

面向内容的立体视频帧重要性区分模型研究

王晓东, 胡彬彬, 王腾飞, 蒋刚毅, 章联军

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 异构网络引起的视频传输失真是影响视频终端感知质量的重要因素, 通常只能在解码端进行评估和补偿. 因此在编码端建立一个面向内容的立体视频传输失真模型, 考虑错误隐藏和错误扩散两种失真, 通过近似估计视频中各帧对终端质量的贡献度, 从而实现每帧重要性的区分. 新模型综合考虑了视频帧编码位置以及表征内容的编码模式、运动矢量、视差矢量等关键信息, 可以作为视频在异构网络中传输策略的重要依据.

关键词: 视频流; 失真模型; 错误隐藏; 错误扩散; 帧重要性

中图分类号: TP391

文献标志码: A

文章编号: 1001-5132 (2015) 01-0030-05

随着科学技术的发展, 人们对立体视频传输相关技术的研究愈加深入^[1-2]. 相比 2D 视频, 立体视频的数据量增加了 1 倍, 然而现实网络具有异构性, 视频在传输过程中会出现延时、丢包、抖动等现象^[3], 造成视频终端感知质量的下降. 由于立体视频中各帧位置和内容的不同, 其对视频终端质量的贡献也不同, 利用这种特性, 在实时传输中可进行帧级区分服务^[4], 以提高视频的终端质量, 是一个值得重点研究的问题.

区分帧重要性最简单的方法是按编码位置进行划分, 文献[5]将 I 帧和 P 帧重要性的比重确定为 1.15:0.35. 文献[6]提出, 在 1 个 GOP 中, 将 I 帧的重要性设为最高, P 帧次之, B 帧的重要性最低, 当网络拥塞时, 可以丢掉部分重要性低的 B 帧, 从而保证视频实时传输. 目前的 H.264/AVC 将视频数据按片编码后也存在严格的位置关系^[7], 当进行可扩展编码时, 将编码数据分为基本层和增强层^[8], 后续增强层必须依赖基本层进行解码. 因此, 认为基本层的重要性比后续增强层的重要性高. 文献[5-8]中, 只按编码位置进行帧重要性区分, 忽略了图像内容的影响, 虽然简单, 但不准确, 也无法保证终端视频的质量. 另外一种是利用视频流媒体

传输失真模型确定每帧的重要性, 周敬利等^[9]介绍了一种在发送端建立视频流传输失真模型的思想, 通过视频特性参数和网络传输质量递归估算出每帧视频终端表现质量. 周圆等^[10]考虑了编码模式、错误传播等能表征内容的因数, 推导出不同情况下多视点视频传输失真的计算公式, 但此过程仍然采用的是递归模型. 基于递归的帧重要性区分方法, 可以预测多视点视频中任何视点帧带来的错误传播, 在随机丢包的情况下, 准确地估计出各帧的失真值, 但这种方法计算量很大, 不利于视频的实时传输.

文献[11]和[12]提出了一种单视点视频传输失真估计模型, 考虑错误隐藏失真和错误扩散两种失真. 在计算错误隐藏失真时, 综合考虑编码模式、运动矢量等参数; 在计算错误扩散失真时, 利用指数模型来表示前端编码帧对后续帧的影响, 计算量相对比较低. 笔者在此基础上, 建立了一个面向内容的视频帧重要性区分模型. 由于立体视频比单视点视频增加了视差和深度等特殊信息, 编码时参考关系更加复杂, 右视点部分帧会参考左视点帧进行编码, 左视点帧丢失带来的扩散失真会更严重, 针对以上问题, 在模型中需综合考虑

收稿日期: 2014-01-09.

宁波大学学报(理工版)网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家科技支撑计划项目(2012BAH67F01); 国家自然科学基金(60832003, 61071120); 浙江省科技厅创新团队自主设计项目(2012R10009-08); 宁波市科技创新团队研究计划项目(2011B81002).

