

# VR 全景视频时空切片传输的流量评估与建模

韩圣千, 娄函, 王君来

(北京航空航天大学电子信息工程学院, 北京 100191)

**摘要:** 为了刻画虚拟现实 (VR) 全景视频的传输流量需求, 针对时空切片传输模式, 首先, 基于 VR 全景视频观看数据集, 对传输流量评估与空间切块方式和时间分段方式之间的关系进行仿真。然后, 基于仿真结果, 对 VR 全景视频的传输流量进行统计建模, 建立传输流量与空间切块数量和时间分段长度之间的函数关系。最后, 基于得到的传输流量模型, 提出一种 VR 全景视频时空切片方式的嵌套优化方法, 其外层对空间切块数量进行遍历寻优, 内层直接计算最优的时间分段长度。研究和仿真结果表明, 传输流量近似服从正态分布, 其均值和方差与时空切片方式有关; 通过对不同类型的视频分别进行传输流量统计建模, 基于所建立的流量模型优化得到的时空切片方式与通过仿真搜索得到的最优切片方式性能接近。

**关键词:** 虚拟现实; 全景视频; 流量模型; 时空切片

**中图分类号:** TN92

**文献标志码:** A

**DOI:** 10.11959/j.issn.1000-436x.2022139

## Traffic evaluation and modeling for spatio-temporal slicing based VR panoramic video transmission

HAN Shengqian, LOU Han, WANG Junlai

School of Electronic Information Engineering, Beihang University, Beijing 100191, China

**Abstract:** To characterize the traffic load of virtual reality (VR) panoramic video under the spatio-temporal slicing transmission mode, firstly, based on a real-world VR video viewing dataset, the relationship between the traffic load and the spatio-temporal slicing strategy was evaluated via simulations. Secondly, on the basis of the simulation results, a statistical traffic model was established, describing the interaction between the traffic load and the number of spatial tiles as well as the duration of temporal chunks. Finally, based on the established model, a hierarchical spatial-temporal slicing strategy optimization method was proposed, where the outer layer found the optimal number of spatial tiles by exhaustive searching while the inner layer could directly compute the optimal duration of temporal chunks. Research and simulation results show that the traffic load approximately follows a normal distribution, where the mean and variance are functions of the spatio-temporal slicing strategy. By establishing the traffic model for each type of videos, the optimized slicing strategy based on the established model performs close to the optimal strategy obtained by searching by simulations.

**Keywords:** virtual reality, panoramic video, traffic model, spatio-temporal slicing

### 0 引言

虚拟现实 (VR, virtual reality) 全景视频被认为是未来移动通信系统高带宽需求的主要应用场景之一<sup>[1-2]</sup>。为了满足良好的用户观看体验, VR 全景视频具有很高的画质要求, 对未来移动通信系统的传

输性能提出了严峻挑战<sup>[3]</sup>。

基于时空切片的 VR 全景视频传输是一种减轻通信系统传输压力的有效手段之一<sup>[4]</sup>。图 1 给出了 VR 全景视频时空切片传输示意, 其中, 空间切块参数  $t=3$ , 时间分段长度  $d=3$ 。其基本思想如下。1) 在空间维, 考虑到用户只会观看全景画面的一部分, 而

收稿日期: 2022-03-28; 修回日期: 2022-06-28

基金项目: 国家自然科学基金资助项目 (No.61871015)

Foundation Item: The National Natural Science Foundation of China (No.61871015)

未被观看的部分可以不传输，为此将每帧画面空间切分为视频块 (Tile)，然后根据用户当前的视点，只传输视域 (FoV, field of vision) 范围内的切块<sup>[5]</sup>。2) 在时间维，为了适应通信系统传输速率的动态变化，视频通常被切分为多个视频段，每一段进行独立的压缩编码，作为一个独立单元进行传输，例如基于 HTTP 的动态自适应流 (DASH, dynamic adaptive streaming over HTTP) 技术。基于时空切片传输，系统只需在每个时间分段里传输视域范围内的空间切块，可以大大减轻系统传输负担<sup>[6]</sup>。

针对空间切块传输，文献[7]研究了切块数量对画面质量的影响，比较了不同切块数量下的画面峰值信噪比 (PSNR, peak signal to noise ratio)。文献[8]观察到用户在观看 VR 视频时很少关注画面的上方和下方，提出了非均匀空间切块方式，结果表明非均匀空间切块可以节省更多的带宽，但会降低视频压缩的性能。文献[9]基于观众对视频画面不同区域的访问概率以及画面内容，提出了一种自适应空间切块方式，可以有效降低系统的带宽需求，但实现复杂度较高。文献[10]考虑用户视点的非理想预测，评估了空间切块传输相对于非切块传输对视角内 PSNR 的提升作用。针对时间分段传输，文献[11]研究了在给定压缩编码率条件下，时间分段的长度对全景视频画面质量的影响。文献[12]针对空间切块传输，对每个时间分段的视频编码率进行了优化。文献[13]在每个时间分段内为视点中心和边缘

的空间切块设置不同的视频编码率，从而降低传输流量。针对传统二维视频，文献[14-15]研究了在给定时间分段方式时如何根据通信条件的动态变化来自适应地选择每个分段的视频质量。

现有文献在研究时空切片传输方式如何影响 VR 全景视频传输流量时，主要考虑了传输画面减少带来的流量节省<sup>[16-17]</sup>。然而，空间切块和时间分段的方式对传输流量有着复杂的影响。首先，空间切块并不一定越多越好。一方面，切块数量越多，实际传输的画面面积越小，这有利于节省流量；另一方面，每个空间切块都是独立压缩编码的，画面切分得越小，则利用画面空间相关性进行视频压缩的效率越低，导致每个切块的压缩率下降。综合考虑两方面因素可知，增加切块数量并不一定导致传输流量下降。其次，时间分段也不一定越长越好。一方面，分段越长，则利用画面时间相关性进行视频压缩的效率越高，有利于降低视频传输流量；另一方面，对于时空切片传输而言，增加时间分段长度可能会导致传输更多的切块，从而增加传输流量。例如，在图 1 中，每个时间分段长度包含 3 帧画面，如果用户在观看整个分段时视点保持不变，如时间分段 1，则只需传输 FoV 包含的 {4,5,7,8} 这 4 个空间切块即可；但是如果用户在观看过程中发生了头动，如时间分段 2，则此时需要传输 {4,5,6,7,8,9} 这 6 个空间切块。时间分段越长，那么用户在一个时间分段内发生头动的概率越大，可见

