

宏块级立体视频传输失真快速估算方法研究

姚 婷, 周 宇*, 王晓东, 章联军, 王腾飞, 胡彬彬

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 鉴于传输丢包会严重影响视频重建质量, 甚至大范围传播错误, 因此通过研究视频通信质量与传输丢包之间的关系, 提出了一种适合网络丢包特性的立体视频传输失真模型. 该失真模型结合了快速宏块级失真估计方法及帧间宏块拷贝和视点间宏块拷贝的错误隐藏方法, 可在丢包信道下快速估计宏块级的立体视频传输失真. 实验对比分析表明: 所提出的宏块级立体视频传输失真模型准确度较好, 可为立体视频传输失真估计提供一定的理论依据与借鉴, 对立体视频通信质量的保障具有一定的实用价值.

关键词: 失真模型; 立体视频; 宏块级

中图分类号: TP391

文献标志码: A

文章编号: 1001-5132 (2015) 02-0019-05

随着三维显示技术的进步, 立体视频技术成为一个广泛研究的热点. 然而被压缩的立体视频由于采用低比特率视频编码方案, 依靠帧间编码提高编码效率, 面对传输错误时非常脆弱. 在易错的分组交换网络中, 不可避免的丢包将会造成重建视频质量的严重下降. 为保证立体视频的服务质量, 需要研究包分组交换网络中传输丢包与视频通信质量之间的关系, 建立适合网络丢包特性的视频传输失真模型.

现有的传输失真估计模型大致可以分为三类: (1)像素级失真建模. 最具有代表性的是 Zhang 等^[1]提出的基于像素递推的失真估计方法(ROPE), 它在特定的码率、丢包环境和错误隐藏方法下递归地计算每个解码重建像素的第一时刻和第二时刻的失真来确定每个像素的总失真. 文献[2-3]对 ROPE 方法进行改进, 使其适用于运动补偿精度为半像素的场合且对 ROPE 中的互相关系数进行简化计算. (2)帧级失真建模. 文献[4-6]提出一种帧级的视频流媒体失真模型, 模型涉及了一些取值在 0~1 之间的随机参数, 在一定程度上影响了模型的准确性. Wang 等^[7]提出一种基于帧级递归的传输失真估计

模型, 模型考虑了丢包率、编码模式、帧内刷新率、亚像素运动矢量以及 H.264 的帧内预测和环路滤波两个新特性, 引入了误差扩散因子, 上述帧级失真模型都是针对单视点提出, 周圆^[8]提出一种面向 IP 网络的多视点视频传输失真模型, 模型从不同的编码模式出发, 详细分析了错误扩散理论和编码技术原理, 分别对每种编码模式下正确接收的帧产生的失真和丢失的帧产生的失真进行计算. (3)宏块级失真建模. 文献[9]分析了 H.264 中的一些关键参数和视频内容特性之间存在的关系, 在此基础上提出一种权重估计的宏块失真模型. 文献[10]也在宏块级估算总失真, 它通过考虑前帧匹配块是否正确接收来确定当前宏块的失真. Zhang 等^[11]围绕单视点视频提出一种快速的宏块级失真估计方法, 该方法计算复杂度较低, 可在编码端快速地估计视频的传输失真.

在上述模型中, 基于像素级的估算模型虽然精确度高, 但极大地增加了计算复杂度; 基于帧级的估算模型计算复杂度虽低, 却不能满足实时性的要求. 计算复杂度与实时性是一对相互矛盾的共同体, 基于宏块级的估算可以较好地解决计算

收稿日期: 2013-10-23.

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

基金项目: 国家自然科学基金(60832003, 61071120); 国家科技支撑计划(2012BAH67F01); 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: 姚婷(1990-), 女, 江西抚州人, 在读硕士研究生, 主要研究方向: 网络通信及图像处理. E-mail: yaoting99@sina.cn

*通信作者: 周宇(1960-), 男, 山东威海人, 教授, 主要研究方向: 网络通信及计算机软件. E-mail: zhouyu@nbu.edu.cn

复杂度和实时性之间的矛盾问题,是一种较好的折中方法.但目前这些宏块级失真估计方法都是针对单视点视频提出来的,且针对立体视频的宏块级传输失真估计模型的研究还未见报道.为此,笔者在详细分析立体视频编码特点的基础上,提出一种宏块级的立体视频传输失真模型,并采用快速近似估计法以降低计算复杂度,最后通过相关实验进行验证.

