

基于 ROI 的多路视频传输码率控制策略

胡伟东 朱道平 毛焕 福州大学物理与信息工程学院

摘要: 针对异构网络下感兴趣区域(ROI)视频流传输的码率控制问题, 本文提出一种宏块级别的比特分配方法。通过综合考虑视频端到端传输时延以及发送缓存区溢出问题, 获得 ROI 编码下的帧比特上界; 以提升符合人眼系统(HVS)的主观视频质量为目标, 在缓冲区大小及时延约束下, 针对感兴趣视频流帧内不同区域的重要性差异, 本文设计了不同的码率控制策略(RRCS)。设计经过试验平台得到证明, 所提方法能够有效提升用户主观视觉质量, 在时延约束 250ms 时, 帧丢失率低于 7%, 加权 PSNR 最高提升 5.6dB。

关键词: 感兴趣区域 HVS 码率控制 加权 PSNR

DOI:10.16045/j.cnki.cattvtec.2021.07.029

1 引言

随着智能手机等移动终端的普及以及视频直播、安防监控等领域的发展, 移动视频流量迅速增长。思科统计数据表明, 在 2021 年, 移动视频流量在互联网网络流量的占比将高达 82%。面对如此庞大的视频数据量, 单一的无线网络已无法满足用户对于视频的实时性和高清需求^[1]。在下一代无线网络中, 移动客户端通过配备多个接口, 同时接入多个无线网络聚合带宽, 降低时延。虽然异构网络^[2]下的视频多路传输方法被提出作为克服网络限制的解决方案, 但带宽的限制和数据帧丢包等问题依旧存在。感兴趣区域(ROI)编码针对图像不同区域对于人类视觉系统(HVS)^[3]的重要差异性, 通过提高非感兴趣区域的量化参数和降低感兴趣区域的量化参数来减少传输比特和节省带宽, 提高主观视频质量。目前, 对于感兴趣区域编码的研究主要针对码率控制和感兴趣区域检测。文献[4]提出了一种基于块的感兴趣提取方法, 实现了基于区域编码的 H.263 编码器。文献[5]提出了一种基于感兴趣区域编码的质量

可调码率控制方法, 并根据码率控制算法预测感兴趣区域 QP 的范围。文献[6]使用 Kinect 相机提供的骨架信息对人体躯干进行感兴趣区域跟踪, 它以牺牲背景质量来传输高质量动态感兴趣区域视频。虽然诸多研究对基于 ROI 的不同码率控制算法做了阐述, 但算法多基于单链路环境。基于 ROI 编码模式进行多路传输, 能够整合两种技术的各自优势, 实现更佳的多维度目标整体优化, 但问题也随之变得更加复杂。与传统单链路的 ROI 编码传输相比, 实现感兴趣区域视频流在异构网络下进行传输有着复杂的因素需要考虑。首先, 多路传输需要一帧数据被接收端完全接收才可开始解码, 故需要以一定的分流策略将视频帧拆分并分流到多条链路中, 以确保视频帧同时到达服务端。因此, 多路特性是感兴趣区域视频流在多链路环境传输需要考虑的首要因素。其次, 在多路传输下, 需要综合考虑信道的时变性、分流策略决策以及时延等多方面因素, 进而确认感兴趣区域和非感兴趣区域的不同目标比特数。

针对以上问题, 本文对基于感兴趣区域(ROI)的视频流传输模式提出

了一种宏块级别的码率控制策略, 使用 YOLO v3 进行目标模型检测来提取感兴趣区域, 对不同区域宏块进行差异化的码率控制。首先, 建立感兴趣视频流多路并行系统, 进而分析视频帧在该系统传输的端到端时延模型, 在最小时延约束及发送缓冲区溢出约束下获得一帧图像中感兴趣及非感兴趣区域的比特数目标和上界; 其次, 基于帧内图像不同区域的重要差异性设计比特资源分配方法; 最后, 在网络仿真环境下, 验证上述所提出的算法。实验数据结果显示, RRCS 算法在 250ms 的时延约束下, 能够保证低于 7% 的视频帧超时丢失率, 并且最高将加权平均信噪比(weighted-PSNR)分别提升 5.6dB。