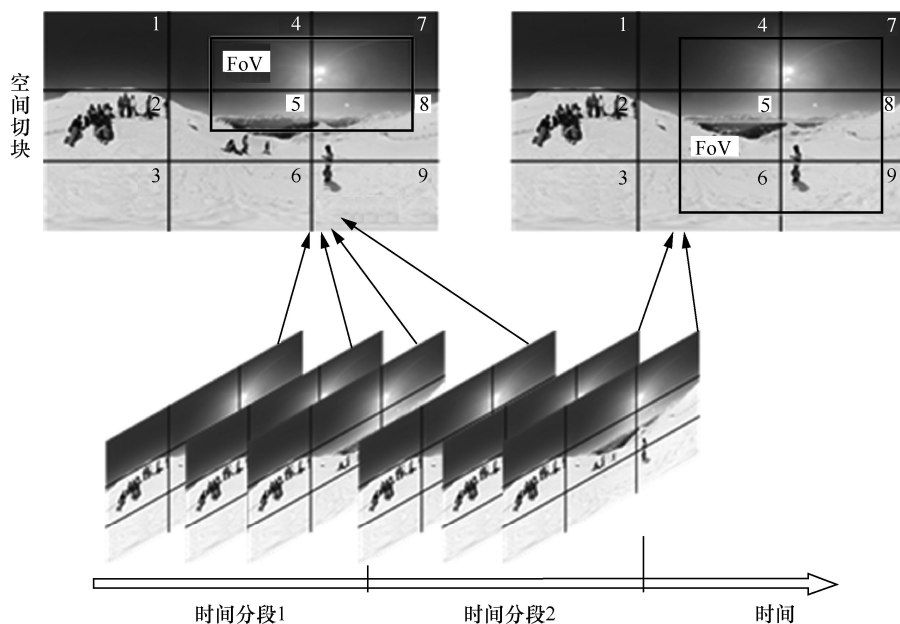


图 1 VR 全景视频时空切片传输示意

时间分段长度并不一定越大越好。

为了分析和刻画 VR 全景视频的传输流量需求与时空切片方式之间的复杂关系, 首先, 基于实际的 VR 全景视频观看数据集, 通过仿真来分析和评估时空切片方式对传输流量需求的影响。然后, 基于仿真结果, 对传输流量需求进行统计建模, 建立统计模型参数与时空切片方式之间的函数关系。最后, 基于建立的模型, 提出一种时空切片方式的优化方法, 最小化 VR 全景视频的传输流量。仿真结果表明, 基于所建立的模型得到的时空切片方式可以获得接近最优的性能。

## 1 VR 全景视频的传输流量需求评估

本文采用开源 VR 全景视频观看数据集进行仿真分析, 文献[18]对相关数据集进行了总结和对比。通过比较各个数据集的规模(定义为用户数 $\times$ 视频数 $\times$ 时间长度), 本文选用其中规模较大的数据集<sup>[19]</sup>。该数据集在各研究中得到广泛应用, 例如文献[20]使用该数据集进行 VR 用户的头动和视点预测; 文献[21]利用该数据集中不同用户观看相同视频的头动相关性来预测用户视点; 文献[22]基于该数据集分析了不同用户观看视频时的头动相关性, 并对提出的头动预测方法进行了性能评估; 文献[23]采用该数据集对提出的基于深度强化学习的全景视频传输策略进行训练和用户体验质量(QoE, quality of experience)评估; 文献[24]提取该数据集中的用户头动信息来训练神经网络预测用户视点。

该数据集包含 3 个表演类视频(编号 1、4、7)、3 个运动类视频(编号 2、5、8)、一个电影类视频(编号 3)和 2 个纪录类视频(编号 6、9)。所有视频均未进行时空切片, 投影方式为等距柱状投影(ERP, equi-rectangular projection), 分辨率为 2 560 像素 $\times$ 1 440 像素, 帧率为 30 帧/秒, 视频时长范围为 120~160 s。每个视频包含 48 个用户的观看头动记录, 每条记录由时间戳、描述头盔显示设备旋转的四元数, 以及描述头盔显示设备平移的空间坐标等元素组成。

为了评估传输流量, 需要对数据集中的 VR 全景视频进行时空切片, 然后对每个时空视频块进行压缩编码, 并在此基础上引入用户头动数据, 最终获得在任意给定时空切片方式下的实际传输流量数据。为此, 本文采用 FFmpeg 的裁剪(Crop)功能<sup>[25]</sup>来实现画面的空间切块。当空间切块参数为  $t$

时, 画面横向和纵向分别被等分为  $t$  块, 得到总切块数为  $t^2$ 。视频的时间分段采用 H264 编码器来实现, 当分段长度为  $d$  帧时, 则指定 H264 编码器<sup>[26]</sup>的画面组(GoP, group of picture)参数为  $d$ 。其中, 对于 VR 全景视频, 为了降低解码时延, 在进行压缩编码时通常不采用 B 帧, 而只采用 I 帧和 P 帧, 因此每个 GoP 由一个 I 帧和  $d-1$  个 P 帧组成。完成编码的时空视频块经过 MP4 Dash 工具<sup>[27]</sup>处理后, 转换为流媒体格式进行存储。对于用户观看每个视频的视点数据, 采用数据集中头盔显示设备的四元数换算得到。首先, 由数据集的四元数换算为空间三维坐标系中的用户视点方向向量; 然后, 根据 ERP 格式全景视频的映射规则得到平面二维坐标系中的用户视点坐标。最后, 根据所得的视点坐标, 从画面中提取对应的空间切块, 用于后续的仿真评估。

### 1.1 空间切块对全景视频压缩率的影响

如前文所述, 增加空间切块数量会减小每个切块画面的大小, 导致利用画面空间相关性进行视频压缩的效率降低。对于任一视频, 给定空间切块参数  $t$ , 分别设置时间分段长度  $d$  为 30 帧, 对该视频进行时空切片并进行压缩编码, 得到压缩后的视频大小。计算该视频在 30 帧分段长度  $d$  下的视频大小的均值, 记为  $Q_t$ 。表 1 给出了运动类和表演类视频切块后 ( $t > 1$ ) 视频大小相对于未切块 ( $t = 1$ ) 视频大小的增长百分比, 即  $\frac{Q_t - Q_1}{Q_1}$ 。

从表 1 可知, 切块后视频的大小相对于未切块视频均有不同程度的增加。当切块参数为  $t = 6$  时, 视频大小增幅最高接近 12%。可以发现, 随着切块数的增加, 视频大小并非单调增加。这是因为不同的切块数下, 每个空间切块包含的画面内容会发生变化。例如,  $t = 4$  时的切块画面不一定包含在  $t = 3$  时的切块画面里。但是, 由于  $t = 4$  时的切块是对  $t = 2$  时的切块的二等分, 可知  $t = 4$  时的切块画面一定包含在  $t = 2$  时的切块画面里, 导致  $t = 4$  时的视频大小一定比  $t = 2$  时大。同理,  $t = 6$  时的视频大小一定比  $t = 3$  时大。

### 1.2 时间分段对视频压缩率和头动次数的影响

通过分析图 1 可知, 增大时间分段有助于提高视频的压缩效率。图 2 给出了不同时间分段长度(即 GoP 长度)的归一化视频大小, 其中归一化因子是分段长度为 1 帧时的视频大小。从图 2 中可以看出, 随着分段长度的增加, 视频的压缩率首先会快速提升, 然后逐渐放缓。

表 1 切块视频相对于未切块视频大小的增长百分比

视频编号	类型	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
视频 2	运动类	0.77%	9.28%	5.70%	2.86%	11.92%
视频 5	运动类	0.58%	9.11%	4.87%	2.52%	10.48%
视频 8	运动类	0.36%	9.55%	4.04%	1.79%	11.54%
视频 1	表演类	0.33%	4.81%	2.97%	1.10%	5.84%
视频 4	表演类	0.31%	4.40%	2.67%	1.07%	5.37%
视频 7	表演类	0.23%	6.87%	2.98%	1.39%	8.61%

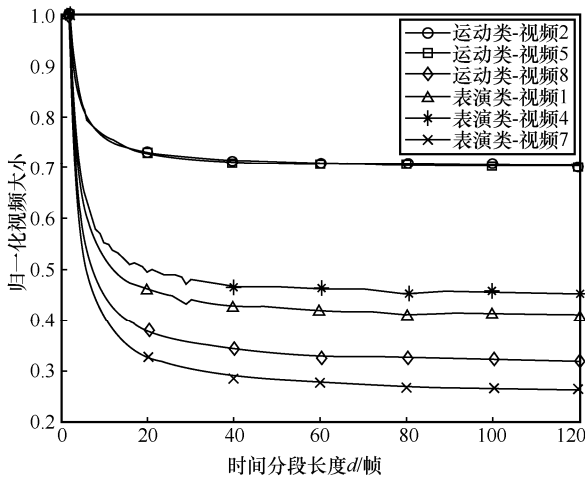


图 2 不同时间分段长度的归一化视频大小

时间分段长度增加的另一方面影响是增加用户在观看一个分段过程中发生头动的概率, 进而导致需要传输的空间切块增多。图 3 对此进行了评估。首先, 在评估中, 头动发生的定义是用户在 1 ms 内视点偏移大于 1°。在此定义下, 给定时间分段长度, 针对每个视频包含的所有时间分段, 提取对应的 48 个用户的观看头动记录, 分别统计所有分段内头动发生的次数, 进而计算头动次数的均值并作为该视频在当前分段长度下的头动次数。

