

编码端帧级立体视频传输失真模型

王晓东 王腾飞 胡彬彬 蒋刚毅 章联军

(宁波大学信息科学与工程学院 宁波 315211)

摘要 通过分析网络传输与立体视频失真的关系,结合立体视频时空相关性、网络性能和终端错误隐藏技术,以及递归思想提出了一种编码端的帧级立体视频传输失真模型,本模型可在编码端估计特定网络条件下的终端立体视频质量。针对不同运动程度和视差大小的立体视频序列,在不同网络环境下进行实验。当用 MSE 表征失真时,平均预测误差为 4.16%;当用 PSNR 表征时,平均预测误差为 0.93%,表明本模型可在编码端精确估计终端的立体视频传输失真。

关键词 失真模型,立体视频,网络传输,视频流,错误隐藏,网络性能

中图分类号 TP391

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2014.10.036

Frame-level Stereoscopic Video Transmission Distortion Model at Encoder

WANG Xiao-dong WANG Teng-fei HU Bin-bin JIANG Gang-yi ZHANG Lian-jun

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

Abstract Taking into account the relationship between network transmission and stereoscopic video, the temporal correlation and spatial correlation of stereo video sequences, channel parameters, as well as terminal error concealment technology, we put forward a frame-level stereoscopic video transmission distortion model at the encoder. We made experiment of stereoscopic video sequences with different motion intensity and disparity in different network conditions. When MSE represents transmission distortion, average distortion error is 4.16%. When PSNR represents transmission distortion, average distortion error is 0.93%. The model can accurately estimate the transmission distortion of stereoscopic video at the encoder.

Keywords Distortion model, Stereoscopic video, Network transmission, Video streaming, Error concealment, Network performance

1 引言

近年来,随着多媒体领域的迅猛发展,人们对立体视频相关技术的研究愈加深入^[1,2]。立体视频数据量巨大,如何进行有效的网络传输引起了广泛研究^[3,4]。网络传输失真估计是立体视频传输系统优化的核心环节,精确的传输失真估计可为立体视频编码、网络传输和终端的错误隐藏技术提供有效的决策依据。

传统单视点传输失真模型可以分为两类,一是基于信道仿真的有损率失真模型,如文献[5,6]在编码端利用信道仿真器仿真信道失真;另一类是基于像素域递归的传输失真模型,如 He 等提出的基于像素域递归的 ROPE 算法^[7];Zhang 等人在此基础上进行了改进,提出了亚像素域相关系数的求取算法^[8];Liang 等提出了一种快速失真估计算法^[9],考虑了帧内编码宏块正确接收时产生失真对视频序列的影响。以上方

法只可以估计单视点视频传输失真,在立体视频方面,Tan 等结合不等错误保护提出了一种立体视频率失真优化模型^[10];考虑到视点间的帧内更新,Xiang 等人在文献[11]中提出了一种率失真模型。上述算法更多和率失真相结合,侧重于某个视点的失真估计,忽略了左视点失真对右视点产生的错误扩散影响。Zhou 等人提出了丢包网络中多视点视频传输失真模型,此模型考虑了丢包网络中视差因素对错误扩散的影响,但需要解码端反馈,在编码端不能进行失真估计^[12,13]。

综上所述,当前的单视点传输失真模型没有考虑立体视频特点,立体视频传输失真模型没有考虑左视点失真对于右视点的影响,而多视点传输失真模型则需要解码端反馈。本文首先分析网络传输对立体视频编码码流的影响,然后针对特定的编码结构,根据立体视频序列特点、编码参数、网络参数和相关解码技术在编码端建立了立体视频传输失真模型。

投稿日期:2013-12-17 返修日期:2014-04-08 本文受浙江省教育厅科研资助项目(Y201327703),浙江省自然科学基金(Y1110161),宁波市自然科学基金(2012A610019),科技部国家科技支撑计划(2012BAH67F01),国家自然科学基金(60832003,61071120),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(2011B81002)资助。