2 ROI 编码传输系统及问题模型

2.1 ROI 编码传输系统框图

本文研究的基于感兴趣区域视频流多路并行传输系统模型如图 1 所示。该模型由客户端(视频发送端)和服务端(视频接收端)两部分组成。客户端由 YOLO v3 对输入的摄像头采集

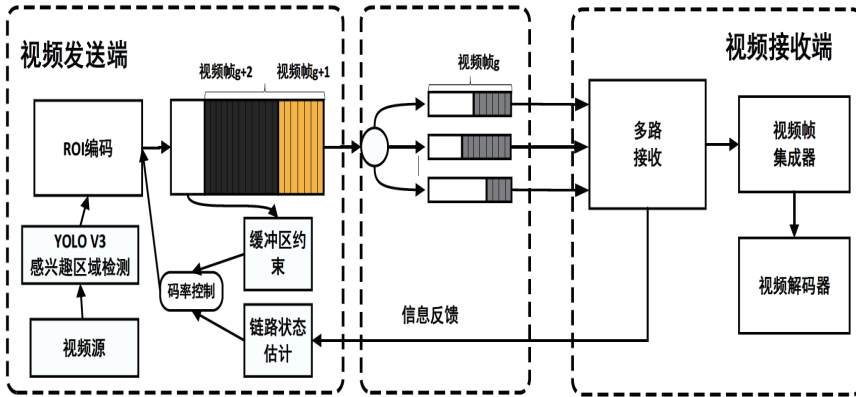


图1 ROI编码传输系统框图

的YUV数据进行感兴趣区域检测提取，之后根据发送缓冲区充盈度和接收端反馈的链路状况（如带宽、往返时延）共同决策调整ROI编码器的码率控制策略并执行。经过ROI编码后的视频码流通过分流策略分发到各条链路，在视频接收端由视频帧集成器合并复原，进而利用视频解码器解码播放。

2.2 ROI编码传输时延模型

在ROI编码传输系统下，多路并行传输视频帧g的端到端时延主要由视频帧g的排队时延、数据发送时延以及传播时延三部分构成，具体如公式（1）所示：

$$D^g = d_q^g + d_s^g + d_e^g \quad (1)$$

式中， D^g 表示ROI编码模式下的视频帧g的端到端时延， d_q^g 表示视频帧g的排队时延， d_s^g 表示发送时延， d_e^g 代表传播时延。其中，发送缓冲区排队时延指视频帧g进入发送缓冲区，到其完全离开的阶段时间大小，等价于前g帧发送时间之和，视频帧的发送时间与各个帧的流量分配 L_n^g 以及当前的网络传输带宽 μ_n 相关，其计算如公式（2）所示：

$$d_s^g = \sum_{n=1}^{g_{\text{step}}} \max\left(\frac{L_n^g}{\mu_n}\right) \quad (2)$$

其中， μ_n 表达链路n的可用带宽，

L_n^g 表示第g帧分流到链路n的比特数。在多链路传输场景下，一帧视频被发送到多个链路进行传输，必须等到所有视频帧数据被完整接收，才能被解码播放，故发送时延和传播时延应取各个阶段的时延最大值。其具体计算如公式（3）所示：

$$d_s^g + d_e^g = \max\left(\frac{L_n^g}{\mu_n} + \frac{RTT_n}{2}\right) \quad (3)$$

其中，链路n的往返时延为 RTT_n ；结合公式（1）、（2）、（3），理论上推导出第g视频帧的端到端时延 D^g 表达式如公式（4）所示：

$$D^g \geq \max\left(\frac{L_n^g}{\mu_n} + \frac{RTT_n}{2}\right) + \sum_{n=1}^{g_{\text{step}}} \max\left(\frac{L_n^g}{\mu_n}\right) \quad (4)$$