第一作者: 王晓东(1970-), 男, 浙江绍兴人, 副教授, 主要研究方向: 网络通信及无线传感网络. E-mail: wangxiaodong@nbu.edu.cn

图像内容、视点间相关性、编码位置等信息。

1 立体视频传输失真模型

不同视频帧对终端视频质量的贡献不同, 通过模型得到各帧的失真值, 将其定义为各帧的重要性, 如果该帧失真值越大, 就表示它的重要性越高。综合考虑编码模式、运动矢量、视差矢量等能表征图像内容的关键信息建立立体视频传输失真模型, 并考虑错误隐藏失真和错误扩散两个方面的失真, 得到帧重要性失真参数(Frame Importance Distortion Parameter, FIDP), 其公式如下:

$$R_{\text{FIDP},k} = \beta I_{\text{EC},k} + (1 - \beta) I_{\text{EP},k}, \quad (1)$$

其中, 假设1个GOP的长度为 N , β 为加权系数, 经验值取0.5。 $I_{\text{EC},k}$ 表示当前GOP中第 K 帧视频的误差隐藏失真, 如果传输时该帧丢失, 采用Direct-copy方法进行误差隐藏恢复。当左视点某块丢失时, 复制左视点前一帧的相应块恢复当前块; 当右视点某块丢失时, 复制该块编码时参考的对应块恢复当前块。 $I_{\text{EP},k}$ 表示由第 K 帧的误差隐藏失真对该GOP后续编码帧带来的误差扩散失真。由于左右视点帧编码位置的不同, 右视点帧失真只对该GOP右视点的后续编码帧产生影响, 而左视点帧的失真则会对左右2个视点后续编码帧都产生影响。

1.1 右视点传输失真模型

对于GOP中的第 K 帧, 因为我们考虑了不同编码模式的信息, 先计算出其中每个宏块丢失带来的误差隐藏失真, 然后将1帧中所有宏块的失真相加, 得到该帧丢失带来的误差隐藏失真 $I_{\text{EC},k}$ 。

1.1.1 帧级误差隐藏失真

对于帧内编码宏块, 笔者认为丢失带来的误差隐藏失真和运动矢量存在关系^[11], 考虑到宏块可能有 16×16 、 16×8 、 8×16 、 8×8 几种编码模式, 用下式来计算整个宏块的运动矢量:

$$R_{\text{MV},\text{MB}} = \sum_{j=1}^{N_{\text{part}}} \frac{L_{j,x} \cdot L_{j,y}}{16 \times 16} \cdot \sqrt{R_{\text{MV},j,x}^2 + R_{\text{MV},j,y}^2}, \quad (2)$$

其中, $R_{\text{MV},j,x}$ 和 $R_{\text{MV},j,y}$ 表示子宏块的运动矢量, $L_{j,x}$ 和 $L_{j,y}$ 表示子宏块的宽和高, 可能包含 $(16,16)$ 、 $(16,8)$ 、 $(8,16)$ 、 $(8,8)$ 等编码, N_{part} 表示子宏块的个数。在立体视频中还要考虑视差矢量DV, 在综合 R_{MV} 和 R_{DV} 的情况下, 则右视点整帧丢失带来的误差隐藏失真 $I_{\text{EC},k}$ 为:

$$I_{\text{EC},k} = \sum_{i=1}^{N_{\text{MVinter}}} p \cdot \sum_{j=1}^{N_{\text{MVpart}}} \frac{L_{j,x} \cdot L_{j,y}}{16 \times 16} \cdot \sqrt{R_{\text{MV},j,x}^2 + R_{\text{MV},j,y}^2} + \sum_{i=1}^{N_{\text{DVinter}}} p \cdot \sum_{j=1}^{N_{\text{DVpart}}} \frac{L_{j,x} \cdot L_{j,y}}{16 \times 16} \cdot \sqrt{R_{\text{DV},j,x}^2 + R_{\text{DV},j,y}^2} + \sum_{i=1}^{N_{\text{intra}_16 \times 16}} p \cdot W_{\text{intra}_16 \times 16} \cdot \sqrt{16^2 + 16^2} + \sum_{i=1}^{N_{\text{intra}_4 \times 4}} p \cdot W_{\text{intra}_4 \times 4} \cdot \sqrt{16^2 + 16^2}, \quad (3)$$