1 宏块级立体视频传输失真模型

假设视点 s 第 t 时刻第 m 个宏块的第 i 个像素的编码重建值记为 $\hat{F}(s, t, m, i)$, 相应的解码重建值为 $\tilde{F}(s, t, m, i)$, 则该宏块的传输失真可表示如下:

$$D(s, t, m) = E\{[\hat{F}(s, t, m, i) - \tilde{F}(s, t, m, i)]^2\}, \quad (1)$$

接着根据(1)式分别针对立体视频的左右视点进行失真建模.文中采用 IPPP 编码结构,且 I 帧中的宏块不会丢失.除 I 帧外,如果其他帧中有宏块丢失,则在解码端采用下述错误隐藏方法进行恢复:当左视点帧中某宏块丢失时,采用帧间宏块拷贝的方式进行错误隐藏;当右视点帧中某宏块丢失时,采用视点间宏块拷贝的方法进行错误隐藏,丢包率用 P 表示.本文算法框图如图 1 所示.

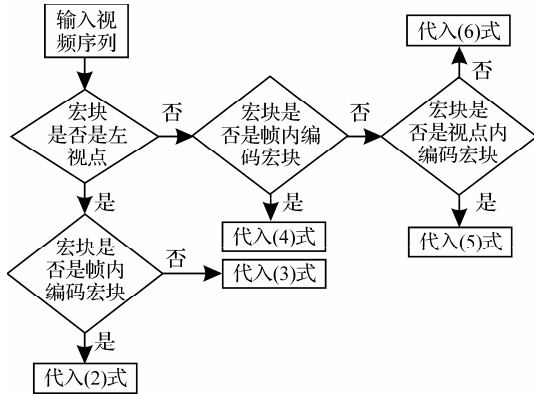


图 1 本文算法框图

1.1 左视点视频传输失真模型

对于左视点中的帧内编码宏块,如果宏块被正确接收,其解码重建值与编码重建值相等,虽然有可能不一定准确,但它们的差别很小,因此本文忽略不计;若宏块丢失,在解码端进行错误隐藏.则左视点帧内编码宏块的传输失真可表示为:

$$D_L^I = pE\{[\hat{F}(s, t, m, i) - \tilde{F}(s, t-1, m, i)]^2\} + (1-p)E\{[\hat{F}(s, t, m, i) - \hat{F}(s, t, m, i)]^2\} =$$

$$pE\{[\hat{F}(s, t, m, i) - \hat{F}(s, t-1, m, i)]^2\} + pE\{[\hat{F}(s, t-1, m, i) - \tilde{F}(s, t-1, m, i)]^2\} = pR_{\text{RFD}}(t, t-1, m) + pD(s, t-1, m), \quad (2)$$

其中, $R_{\text{RFD}}(t, t-1, m)$ 表示编码重建的第 t 帧和第 $t-1$ 帧图像中第 m 个宏块之间的差别,该值在编码端很容易计算. $D(s, t-1, m)$ 是 $t-1$ 帧中第 m 个宏块的传输失真.由于 I 帧宏块不丢失,左视点第一帧中的宏块初始值为 $D(0, 0, m) = 0$ (视点编号和宏块编号都从 0 开始),则 $D(s, t-1, m)$ 的值可以递归得到.