2.3 ROI编码码率控制模型

为了实现宏块级别的码率控制，需要对视频帧ROI及非ROI区域与各自量化步长之间的关系进行建模。根据H.264/AVC标准，量化步长与编码后的视频帧比特数呈反比关系，本文基于传统量化参数 Q_{step} 与比特数L的关系，利用ROI对于帧的占比P和量化步长 Q_{step} 二者作为参数，建立函数关系式，如公式（5）、（6）所示：

$$L = f(Q_{\text{step}}, P) \quad (5)$$

$$PL = P(\alpha Q_{\text{step}} + \beta Q_{\text{step}} + \gamma) \quad (6)$$

其中 Q_{step} 代表量化步长， α 和 β 是常数系数，可以通过大量实验和最小二乘线性拟合得出。同时，量化参数和

量化步长之间的互推关系由公式（7）给出：

$$Q_{\text{step}} = (2^{\frac{1}{6}})^{QP-4} \quad (7)$$

式中QP是量化参数，该参数用于确定ROI及非ROI的量化步长 Q_{step} ，其可被X264编码器通过指定光栅顺序下的宏块位置，直接用于控制视频帧大小。

2.4 问题定义

本文的最终目的是为每个视频帧的感兴趣区域和非感兴趣区域提供最佳的QP决策和数据分配，在目标端到端时延和发送缓冲区充盈度约束下，最大化每帧图像的加权平均峰值信噪比，以提高每个视频帧的主观质量，具体表示如公式（8）所示：

$$\begin{aligned} \text{Optimize } \{QP_{\text{roi}}^g, QP_{\text{n-roi}}^g\} &= \arg \max P^g \\ \text{s.t.} \quad & C1: \sum_{n=1}^N L_n^g = L^g, \forall g \\ & C2: L^g \leq L_{\text{max}}^g, \forall g \\ & C3: L_{\text{roi}}^g \geq L_{\text{min}}^g, \forall g \\ & C4: L_n^g \geq 0, \forall n, g \end{aligned} \quad (8)$$

式中， QP_{roi}^g 、 $QP_{\text{n-roi}}^g$ 分别表示视频帧感兴趣区域和非感兴趣区域的量化参数值， P^g 表示ROI编码模式下视频帧g的加权平均PSNR值。传统PSNR在计算中同等对待同一帧图像中的各个像素点，并不适合ROI编码模式下的视频质量衡量。本文由传统PSNR推广到ROI编码模式下而得到：

$$P^g = k \log \frac{255^2}{MSE_R} + (1-k) \log \frac{255^2}{MSE_{NR}}$$

MSE_R 、 MSE_{NR} 分别代表感兴趣区域和非感兴趣区域的均方误差，系数k为协调因子^[6]，可以根据感兴趣重要程度的不同设定，以分配更多权重给感兴趣区域，本文取0.7。

L_{max}^g 表示满足时延约束和缓冲区约束的视频帧比特上界。为了保证

ROI 区域的解码质量, ROI 区域的比特数应该满足一个下限值 L_{\min}^g , 本文将该值取为量化 26 时对应的比特数, 此时视频质量良好^[7]。优化问题 OP1 难以通过凸优化进行最优求解, 因此, 本文设计一种启发式算法进行优化设计。

3 算法实现

为了求解优化问题 OP1, 本文首先根据 ROI 编码模式下视频流传输的端到端时延约束以及发送缓冲区充满度, 获取最佳视频帧大小。其次, 通过分析视频帧内不同区域对于人眼系统 (HVS) 的重要程度, 针对感兴趣区域和非感兴趣区域分别提出了不同的宏块级码率控制策略。最后, 通过一种链路状态估计方法, 实时预测链路数据 (如带宽、往返时延), 为算法实施提供支持。

3.1 ROI 编码帧比特上界决策

时延性能是实时视频应用需要考虑的首要对象。因此, 本文针对多链路环境下的最小时延约束, 参考文献[8]中的“注水算法”来求解 ROI 编码模式下的视频帧大小。将异构网络下的链路建模成具有横截面积的水桶, 链路本身的固定时延用水桶内预装水量表示, 各个桶的宽度代表各条链路当