其中, N_{MVinter} 和 N_{DVinter} 分别表示该帧中采用运动矢量和视差矢量编码的宏块数; N_{MVpart} 和 N_{DVpart} 分别表示某个宏块中采用运动矢量和视差矢量编码的子宏块数; $N_{\text{intra}_16 \times 16}$ 和 $N_{\text{intra}_4 \times 4}$ 分别表示该帧中采用帧内 16×16 和 4×4 编码的宏块数, $W_{\text{intra}_16 \times 16}$ 和 $W_{\text{intra}_4 \times 4}$ 表示采用帧内 16×16 和 4×4 块编码的系数, 因为帧内编码宏块的重要性很高, 可通过离线实验进行统计, 我们一般把运动矢量2个方向上的分量都赋值为16; 而 p 表示误差隐藏失真系数。

1.1.2 帧级误差扩散失真

令 $I_{\text{EP},k}(i)$ 为 $I_{\text{EC},k}$ 对GOP中右视点后续第 i 个编码帧的误差扩散影响, 则第 K 帧给整个GOP带来的失真 $I_{\text{EP},k}(i)$ 可以表示为:

$$I_{\text{EP},k} = \sum_{i=1}^{N-k+1} I_{\text{EC},k}(i), \quad (4)$$

$I_{\text{EC},k}$ 带来的误差扩散只会影响到帧间预测编码的宏块, 并对后面帧的误差扩散影响会随 i 增大而逐渐减弱, 可以用指数函数来近似表示这种扩散:

$$I_{\text{EP},k}(i) = I_{\text{EC},k}(i-1) \cdot \alpha \cdot R_{k+i} \quad I_{\text{EP},k} = \sum_{i=1}^{N-k+1} I_{\text{EC},k}(i) = I_{\text{EC},k} \cdot \alpha^i \cdot \prod_{t=1}^i R_{k+t}, \quad (5)$$

其中, α 为扩散因子, 一般取0~1之间的数值, 文中取0.96; R_{k+t} 表示右视点后续第 $k+t$ 帧中的帧间编码宏块比例。最后利用(1)式, 将2种失真相加, 得到右视点某帧的总失真。

1.2 左视点传输失真模型

与右视点失真类似, 左视点也是考虑误差隐藏和误差扩散两方面失真, 用来表征各帧的重要性, 但是左视点图像失真会对左右2个视点都产生失真。因此, 模型中分别估计对2个视点的影响。

1.2.1 帧级误差隐藏失真

由于左视点帧只参考该视点的帧进行编码, 没有视差矢量, 故将(3)式简化, 得到左视点帧的

错误隐藏失真 $I_{EC,k}$ 为:

$$I_{EC,k} = \sum_{i=1}^{N_{inter}} p \cdot \sum_{j=1}^{N_{part}} \frac{L_{j,x} \cdot L_{j,y}}{16 \times 16} \cdot \sqrt{R_{MV,j,x}^2 + R_{MV,j,y}^2} + \sum_{i=1}^{N_{intra_16 \times 16}} p \cdot W_{intra_16 \times 16} \cdot \sqrt{16^2 + 16^2} + \sum_{i=1}^{N_{intra_4 \times 4}} p \cdot W_{intra_4 \times 4} \cdot \sqrt{16^2 + 16^2}, \quad (6)$$

上式中参数意义和右视点模型中的一样。

1.2.2 帧级错误扩散失真

由于右视点部分帧是参考左视点进行编码, 所以左视点帧的错误扩散失真要同时考虑对左右视点后续编码帧的影响, 将(4)式和(5)式进行变化, 得到左视点帧的错误扩散失真 $I_{EP,k}(j)$ 为:

$$I_{EP,k} = \sum_{j=1}^{N-k+1} (I_{EPL,k}(j) + I_{EPR,k}(j)), \quad (7)$$

其中, $I_{EPL,k}$ 和 $I_{EPR,k}$ 分别表示左视点某帧对左右视点后续帧的错误扩散失真。 $I_{EPL,k}$ 的计算方法和右视点一样, 通过(4)式和(5)式计算可得, 而 $I_{EPR,k}$ 则通过下式计算:

$$I_{EPR,k}(j) = I_{EPR,k}(j-1) \cdot R_{k+j} = I_{EC,k} \cdot \prod_{t=1}^j \frac{N_{DV,k+t}}{N_{MV,k+t} + N_{DV,k+t}}, \quad (8)$$