王晓东(1970—),男,副教授,主要研究方向为网络通信、无线传感网络、信号处理,E-mail: wangxiaodong@nbu.edu.cn;王腾飞(1989—),男,硕士生,主要研究方向为立体视频编码和网络传输;胡彬彬(1989—),男,硕士生,主要研究方向为立体视频传输和错误隐藏;蒋刚毅(1964—),男,教授,主要研究方向为多媒体处理与通信体信息、信息处理、立体视频传输;章联军(1980—),男,主要研究方向为立体视频编码、网络传输、信息隐藏。

2 端到端传输失真分析

由于立体视频流的平均码率大,传输时间较长,需要稳定且充足的带宽资源作为保障,但实际的网络带宽是动态变化的,这使得网络有效带宽不能实时满足传输要求;当前的IP网络提供尽力而为的服务,这种服务可能导致传输差错、网络延迟和数据包丢失。传输差错会降低终端质量甚至无法解码,网络延迟导致丢包,降低终端感知质量。

在立体视频流传输过程中,带宽不足、传输延迟和时延抖动最终表现为数据包丢失,终端视频质量降低。立体视频编码码流之间存在着强烈的时空相关性,某帧或某宏块丢失不仅使当前帧产生失真,还会造成错误扩散。网络传输中的数据包丢失是影响终端立体视频失真的重要因素。

下文将从数据丢包的角度出发,研究丢包率对于传输失真的影响。

3 立体视频传输失真模型

立体视频流传输中的数据包丢失是不可避免的,我们称网络传输丢包产生的失真为传输失真。假设编码端重建的 s 视点 t 时刻像素 i 的值记为 $F_i(s, t)$,解码端重建的对应像素的值记为 $\bar{F}_i(s, t)$,则 s 视点 t 时刻帧的传输失真可用 $D(s, t)$ 表示:

$$D(s, t) = E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} \quad (1)$$

通过上式可知,立体视频某帧的传输失真可表示为编码端重建帧与解码端重建帧所有对应像素的均方差期望。下文将讨论丢包率对立体视频帧的影响。假定立体视频编码格式为IPPP结构,立体视频流的丢包率为 p ($0 < p < 1$)。假设某个宏块丢失,在解码端采用下面错误隐藏方法进行恢复:左视点帧宏块丢失,直接拷贝同一视点上一时刻帧对应位置宏块;右视点帧宏块丢失,则直接拷贝同一时刻上一视点对应位置宏块。本文假设左视点第一帧(I帧)不会产生数据丢失。

3.1 左视点帧传输失真模型

立体视频左视点帧中的宏块可分为帧内编码宏块和帧间编码宏块。帧内编码宏块的传输失真可表示为:

$$\begin{aligned} D_L^I &= p * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} + (1-p) * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} \\ &= p * E\{[F_i(s, t) - F_i(s, t-1) + F_i(s, t-1) - \bar{F}_i(s, t-1)]^2\} \end{aligned} \quad (2)$$

文献[9]证明帧内编码宏块正确接收时的错误扩散可忽略不计,所以正确接收部分的传输失真为零。文献[14]的实验表明,编码器引起的失真和信道传输引起的失真不相关,即:

$$\begin{aligned} D_L^I &= p * E\{[F_i(s, t) - F_i(s, t-1)]^2\} + p * E\{[F_i(s, t-1) - \bar{F}_i(s, t-1)]^2\} \\ &= p * M(t, t-1) + p * D(s, t-1) \end{aligned} \quad (3)$$

其中, $M(t, t-1)$ 指编码端重建的同一视点相邻时刻帧的MSE, $D(s, t-1)$ 指 s 视点 $t-1$ 时刻帧的失真。

对于帧间预测的宏块,如果正确接收,则解码端重建像素为 $e_i(s, t) + \bar{F}_j(s, t-1)$, s 视点 $t-1$ 时刻帧图像的像素 j 指参与预测 s 视点 t 时刻像素 i 的像素集合, $e_i(s, t)$ 指编码残差,则左视点帧间编码宏块失真 D_L^I 为:

$$D_L^I = p * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} + (1-p) * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\}$$