前的可用带宽。由于多链路环境下视频帧被完整传输完后才能开始解码播放, 所以为了确保最小时延, 应该使 g 帧大小的数据流在填充到各个链路之后, 所有链路的时延相等且尽可能低。

假设 L^g 单位的水填充到各个水桶之后的“水线”, 最小时延为 D_1^g , 当 $D_1^g > \max_{n \in N} \{\frac{RTT_n}{2}\}$, 系统模型如图 2 左侧所示。此时的 $L_n^g > 0, \forall n, g$, 故在此约束下的视频帧比特上界可以计算得到, 如公式 (9) 所示:

$$L_1^g = \sum_{n=1}^N \mu_n (D_1^g - \frac{RTT_n}{2} - \sum_{g=1}^{g_{\max}} \max \frac{L_n^g}{\mu_n}) \quad (9)$$

当 $D_1^g \leq \max_{n \in N} \{\frac{RTT_n}{2}\}$, 系统模型如图 2 右侧所示。此时 $L_n^g = 0, \in n, g$, 在此约束下的视频帧比特上界可以计算得到, 如公式 (10) 所示:

$$L_1^g = \sum_{n=1}^{N-1} \mu_n (D_1^g - \frac{RTT_n}{2} - \sum_{g=1}^{g_{\max}} \max \frac{L_n^g}{\mu_n}) \quad (10)$$

此外, 实时传输系统中必须将缓冲区控制在较小范围以防止造成过大时延。然而较小的缓冲往往造成视频发送缓冲区溢出问题, 通常会造成丢帧现象。为了保护 ROI 在系统中传输, 应尽可能避免丢帧, 虽然通过增大缓冲区可以有效防止丢帧现象产生, 但与此同时, 根据公式 (2), 时延将会变大, 在设计低时延的码率控制算法时, 为了在缓冲区有限的前提下确保不会导致码流在缓冲区中溢出和下溢, 我们改进了算法比特数, 同时还考虑了

缓冲区现在的状态, 具体如公式 (11) 所示。

$$L_2^g < B_e - F(t) \quad (11)$$

根据上述分析可知, 由公式 (9)、(10)、(11) 可以获得满足视频帧发端缓存不上溢条件下 L_2^g 的上界, 以及最小端到端时延约束下的视频帧上界 L_1^g , 二者共同决策出视频帧比特上界 L_{\max}^g , 具体如公式 (12) 所示。

$$L_{\max}^g = \min(L_1^g, L_2^g) \quad (12)$$

3.2 ROI 与非 ROI 比特资源分配

在 ROI 编码模式下, 感兴趣区域和非感兴趣区域的重要程度不同, 人们的关注点多集中在感兴趣区域。因此, 本文定义了 C3, 确保感兴趣区域有足够的比特资源预算, 用以保证感兴趣区域的解码质量。同时, 考虑到实时视频的低时延要求, 本文定义了 C2, 以目标端到端时延约束了帧大小。为了在低时延下最大化用户的主观视觉质量, 本文通过公式 (2) 获得帧平均 QP_{ave} , 以衡量当前资源预算充足情况。综合考虑人眼特性和约束条件 C2、C3, 提出一种基于感兴趣区域的比特资源分配方法。

(1) 当带宽及缓冲区等条件足够满足时延约束, 且帧比特资源充足, 根据计算的帧平均 QP_{ave} , 感兴趣区域能够满足约束 C3, 或能获得最大的视觉质量, 此时再耗费比特资源去提高感兴趣的质量则意义不大。因此需要通过合理的帧内比特资源分配策略, 确保链路带宽够大且低时延状态下, 比特资源不会浪费在感兴趣区域内视觉上察觉不到的质量改善上, 而是用于增强非感兴趣区域, 提高图像质量。此时可通过将剩余比特资源分配给非 ROI 区域, 提高视频整体质量, 即