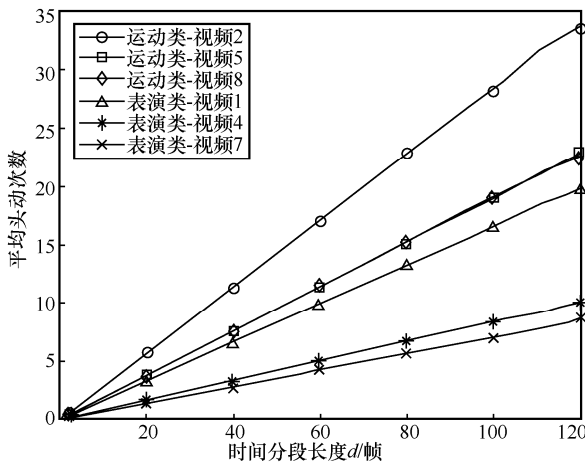


图 3 不同时间分段长度内的平均头动次数

从图 3 可以看出, 对于所有的视频, 一个时间分段内的平均头动次数随着时间分段长度的增加线性增长, 说明需要传输的空间切块数随着时间分段长度的增加相应增长。对于表演类视频, 用户视点主要集中在舞台上的表演, 用户发生头动的频次较少。相比而言, 用户在观看运动类视频时, 由于拍摄画面切换得更加频繁, 用户头动次数明显增多。

### 1.3 时空切片下全景视频传输流量的评估

从以上的评估分析可以看出, 空间切块虽然可以减少传输的画面内容 (仅传输视域内的画面), 但会导致视频压缩效率降低, 而增大时间分段长度虽然可以提高视频压缩效率, 但同时也会增加需要传输的空间切块数。下面对全景视频在时空切片模式下的传输流量进行仿真评估。

考虑到用户在观看不同类型的视频时具有不同的头动响应, 而用户头动直接影响时空切片模式下传输的切片数量, 因此在仿真中使用 4 个不同类型的视频, 即运动类视频 2、表演类视频 7、电影类视频 3 和纪录类视频 6。

图 4 给出了数据集中 48 个用户在观看运动类视频 2 时, 视频采用不同的空间切块参数 ( $t=1 \sim 11$ ) 和不同的时间分段长度 ( $d=1 \sim 60$ ) 进行时空切片时的平均传输流量。从图 4 中可知, 随着  $t$  的增加, 传输流量并非一致增加。这与第 1.1 节的分析一致, 说明当  $t$  较小时, 空间切块带来的传输画面减少起主导作用, 而当  $t$  较大时, 空间切块导致的压缩率下降起主导作用。对于时间分段, 可以看出当空间切块参数较大 (如  $t \geq 2$ ) 时, 传输流量随着时间分段长度的增加先下降后上升。这是因为当分段较短时, 分段长度对压缩率的影响显著, 如图 2 所示; 而当分段较长时, 分段长度增加导致的传输切块数增多的影响起主导作用。从图 4 中可以看出, 运动类视频 2 在  $t=10, d=11$  时可以获得最低的平均传输流量, 相比于未进行时空切片 ( $t=1, d=1$ ) 的传输流量节省 77.1%。

图 5 给出了表演类视频 7 的平均传输流量。根据对图 3 的分析结果, 用户在观看表演类视频时的头动频率低于运动类视频。这直接导致在给定时间分段长度内, 表演类视频需要传输更少的空间切块, 带来比表演类视频更低的传输流量。从图 5 中可以看出, 表演类视频在  $t=10, d=27$  时可以获得最低的平均传输流量, 相比于未进行时空切片 ( $t=1, d=1$ ) 的传输流量节省 90.5%。

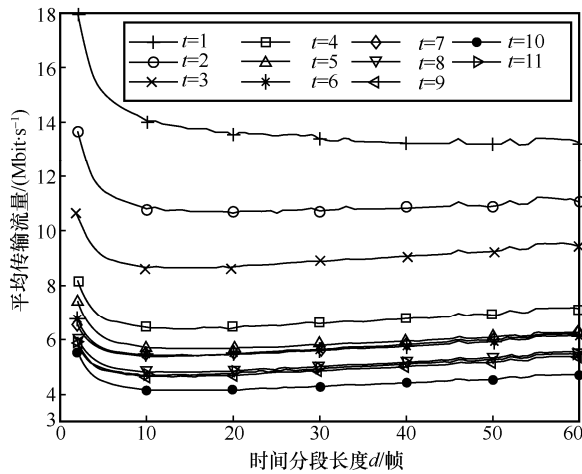


图 4 运动类视频 2 的平均传输流量

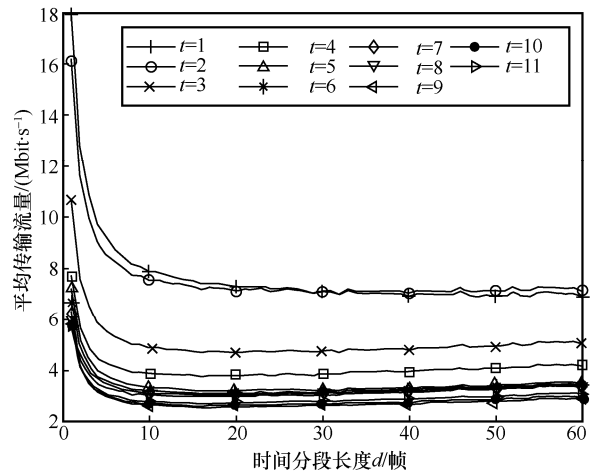


图 7 纪录类视频 6 的平均传输流量

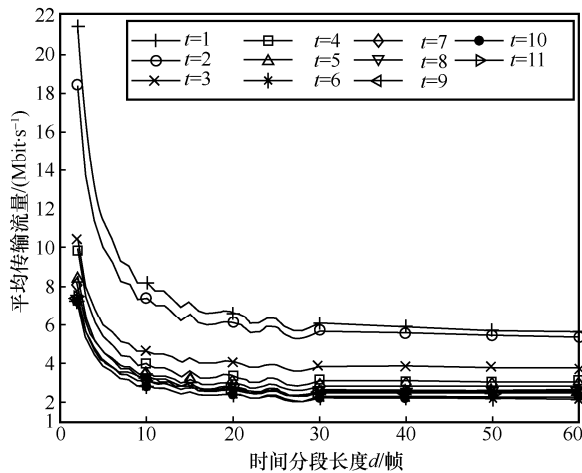


图 5 表演类视频 7 的平均传输流量

图 6 和图 7 给出了电影类视频 3 和纪录类视频 6 的平均传输流量，从图 6 和图 7 中可以看出与表演类和运动类视频相似的规律，即  $t$  和  $d$  并非越大越好。例如，纪录类视频 6 的最优时空切片参数是  $t = 9$ ， $d = 17$ 。

