据证明了第 1 个推论的准确性。基于这一推论,权衡方法复杂度与编码效率,本文后续研究只探究 $\Delta_{1,opt}$ 和 $\Delta_{2,opt}$ 的选择, Δ_3 和 Δ_4 则按照 JCTVC-X0038 QPC 方法进行默认设置。

2.2 视频内容特性对 QP 选择的影响

视频作为一幅幅图像在时间维度上的集合,其内容特性可以从空域和时域两个角度评价^[21]。空域通常指的是图像纹理复杂度,时域通常指的是视频内容的运动快慢^[21]。第 2.1 小节研究内容说明了参考图像会影响编码图像的编码效率,具体影响程度会与本小节关注的视频空时域内容特性密切相关。通常,视频纹理越简单或运动越慢时,相邻图像内容变化较小,其相关性较强,此时参考图像会对编码图像产生较大的影响,反之亦然。下面通过实验验证上述分析。

选择的测试序列、编码器、编码结构、档次、 Q_0 等都与第 1 节描述的实验设置一致,但本实验只编码序列的前 2 帧,第 2 帧编码时会第 1 帧作为参考图像。相比于 MeetingRoom 序列, DormitoryHall 序列的纹理更加简单运动更慢。图 3 所示为 Q_0 分别为 22、27、32、37 时第 2 帧对应的预测残差概率密度分布。由图中数据易看出,不论 Q_0 取值为多少, DormitoryHall 序列对应的取值接近于 0 的残差明显更多,概率密度分布曲线的形状更加陡峭,能量更加集中。这表明相比于 MeetingRoom 序列, DormitoryHall 序列对应预测技术的效率更高,即第 1 帧对于第 2 帧的影响更大。这一实验数据验证了之前的分析,即对于纹理越简单或运动越慢的视频,参考图像对编码图像的影响更大。

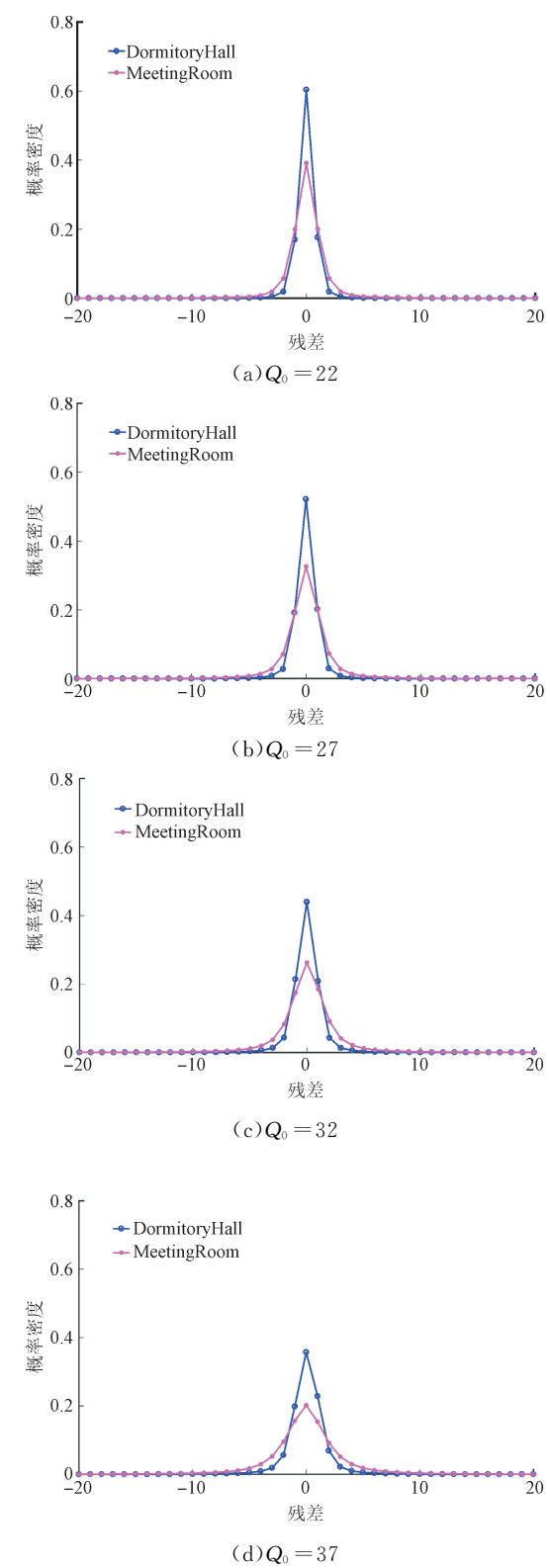


图 3 第 2 帧预测残差的概率密度分布

Fig. 3 Probability density distribution of prediction residual values in the second frame

