基于内容丢帧的视频自适应传输优化策略①

胡胜红, 吴保荣, 贾玉福

(湖北经济学院 信息管理学院, 武汉 430205)

摘 要: 根据用户对视频内容的个性化偏好, 提出结合语义事件与精彩度的视频内容分级方法, 构建时域内基于多粒度语义内容的统一丢帧模型, 在 RTP/RTSP 流媒体服务器中基于语义丢帧算法设计了视频自适应传输优化策略. 实验结果表明, 本文提出的基于内容的丢帧策略在网络性能、语义质量等方面优于传统基于帧类型的丢帧策略.

关键词: 自适应传输; 基于内容丢帧; 足球视频

Content-Based Frame Dropping in Optimized Strategy of Video adaptive Transmission

HU Sheng-Hong, WU Bao-Rong, JIA Yu-Fu

(School of Information Management, Hubei University of Economics, Wuhan 430205, China)

Abstract: According to user's preference to video content, a measurement for importance of video content with semantic event and exciting intensity is proposed. A multiple granularity of semantic content based frame dropping model in temporal domain is constructed. The video adaptive transmission strategy based on semantic frame dropping is designed. Results of the experiments indicate that the proposed strategy achieves much more improvement in network performance and semantic quality than traditional approach of frame-type based frame dropping.

Key words: adaptive transmission; content-based frame dropping; soccer video

随着 YouTube、优酷网等在线媒体网站的迅猛发展, 海量视频内容的分享与消费成为 Internet 热门应用之一. 然而, 无线网络中视频内容的自适应传输不仅受到网络信道质量的约束, 还受到理性用户的偏好多样性等主观因素约束, 导致用户体验质量始终难以有效提高. 传统的实时视频流技术不考虑内容重要性, 仅仅解决信号层错误, 无法满足端到端的用户体验需求. 因为用户偏好体现出个性化语义事件有特定的质量要求, 即视频内容因用户的偏好不同而赋予了不同的重要性, 重要的内容应该在传输前被标注并且在传输中使用信号层技术予以有效保护, 而不重要的内容可以在网络质量较差时可被裁减甚至有丢弃, 使得网络资源能得到有效利用. 因此, 在应用层确定内容感知的自适应传输策略, 在信号层智能处理自适应传输内容的基于内容自适应传输技术成为视频流中优化用户体验质量的制胜法宝[1,2].

基于 RTP/RTSP 流传输技术因为较低的客户端开销和良好的音视频同步技术, 成为当前主流的实时视频流传输技术之一. 在 RTP 流传输系统中, 各帧图像被独立封包发送, 当网络出现拥塞时, 丢掉非参考帧可以达到缓解拥塞和改善率失真质量的目的, 这是基于传统 RTP 流自适应传输流的研究热点[3,4]. 近年来, 由于视频内容分析技术得到快速发展, 视频数据中的语义特征能够使用领域知识和机器学习相结合的方法被准确识别出来, 内容重要性标注以及对主观质量的优化策略便成为当前自适应视频传输的研究热点. 于[5], Khan[6]等基于运动强度特征研究低比特率环境中视频自适应传输策略时均考虑了内容重要性分级, 针对运动强度较高的节目内容优化了传输质量. 然而, 这些自适应传输方案仍未考虑高层语义内容的重要性

① 基金项目:湖北省自然科学基金(2012FFC01301);湖北省教育厅重点项目(D20132205)
收稿时间:2013-11-30;收到修改稿时间:2014-01-03

与用户偏好无关, 不能直接有效地提高用户访问媒体的主观满意度. 针对这一问题, 本文提出基于内容丢帧(content-based frame dropping, CBFD)的实时自适应视频传输算法, 根据用户偏好和内容特征标注事件重要性级别, 在不同语义层次上制定丢帧策略, 适配可变网络带宽同时优化用户体验质量.