$$\begin{aligned} &= p * E\{[F_i(s, t) - F_i(s, t-1) + F_i(s, t-1) - \bar{F}_i(s, t-1)]^2\} + (1-p) * E\{[F_i(s, t) - e_i(s, t) - \bar{F}_j(s, t-1)]^2\} \\ &= p * M(t, t-1) + p * D(s, t-1) + (1-p) * E\{[F_j(s, t-1) - \bar{F}_j(s, t-1)]^2\} \\ &= p * M(t, t-1) + p * D(s, t-1) + (1-p) * b * D(s, t-1) \end{aligned} \quad (4)$$

b 指某帧正确接收宏块失真占该帧总失真的比例, b 与序列的运动剧烈程度有关,在运动平稳的序列中, b 是一个大于1小于0的常数。如果 s 视点 t 时刻帧的帧内编码比例为 β ,则该帧的传输失真可表示为:

$$\begin{aligned} D(s, t) &= b * D_L^I + (1-b) * D_L^I \\ &= b * [p * M(t, t-1) + p * D(s, t-1)] + (1-b) * [p * M(t, t-1) + p * D(s, t-1) + (1-p) * b * D(s, t-1)] \\ &= p * M(t, t-1) + [p + (1-b) * (1-p) * b] * D(s, t-1) \\ &= 0, D(0, 0) = 0 \end{aligned} \quad (5)$$

从式(5)可以看出,左视点帧的传输失真模型是一个递归公式,它与相邻时刻帧均方差及相同视点上一时刻帧失真相关。如果得到初始帧的传输失真和视频序列的相关参数,即可求出后续帧的传输失真。

3.2 右视点帧传输失真模型

相比于左视点,右视点帧的帧间预测又可分为视点内预测和视点间预测。假设右视点帧中帧内预测宏块比例为 β ,视点内预测宏块比例为 γ ,数据丢失时采用拷贝上一视点对应位置的错误隐藏方法进行恢复。则帧内编码宏块失真 D_R^I 为:

$$\begin{aligned} D_R^I &= p * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} + (1-p) * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} \\ &= p * E\{[F_i(s, t) - F_i(s-1, t) + F_i(s-1, t) - \bar{F}_i(s-1, t)]^2\} \\ &= p * E\{[F_i(s, t) - F_i(s-1, t)]^2\} + p * E\{[F_i(s-1, t) - \bar{F}_i(s-1, t)]^2\} \\ &= p * M(s, s-1) + p * D(s-1, t) \end{aligned} \quad (6)$$

$M(s, s-1)$ 指编码端重建的同一时刻相邻视点帧的MSE, $D(s-1, t)$ 指 t 时刻上一视点帧的传输失真。帧内编码宏块失真与相邻视点帧均方差及相同时刻上一视点帧失真有关。

帧间编码宏块传输失真 D_R^I 为:

$$\begin{aligned} D_R^I &= p * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} + (1-p) * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} \\ &= p * E\{[F_i(s, t) - F_i(s-1, t) + F_i(s-1, t) - \bar{F}_i(s-1, t)]^2\} + (1-p) * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} \\ &= p * M(s, s-1) + p * D(s-1, t) + (1-p) * E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} \end{aligned} \quad (7)$$

由于预测关系不同,正确接收宏块传输失真可分为视点内预测失真 $D_R^{I'}$ 和视点间预测失真 $D_R^{I''}$,则 $E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\}$ 可描述为:

$$\begin{aligned} E\{[F_i(s, t) - \bar{F}_i(s, t)]^2\} &= \gamma * D_R^{I'} + (1-\gamma) * D_R^{I''} \end{aligned}$$

$$\begin{aligned}
&= \gamma * E\{[F_i(s, t) - e_i(s, t-1) - \bar{F}_j(s, t-1)]^2\} + (1 - \gamma) * E\{[F_i(s, t) - e_i(s-1, t) - \bar{F}_j(s-1, t)]^2\} \\
&= \gamma * E\{[F_j(s, t-1) - \bar{F}_j(s, t-1)]^2\} + (1 - \gamma) * E\{[F_j(s-1, t) - \bar{F}_j(s-1, t)]^2\} \\
&= \gamma * b * D(s, t-1) + (1 - \gamma) * c * D(s-1, t) \quad (8)
\end{aligned}$$