假定整个视频的编码总码率为 R_c , 分配给低时间层图像和高时间层图像的码率分别为 R_l 和 R_u 。首先,对于纹理简单或运动慢的视频,参考图像对编

码图像的影响更大,其对于整个视频的编码效率的影响也更大。其次,由第 2.1 小节分析及图 1 易知,低时间层图像会被更多图像做参考图像。结合上述两点可知,对于纹理简单或运动慢的视频,在实际编码时应增大 R_l ,用更小的 QP 编码低时间层图像,以优先保证编码质量。在 R_c 近似不变的情况下,增大 R_l 意味着减小 R_u ,即其他高时间层图像对应的 QP 会增大。这最终会导致时间层图像间的 QP 差值增大。综上,得出第 2 个推论:视频的空时域内容特性对 QP 的选择具有重要影响,且纹理越简单运动越慢的视频,其对应的 $\Delta_{1,opt}$ 和 $\Delta_{2,opt}$ 会越大。

下面进行实验证明上述第 2 个推论的准确性。测试序列、编码器、编码结构、档次、 Q_0 、时间 0 层图像编码 QP 等都与第 2.1 小节描述的实验设置一致, Q_i 按照式(1)设置,其中, Δ_1 和 Δ_2 都分别取 1~10 之间的整数, Δ_3 和 Δ_4 被分别设置为由 JCTVC-X0038 QPC 方法得出的值,即 0 和 1。然后,以 JCTVC-X0038 QPC 方法为基准方法计算不同 Δ_1 和 Δ_2 组合对应的 ΔR 。

图 4 为不同 Δ_1 和 Δ_2 组合对应的 ΔR 。如图 4(a) 中红点所示,当 Δ_1 和 Δ_2 分别为 3 和 6 时,其对应的 ΔR 最小,为 -3.58% 。所以,MeetingRoom 序列对应的 $\Delta_{1,opt}$ 和 $\Delta_{2,opt}$ 分别为 3 和 6。同理,如图 4 (b) 中红点所示,DormitoryHall 序列对应的 $\Delta_{1,opt}$ 和 $\Delta_{2,opt}$ 分别为 4 和 8。这一实验结果验证了第 2 个推论的准确性。另外,图 4 和表 3 的结果也直观地证明了 JCTVC-X0038 QPC 方法并不是最优的。

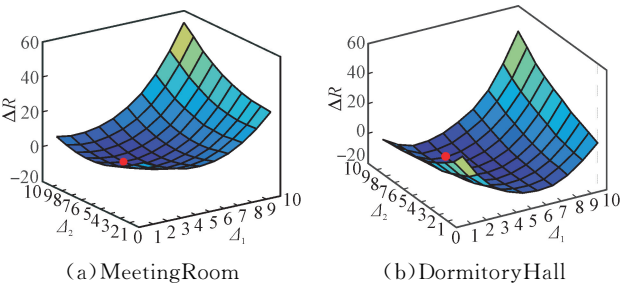


图 4 不同 Δ_1 和 Δ_2 组合对应的 ΔR

Fig. 4 ΔR of different combinations of Δ_1 and Δ_2

3 内容自适应的 QPC 方法

本文提出的 SCA-QPC 方法也按照式(1)的方案设置每个时间层的 QP。进一步基于第 2 节证明的第 1 个推论,SCA-QPC 方法只探究 $\Delta_{1,opt}$ 和 $\Delta_{2,opt}$ 的选择, Δ_3 和 Δ_4 则按照 JCTVC-X0038 QPC 方法进行默认设置。 $\Delta_{1,opt}$ 和 $\Delta_{2,opt}$ 的选择方法将在第 3.2 小节描述。另外,基于第 2 节证明的第 2 个推论,在

构建 $\Delta_{1,\text{opt}}$ 和 $\Delta_{2,\text{opt}}$ 的选择方法时,充分考虑视频空时域内容特性的显著影响。视频空时域内容特性的量化评价将在第 3.1 小节描述。