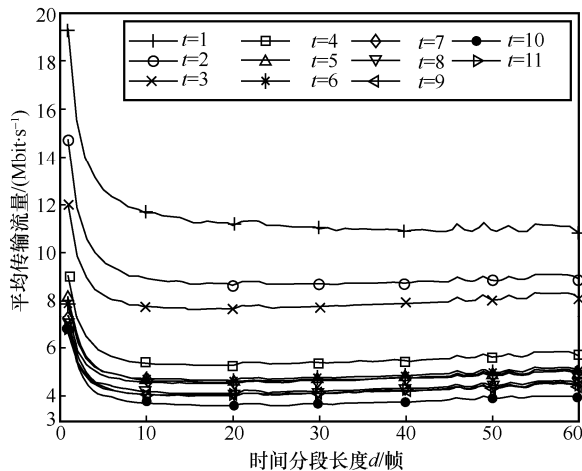


图 6 电影类视频 3 的平均传输流量

## 2 时空切片传输下的流量建模

基于 1.3 节的传输流量仿真评估，本节研究全景视频时空切片传输模式下的流量统计建模。传输流量的建模对象是实际传输给用户的视角范围内的时空切块大小，例如图 1 左上方的视频帧只需传输 4、5、7、8 切块。可以观察到，传输的切块总大小与 2 个因素有关，一是完整视频的大小，二是传输给用户的视角内切块大小占完整视频的比例；两者的乘积即实际传输给用户的业务流量。基于这一观察，下面采用 2 个步骤进行建模。首先，针对不同的空间切块数量和时间分段长度，对压缩编码之后的完整视频文件大小进行建模。其次，针对不同的空间切块数量和时间分段长度，根据用户的头动数据，得到用户在实际观看过程中视域范围内的切块大小占完整视频文件大小的比例，并对比例进行统计建模。最后，把完整视频文件大小的模型与视域内切块大小占比的模型相乘，即可得到实际传输给用户的视频切块大小的模型，即传输流量模型。

需要说明的是，根据 1.3 节的评估结果，不同类型的视频具有不同的用户头动规律以及不同的最优时空切片方式，因此在传输流量建模时有必要针对每类视频单独进行建模。本节提出的传输流量建模方法适用于任意类型的视频，在应用时只需根据不同类型视频的流量样本数据拟合出对应的模型参数。为了评估提出的建模方法的合理性，本节将以运动类视频（其数据集规模大）为例对模型的准确性进行评估。

### 2.1 视频大小建模

考虑数据集中的某一类视频，在上述仿真过程

中已经采用了不同的时空切片方式对这些视频进行了切分、编码、存储，由此可以得到任一时空切片方式下 I 帧和 P 帧的平均大小。这里的平均是指对所有同类视频的所有 I 帧或所有 P 帧进行平均。

定义  $\bar{S}_I(t)$  和  $\bar{S}_P(t, d)$  分别表示当空间切块参数为  $t$ 、时间分段长度为  $d$  时 I 帧和 P 帧的平均大小。考虑到 I 帧是单帧独立压缩编码的，因此  $\bar{S}_I(t)$  与分段长度  $d$  无关。由于每个时间分段内包含一个 I 帧和  $d-1$  个 P 帧，因此视频的平均大小可以表示为

$$\bar{S}(t, d) = \frac{L}{d} (\bar{S}_I(t) + (d-1)\bar{S}_P(t, d)) = L\bar{S}_I(t) \left( \frac{1}{d} + \frac{d-1}{d\bar{W}(t, d)} \right) \quad (1)$$

其中， $L$  表示视频包含的总帧数， $\frac{L}{d}$  表示视频包含的时间分段数， $\bar{W}(t, d)$  表示 I-P 帧平均压缩比，定义为

$$\bar{W}(t, d) = \frac{\bar{S}_I(t)}{\bar{S}_P(t, d)} \quad (2)$$

图 8 给出了不同时空切片方式下运动类视频的平均压缩比  $\bar{W}(t, d)$ ，可以看出当时间分段长度  $d$  改变时， $\bar{W}(t, d)$  基本保持不变，因此可以从  $\bar{W}(t, d)$  中去掉因变量  $d$ ，记为  $\bar{W}(t)$ 。同时，从图 8 中可知  $\bar{W}(t)$  与  $t$  之间近似满足线性关系，因此可以采用如下的线性模型来拟合  $\bar{W}(t)$

$$\bar{W}(t) = p_0 t + p_1 \quad (3)$$

其中，系数  $p_0$  和  $p_1$  可以由现有的线性拟合算法得到。表 2 给出了运动类视频的拟合系数。

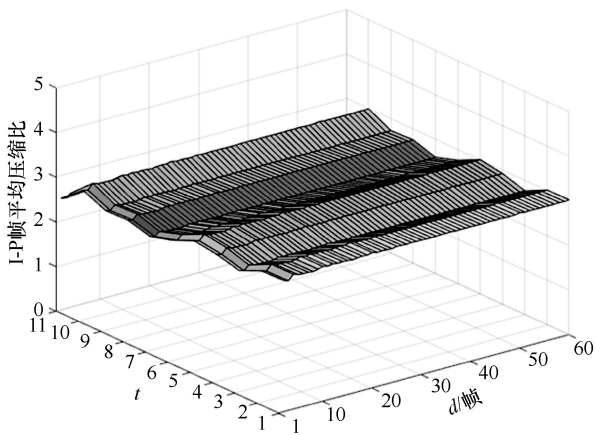


图 8 不同时空切片方式运动类视频的平均压缩比

表 2 运动类视频的拟合系数

系数	数值	系数	数值
$p_0$	-0.15	$p_{\mu,5}$	$3.39 \times 10^{-4}$
$p_1$	3.723	$p_{\mu,6}$	$-2.68 \times 10^{-5}$
$p_{\mu,-3}$	7.217	$p_7$	$-1.97 \times 10^{-4}$
$p_{\mu,-2}$	-17.17	$p_8$	0.01
$p_{\mu,-1}$	14.81	$p_9$	0.077
$p_{\mu,0}$	-4.62	$p_{10}$	0.32
$p_{\mu,1}$	0.81	$p_{11}$	-0.41
$p_{\mu,2}$	-0.066	$p_{12}$	$1.48 \times 10^{-4}$
$p_{\mu,3}$	0.002	$p_{13}$	$2.41 \times 10^{-5}$
$p_{\mu,4}$	$3.69 \times 10^{-4}$	—	—

将式(3)代入式(1)，可得视频的平均大小为

$$\bar{S}(t, d) = L\bar{S}_I(t) \left( \frac{1}{d} + \frac{d-1}{d(p_0 t + p_1)} \right) \quad (4)$$

在式(4)中， $\bar{S}_I(t)$  表示空间切块参数为  $t$  时的 I 帧平均大小。如表 1 所示，不同的切块参数  $t$  可能会导致切块内的画面内容大幅变化，使视频压缩率与切块参数  $t$  之间的关系具有较强的随机性，难以刻画  $\bar{S}_I(t)$  与  $t$  的函数关系。为此，采用查表法对  $\bar{S}_I(t)$  进行建模，即把每一类视频在不同  $t$  时的 I 帧平均大小以表格方式存储，如表 3 所示。

表 3 不同空间切块参数下 I 帧平均大小  $\bar{S}_I(t)$

切块参数 $t$	数值/KB
1	763
2	766
3	774
4	777
5	778
6	788
7	805
8	803
9	806
10	808
11	828

为了验证所建立模型的准确性，图 9 比较了式(4)中括号部分的理论计算结果和实际仿真结果。括号部分可以理解为归一化平均视频大小，其中归一化因子是  $L\bar{S}_I(t)$ ，即当所有帧均为 I 帧时的视频大小。从图 9 中可以看出，对于所有的切块参数  $t$ ，当分段