$$QP_{roi} = QP_{n-roi} = QP_{ave}$$

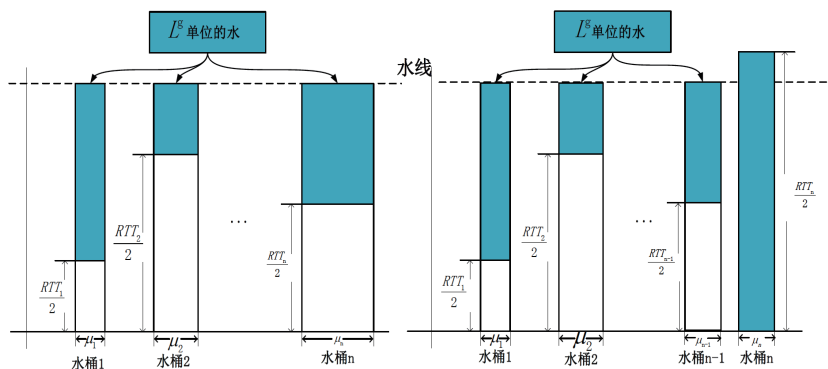


图 2 不同注水情况

RRCS 算法:

输入: $d_{target}, \{RTT_n, \mu_n\}_{n \in N}, k_{left}$

输出: QP_{roi}^g, QP_{n-roi}^g

计算 ROI 编码帧比特上界

```

1)  FOR  n = 1  To  N
2)      FOR  g = 1  TO  gleft, DO
3)           $D^g \geq \max(\frac{L_n^g}{\mu_n} + \frac{RTT_n}{2}) + \sum_{g=1}^{g=g_{left}} \max \frac{L_n^g}{\mu_n}$ 
4)      END FOR
5)   $L_1^g = \sum_{n=1}^N \mu_n (D^g - \frac{RTT_n}{2} - \sum_{g=1}^{g=g_{left}} \max \frac{L_n^g}{\mu_n})$ 
6)   $L_2^g = B(e) - F(t)$ 
7)   $QP_{ave} \leftarrow L_{max}^g = \text{Min}(L_1^g, L_2^g), L_{s1} = f(QP = 51, 1 - P)$ 

```

ROI 和非 ROI 比特资源分配

```

8)  IF  QPave ≤ 26 THEN
9)      QProi = QPn-roi = QPave
10) ELSE IF  QPave ≥ 51 THEN
11)     QProi = QPn-roi = 51
12) ELSE
13)     QProi = 26
14)     WHILE  Lroi + Ls1 > Lmaxg
15)         QProi ++
16)         Lroi = f(QProi, P)
17)     END
18)     Ln-roi = Lk - L → QPn-roi
19) END IF
20) END FOR

```

图 3

(2) 当帧比特资源有限, 即 $QP_{ave} \geq 26$ 时, 则需要牺牲非感兴趣区域的比特资源保证感兴趣区域的资源, 以满足时延约束, 并保证 ROI 质量。基于优先级, 首先预设 ROI 的 QP 为 26, 并判断 QP 为 26 时所对应的比特数 $L_{26} = f(Q_{step}, P)$ 与 L_{max}^g 的大小关系, 当 $L_{26} > L_{max}^g$, 为了满足时延和带宽约束, 需要放弃约束 C2, 动态提高

QP_{roi} , 直至满足 L_{max}^g 的限制, 进而将剩余比特由公式 (1) 映射得到非 ROI 的 QP 大小 QP_{n-roi} 。算法伪代码如图 3 所示。

4 实验分析

4.1 实验环境

本文在真实环境中搭建测试平台

验证 ROI 编码模式视频流传输性能, 测试平台包含网络仿真器、视频编码器、视频解码器三部分。网络仿真器采用 HoloWAN 网络模拟器构建异构网络下的 ROI 编码模式视频流传输环境, HoloWAN 将记录的链路状态信息以文件方式写入、复现实际网络情况; 视频编码器利用 X264 编码器将 $1920 \times 1080P$ YUV 格式视频源以每秒 25 帧速率编码, 关键帧间隔设置为 24, 每帧的延迟截止时间为 250ms; 解码端基于 FFMpeg 完成实现视频解码器。