其中, R_{k+j} 表示右视点所有帧采用视差矢量进行编码的平均比例; $N_{DV,k+t}$ 和 $N_{MV,k+t}$ 表示右视点所有帧中, 分别采用视差矢量和运动矢量编码的宏块个数。其中, $j=1, 2, 3, \dots, N-k-1$, 最后利用(1)式将 2 种失真相加, 得到左视点某帧总失真。

2 实验

为验证模型准确性, 进行以下对比实验。编码格式采用 IPPP, GOP 的帧数设为 N , 在立体视频中左右视点各有 N 帧, 因此 1 个 GOP 中实际包含 $2N$ 帧, 左视点第 1 帧为 I 帧, 且假设其重要性最高, 重要性权重设为 1。模型方法如图 1 所示, 分别由视频每帧的错误隐藏失真和错误扩散失真加权求和得到各帧传输总失真, 然后在 1 个 GOP 中利用归一化方法得到各帧重要性权重。归一化公式如下:

$$\text{帧重要性权重} = \frac{R_{FIDP,k} - R_{FIDP,min}}{R_{FIDP,max} - R_{FIDP,min}}, \quad (9)$$

其中, $R_{FIDP,k}$ 表示当前 GOP 中第 k 帧的总失真; $R_{FIDP,max}$ 和 $R_{FIDP,min}$ 分别为当前 GOP 除 I 帧以外的

$2N-1$ 帧中总失真的最大值和最小值。图 2 是基于实验方法, 在视频传输中, 假设当前 GOP 中的各帧依次丢失, 重复 $2N-1$ 实验, 解码端采用帧复制的方法, 然后得到视频解码值和视频原始值的差值。文中采用 MSE 进行表征, 假设 I 帧的重要性仍然为 1, 在剩下的 $2N-1$ 中, 分别求出 $R_{MSE,max}$ 和 $R_{MSE,min}$, 同样利用(9)式中的思想进行归一化, 得到各帧重要性权重。

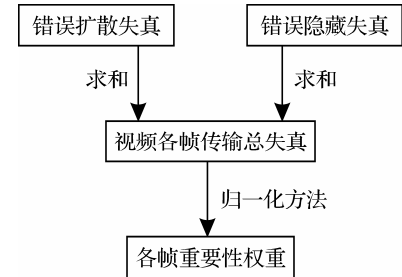


图1 基于失真模型帧重要性权重区分

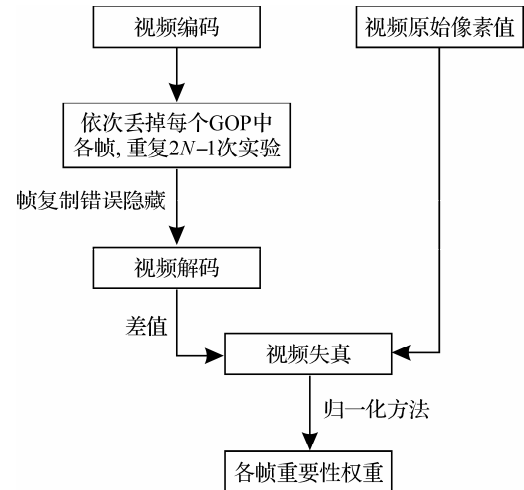


图2 基于实验帧重要性权重区分

实验中, 选取运动缓慢的 book、运动稍剧烈的 aquarium 序列和运动剧烈的 flame 序列, 实验 2 个 GOP, 共 40 帧, R_{QP} 设为 28, GOP 长度 $N=10$, 1~20 和 21~40 分别为单个立体 GOP, 在单个 GOP 中, 假设 I 帧的重要性为 1(即第 1 帧和第 21 帧), 分别利用模型值和实验值归一化, 得到各帧重要性权重。

由图 3~图 5 的对比可知, 模型值得到的帧重要性权重和实验值得到的结果基本一致, 第 1 帧和第 21 帧为 I 帧, 将其重要性权重赋值为 1。通过实验可以统计得到 book、aquarium、flame 视频序列右视点 1 帧(即第 2 帧)只能参考左视点进行编码, 但图像左右视点相关性不强, 所以视差矢量就会