1 CBFD实时自适应视频传输系统

CBFD 实时自适应视频传输系统被设计为内容标注、CBFD 决策、QoS 探测和实时传输引擎等四个主要组件. 内容标注通过特征分析和结构分析自动检测视频流中的语义内容, 根据用户偏好的语义相关性确定重要性级别. 以足球视频为例, 情感分析方法被广泛用于识别精彩事件片段[7], 而多模融合方法能标注出视频内容中详细的语义描述, 包括比赛事件、比赛对象、比赛结果等[8]. QoS 探测使用协议相关的方法检测拥塞状况, 量化为不同网络质量级别. 实时传输引擎根据 CBFD 策略参数在不同粒度语义单元内丢弃一定比率的非重要帧, 同时兼顾 GoP 结构内编码帧的参考顺序, 自适应调节发送速率避免拥塞加剧, 如图 1 所示.

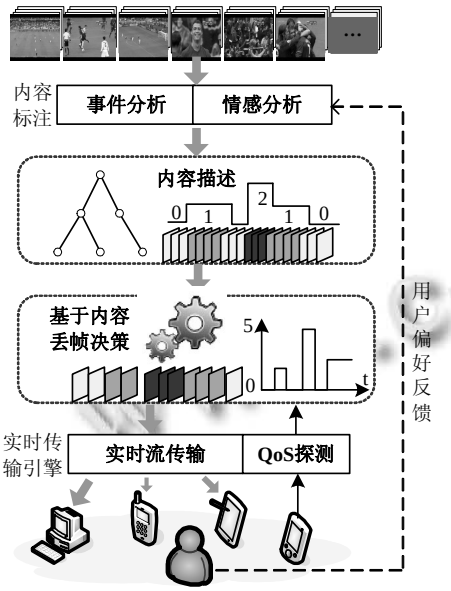

图 1 基于内容丢帧的自适应传输系统框图

2 融合用户偏好的内容重要性标注

足球比赛节目中, 用户对视频内容的偏好通常体现为两类不同的需求, 一类用户对确定性事件如进球、射门、角球、任意球、犯规、红黄牌等比较关注; 另一类则表明对精彩事件(highlight)感兴趣, 可以用情感激励原理计算的精彩度值表征. 然而, 用户的选择总是很模糊, 如果用户偏好的可播放事件范围较大, 所有事件之间难以直接区分其精彩性程度; 而以情感激励因子计算的精彩度阈值虽能较客观地反映精彩程度, 但无法准确描述事件类别.

融合事件偏好和情感特征的语义内容分级方法将足球视频内容以事件为单元标注为 3 个重要性级别.

1)进球事件是所有用户都关注的事件, 具有最高级别;

2)被用户选择的其它事件如未进球的射门、角球以及红黄牌犯规事件若精彩值大于阈值, 也归于最高级别事件, 否则作为重要性级别较低的事件;

3)未被用户选择的事件片段都归为不重要事件.

由于被用户选择的事件通常被用户认为是精彩的, 且实际比赛进程中又在精彩程度上有不同差异, 那么该标注方法能够有效降低用户主观意图的模糊性, 可表示为式(1).

$$Event=\{e_i \mid goal, shot, corner, freekick, foul, card, else\}$$
$$r(i)=\begin{cases}2, pick(e_i)=1 \vee e_i=goal\\ 1, pick(e_i)=1 \wedge highlight(e_i)\ge TH_{highlight}\\ 0, else\end{cases} \quad (1)$$

其中, highlight(ei)表示事件 ei 的精彩度值, 由运动强度 MA(ei)、镜头切换率 SCR(ei)和声音短时能量 STE(ei)等特征融合而来.

$$highlight(e_i)=\frac{1}{J}\sum_{j\in e_i}^{J}\left[MA_j(e_i)+SCR_j(e_i)+STE_j(e_i)\right]/3 \quad (2)$$

因此, 视频数据可标注为一系列重要性不等但时间上连续的语义单元, 级别 2 表示最重要, 级别 1 表示重要, 级别 0 表示不重要. 用户在连接到视频节目前, 被要求对偏好的事件进行选择和对精彩度阈值进行设置. 如果用户略过该步骤, 进球事件被默认为最重要事件(级别 2), 其它事件只有大于默认精彩度阈值的才被认为是重要事件(级别 1).

3 实时视频流自适应传输决策

视频内容在时域上可以表示为语义层和信号层. 语义层结构单元包括事件、镜头、关键帧; 信号层结构单元包括 GoP、帧等. 事件是粒度最大的语义单元并且已经被标注了重要性级别, 构成事件的低层语义

单元也继承了事件的重要性，自适应传输就是根据实时 QoS 条件以丢帧的操作方式丢弃不同粒度语义单元. 由于丢弃不同粒度的语义单元对拥塞状况的改善和对用户体验质量的影响都存在差异性，那么一定存在优化的丢帧策略.