长度  $d$  较大时 (例如  $d > 15$ ), 由式(4)得到的计算结果与仿真结果很接近, 说明了模型的有效性。

### 2.2 传输的切块大小占比建模

在前述仿真中, 通过引入用户的头动数据, 得到了在任一时空切片方式下, 用户在观看每个视频时, 在每个时间分段内传输的视域内空间切块的大小。由此, 可以得到任一用户在观看所有同类视频时, 传输的空间切块大小与视频文件总大小之间的平均比例, 记为  $\eta(t, d)$ , 可以表示为

$$\eta(t, d) = \frac{1}{V} \sum_{v=1}^V \frac{\sum_{i=1}^{\frac{L}{d}} \sum_{j=1}^{t^2} \alpha_{v,ij} S_{v,ij}}{\sum_{i=1}^{\frac{L}{d}} \sum_{j=1}^{t^2} S_{v,ij}} \quad (5)$$

其中,  $V$  表示同类视频的个数,  $\alpha_{v,ij} \in \{0,1\}$  表示视频  $v$  的第  $i$  个时间分段内第  $j$  个空间切块是否位于用户的视域内,  $S_{v,ij}$  表示视频  $v$  的第  $i$  个时间分段内第  $j$  个空间切块的大小, 因此等号右侧分子项表示需要传输给用户的视域内所有切块的总大小, 分母项表示视频文件的总大小。

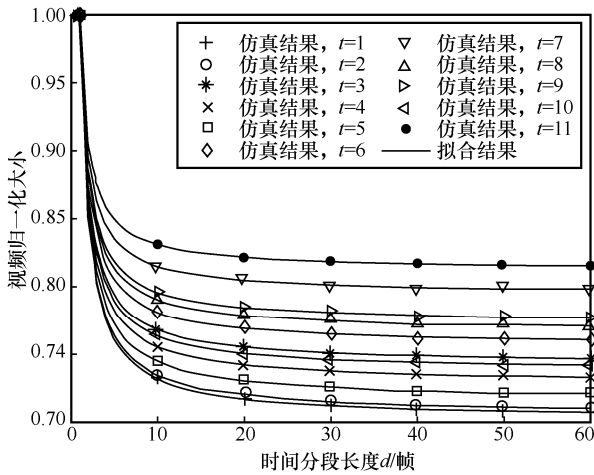


图 9 归一化视频平均大小的模型准确性

图 10 和图 11 给出了数据集中 48 个用户的平均传输比例  $\eta(t, d)$  的频率直方图和拟合结果, 其中前者固定  $d = 15$  而考虑不同的  $t$ , 后者固定  $t = 3$  而考虑不同的  $d$ 。从图 10 和图 11 中可以看出,  $\eta(t, d)$  的统计分布接近正态分布。因此, 采用如下的正态分布对  $\eta(t, d)$  进行建模

$$\eta(t, d) \sim \mathcal{N}(\mu(t, d), \sigma^2(t, d)) \quad (6)$$

其中, 均值和方差均与时空切片方式有关。

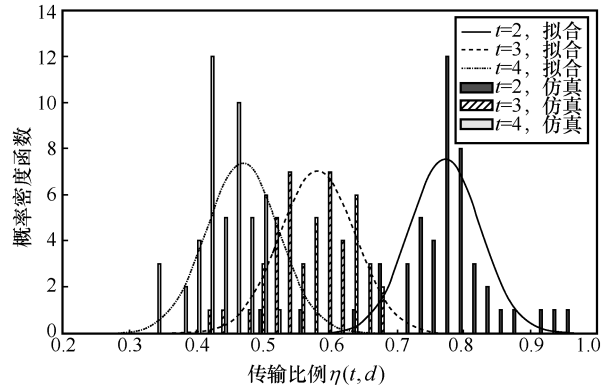


图 10 传输比例  $\eta(t, d)$  的频率直方图和拟合结果 ( $d = 15$ )

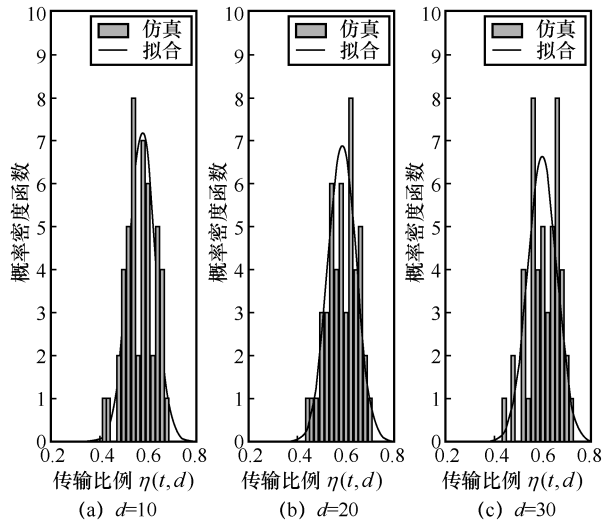


图 11 传输比例  $\eta(t, d)$  的频率直方图和拟合结果 ( $t = 3$ )

为了刻画均值和标准差与时空切片方式的函数关系, 针对  $t$  和  $d$  的不同取值, 通过正态分布拟合找到对应的均值和标准差。均值拟合结果如图 12 所示, 标准差拟合结果如图 13 所示。其中, 黑点代表数据集的仿真结果, 曲面代表拟合结果。

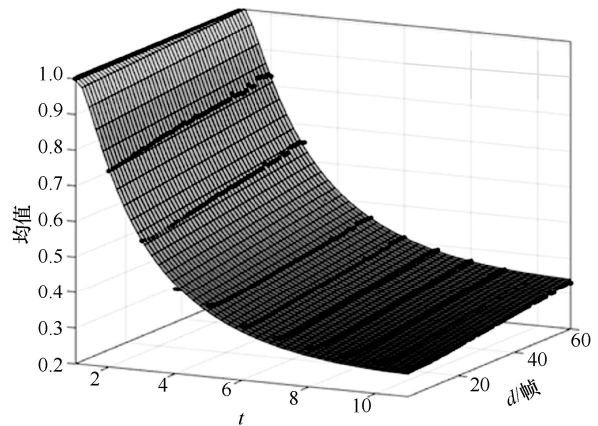


图 12 均值拟合结果

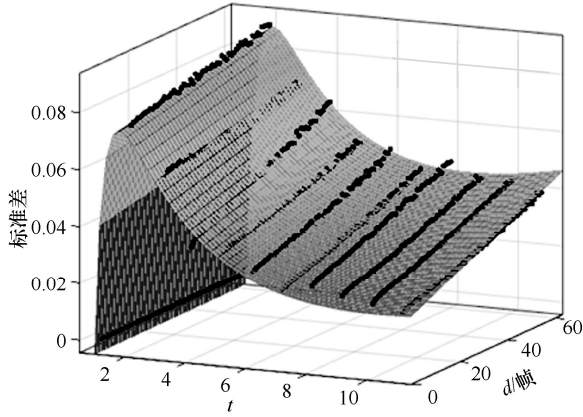


图13 标准差拟合结果

通过对图12和图13进行二维函数拟合,得到均值和标准差与时空切片方式可以近似表示为

$$\mu(t, d) \approx \Omega_1(t) + \Omega_2(t)d \quad (7)$$

$$\sigma(t, d) \approx \Omega_3(t) + \Omega_4(t)d \quad (8)$$

其中,

$$\Omega_1(t) = \sum_{i=-3}^3 p_{\mu,i} t^i \quad (9)$$

$$\Omega_2(t) = p_{\mu,4} + p_{\mu,5}t + p_{\mu,6}t^2 \quad (10)$$

$$\Omega_3(t) = p_7t + p_8 + p_9t^{-1} + p_{10}t^{-2} + p_{11}t^{-3} \quad (11)$$

$$\Omega_4(t) = p_{12} + p_{13}t \quad (12)$$

式(9)~式(12)中的拟合系数可以使用 MATLAB 曲线拟合工具得到,具体内容如表2所示。

将式(7)和式(8)代入式(6),可以得到传输比例  $\eta(t, d)$  的统计分布。从图10和图11中可以看出,所建立的统计模型的拟合结果与频率直方图之间具有较高的匹配度,验证了模型的准确性。