b 与序列的运动剧烈程度有关。 c 指右视点某帧对应的上一视点帧中正确接收部分失真与该帧总失真的比例, 与立体视频空间相关性有关, 在一个变化稳定的序列中, b 和 c 都是一个大于 1 小于 0 的常数。帧间编码宏块传输失真 D_R^k 为:

$$D_R^k = p * M(s, s-1) + p * D(s-1, t) + (1-p) * [\gamma * b * D(s, t-1) + (1-\gamma) * c * D(s-1, t)] \quad (9)$$

帧间编码宏块传输失真受相邻视点帧均方差、上一时刻帧失真及上一视点帧失真的影响。

右视点帧的传输失真模型为:

$$\begin{aligned}
D(s, t) &= \beta * D_R^k + (1-\beta) * D_R^k \\
&= \beta * [p * M(s, s-1) + p * D(s-1, t)] + (1-\beta) * [p * M(s, s-1) + p * D(s-1, t) + (1-p) * [\gamma * b * D(s, t-1) + (1-\gamma) * c * D(s-1, t)]] \\
&= A1 * M(t, t-1) + A2 * D(s, t-1) + A3 * D(s-1, t) \quad (10)
\end{aligned}$$

其中, $A1 = p$, $A2 = (1-\beta) * (1-p) * b * \gamma$, $A3 = p + (1-\beta) * (1-p) * c * (1-\gamma)$ 。右视点帧的传输失真与相邻时刻帧的 MSE、上一视点帧失真及上一时刻帧失真有关。只要得到初始化传输失真和视频序列的相关参数, 即可求出右视点帧的传输失真。

4 实验与讨论

本文在 JM 平台对所提模型进行测试。码流丢包处理后在终端进行错误隐藏恢复, 计算实际失真; 同时在编码端通过模型得到理论失真。选取 book 和 crowd 序列在 10% 丢包率下进行实验。为了减小实验随机性, 在相同条件下取 100 次实验的平均值作为最终结果。实验结果如图 1—图 4 所示。

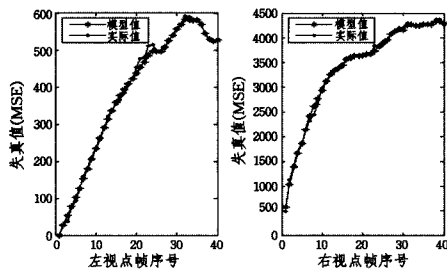


图 1 book 序列传输失真预测对比结果(MSE)

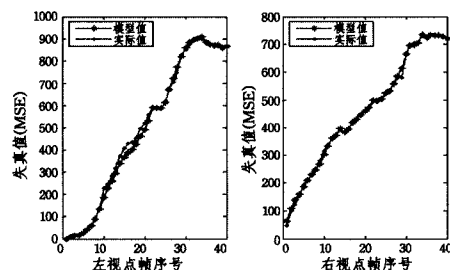


图 2 crowd 序列传输失真预测对比结果(MSE)

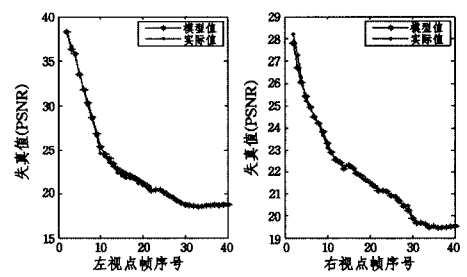


图 3 book 序列传输失真预测对比结果(PSNR)

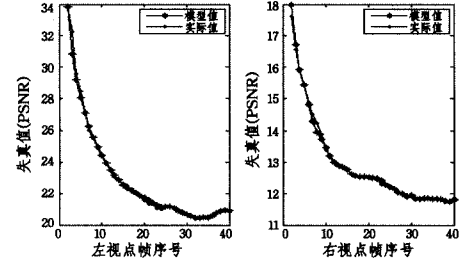


图 4 crowd 序列传输失真预测对比结果(PSNR)