3.1 基于不同语义单元的足球视频丢帧策略

3.1.1 镜头级丢帧策略

足球视频中，镜头可以进一步分类为长镜头、中镜头和特写镜头[9]，不同的镜头类型承载了拍摄者的不同语义. 其中，长镜头被用来跟踪比赛进程，娱乐性较强. 中镜头被用来特写球员动作或回放重要事件，展示比赛的细节信息，娱乐性和信息性兼而有之. 而特写镜头则通常在进球或犯规后用来特写球员表情，以展示特定信息内容为主. 用户观看视频的平均满意度(Mean Opinion Score, MOS)通常跟内容的信息娱乐性相关[10]，信息性强的语义内容可以较大幅度地降低帧率却对用户的主观需求影响不大，例如通过几个关键帧就可以让用户了解到主要的信息. 而娱乐性强的内容则需要较高的帧率使得观众始终能看到平滑的比赛进程. 据此，我们提出了基于镜头类型的丢帧策略:

1)如果该镜头属于长镜头，对于重要镜头至多丢弃全部的 B 帧，而对于非重要镜头直接丢掉剩余的全部帧，将带宽节省下来传输其它重要性更高的镜头;

2)如果该镜头属于中镜头，则对于重要镜头可以丢弃一部分 P 帧，而非重要镜头可以只传输关键帧;

3)如果该镜头属于特写镜头，则无论重要镜头还是不重要镜头，都可以在网络变差时直接丢弃全部 P 和 B 帧，而只保留关键帧.

3.1.2 GoP 级丢帧策略

GoP 级丢帧策略主要是为了防止解码错误而对信号层编码参考性的折中考虑. GoP 级丢帧策略设计如下:

1)因为丢 B 帧不会引起解码错误, B 帧总是以固定偏移方式丢弃以保证回放效果的平滑性;

2)当 P 帧被丢弃的时候，丢弃顺序应该与解码顺序相反，即 GoP 中位置排在后面的 P 帧先被丢弃;

3)每个 GoP 中的 I 帧作为关键帧一般不丢弃，但如果被丢弃，则整个 GoP 内其它帧全部丢弃.

3.2 实时网络质量估计

实时视频流自适应传输设计的另一个重要环节是估测网络质量级别. 在 Apple DSS 服务器[11]中估测网络质量可利用其在传输层实现的 Reliable UDP 协议. Reliable UDP 以 TCP vega 方式改进了不可靠的 UDP 协议运载 RTP 包，每个在发送窗口中的 RTP 包能否顺利发送都依赖于上一次确认包到达，如果发生超时重传，该包的实际发送时延将大大提高. 在网络拥塞从出现到恶化的各阶段，拥塞窗口中待发送包的发送时延越来越大，一旦超出设定阈值，即出现丢包现象. 因此，网络质量状况可以通过对往返时延的观测值判断得出. 根据当前包发送时延的大小 d_k，以及设定的时延阈值 $\delta_x(x = 0,1,2,3,4)$，可以估计出实时网络质量等级. 网络质量被量化为 6 个级别，其赋值条件可表示为式(3).

$$q_k=\begin{cases}0, d_k \le \delta_0 \ OR\ (d_k \le \delta_1 \ AND\ q_{k-1} \le 2)\\ q_{k-1}--, d_k \le \delta_1 \ AND\ q_{k-1}>0 \ AND\ d_k < d_{k-1}\\ q_{k-1}++, d_k > \delta_2 \ AND\ d_k > d_{k-1}\\ q_{k-1}++, d_k > \delta_3\\ 5, d_k \ge \delta_4\end{cases} \tag{3}$$

根据式(3)设定合适的五个阈值 $\delta_{0\sim4}$ 可以有效地将带宽变化设置在表 1 所示的区间内.