将视频平均大小模型式(4)和切块大小占比模型式(6)相乘,最终得到时空切片模式下全景视频传输流量的统计模型。

本文采用统计建模方法建立了所有用户观看所有同类视频时的传输流量模型。考虑到近期深度学习在 VR 全景视频的用户头动和视点预测<sup>[28]</sup>、QoE建模<sup>[29]</sup>、传输策略优化<sup>[30]</sup>等方面的成功应用,未来本文将研究基于深度学习的传输流量建模方法,采用神经网络挖掘用户的视频观看偏好和视频画面内容的特征,有望更加准确地预测特定用户在观看特定视频时的传输流量。

### 3 基于传输流量模型的时空切片优化

本节考虑一个未进行时空切片的 VR 全景视频,

研究如何利用建立的传输流量统计模型,对其时空切片方式(即空间切块参数  $t$  和时间分段长度  $d$ )进行优化。

#### 3.1 时空切片方式优化

根据待切分视频类别,找到对应类别或接近类别的传输流量模型参数。在“流量建模”和“时空切片优化”过程中需要已知视频类别,其中前者可直接从数据集中获取视频类别,而后者通常可以从视频的元数据中获取,例如典型视频网站在发布视频或用户在上传视频时通常会被要求标注类型等信息。如果待切分视频是未经压缩的原始视频,则可以直接使用已建立的同类视频模型。如果待切分视频已经在未进行空间切块( $t=1$ )的情况下进行了压缩编码,则可以统计出当前视频的 I-P 帧压缩比,记为  $W$ 。据此,可以对统计模型中的 I-P 帧平均压缩比  $\bar{W}(t)$  进行修正,以确保当  $t=1$  时,  $\bar{W}(t)$  等于  $W$ 。修正后的  $\bar{W}(t)$  可以表示为

$$\bar{W}(t) = p_0t + p_1 + A \quad (13)$$

其中,  $A = W - p_0 - p_1$ , 以保证  $\bar{W}(1) = W$ 。

将式(13)代入式(1),可得待切分视频在时空切片后的大小  $\bar{S}(t, d)$ 。进而乘以传输的切块大小占比  $\eta(t, d)$ , 可得用户观看该视频时需要传输的流量满足如下正态分布

$$T(t, d) = \bar{S}(t, d)\eta(t, d) \sim \mathcal{N}\left(\bar{S}(t, d)\mu(t, d), \bar{S}^2(t, d)\sigma^2(t, d)\right) \quad (14)$$

为了区分空间切块参数  $t$  和时间分段长度  $d$  的影响,将式(1)给出的  $\bar{S}(t, d)$  表示为

$$\bar{S}(t, d) = \Omega_5(t)\frac{1}{d} + \Omega_6(t) \quad (15)$$

其中,  $\Omega_5(t) = L\bar{S}_1(t)\left(1 - \frac{1}{\bar{W}(t)}\right)$ ,  $\Omega_6(t) = \frac{L\bar{S}_1(t)}{\bar{W}(t)}$ 。

基于式(7)、式(8)和式(15),均值  $\bar{S}(t, d)\mu(t, d)$  和标准差  $\bar{S}(t, d)\sigma(t, d)$  可以分别表示为

$$\bar{S}(t, d)\mu(t, d) = \Omega_5(t)\Omega_1(t)\frac{1}{d} + \Omega_6(t)\Omega_2(t)d + \Omega_5(t)\Omega_2(t) + \Omega_6(t)\Omega_1(t) \quad (16)$$

$$\bar{S}(t, d)\sigma(t, d) = \Omega_5(t)\Omega_3(t)\frac{1}{d} + \Omega_6(t)\Omega_4(t)d + \Omega_5(t)\Omega_4(t) + \Omega_6(t)\Omega_3(t) \quad (17)$$

从式(16)和式(17)可以看出,首先,空间切块参数

$t$  对模型参数有着复杂的影响, 体现在  $\Omega_1(t) \sim \Omega_6(t)$  等多个函数上; 然后, 在给定空间切块参数  $t$  时, 时间分段长度  $d$  与模型参数的关系式相对简单, 均值和标准差都具有  $\frac{\theta_1}{d} + \theta_2 d + \theta_3$  的形式, 其中系数  $\theta_1 \sim \theta_3$  可由式(16)和式(17)得到; 最后, 均值和标准差都与空间切块参数  $t$  和时间分段长度  $d$  的交叉项有关 (即式(16)和式(17)等号右侧的前两项), 因此它们对模型参数的影响不是简单的叠加, 而存在相互耦合的关系。

对于随机传输流量  $T(t, d)$ , 其置信度为  $\rho \in [0, 1]$  的上界 (记为  $T_\rho(t, d)$ ), 定义为

$$P(T(t, d) \leq T_\rho(t, d)) = \rho \quad (18)$$

基于式(14), 可以得到  $T_\rho(t, d)$  的表达式为

$$T_\rho(t, d) = \bar{S}(t, d)(\sigma(t, d)\phi(\rho) + \mu(t, d)) \quad (19)$$

其中,  $\phi(\rho)$  是标准正态分布的  $\rho$  分位点, 其数值可以通过查表或数值计算等方式得到。

将式(16)和式(17)代入式(19), 并进行同类项合并, 则  $T_\rho(t, d)$  可以表示为

$$T_\rho(t, d) = \Delta_1(t)\frac{1}{d} + \Delta_2(t)d + \Delta_3(t) \quad (20)$$

其中,  $\Delta_1(t) = \phi(\rho)\Omega_3(t)\Omega_3(t) + \Omega_3(t)\Omega_4(t)$ ,  $\Delta_2(t) = \phi(\rho)\Omega_6(t)\Omega_4(t) + \Omega_6(t)\Omega_2(t)$ ,  $\Delta_3(t) = \phi(\rho)\Omega_5(t)\phi(\rho)\Omega_6(t)v_3(t) + \Omega_5(t)\Omega_2(t) + \Omega_6(t)\Omega_1(t)$ 。

时空切片方式的优化目标是 minimized 传输流量上界  $T_\rho(t, d)$ , 可以建模为

$$\begin{aligned} \min \quad & T_\rho(t, d) \\ \text{s.t.} \quad & 1 \leq t \leq t_{\max}, t \in \mathbb{Z} \\ & 1 \leq d \leq d_{\max}, d \in \mathbb{Z} \end{aligned} \quad (21)$$

其中,  $t_{\max}$  和  $d_{\max}$  是系统允许的空间切块和时间分段数的最大值。

由于  $T_\rho(t, d)$  与优化变量  $t$  的关系式很复杂, 同时考虑到空间切块参数  $t$  的取值范围一般不会太大, 因此可以采用遍历的方式对  $t$  进行搜索寻优。对于任意给定的参数  $t$ , 下面计算最优的参数  $d$ 。

由式(20)可得  $T_\rho(t, d)$  关于参数  $d$  的二阶导数为  $\ddot{T}_\rho(t, d) = \frac{2\Delta_1(t)}{d^3}$ , 其正负号在定义域  $1 \leq d \leq d_{\max}$  内只与系数  $\Delta_1(t)$  有关。下面针对  $\Delta_1(t)$ , 分 3 种情况讨论参数  $d$  的最优解。