对于帧间编码宏块,如果它被正确接收,解码端重建值可表示为 $e(s, t, m, i) + \tilde{F}(s, t-1, m', i)$, 而 $\tilde{F}(s, t-1, m', i)$ 是 $\tilde{F}(s, t, m, i)$ 的预测参考值, $e(s, t, m, i)$ 表示预测残差;如果宏块丢失,在解码端进行错误隐藏.因此帧间编码宏块的传输失真表示为:

$$D_L^P = PE\{[\hat{F}(s, t, m, i) - \tilde{F}(s, t-1, m, i)]^2\} + (1-P)E\{[\hat{F}(s, t, m, i) - e(s, t, m, i) - \tilde{F}(s, t-1, m', i)]^2\} = pE\{[\hat{F}(s, t, m, i) - \hat{F}(s, t-1, m, i)]^2\} + pE\{[\hat{F}(s, t-1, m, i) - \tilde{F}(s, t-1, m, i)]^2\} + (1-P) \cdot E\{[\hat{F}(s, t-1, m', i) - \tilde{F}(s, t-1, m', i)]^2\} = pR_{\text{RFD}}(t, t-1, m) + pD(s, t-1, m) + (1-P) \times D(s, t-1, m'), \quad (3)$$

其中, $D(s, t-1, m')$ 表示 t 帧第 m 个宏块进行帧间预测的参考宏块的传输失真.由于要精确地在宏块级计算 $D(s, t-1, m')$ 非常困难,笔者采用文献[11]中的快速算法进行近似估计.1 个宏块可以分割成更小的子块,为了寻找当前宏块的参考匹配块,使用所有子块运动矢量的平均值作为当前宏块的运动矢量,并根据宏块的运动矢量在 $t-1$ 帧中找到匹配块.此匹配块可能由 4 个宏块组成(图 2).那么 $D(s, t-1, m')$ 等于匹配块所占的 4 个宏块的传输失真加权求和,各宏块的权重因子由用于参考的像素数目与宏块像素总数之比确定,这样就快速地估计了 $D(s, t-1, m')$ 的值.

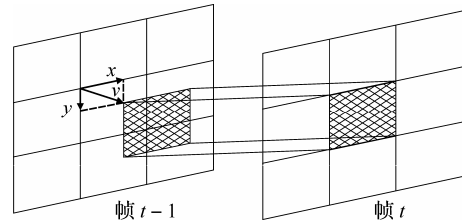


图 2 帧间编码宏块的匹配块示意图

1.2 右视点视频传输失真模型

右视点的帧内编码宏块的失真模型与左视点帧内编码宏块的失真模型类似, 唯一不同之处在于采用的错误隐藏方法. 与左视点不同, 右视点宏块丢失时, 采用视点间宏块拷贝的方法进行错误隐藏. 右视点帧内编码宏块的失真可表示为:

$$\begin{aligned} D_R^I &= pE\{[\hat{F}(s,t,m,i) - \tilde{F}(s-1,t,m,i)]^2\} + \\ &\quad (1-P)E\{[\hat{F}(s,t,m,i) - \hat{F}(s,t,m,i)]^2\} + \\ &\quad pE\{[\hat{F}(s,t,m,i) - \hat{F}(s-1,t,m,i)]^2\} + \\ &\quad pE\{[\hat{F}(s-1,t,m,i) - \tilde{F}(s-1,t,m,i)]^2\} = \\ &\quad pR_{\text{RFD}}(s,s-1,m) + pD(s-1,t,m), \end{aligned} \quad (4)$$

其中, $R_{\text{RFD}}(s,s-1,m)$ 表示编码重建的相邻视点同一时刻帧的第 m 个宏块间的差别, 可在编码端计算得到. $D(s-1,t,m)$ 代表 $s-1$ 视点 t 时刻帧中第 m 个宏块的传输失真, 该值可通过递归计算得到.

右视点中的帧间预测又可分为视点内帧间预测和视点间帧间预测. 视点内帧间预测指参考当前视点前时刻帧进行预测, 如果宏块被正确接收, 则解码重建像素值可表示为 $e(s,t,m,i) + \tilde{F}(s,t-1,m',i)$, 而 $\tilde{F}(s,t-1,m',i)$ 是 $\tilde{F}(s,t,m,i)$ 的预测参考值, $e(s,t,m,i)$ 表示预测残差; 如果宏块丢失, 在解码端进行错误隐藏. 经过以上分析, 右视点的视点内帧间预测宏块的传输失真可表示为:

$$\begin{aligned} D_R^{PI} &= PE\{[\hat{F}(s,t,m,i) - \tilde{F}(s-1,t,m,i)]^2\} + \\ &\quad (1-P)E\{[\hat{F}(s,t,m,i) - e(s,t,m,i) - \\ &\quad \tilde{F}(s,t-1,m',i)]^2\} = pE\{[\hat{F}(s,t,m,i) - \\ &\quad \hat{F}(s-1,t,m,i)]^2\} + pE\{[\hat{F}(s-1,t,m,i) - \\ &\quad \tilde{F}(s-1,t,m,i)]^2\} + (1-P) \cdot \\ &\quad E\{[\hat{F}(s,t-1,m',i) - \tilde{F}(s,t-1,m',i)]^2\} = \\ &\quad pR_{\text{RFD}}(s,s-1,m) + pD(s-1,t,m) + \\ &\quad (1-P)D(s,t-1,m'). \end{aligned} \quad (5)$$

其中, $D(s,t-1,m')$ 估计方法与(3)式估计方法类似.

视点间帧间预测指参考前一视点同时刻的帧进行预测, 如果宏块正确接收, 则解码重建像素值可表示为 $e(s,t,m,i) + \tilde{F}(s-1,t,m',i)$, 而 $\tilde{F}(s-1,t,m',i)$ 是 $\tilde{F}(s,t,m,i)$ 预测参考值, $e(s,t,m,i)$ 表示预测残差; 如果宏块丢失, 在解码端进行错误隐藏. 则右视点的视点间预测宏块的传输失真为:

$$\begin{aligned} D_R^{PV} &= PE\{[\hat{F}(s,t,m,i) - \tilde{F}(s-1,t,m,i)]^2\} + \\ &\quad (1-P)E\{[\hat{F}(s,t,m,i) - e(s,t,m,i) - \\ &\quad \tilde{F}(s-1,t,m',i)]^2\} = pE\{[\hat{F}(s,t,m,i) - \\ &\quad \hat{F}(s-1,t,m,i)]^2\} + pE\{[\hat{F}(s-1,t,m,i) - \\ &\quad \tilde{F}(s-1,t,m,i)]^2\} + (1-P) \cdot \\ &\quad E\{[\hat{F}(s-1,t,m',i) - \tilde{F}(s-1,t,m',i)]^2\} = \\ &\quad pR_{\text{RFD}}(s,s-1,m) + pD(s-1,t,m) + \\ &\quad (1-P)D(s-1,t,m'). \end{aligned} \quad (6)$$

同理, $D(s-1,t,m')$ 的估算方法也与(3)式中 $D(s,t-1,m')$ 的估算方法类似.

2 实验验证及分析

2.1 实验验证

为验证笔者所提出的宏块级立体视频传输失真模型的有效性, 在 JM 平台上对 Book、Crowd、Xmas 等立体视频序列在不同丢包率(分别取 5%和 10%)下进行了测试. 其中, Book 序列运动平缓, Crowd 和 Xmas 序列运动剧烈, 图 3 给出了该 3 个测试序列的第 1 帧图像. 实验采用的立体视频编码结构为 IPPP, 且 I 帧不丢包, 其他帧采用随机丢宏块的方式进行丢包. 实验结果以每帧的和方差 SSE 值作为比较依据(图 4).