3.1 视频空时域内容特性评价

本文采用梯度 σ 衡量视频的纹理复杂度^[22], σ 越大表示视频的纹理越复杂,计算式为

$$\sigma = \frac{1}{v} \sum_{i=1}^v \left(\frac{1}{wh} \sum_{x=1}^w \sum_{y=1}^h G_p(i, x, y) \right) \quad (4)$$

式中: v 表示视频包含的图像数; w, h 分别表示图像的宽和高; $G_p(i, x, y)$ 表示第 i 幅图像第 x 行第 y 列像素 4 个纹理方向对应的梯度最大值,计算式为

$$G_p(i, x, y) = \max_{k=1,2,3,4} \left\{ \left| \frac{1}{16} \sum_{m=1}^5 \sum_{n=1}^5 (I(i, x-3+m, y-3+n) g_k(m, n)) \right| \right\} \quad (5)$$

其中, $I(i, x-3+m, y-3+n)$ 表示第 i 幅图像第 $x-3+m$ 行第 $y-3+n$ 列像素的亮度值, $g_k(m, n)$ 表示第 k 个纹理方向对应的加权滤波器,其取值如图 5 所示。

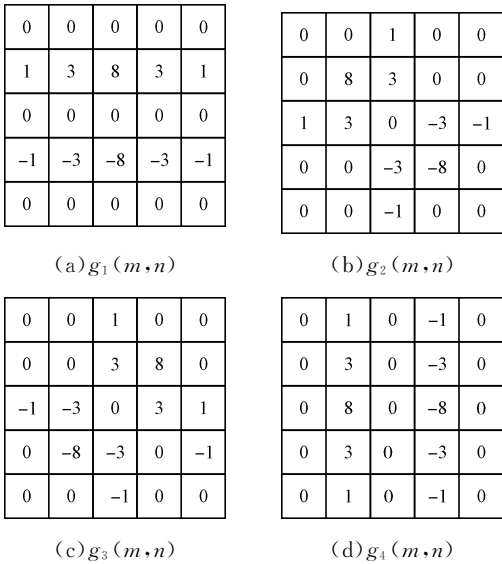


图 5 加权滤波器 $g_k(m, n)$ ($k=1, 2, 3, 4$) 的取值
Fig. 5 Weighted filter $g_k(m, n)$ ($k=1, 2, 3, 4$)

采用背景帧差 d 衡量视频的运动快慢^[21], d 越大表示视频的运动越快,计算式为

$$d = \frac{1}{v} \sum_{i=2}^v \left(\frac{1}{wh} \sum_{x=1}^w \sum_{y=1}^h ((I(i, x, y) - I(i-1, x, y) + I_b(i, x, y) - I_b(i-1, x, y))/2) \right) \quad (6)$$

式中: $I_b(i, x, y)$ 表示第 i 幅图像第 x 行第 y 列像素的背景亮度值,计算式为

$$I_b(i, x, y) = \frac{1}{32} \sum_{m=1}^5 \sum_{n=1}^5 (I(i, x-3+m, y-3+n) a(m, n)) \quad (7)$$

其中, $a(m, n)$ 表示加权滤波器,其取值由图 6 所示。

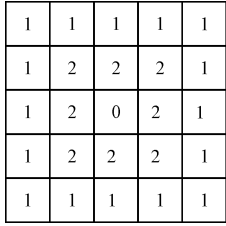


图 6 加权滤波器 $a(m, n)$ 的取值
Fig. 6 Weighted filter $a(m, n)$

表 5 为 DormitoryHall 和 MeetingRoom 两个序列对应的 σ 与 d 。相比于 DormitoryHall 序列, MeetingRoom 序列的纹理更加复杂运动越快,而其对应的 σ 与 d 取值也更大。所以, σ 与 d 能够比较准确地评价视频的空时域内容特性。