4.2 ROI 划分

在视频序列中, 感兴趣区域 (ROI) 是人们关注的重点, 通常可以使用目标检测算法进行检测。YOLO v3 是目前比较流行的目标检测算法之一。本文以 YOLO v3 作为目标检测的框架, 在 COCO 数据集上训练完成之后, 读取视频 YUV 序列, 然后将图片逐帧送进 YOLO 目标模型进行目标检测, 获取每一帧的感兴趣区域的像素坐标。

4.3 对比

本文将上述提出的基于 ROI 的码率控制算法、RRCS 算法与图像组 (GOP) 级别的码率控制算法^[9]以及帧级别 (Frame) 的码率控制算法在端到端时延、超时丢失率和加权 PSNR 等衡量指标下进行性能对比。

(1) GOP 级别码率控制算法

GOP 级别算法是根据多路传输下各个链路的传输总量, 在图像组级别进行量化参数的设定, 本文将该方法称之为 GRCS。

(2) Frame 级别码率控制算法

Frame 级别算法利用链路各个参数获得帧比特上界, 利用多链路下多个参数对帧大小进行决策, 进而映射

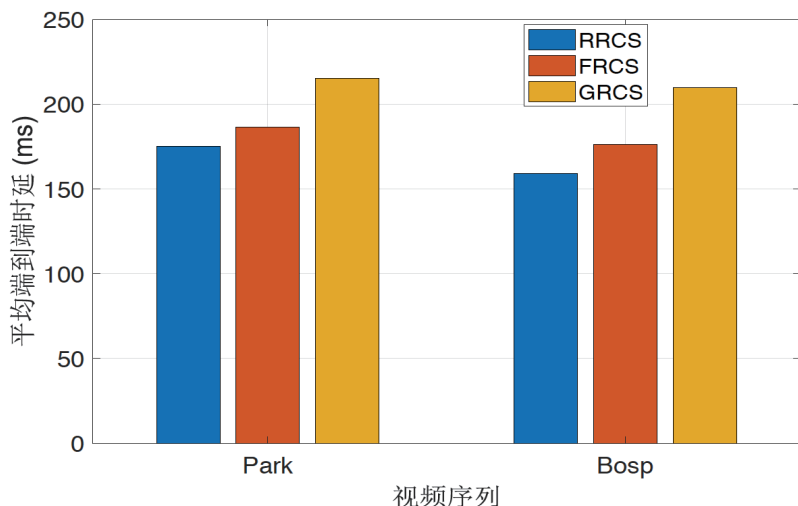


图4 Park、Bosp序列中平均端到端时延对比

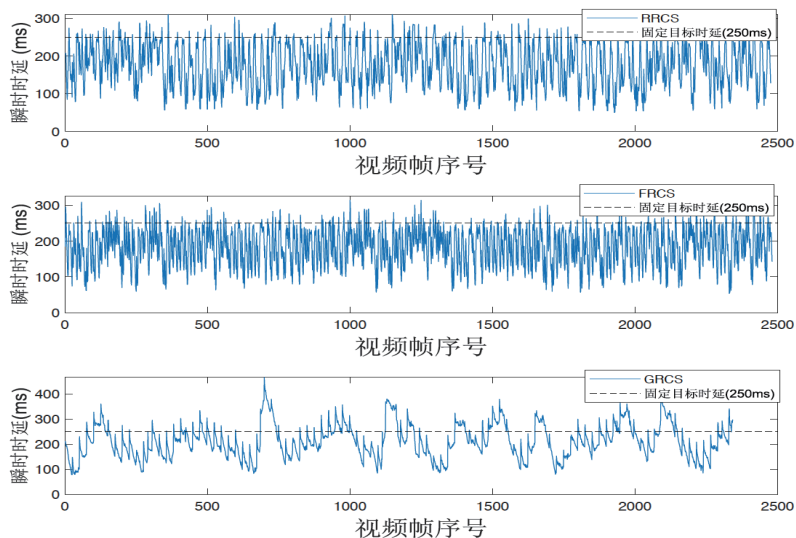


图5 0~2500帧Park序列端到端时延对比

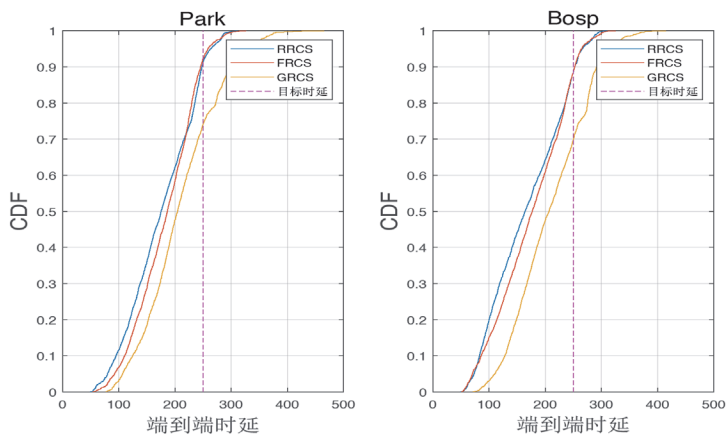


图6 Park序列及Bosp序列CDF图