2.2 分析与讨论

左视点的第 1 帧为 I 编码帧, 本中设定 I 帧不会产生丢包, 因此左视点第 1 帧的失真都为 0. 从图中可以看出, 各序列在 5%和 10%丢包率下解码



(a) Book 序列第 1 帧



(b) Crowd 序列第 1 帧



(c) Xmas 序列第 1 帧

图 3 测试序列的第 1 帧图像

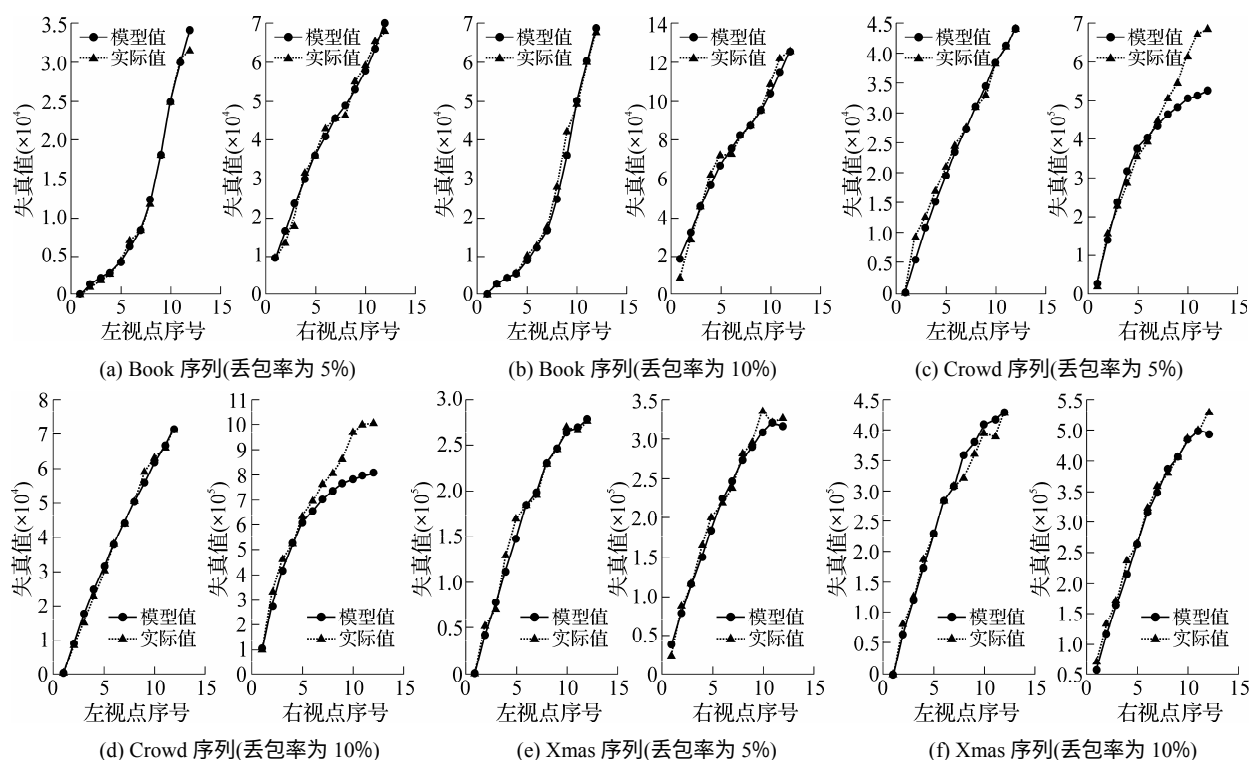


图4 模型失真值与实际失真值实验对比

图像的模型值与 SSE 实际值基本吻合, 略显不足的是, 对于运动较剧烈的 Crowd 序列和 Xmas 序列, 它们的估计精度低于运动平缓的 Book 序列, 这方面还有待改进; 同时由于右视点的失真不仅受左视点帧的影响, 而且受到右视点前面帧的影响, 导致误差累积较大, 失真快速扩散, 所以右视点的效果要劣于左视点。

为进一步从整体上比较算法对传输失真估计效果, 笔者定义序列传输失真估计误差 E 为:

$$E = \frac{1}{N} \times \sum_{n=1}^N \left| \frac{D(s,t) - D_0(s,t)}{D_0(s,t)} \right| \times 100\%, \quad (7)$$

其中, s 为 0 表示左视点, s 为 1 表示右视点. N 为序列帧数, $D(s,t)$ 为本文模型估计的传输失真, $D_0(s,t)$ 为实际传输失真。

为验证丢包率对视频传输失真的稳定性, 定义不同丢包率下的传输失真波动 H 为:

$$H = E_{\max} - E_{\min}, \quad (8)$$

式中, $E_{\max}$ 为同一序列不同丢包率下的最大传输失真估计误差, $E_{\min}$ 为相应的最小平传输失真估计误差。

根据(7)式进行计算, 可分别得到 Book 序列、Crowd 和 Xmas 序列在 3%、5% 和 10% 丢包率下的

传输失真估计误差及不同丢包率下的传输失真波动, 结果见表 1。由表 1 可以得出, 对于以上 3 种测试序列, 使用本文失真度模型估算出来的失真值与实际解码得出的真实的失真值相比, 估算误差基本能够控制在 8% 之内(平均误差为 7.11%), 在不同的视频序列和丢包率下都比较稳定, 从而证明了提出的传输失真模型的有效性。