表 5 视频评价指标 σ 与 d 的取值

测试序列	σ	d
DormitoryHall	3.16	0.12
MeetingRoom	10.27	2.99

3.2 基于反向传播神经网络的 $\Delta_{t,\text{opt}}$ ($t=1, 2$) 预测模型

由于良好的非线性映射能力和低复杂度特性,反向传播神经网络(BPNN)被广泛应用于回归预测^[23-26]。权衡方法准确率和复杂度,本文构建了基于轻量型 BPNN 的 $\Delta_{t,\text{opt}}$ ($t=1, 2$) 预测模型,如图 7 所示。特别说明,本文构建的 $\Delta_{1,\text{opt}}$ 与 $\Delta_{2,\text{opt}}$ 的预测模型是独立的,但是二者的网络结构是类似的。只是在训练时使用了不同的训练数据,以使其训练好的网络模型和对应的参数组合能够分别适用于 $\Delta_{1,\text{opt}}$ 和 $\Delta_{2,\text{opt}}$ 的预测。

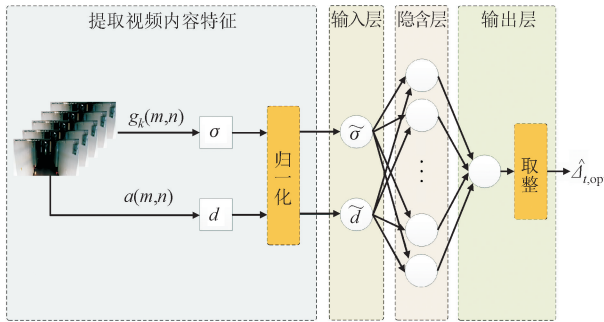


图 7 $\Delta_{t,\text{opt}}$ ($t=1, 2$) 的预测模型
Fig. 7 Prediction models of $\Delta_{t,\text{opt}}$ ($t=1, 2$)

对于输入的视频,首先分别根据式(4)和式(6)计算 σ 和 d 。然后,将 σ 和 d 归一化后作为BPNN的输入。BPNN的输出为 $\Delta_{t,\text{opt}}(t=1,2)$ 的预测值,定义为 $\hat{\Delta}_{t,\text{opt}}(t=1,2)$ 。隐含层包含1层神经元,神经元的数量按照文献[25]中的经验公式进行设置。综合考虑预测准确度与方法复杂度,对于 $\Delta_{1,\text{opt}}$ 和 $\Delta_{2,\text{opt}}$ 的预测模型,其对应的隐含层神经元数量分别设置为8和10。隐含层和输出层的激励函数分别为Tansig和Purelin。具有不同分辨率与场景内容的40个监控视频组成视频数据集。随机选择30个视频组成训练数据集,用于训练BPNN。剩余的10个视频组成测试数据集,用于测试提出SCA-QPC方法的性能。每一个视频对应的 $\Delta_{t,\text{opt}}(t=1,2)$ 通过第2.2小节描述的编码实验获得。使用Trainlm作为训练函数,通过正向传播和误差反向传播不断修正BPNN中各层的权重和阈值,直到 $\Delta_{t,\text{opt}}$ 和 $\hat{\Delta}_{t,\text{opt}}$ 之间的误差小于目标误差。

最终,SCA-QPC方法的具体实施步骤如下。

(1)对于待编码的视频,分别根据式(4)和式(6)计算其对应的 σ 和 d 。

(2)将 σ 和 d 作为输入,使用训练好的如图7所示的预测模型,预测得到 $\hat{\Delta}_{t,\text{opt}}(t=1,2)$ 。

(3)设置时间0层图像的实际编码QP为 $Q_0-\hat{\Delta}_{1,\text{opt}}$,其他所有时间层图像对应的QP设置为

$$Q_t = \begin{cases} Q_{t-1} + \hat{\Delta}_{t,\text{opt}}, & t = 1, 2 \\ Q_{t-1} + 0, & t = 3 \\ Q_{t-1} + 1, & t \geq 4 \end{cases} \quad (8)$$

如表3所示,当大于等于时间3层即 $t \geq 3$ 时,图像对应的 Δ_t 按照JCTVC-X0038 QPC方法得出的值进行设置。

(4)根据第(3)步设置的QP编码视频,完成编码操作。

4 实验与分析

如式(8)所示,SCA-QPC方法的性能与 $\hat{\Delta}_{t,\text{opt}}(t=1,2)$ 的预测准确度密切相关。所以,本节首先评估 $\hat{\Delta}_{t,\text{opt}}(t=1,2)$ 预测模型的准确性,然后再评估SCA-QPC方法的率失真性能。