得到帧的量化参数大小，本文将该方法称之为FRCS。

4.4 实验结果与分析

(1) 端到端时延

图4展示了Park序列、Bosp序列在不同算法的端到端时延变化情况。横坐标为序列名称，纵坐标为视频帧端到端时延取得的平均时延数值。可以看出，本文提出的ROI算法和FRCS算法在端到端时延性能上要优于GRCS，GRCS在GOP级别下进行码率控制，其编码出的关键帧尺寸可能较大，导致更大的端到端时延。图5展示了Park序列0~2500帧的瞬时端到端时延情况，在固定目标时延250ms下，本文提出的算法能够保持较好的性能。图6是端到端时延的CDF图，可以看到，在Park序列中，ROI保证了接近93%的视频帧准时到达率，而GRCS仅仅为70%。

(2) 超时丢失率

图7绘制了Park序列和Bosp序列在各个算法下的平均超时丢失率。

(3) PSNR

图8给出了Park序列和Bosp序列在0~2500帧的加权平均PSNR性能。可以看出，不管是在Park序列下还是Bosp序列下，RRCS算法总是能获得最优加权PSNR。

(4) Park序列原始帧及各种码率

控制算法下的接收帧视觉效果如图9所示。与左上角原始视频帧相比，本文所提出的RRCS码率控制算法的接收帧具有良好的视觉质量，虽然非ROI背景存在一些失真，但人们重点关注的ROI区域（自行车、人物）能够被清晰的显示；FRCS算法下人物手臂及面部表情失真严重；GRCS算法则产生了更加严重的失真。因此，RRCS的主观视频质量优于其他两种算法。