1)  $\Delta_1(t) = 0$ 。此时  $T_\rho(t, d) = \Delta_2(t)d + \Delta_3(t)$ , 是  $d$

的单调函数。如果  $\Delta_2(t) \geq 0$ , 则  $d$  的最优解为 1, 否则为  $d_{\max}$ 。

2)  $\Delta_1(t) > 0$ 。此时  $T_\rho(t, d)$  是  $d$  的严格凸函数。如果  $\Delta_2(t) \leq 0$ , 则不难得到  $T_\rho(t, d)$  是  $d$  的单调减函数, 因此  $d$  的最优解为  $d_{\max}$ 。如果  $\Delta_2(t) > 0$ , 则  $T_\rho(t, d)$  的最小值位于一阶导  $\dot{T}_\rho(t, d) = 0$  处, 即  $d^* = \sqrt{\frac{\Delta_1(t)}{\Delta_2(t)}}$ 。由于  $T_\rho(t, d)$  是  $d$  的严格凸函数, 可知当  $d < d^*$  时,  $T_\rho(t, d)$  单调递减, 而当  $d > d^*$  时,  $T_\rho(t, d)$  单调递增。由此可知, 如果  $d^* < 1$ , 则  $d$  的最优解为 1; 如果  $d^* \geq d_{\max}$ , 则  $d$  的最优解为  $d_{\max}$ ; 如果  $1 \leq d^* \leq d_{\max}$ , 则分别对  $d^*$  向上取整和向下取整, 计算两者对应的  $T_\rho(t, d)$  值, 其中较小者即  $d$  的最优解。

3)  $\Delta_1(t) < 0$ 。此时  $T_\rho(t, d)$  是  $d$  的严格凹函数。如果  $\Delta_2(t) \geq 0$ , 则不难得到  $T_\rho(t, d)$  是  $d$  的单调增函数, 因此  $d$  的最优解为 1。如果  $\Delta_2(t) < 0$ , 则  $T_\rho(t, d)$  的最小值位于  $d^* = \sqrt{\frac{\Delta_1(t)}{\Delta_2(t)}}$ 。由于  $T_\rho(t, d)$  是  $d$  的严格凹函数, 可知当  $d < d^*$  时,  $T_\rho(t, d)$  单调递增, 而当  $d > d^*$  时,  $T_\rho(t, d)$  单调递减。由此可知, 如果  $d^* < 1$ , 则  $d$  的最优解为  $d_{\max}$ ; 如果  $d^* \geq d_{\max}$ , 则  $d$  的最优解为 1; 如果  $1 \leq d^* \leq d_{\max}$ , 则分别对  $d^*$  向上取整和向下取整, 计算两者对应的  $T_\rho(t, d)$  值, 其中较小者即  $d$  的最优解。

综上所述, 可以采用内外嵌套方法来求解优化问题式(21)。在外层, 对空间切块参数  $t$  进行遍历搜索; 在内层, 针对给定的外层参数  $t$ , 采用上述优化方法直接计算参数  $d$  的最优值, 进而得到在当前外层参数  $t$  下的函数值, 用于遍历寻优。

### 3.2 仿真结果

本节对所提出的时空切片优化方法的性能进行评估, 采用第 2 节提出的建模方法, 分别使用数据集中的运动类视频和表演类视频来建立传输流量统计模型, 针对四类视频分别进行性能评估, 包括运动类视频 2、电影类视频 3、纪录类视频 6 和表演类视频 7。评估中, 令  $t_{\max} = 11$ ,  $d_{\max} = 60$ ; 置信度选择为  $\rho = 0.9$ , 通过数值计算可得  $\rho$  分位点  $\phi(\rho) = 1.28$ 。

为了评估由模型优化得到的时空切片方式与实际最优方式的差距, 本节通过仿真得到待切分视

频的最优切片方式。具体地, 首先, 把待切片视频分别按照  $t=1,2,\dots,11, d=1,2,\dots,60$  的配置进行切片和压缩编码; 然后, 导入用户的头动数据, 获得每一种切片配置下 48 个用户观看时的传输流量; 最后, 绘制 48 个用户的传输流量的累积分布函数曲线, 找到  $\rho=0.9$  时的分位点, 即  $T_\rho(t,d)$ 。通过遍历所有的  $t$  和  $d$  组合, 找到使  $T_\rho(t,d)$  最小的最优时空切片方式。最终得到运动类视频 2 的最优方式是  $t=10, d=11$ ; 电影类视频 3 的最优方式是  $t=10, d=15$ ; 纪录类视频 6 的最优方式是  $t=9, d=15$ ; 表演类视频 7 的最优方式是  $t=10, d=27$ 。

在图 14 中, 本节采用基于运动类视频类建立传输流量统计模型, 使用所提出的优化方法求解问题式(21), 得到使目标函数最小的时空切片方式为  $t=10, d=14$ 。图 14 中虚线给出了 4 个视频在模型优化方式下的归一化传输流量的累积概率分布, 实线给出了 4 个视频通过仿真搜索得到的最优切片方式的结果。可以看出, 对于运动类视频 2 和电影类视频 3, 模型优化的结果与仿真搜索的最优结果很接近。这是因为本文使用了由运动类视频建立的传输流量统计模型, 而电影类视频与运动类视频具有类似的用户头动规律。对于纪录类视频 6 和表演类视频 7, 由于其视频类型与传输流量统计模型失配, 可以看出模型优化的结果与仿真搜索的最优结果存在一定的差距。表演类视频 7 的差距比较明显, 其原因是用户在观看表演类视频时, 视线主要集中在舞台上的表演, 发生头动的频次较少, 这与观看运动类视频时的头动规律差别很大, 如图 3 所示。这一结果说明对于用户头动规律差别较大的视频类型有必要进行单独的传输流量统计建模。

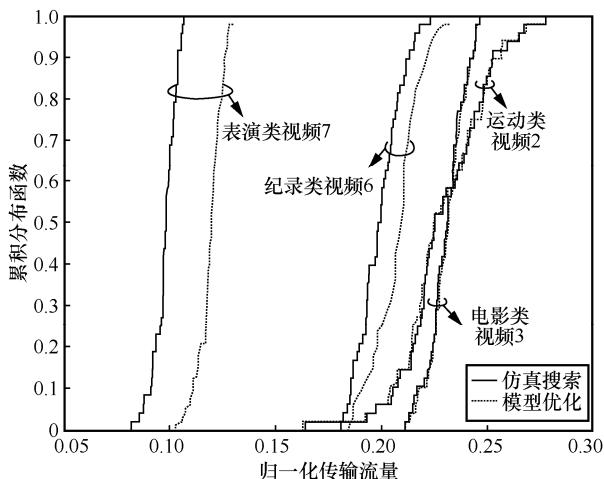


图 14 基于运动类视频模型优化的时空切片方式的性能

在图 15 中, 本节采用表演类视频来建立传输流量统计模型, 进而基于此模型求解优化问题式(21), 得到使目标函数最小的时空切片方式为  $t=10, d=24$ 。通过对比图 14 和图 15 中的表演类视频 7, 可以看出模型优化的结果与仿真搜索的最优结果之间的差距减小。其他 3 类视频由于与传输流量模型失配, 导致模型优化的结果与仿真搜索的最优结果之间的差距变大。