表 1　网络质量取值与带宽波动区间

网络质量级别	0~1	2~3	4~5
带宽波动区间	[0, 30%)	[30%, 50%)	[50%, 100%]

3.3 基于内容的实时自适应丢帧算法设计

在定义好网络质量后，基于内容的丢帧决策参数已经全部确定，现在根据实时流媒体传输特性设计基于内容的丢帧决策算法. 在实际传输中，丢不同类型 I/P/B 帧对网络拥塞改善的程度是不一样的. 如图 2 所示，实际调查结果表明 I、P、B 三种帧在不同质量编码的视频节目中大多数情况下都接近于 15%、50%、35%的数据比率.

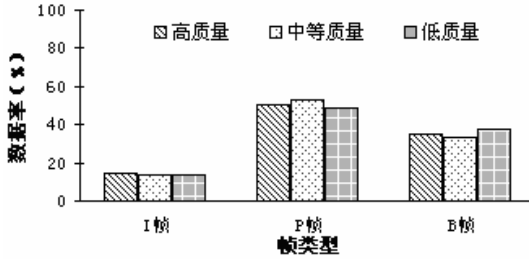

图 2　不同类型帧在视频数据中所占比率

由于现有编码标准中所有视频数据都由若干个 GoP 构成，每个 GoP 结构相同，以 GoP 为单位设计 I/P/B 三种帧的传输比例，使表 1 所示的网络质量和带宽波动区间可以近似地表示为表 2 所设置的传输比例.

表 2 丢帧比率与带宽有效占用比

帧比率	I	I+25%P	I+75%P	I+全 P	I+P+B
带宽率	15%	30%	50%	65%	100%

视频内容的描述文件存储在流媒体服务器中，取包函数依次读取每一帧关联的三种内容特征值：帧号 i、内容级别 r_i、镜头类型 st_i. 再根据当前质量 q_i，联合确定当前帧是否应该被传输. 如果某帧被确定为必须传输，则使用取包函数 GetPackets(i)取出所有帧 i 的包，否则跳过该帧，详细的决策算法流程如下：

算法 1 基于实时自适应传输的 CBFD 算法

(1)读取实时判断的网络质量级别 q_i;

(2)读取第 i 帧的内容描述向量(i, t_i, r_i, st_i);

(3)**如果**当前 q_i>=4

如果 r_i>0，只有当前帧是关键帧才传输;

否则，丢弃整个镜头，读取下一镜头首关键帧;

否则，如果 q_i>=2，根据 r_i 和 st_i 执行丢帧操作:

(3.1)如果 st_i 为长镜头且 r_i>0，跳过当前镜头中的 B 帧而传输所有 I/P 帧，否则只传输当前镜头中每个 GoP 的关键帧和前 75%P 帧;

(3.2)如果 st_i 为中镜头且 r_i>0，只传输当前镜头中每个 GoP 的关键帧和前 75%P 帧，否则只传输当前镜头中每个 GoP 的关键帧和前 25%P 帧;

(3.3)如果 st_i 为特写镜头且 r_i>0，只传输当前镜头中的关键帧和前 25%P 帧，否则只传输当前镜头中的关键帧;

(3.4)否则，跳过当前 GoP 内剩余的所有帧.

否则，如果 q_i>0 且 r_i<2，只要 t_i 为 B 帧就跳过;

否则传输全部帧.

(4)每取一帧都重复执行(1)-(3).

其中，网络质量 q_i 值从 0 到 5 变化，表明网络质量从最好变为最差. 内容级别 r_i 值取 0 至 2 表明内容重要性从低到高. 既然内容特征是离线分析并存储在标注文件中，传输前已预取到内存，而网络质量也是实时计算得到的，该算法的时间复杂度为 O(n)，完全符合实时流传输的要求.

4 实验结果与分析

本实验系统使用 C/S 方式构建，服务端安装基于 CBFD 原理重新设计的 DSS5.5.5 流媒体服务器，客户端使用 Quicktime 7.5 点播视频流，服务器和客户端之间的网络带宽通过安装了 Nistnet2.0.12b[12]的 Linux 路由器控制在 380kbps，512kbps，760kbps 等三个级别内变化，所有的帧发送记录、丢包记录和网络质量变化均通过 DSS 服务器日志采集.