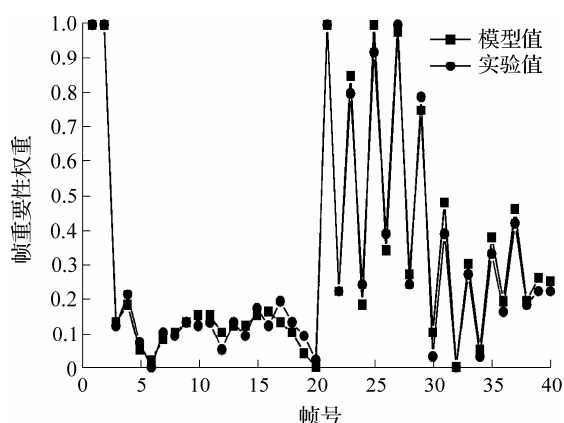


图 3 book 序列帧重要性权重区分

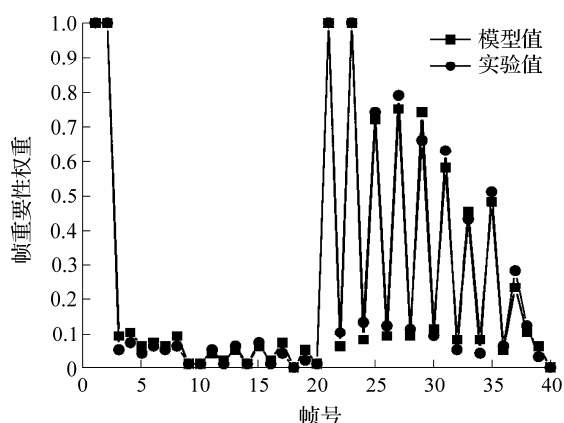


图 4 aquarium 序列帧重要性权重区分

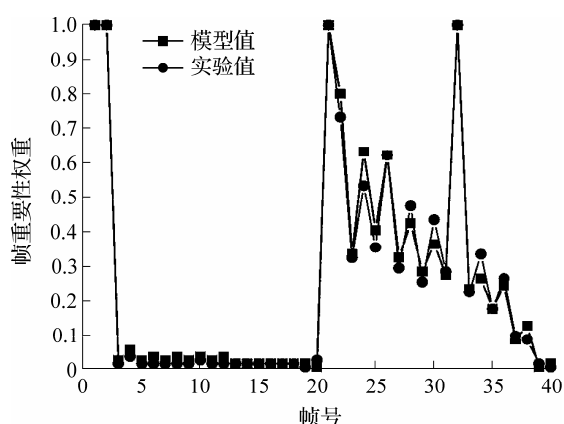


图 5 flame 序列帧重要性权重区分

很大, 因此此帧丢失带来的总失真就会很大, 在归一化后, 帧重要性权重就会很高. 图 5 的 flame 序列第 1 个 GOP 当中, I 帧和右视点第 1 帧的重要性等级最高, 属于 DPA 部分, 后续编码帧的重要性权重较低, 通过实验值可验证模型的区分较准确. 对于 flame 序列第 2 个 GOP, 即第 20 帧到 40 帧, 视频中的对象正在做一个剧烈的旋转运动, 每帧的

运动矢量都较大, 与右视点第 1 帧的视差矢量数值相差不大, 因而结果与第 1 个 GOP 不同. 与传统没有考虑编码模式、运动和静止等内容因子的方法相比, 更加能够提高区分的准确性, 反映人眼的感知.

为验证模型适用性和更好表征帧重要性, 笔者对不同运动剧烈程度序列, 分别进行实验根据模型值和实验值. 定义(10)式来计算视频帧重要性预测误差,

$$E = \frac{\sum_{k=1}^{N_{\text{total}}} |R_{\text{FIDP},k} - R_{\text{MSE},k}|}{\sum_{k=1}^{N_{\text{total}}} R_{\text{MSE},k}} \times 100\%, \quad (10)$$

其中, $R_{\text{FIDP},k}$ 和 $R_{\text{MSE},k}$ 分别表示第 k 帧传输失真的模型值和实验值, N_{total} 为单个序列实验总帧数, 文中实验 N_{total} 取 40.