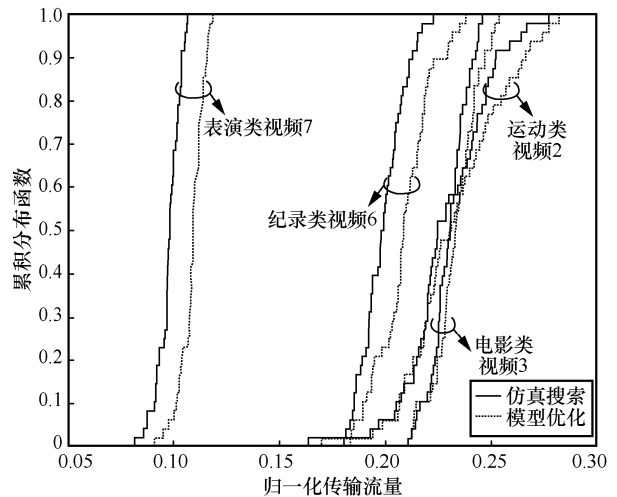


图 15 基于表演类视频模型优化的时空切片方式的性能

#### 4 结束语

本文针对 VR 全景视频传输, 研究了时空切片方式对传输流量的影响。首先, 基于实际的 VR 全景视频观看数据集, 通过仿真分析了空间切块数量和时间分段长度对视频大小的影响, 并通过引入用户观看头动数据, 评估了时空切片方式对全景视频传输流量的影响。然后, 基于仿真数据, 对时空切片模式下的传输流量进行了统计建模, 结果表明传输流量服从正态分布, 并建立了均值和标准差与时空切片方式之间的函数关系。最后, 基于所建立的统计模型, 提出了一种时空切片方式的嵌套优化方法, 外层对空间切块数量进行遍历寻优, 内层可直接计算最优的时间分段长度。评估结果表明, 通过对不同类型的视频分别进行传输流量统计建模, 基于所建立的流量模型优化得到的时空切片方式与通过仿真搜索得到的最优切片方式性能接近。

#### 参考文献:

[1] 华为 iLab. 面向 VR 业务的承载网络需求白皮书[R]. 2016. HUAWEI iLab. White paper on bearer network requirements for VR

- services[R]. 2016.
- [2] 3GPP. Virtual reality (VR) media services over 3GPP: TR 26.918[S]. 2017.
- [3] ZINK M, SITARAMAN R, NAHRSTEDT K. Scalable 360° video stream delivery: challenges, solutions, and opportunities[J]. Proceedings of the IEEE, 2019, 107(4): 639-650.
- [4] HOSSEINI M, SWAMINATHAN V. Adaptive 360 VR video streaming based on MPEG-DASH SRD[C]//Proceedings of 2016 IEEE International Symposium on Multimedia. Piscataway: IEEE Press, 2016: 407-408.
- [5] GADDAM V R, RIEGLER M, EG R, et al. Tiling in interactive panoramic video: approaches and evaluation[J]. IEEE Transactions on Multimedia, 2016, 18(9): 1819-1831.
- [6] MISRA K, SEGALL A, HOROWITZ M, et al. An overview of tiles in HEVC[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(6): 969-977.
- [7] ALFACE P R, MACQ J F, VERZIJP N. Interactive omnidirectional video delivery: a bandwidth-effective approach[J]. Bell Labs Technical Journal, 2012, 16(4): 135-147.
- [8] ZARE A, AMINLOU A, HANNUKSELA M M, et al. HEVC-compliant tile-based streaming of panoramic video for virtual reality applications[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York: ACM Press, 2016: 601-605.
- [9] TU Z Z, ZONG T Y, XI X L, et al. Content adaptive tiling method based on user access preference for streaming panoramic video[C]//Proceedings of 2018 IEEE International Conference on Consumer Electronics. Piscataway: IEEE Press, 2018: 1-4.
- [10] NGUYEN D V, TRAN H T T, THANG T C. An evaluation of tile selection methods for viewport-adaptive streaming of 360-degree video[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2020, 16(1): 1-24.
- [11] SINGLA A, GÖRING S, RAAKE A, et al. Subjective quality evaluation of tile-based streaming for omnidirectional videos[C]//Proceedings of the 10th ACM Multimedia Systems Conference. New York: ACM Press, 2019: 232-242.
- [12] AN C L, YEN S C, HUANG C Y, et al. On the optimal encoding ladder of tiled 360° videos for head-mounted virtual reality[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(4): 1632-1647.
- [13] CHEN X L, WU D, AHMAD I. Optimized viewport-adaptive 360-degree video streaming[J]. CAAI Transactions on Intelligence Technology, 2021, 6(3): 347-359.
- [14] BETHANABHOTLA D, CAIRE G, NEELY M J. WiFlix: adaptive video streaming in massive MU-MIMO wireless networks[J]. IEEE Transactions on Wireless Communications, 2016, 15(6): 4088-4103.
- [15] HAN S Q, SU H T, YANG C Y, et al. Proactive edge caching for video on demand with quality adaptation[J]. IEEE Transactions on Wireless Communications, 2020, 19(1): 218-234.
- [16] SUN Y P, CHEN Z Y, TAO M X, et al. Communications, caching, and computing for mobile virtual reality: modeling and tradeoff[J]. IEEE Transactions on Communications, 2019, 67(11): 7573-7586.
- [17] WEI X, YANG C Y, HAN S Q. Prediction, communication, and computing duration optimization for VR video streaming[J]. IEEE Transactions on Communications, 2021, 69(3): 1947-1959.
- [18] XU M, LI C, ZHANG S Y, et al. State-of-the-art in 360° video/image processing: perception, assessment and compression[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 5-26.
- [19] WU C L, TAN Z H, WANG Z, et al. A dataset for exploring user behaviors in VR spherical video streaming[C]//Proceedings of the 8th ACM on Multimedia Systems Conference. New York: ACM Press, 2017: 193-198.
- [20] YAQOUB A, MUNTEAN G M. A combined field-of-view prediction-assisted viewport adaptive delivery scheme for 360° videos[J]. IEEE Transactions on Broadcasting, 2021, 67(3): 746-760.
- [21] BAN Y X, XIE L, XU Z M, et al. CUB360: exploiting cross-users behaviors for viewport prediction in 360 video adaptive streaming[C]//Proceedings of 2018 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2018: 1-6.
- [22] XIE L, ZHANG X G, GUO Z M. CLS: a cross-user learning based system for improving QoE in 360-degree video adaptive streaming[C]//Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM Press, 2018: 564-572.
- [23] ZHANG Y X, ZHAO P Y, BIAN K G, et al. DRL360: 360-degree video streaming with deep reinforcement learning[C]//Proceedings of IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2019: 1252-1260.
- [24] NGUYEN A, YAN Z S, NAHRSTEDT K. Your attention is unique: detecting 360-degree video saliency in head-mounted display for head movement prediction[C]//Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM Press, 2018: 1190-1198.
- [25] FFMPEG Team. Ffmpeg software[EB]. 2022.
- [26] RICHARDSON I E. G. H.264/MPEG-4 part 10 white paper[R]. 2003.
- [27] BENTO4. Bento4 MP4, DASH, HLS, CMAF SDK and tools[EB]. 2020.
- [28] PERFECTO C, ELBAMBY M S, SER J D, et al. Taming the latency in multi-user VR 360°: a QoE-aware deep learning-aided multicast framework[J]. IEEE Transactions on Communications, 2020, 68(4): 2491-2508.
- [29] FEI Z S, WANG F, WANG J, et al. QoE evaluation methods for 360-degree VR video transmission[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(1): 78-88.
- [30] ZHENG C, LIU S H, HUANG Y M, et al. Hybrid policy learning for energy-latency tradeoff in MEC-assisted VR video service[J]. IEEE Transactions on Vehicular Technology, 2021, 70(9): 9006-9021.

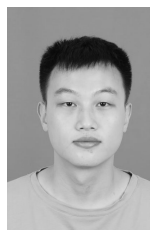
## [作者简介]



韩圣千（1981-），男，山东威海人，博士，北京航空航天大学副教授、博士生导师，主要研究方向为无线智能通信、超可靠低时延通信、多天传输等。



娄函（1997-），男，云南文山人，北京航空航天大学硕士生，主要研究方向为VR全景视频传输。



王君来（1997-），男，湖北十堰人，北京航空航天大学硕士生，主要研究方向为VR全景视频传输。