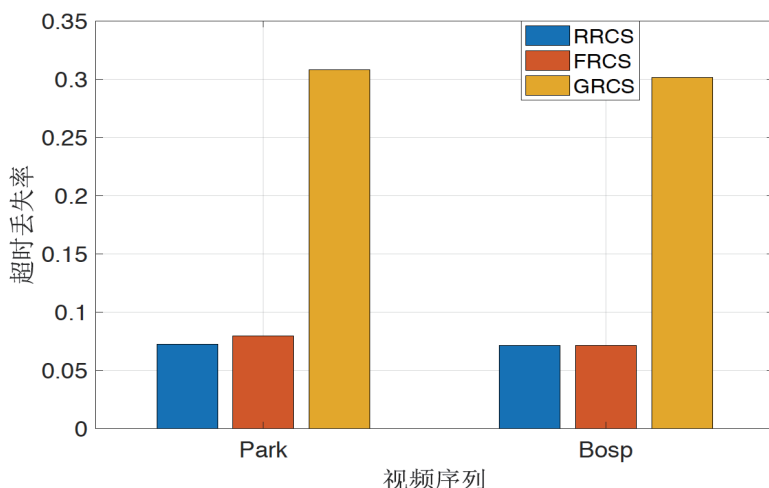


图7 Park 序列及 Bosp 平均超时丢失率

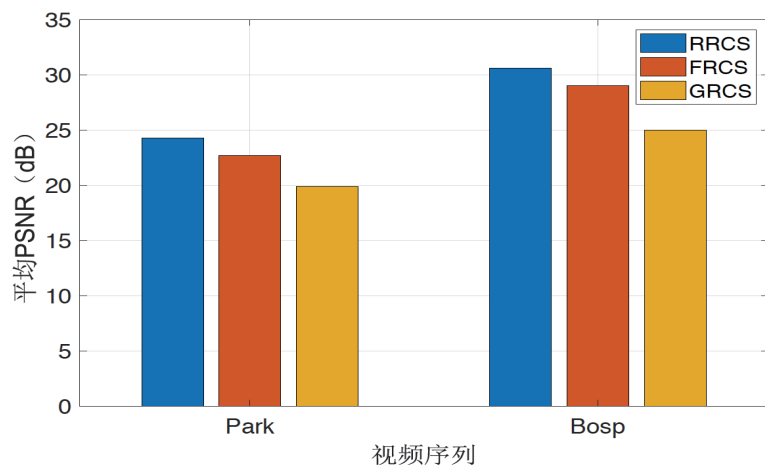


图8 平均 PSNR 图



图9 视觉效果图

5 结束语

为了保证用户对于实时视频传输高清、低延时需求,本文提出了一种基于 ROI 的码率控制方法。通过建模和提取感兴趣区域,根据异构网络各个链路参数,实现了 ROI 编码模式下对不同区域宏块的比特资源的差异分配,采用注水方法将感兴趣区域视频流分发到多个链路,降低多路同传的端到端时延,并将提出的算法在感兴趣区域视频传输系统中进行测试,验证所设计的算法在时延、超时丢失率、加权 PSNR 等方面性能,实验结果表明所提方法各项指标均优于现有方案。

参考文献

- [1]Ye Y,Andrion P.The Scalable Extensions of HEVC for Ultra-High-Definition Video Delivery[J].IEEE Multimedia,2014,21(3):58-64.
- [2]Mahdian M,Yeh E M.Throughput and Delay Scaling of Content-Centric Ad Hoc and Heterogeneous Wireless Networks[J].IEEE/ACM Transactions on Networking,2017,25(5):3030-3043.
- [3]张元东.基于感兴趣区域的高性能视频编码算法优化研究[D].吉林:东北电力大学,2020.
- [4]H. Song,C.-C. J.Kuo.A Region-Based H.263+ Codec and Its Rate Control for Low VBR Video[J]. IEEE Transactions on Multimedia,2004,6(3):489 - 500.
- [5]Redmon J ,Farhadi A.YOLOv3:An Incremental Improvement[J].arXiv e-prints,2018,4.
- [6]L. Yang,L. Zhang,S. Ma,et al.A ROI quality adjustable rate control scheme for low bitrate video coding[C].Chicago,IL,USA:IEEE,2009,5.
- [7]Zhang Z,Jing T,Han J,et al.ROI-Based Video Transmission in Heterogeneous Wireless Networks With Multi-Homed Terminals[J].IEEE Access,2017(5):26328-26339.
- [8]Wu J,Yuen C,Cheung N M,et al.Delay-Constrained High Definition Video Transmission in Heterogeneous Wireless Networks with Multi-Homed Terminals[J].IEEE Transactions on Mobile Computing,2016,15(3):641-655.
- [9]吴冀衍,程渤,南国顺,等.面向异构无线网移动视频传输的联合信源信道编码方式[J].计算机学报,2015,38(2):439-454.

RTN