由表 1 的对比数据可见, 3 个序列的预测误差在 7.56%~8.46%, 平均误差为 8.03%, 模型准确度较高.

表 1 帧重要性平均预测误差

测试项目	序列名称		
	book	aquarium	flame
误差/%	7.56	8.46	8.08

3 结论

传统的帧重要性权重区分基本都只考虑了位置信息, 笔者在此基础上加入编码模式、运动矢量、视差矢量等表征图像内容的影响因子, 提出一种面向内容的立体视频帧重要性区分模型. 选取运动缓慢的 book 序列、运动稍剧烈的 aquarium 序列和运动剧烈的 flame 序列, 帧的重要性权重区分效果与实验结果很近似. 此外, 相比以往的方法, 模型中没有采用递归思想, 计算量相对较低, 而且考虑了影响终端视频感知质量的内容信息. 笔者所提出的模型可以在编码端估计各帧对终端视频感知质量的贡献度, 指导视频传输过程中的拥塞控制策略和差错控制策略, 提高接受端视频感知质量, 具有重要的实际意义.

参考文献:

- [1] Zilly F, Kluger J, Berlin G, et al. Production rules for stereo acquisition[J]. Proceedings of the IEEE, 2011, 99

- (4):590-606.
- [2] 高昂, 许爽, 李增智, 等. P2P 流媒体分发模型及在视频会议系统中的应用[J]. 西南交通大学学报, 2013, 48(1):88-93.
- [3] 周圆, 侯春萍. 网络丢包对多视点 3D 视频质量的影响[J]. 吉林大学学报: 工学版, 2013, 43(S1):419-423.
- [4] Gurler, C G, Koc Univ E, Istanbul. Flexible transport of 3-D video over networks[J]. Proceedings of the IEEE, 2011, 99(4):694-707.
- [5] Hartanto F, Sirisena H R. Hybrid error control mechanism for video transmission in the wireless IP networks[C]. 10th IEEE Workshop on Local and Metropolitan Area Networks, 1999:126-132.
- [6] Yueh L C, Ting L L, Pamela C C. Network-based H.264/AVC whole-frame loss visibility model and frame dropping methods[J]. IEEE Transactions on Image processing, 2012, 21(8):3353-3363.
- [7] Ohm J R, Sullivan G J. High efficiency video coding: the next frontier in video compression (standards in a nutshell) [J]. IEEE Signal Processing Magazine, 2013, 30(1):152-158.
- [8] Djamah M, O'Shaughnessy D. Fine granularity scalable speech coding using embedded tree-structured vector quantization[J]. Speech Communication, 2012, 54(1):23-39.
- [9] 周敬利, 马志龙, 范晔斌. 视频流媒体传输失真研究[J]. 小型微型计算机系统, 2006, 27(2):264-267.
- [10] 周圆. 面向 IP 网络的多视点立体视频传输失真分析与建模[D]. 天津: 天津大学, 2011.
- [11] 孙钢. 基于内容感知的视频传输质量评价及资源分配机制研究[D]. 杭州: 浙江大学, 2009.
- [12] Sun Gang, Wei Xin, Lu Dongming. Low-complexity frame importance modeling and resource allocation scheme for error-resilience H.264 video streaming[C]. Proceedings of the 2008 IEEE 10th Workshop on Multimedia Signal Processing (MMSP), 2008:819-823.

Differentiating Importance of Frame in Content-oriented Stereoscopic Video Model: Research

WANG Xiao-dong, HU Bin-bin, WANG Teng-fei, JIANG Gang-yi, ZHANG Lian-jun
(College of Information Science and Technology, Ningbo University, Ningbo 315211, China)

Abstract: The video transmission distortion caused in the heterogeneous network video has an important impact on the quality of the terminal, which is usually evaluated and compensated only at the decoder side. In terms of the content, this thesis proposes a stereoscopic video transmission distortion model at the encoder with consideration of error concealment and error propagation distortion. The proposed approach can estimate the contribution of each frame to the video quality of the terminal by approximately distinguishing the importance of every frame, the video encoding position, encoding mode, motion vector, disparity vector, and other factors that can represent content in the model.

Key words: video stream; distortion model; the importance of frame; error concealment; error propagation

(责任编辑 章践立)