如图 3 所示，从 2006 世界杯 “意大利 vs 乌克兰(1)”和 2010 世界“西班牙 vs 荷兰(2)”两场比赛中各选取一段 3 分钟视频验证我们提出的 CBFD 算法，其中 clip1 包含两个重要的事件：任意球和进球，clip2 包含两个重要事件：射门和进球. 两段视频均为 QuickTime 线索化(hinted)的 h.264 格式码流，帧率为 25 帧每秒，GoP 长度为 30，平均比特率为 1300kbps. 基于帧类型丢帧(简记为 FD)的自适应视频传输策略仅仅考虑从 I、P、B 帧优先级的各种时域组合设计丢帧算法，被用来与本文提出的 CBFD 丢帧算法在网络吞吐量、语义质量以及视频质量方面进行性能比较. 模拟用户偏好，对 clip1 将选择进球和射门两个重要事件，并将精彩度阈值设定为 0.65，视频内容分级结果如图 3(1)所示；对 clip2 将选择进球和任意球两个重要事件，并将精彩度阈值设定为 0.55，视频内容分级结果如图 3(2)所示.

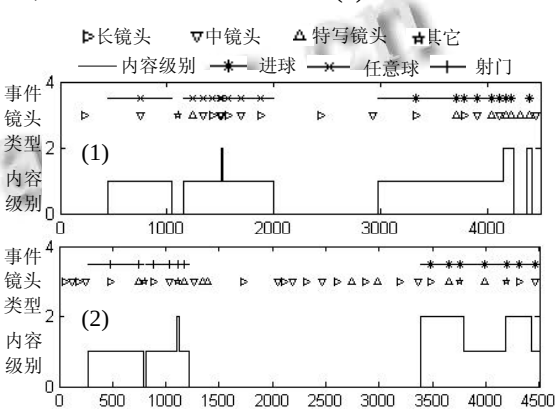

图 3 视频内容重要性标注结果

4.1 网络质量分析

CBFD 算法和基于帧类型的 FD 算法都能够根据预测的网络质量丢帧，但 FD 算法对网络质量的反应更为迟缓，不同镜头和事件中的丢帧概率相同. 由于重要事件往往比不重要事件的语义质量要求高，因此重要事件应尽可能少丢帧，又由于一个重要事件中通

常穿插一些不太重要的观众镜头和特写镜头，CBFD 算法可以丢弃这些镜头中更多的帧为同一事件的其它镜头预留缓存等网络资源.

CBFD 算法根据内容特征判断不同镜头类型使用不同的丢帧概率，短镜头丢帧概率高于长镜头和中镜头，有时候仅仅只传输关键帧，因此为其它两种镜头预留了传输资源，使得重要事件中对用户体验影响最大的长镜头和中镜头传输了更多的帧. 图 4 表明，CBFD 算法根据不同内容特征调节不同粒度内容的丢帧比率，延长了进入拥塞避免的时间，使得拥塞窗口变化平缓，保持了较大的吞吐量.

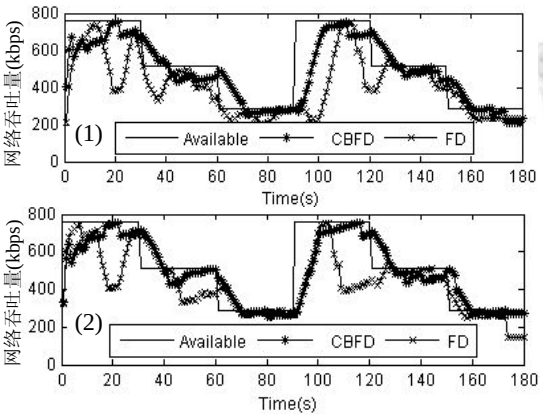

图 4　网络吞吐量比较

4.2 语义质量分析

基于内容的自适应视频传输系统中，语义质量显然是一个重要的评判依据，语义质量通常被定义成语义单元数据量的正相关函数[13]，语义单元中帧率越高，语义质量也越大. 在本文实验中，不同类型镜头被丢帧后，尽管原始帧率一样，但播放帧率将变得不一样，因此本文提出使用镜头帧率来比较语义质量大小，避免直接计算语义质量，如式(4)所示.

$$\text{镜头帧率}=\frac{\sum \text{镜头内被发送帧数}}{\text{镜头持续时间}} \tag{4}$$

图 5(1)-(2)中，镜头帧率因为丢帧操作导致不同镜头帧率出现较大差异. 长镜头和中镜头比其它类型镜头有较高的帧率，从而运动平滑性更好，娱乐性损伤幅度较小. 在特写镜头中，镜头帧率较低，由于主要用于显示信息内容，且观众仍然能观察到特写球员的表情变化和动作变化，信息内容损伤幅度较小. 因此，针对不同类型镜头设定丢帧比率能有效改善语义质量.