由图可知, 两个序列失真都逐帧递增, 这是因为数据丢失不仅使本帧产生失真, 还会向后进行错误扩散, 错误积累会导致后续帧的失真越来越大。crowd 序列比 book 序列运动剧烈, 帧内编码宏块较多, 这会导致更多的帧内编码宏块丢失; 且运动剧烈序列的运动矢量较大, 再加上终端采用帧拷贝方式进行错误隐藏恢复, 以上原因使得运动剧烈序列的失真大于运动缓慢序列失真, 这点可以由图得到反映。视差大小决定了左右视点失真差异情况, 视差越小, 视点间相关性越强, 采用本文错误隐藏方法恢复以后的左右视点帧失真差异就越小。book 序列的平均视点内编码比率为 74%, 而 crowd 序列为 96%, 由此可见 crowd 序列的视点间相关性较弱, 所以 crowd 序列左右视点失真差值大于 book 序列。

本文测试不同特性序列以证明模型适用性。为验证模型的准确性, 定义视点平均预测误差:

$$E(s) = \frac{\sum_{t=0}^N |D(s, t) - MSE(s, t)|}{\sum_{t=0}^N MSE(s, t)} \quad (s=0, 1) \quad (11)$$

$D(s, t)$ 指 s 视点 t 时刻帧的模型失真, $MSE(s, t)$ 指 s 视点 t 时刻帧的实际失真。

表 1 视频序列左右视点平均预测误差(单位: %)

序列丢包率	视点方向		视点方向	
	左视点	右视点	左视点	右视点
	(MSE)	(MSE)	(PSNR/dB)	(PSNR/dB)
book	3%	4.82	5.72	0.71
	5%	2.11	3.64	0.90
	10%	2.58	1.23	0.76
crowd	3%	1.23	3.41	0.21
	5%	1.05	2.64	0.17
	10%	1.49	0.99	0.51
flamenco	3%	13.25	3.14	1.44
	5%	1.89	2.81	0.43
	10%	5.84	3.14	1.57
xmas	3%	12.13	13.47	1.53
	5%	2.40	7.30	1.23
	10%	3.54	7.10	0.65
flamenco2	3%	3.83	3.55	0.46
	5%	2.52	1.40	0.55
	10%	2.95	3.68	1.24
平均	4.11	4.21	0.82	1.04

表 1 给出了不同特性序列在不同网络环境下的平均预测误差。其中 book 序列为运动缓慢序列, 相机间距 6.5cm; xmas 为中等运动剧烈序列, 相机间距 3mm; crowd 为运动剧烈序列, 相机间距 20cm; flamenco2 和 flamenco 为旋转运动序列, 相机间距 20cm。由表可知, 用 MSE 表征时, 左视点平均失真误差 4.11%, 右视点平均失真误差 4.21%; 用 PSNR 表征时, 左视点平均失真误差 0.82%, 右视点平均失真误差 1.04%, 证明本文所提模型适用于不同丢包率下不同视差大小和运动剧烈程度序列。

结束语 本文提出了一种递归式立体视频传输失真模型。模型能够精确预测终端的立体视频传输失真, 适用于不同运动剧烈程度和视差大小的视频序列, 且在不同的网络传输环境下均表现良好。模型可用于确定立体视频编码参数, 提高率失真优化性能, 优化立体视频传输机制(如拥塞控制), 指导信道的差错控制和终端错误隐藏技术, 为丢包网络环境下立体视频传输系统的进一步优化提供了有力保障。