4.1 $\Delta_{t,\text{opt}}(t=1,2)$ 的预测准确度

测试集由10个包含不同分辨率与场景内容的监控视频组成, $\Delta_{t,\text{opt}}(t=1,2)$ 的对测试集的预测结果如表6所示。测试序列对应的 $\hat{\Delta}_{1,\text{opt}}$ 与 $\hat{\Delta}_{2,\text{opt}}$ 分别由第3.2小节训练好的模型预测得到。从表6数据易看出,对于大部分序列, $\hat{\Delta}_{t,\text{opt}}$ 与 $\Delta_{t,\text{opt}}$ 都相等,有偏差的

也仅相差1。对于 $\Delta_{1,\text{opt}}$ 和 $\Delta_{2,\text{opt}}$,提出模型对应的均方误差都很小,只有0.2。综上可知,本文构建的 $\Delta_{t,\text{opt}}(t=1,2)$ 预测模型可以较准确地获得预测结果。

表 6 $\Delta_{t,\text{opt}}(t=1,2)$ 的预测结果
Table 6 Prediction results of $\Delta_{t,\text{opt}}(t=1,2)$

测试序列	$\Delta_{1,\text{opt}}$	$\hat{\Delta}_{1,\text{opt}}$	$\Delta_{2,\text{opt}}$	$\hat{\Delta}_{2,\text{opt}}$
Office	4	4	7	7
Basketball	1	1	7	7
Bookstore	4	4	6	6
Carpark	4	4	6	5
Community	5	4	6	6
DiningHall	4	4	6	6
ElectricBicyclePark	1	1	4	4
Laboratory	4	4	5	6
LivingQuarters	2	2	7	7
MeetingRoom	3	4	6	6
均方误差	0.2		0.2	

4.2 率失真性能

率失真性能是衡量QPC方法编码效率的最关键指标,且通常使用 ΔR 作为测度^[9,14,17-20]。目前,JCTVC-X0038 QPC方法是唯一面向VVC标准提出的QPC方法,且被VTM集成。所以,本小节将JCTVC-X0038 QPC方法作为对比方法,以评估SCA-QPC方法的率失真性能。

所有测试序列分别使用JCTVC-X0038 QPC方法和SCA-QPC方法编码,获得对应的编码码率和失真数据。编码器、编码结构、档次、GOP大小、 Q_0 等与第1节描述的实验设置一致。以JCTVC-X0038 QPC方法为基准,计算SCA-QPC方法对应的 ΔR ,结果如表7所示。对于所有测试序列,SCA-QPC方法对应的平均 ΔR 为-4.26%,即相比于JCTVC-X0038 QPC方法,SCA-QPC方法在获得同等质量重建视频的前提下,可以节省4.26%的编码码率。相比于JCTVC-X0038 QPC方法,SCA-QPC方法显著提高了监控视频的编码效率。图8给出了Carpark和DiningHall两个序列的率失真曲线,以直观地感知SCA-QPC方法对应的率失真性能提升。

如表7所示,对于纹理简单运动慢的序列,SCA-QPC方法可以提升较高的率失真性能,例如DiningHall序列对应的 ΔR 为-8.05%。对于纹理复杂运动快的序列,SCA-QPC方法提升的率失真性能有限,例如ElectricBicyclePark序列对应的 ΔR 为-0.43%。这主要是因为,如第2.2小节分析,相比于纹理简单运动慢的视频,纹理复杂运动快的视

频其对应的 $\Delta_{i,\text{opt}} (t = 1, 2)$ 较小,更加接近于 JCTVC-X0038 QPC 方法得出的 $\Delta_i (t = 1, 2)$ 。因此,相比于纹理简单运动慢的视频,SCA-QPC 方法对于纹理复杂运动快视频的编码性能提升空间有限,所以其对应的 ΔR 较小。

表 7 SCA-QPC 对应的 ΔR

Table 7 ΔR of SCA-QPC

测试序列	$\Delta R/\%$
Office	-5.76
Basketball	-1.10
Bookstore	-6.01
Carpark	-7.58
Community	-7.25
DiningHall	-8.05
ElectricBicyclePark	-0.43
Laboratory	-2.62
LivingQuarters	-0.30
MeetingRoom	-3.46
平均值	-4.26

5 结 论

本文面向监控视频提出了一种适用于最新 VVC 标准考虑视频空时域复杂度的 QPC 方法,实验结果证明了提出方法可以有效提高监控视频的编码效率。本文研究得到如下有意义结论:

(1)从率失真性能角度看,VVC 标准当前推荐的 JCTVC-X0038 QPC 方法不是最优的,其性能还有较大提升空间;

(2)在 VVC 低时延结构中,低时间层图像对于整个视频编码效率的影响更大;

(3)视频的时空域内容特性是影响 VVC 量化参数选择的重要因素,且对于纹理简单运动快的视频,其对应的时间层之间最优 QP 的差值越大;

(4)BPNN 可以有效应用于 VVC 标准的 QP 选择中。

后续有价值的研究方向包括:适用于屏幕内容视频、360°全景视频等其他类型视频的 QPC 方法;适用于 VVC 标准随机接入结构的 QPC 方法。

参考文献:

[1] SCHWARZ H, MARPE D, WIEGAND T. Overview of the scalable video coding extension of the H. 264/AVC standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(9): 1103-1120.

[2] 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.

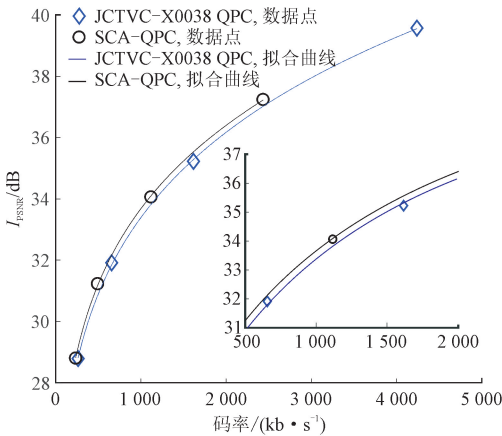
[3] 李维, 任鹏, 赵凡, 等. 高效视频编码中帧内码率控制改进算法 [J]. 西安交通大学学报, 2019, 53(4): 79-84.

LI Wei, REN Peng, ZHAO Fan, et al. A modified method for intra-coding rate control in high efficiency video coding [J]. Journal of Xi'an Jiaotong University, 2019, 53(4): 79-84.

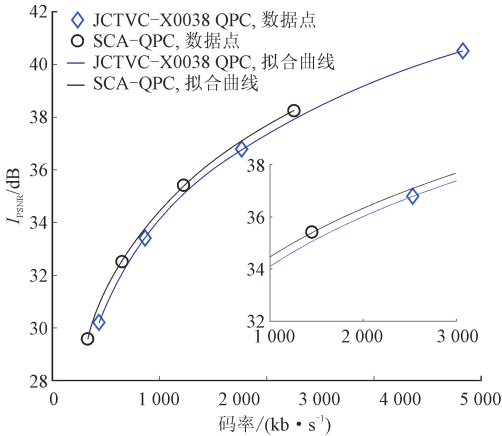
[4] BROSS B, WANG Yekui, YE Yan, et al. Overview of the versatile video coding (VVC) standard and its applications [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3736-3764.

[5] SCHWARZ H, MARPE D, WIEGAND T, et al. Hierarchical B pictures [EB/OL]. (2005-07-23) [2023-05-12]. <https://www.itu.int/wftp3/av-arch/jvt-site/JVT-P014.doc>.

[6] MCCANN K, ROSEWARNE C, BROSS B, et al.