表 1 各序列的传输失真估计误差及丢包波动 %

丢包率	Book 序列		Crowd 序列		Xmas 序列	
	左视点	右视点	左视点	右视点	左视点	右视点
3%	6.60	7.15	7.12	7.47	7.18	7.39
5%	6.32	7.15	7.08	7.34	7.23	7.32
10%	6.41	7.18	7.06	7.52	7.18	7.32
丢包波动	0.28	0.03	0.06	0.18	0.05	0.07

3 结论

通过研究立体视频编码的特点, 提出了一个快速的宏块级立体视频传输失真模型。该模型能够在各种丢包率下对运动平缓和运动较剧烈的立体视频序列进行传输失真估计, 估计误差在 6%~8% 之间, 丢包率波动最大不超过 0.3%, 说明模型

的估计精度在不同的视频序列和丢包率下都比较稳定. 3种视频序列的平均估计误差为7.11%, 可以较精确地估计立体视频的传输失真, 且模型的计算复杂度适中. 模型对于运动较剧烈的视频序列的估计精度略低于运动缓慢的视频序列, 这方面还有待改进. 在接下来的工作中, 将根据此模型指导宏块重要性区分, 以优化资源, 提高视频服务质量.

参考文献:

- [1] Zhang R, Regunathan S L, Rose K. Video coding with optimal intra/inter mode switching for packet loss resilience[J]. IEEE Journal on Selected Areas in Communications, 2000, 18(6):966-976.
- [2] Yang H, Rose K. Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(7):845-856.
- [3] Chen Z F, Pahalawatta P V, Tourapis A M, et al. Improved estimation of transmission distortion for error-resilient video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(4):636-647.
- [4] 周敬利, 马志龙, 范晔斌, 等. 视频流媒体传输失真研究[J]. 小型微型计算机系统, 2006, 27(2):264-266.
- [5] 游敏慧, 王兵. 无线网络中视频传输的信道失真分析[J]. 无线通信技术, 2003(3):44-47.
- [6] 赵峰, 仲丽媛, 荆涛, 等. 基于端到端失真优化的无线视频联合编码方案[J]. 北京交通大学学报, 2010, 34(6):71-75.
- [7] Wang Y, Wu Z Y, Boyce J. Modeling of transmission-loss-induced distortion in decoded video[J]. IEEE Transaction on Circuits and Systems for Video Technology, 2006, 16(6):716-732.
- [8] 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.
- [9] 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.
- [10] Cote G, Shirani S, Kossentini F. Optimal mode selection and synchronization for robust video communications over error-prone networks[J]. IEEE Journal on Selected Areas in Communications, 2000, 18(6):952-965.
- [11] Zhang D, Liang J, Singh I. Fast transmission distortion estimation and adaptive error protection for H.264/AVC-based embedded video conferencing systems[J]. Signal Processing: Image Communication, 2013, 28(5):417-429.

Fast Modeling of MB-level Transmission Distortion for Stereo Video

YAO Ting, ZHOU Yu^{*}, WANG Xiao-dong, ZHANG Lian-jun, WANG Teng-fei, HU Bin-bin

(Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China)

Abstract: Packet losses during transmission severely affect reconstruction quality of video, even worse when it propagates errors at a large scale. Aiming at evaluating the quality of stereo video, in this paper a transmission distortion model is presented through studying the correlation between video quality and packet loss. The model combines a fast MB-level distortion estimation approach with error concealment strategy of both inter-frame and inter-view MB copy. It can fast estimate the MB-level distortion for stereo video over the channel with packet losses. Experimental results suggest the accuracy of the model, which is expected to shed some technical insights for transmission distortion modeling of stereo video.

Key words: distortion model; stereo video; MB-level

(责任编辑 章践立)