参考文献

- [1] Zilly F, Kluger J, Berlin, et al. Production rules for stereo acquisition[J]. Proceedings of the IEEE, 2011, 99(4): 590-606
- [2] Urey H, Erden E, Surman P. State of the art in stereoscopic and autostereoscopic displays[J]. Proceedings of the IEEE, 2011, 99(4): 540-555
- [3] 闫庆森, 李临生, 徐晓峰, 等. 视频跟踪算法研究综述[J]. 计算机科学, 2013, 40(Z6): 204-209
- [4] Gurler, C G, Koc Univ E, et al. Flexible transport of 3-D video over networks[J]. Proceedings of the IEEE, 2011, 99(4): 694-707
- [5] Thomas S, Tomas W, Stephan W. Optimized transmission of H. 26L/JVT coded video over packet-lossy network[C]//IEEE International Conference on Image Processing. NY, 2002: 173-176
- [6] 唐振华, 刘文予, 赵爽, 等. 基于视频编码流的视频传输端到端失真度估算[J]. 华中科技大学学报: 自然科学版, 2008, 36(7): 25-29

(上接第 133 页)

结束语 本文介绍了一种基于嵌入式平台的面向对象的 P2P VOD 缓存策略。由于嵌入式系统的性能和资源限制, 无法缓存大的 VOD 节目片段, PC 上的 VOD 系统中一些常用的 P2P 缓存策略效果不明显。本文提出的这个动态缓存机制结合了嵌入式平台资源有限的特点, 根据 VOD 系统中不同节目的热度来动态调整缓存结构, 最大限度地保障了 VOD 系统中热度高的节目资源尽可能多, 同时也让低热度的节目也有一个大小比较合理的节目缓冲区。该动态缓存机制较好地提高了 P2P VOD 系统的性能。

参考文献

- [1] Liu J, Rao S G, Li B, et al. Opportunities and challenges of peer-to-peer Internet video broadcast[J]. Proc. IEEE, 2007, 96(1): 11-24
- [2] PPLive. Huazhong Univ. Sci. Technol., Wuhan, China [OL]. http://www.pplive.com/
- [3] UUSee. UUSee, Inc., Beijing, China [OL]. http://www.uusee.

- [7] He Zhi-hai, Cai Jiang-fei, Chen Chang-wen. Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(6): 511-523
- [8] Zhang Yuan, Gao Wen, Lu Yu, et al. Joint source-channel rate-distortion optimization for H. 264 video coding over error-prone networks[J]. IEEE Transactions on Multimedia, 2007, 9(3): 445-454
- [9] Zhang Dong, Liang Jie, 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
- [10] Tan A S, Aksay A, Akar G B, et al. Rate-distortion optimization for stereoscopic video streaming with unequal error protection[J]. EURASIP Journal on Applied Signal Processing, 2009, 2009: 7
- [11] Xiang Xin-guan, Zhao De-bin, Ma Si-wei, et al. Rate-distortion optimization with inter-view refreshment for stereoscopic video coding over error-prone networks[C]//IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2009: 72570K-72570K-8
- [12] Zhou Yuan, Hou Chun-ping, Pan R, et al. Distortion analysis and error concealment for multi-view video transmission[C]//2010 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting. Shanghai, 2010: 1-5
- [13] Zhou Yuan, Hou Chun-ping, Xiang Wei. Modeling of transmission distortion for multi-view video in packet lossy networks[C]//Global Telecommunications Conference (GLOBECOM 2010). Maimi, 2010: 1-5
- [14] Wang Yao, Wu Zhen-yu, Boyce J M. Modeling of transmission-loss-induced distortion in decoded video[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(6): 716-732

com/

- [4] Jiang X, Dong Y, Xu D Y, et al. GnuStream: a P2P Media Streaming System Prototype[C]//Proceedings of International Conferences on Multimedia & Expo. (ICME'03). Maryland, USA, 2003
- [5] Ripeanu M, Foster I. Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design[J]. IEEE Internet Computing Journal, 2002, 6(1): 50-57
- [6] Hefeeda M M, Bhargava B K, Yau D K Y. A hybrid architecture for cost-effective on-demand media streaming[J]. Computer Networks, 2004, 44(3)
- [7] Cui Y, Li B, Nahrstedt K. oStream: asynchronous streaming multicast in application-layer overlay networks[J]. IEEE Journal on Selected Areas in Communications, 2004, 22(1): 91-106
- [8] Guo L, Chen S, Zhang X. Design and evaluation of a scalable and reliable P2P assisted proxy for on demand streaming media delivery[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(5): 669-682