(a)Carpark



(b)DiningHall

图 8 率失真曲线

Fig. 8 Rate distortion curves

- High efficiency video coding (HEVC) test model 16 (HM16) improved encoder description [EB/OL]. (2014-10-17) [2023-05-12]. <http://phenix.int-evry.fr/jct/JCTVC-S1002.doc>.
- [7] LI Bin, ZHANG Dong, LI Houqiang, et al. QP determination by Lambda value [EB/OL]. (2012-05-07) [2023-05-12]. <http://phenix.int-evry.fr/jct/JCTVC-I0426.doc>.
- [8] YANG Yuan, WAN Shuai, GONG Yanchao, et al. Adaptive quantization parameter cascading for random-access prediction in H. 265/HEVC based on dependent R-D models [C]//2016 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2016: 4235-4239.
- [9] GONG Yanchao, WAN Shuai, YANG Kaifang, et al. Rate-distortion-optimization-based quantization parameter cascading technique for random-access configuration in H. 265/HEVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(6): 1304-1312.
- [10] ZHAO Tiesong, WANG Zhou, CHEN Changwen. Adaptive quantization parameter cascading in HEVC hierarchical coding [J]. IEEE Transactions on Image Processing, 2016, 25(7): 2997-3009.
- [11] XU Yiwen, LI Qi, LI Xia, et al. Efficient QP cascading in H. 265/HEVC low-delay prediction [C]//2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Piscataway, NJ, USA: IEEE, 2017: 151-156.
- [12] HE Jing, YANG Enhui, YANG Fuzheng, et al. Adaptive quantization parameter selection for H. 265/HEVC by employing inter-frame dependency [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(12): 3424-3436.
- [13] AMER H, YANG Enhui. Adaptive quantization parameter selection for low-delay HEVC via temporal propagation length estimation [J]. Signal Processing: Image Communication, 2020, 84: 115826.
- [14] GONG Yanchao, YANG Kaifang, LIU Ying, et al. Quantization parameter cascading for surveillance video coding considering all inter reference frames [J]. IEEE Transactions on Image Processing, 2021, 30: 5692-5707.
- [15] BROWNE A, YE Yan, KIM S. Algorithm description for versatile video coding and test model 16 (VTM16) [EB/OL]. (2022-01-12) [2023-05-12]. <https://jvet-experts.org/JVET-Y2002.docx>.
- [16] ANDERSSON K, WENNERSTEN P, SAMUELS-SON J, et al. AHG3 recommended settings for HM [EB/OL]. (2016-06-01) [2023-05-12]. <http://phenix.int-evry.fr/jct/JCTVC-X0038.doc>.
- [17] 公衍超, 吕显龙, 杨楷芳, 等. 考虑全帧间参考依赖性的监控视频高效视频编码码率控制算法 [J]. 西安交通大学学报, 2021, 55(3): 72-80.
- GONG Yanchao, LÜ Xianlong, YANG Kaifang, et al. Rate control for high efficiency video coding coded surveillance videos considering all of the inter-frame reference dependency [J]. Journal of Xi'an Jiaotong University, 2021, 55(3): 72-80.
- [18] YANG Kaifang, GONG Yanchao, MA Miao, et al. An efficient rate-distortion optimization method for low-delay configuration in H. 265/HEVC based on temporal layer rate and distortion dependence [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(4): 1230-1236.
- [19] MAO Yunhao, WANG Meng, WANG Shiqi, et al. High efficiency rate control for versatile video coding based on composite Cauchy distribution [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(4): 2371-2384.
- [20] LIU Hewei, ZHU Shuyuan, ZENG Bing. Inter-frame dependency-based rate control for VVC low-delay coding [J]. IEEE Signal Processing Letters, 2022, 29: 2727-2731.
- [21] GONG Yanchao, WAN Shuai, YANG Kaifang, et al. Temporal-layer-motivated lambda domain picture level rate control for random-access configuration in H. 265/HEVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(1): 156-170.
- [22] YANG Xiaokang, LIN Weisi, LU Zhongkang, et al. Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(6): 742-752.
- [23] WANG Wei, ZHU Qingshan, WANG Zhaoba, et al. Research on indoor positioning algorithm based on SA-GA-BP neural network [J]. IEEE Sensors Journal, 2022, 22(4): 3736-3744.
- [24] 金成, 王久洪, 张成, 等. 结合传感器阵列和 BP 神经网络的挥发性有机物定量检测系统 [J]. 西安交通大学学报, 2021, 55(10): 106-113.
- JIN Cheng, WANG Jiuhong, ZHANG Cheng, et al. A quantitative detection system of volatile organic compounds based on combining sensor array and BP neural network [J]. Journal of Xi'an Jiaotong University, 2021, 55(10): 106-113.

- [25] 文必洋, 汤文成, 田应伟. 基于 BP 神经网络的高频雷达有效浪高场反演 [J]. 华中科技大学学报(自然科学版), 2021, 49(4): 114-119.
- WEN Biyang, TANG Wencheng, TIAN Yingwei. Significant wave height field inversion of high frequency radar based on BP neural network [J]. Journal of

Huazhong University of Science and Technology (Nature Science Edition), 2021, 49(4): 114-119.

- [26] LIU Jiguo, HUANG Jian, SUN Rui, et al. Data fusion for multi-source sensors using GA-PSO-BP neural network [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(10): 6583-6598.

(编辑 陶晴 亢列梅)