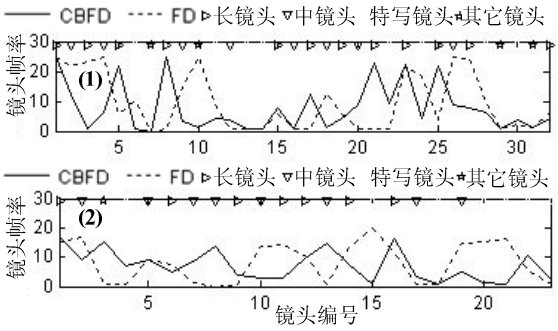

图 5　网络吞吐量比较

4.3 用户体验分析

用户体验主要通过主观实验分析用户对回放视频的 MOS 评分[14]，得分值越高说明用户对回放视频的传输效果越满意. 我们使用 DSIS (Double Stimulus Impairment Scale)法设计评价实验，15 个邀请者分别观看原始视频、基于帧类型丢帧(FD)方法得到的自适应视频，CBFD 方法得到的自适应视频，然后从清晰性(Crispness)、平滑性(Motion-smoothness)、内容可视性(Content visibility)[15]等三个方面给出分值. 其中清晰性指画面是否清晰无失真；平滑性指画面播放是否平滑连续；内容可视性指内容信息是否容易被观察获取. 分值设计为 5 分制: 5-非常满意、4-总体满意、3-基本满意，能接受、2-不满意、1-完全不满意.

表 3　用户体验评测结果

评价指标	清晰性	平滑性	内容可视性	平均值
CBFD	4.2	3.9	4.2	4.1
FD	2.7	2.4	2.8	2.63

最后对所有打分结果计算平均值，从表 3 的结果可以看出, CBFD 比 FD 能取得令用户更满意的主观质量.

5　结语

本文设计的基于内容实时自适应传输系统立足于用户偏好内容的重要性分级，制定多粒度 CBFD 传输策略解决了动态带宽约束条件下的实时自适应视频传输问题，在精彩事件中区分镜头类型重要性，通过特写镜头和观众镜头的丢弃使得精彩事件内娱乐性更强的长镜头和中镜头有了更多的传输资源而降低了丢帧

操作对整个事件的语义质量损伤, 比基于帧类型丢帧的自适应传输方法更好地优化了重要事件传输时的丢帧比率, 智能地保证了用户需求得到最大满足.

参考文献

1 Chang SF, Vetro A. Video adaptation: Concepts, technologies and open issues. Proceedings of the IEEE, 2005,93(1):148−158.

2 Xu M, He XJ, Peng Y, et al. Content on demand video adaptation based on MPEG-21 digital item adaptation. EURASIP J. Wireless Comm. and Networking, 2012. 104−119.

3 梁永生,陈旭,柳伟,等.基于内容感知的视频流媒体渐进式流传输方式.计算机工程,2010,36(24):220−222.

4 李晓城,钱松荣.一种自适应的 3G 网络流媒体速率控制算法.小型微型计算机系统,2012,33(7):1429−1432.

5 于俊清,刘冲,等.利用运动强度自适应传输视频内容.计算机辅助设计与图形学学报,2009,21(6): 847−852.

6 Khan A, Sun LF, Ifeachor EC. QoE Prediction Model and its Application in Video Quality Adaptation Over UMTS Networks. IEEE Trans. on Multimedia, 2012, 14(2): 431−442.

7 于俊清,何欢欢,何云峰.利用情感激励提取足球视频精彩镜头.计算机研究与发展,2010,47(10):1823−1831.

8 D'Orazio T, Leo M. A review of vision-based systems for soccer video analysis. Pattern Recognition, 2010, 43(8): 1−16.

9 童晓峰,刘青山,卢汉清.体育视频分析.计算机学报,2008, 31(7):1242−1251.

10 Ghinea G, Chen SY. Measuring quality of perception in distributed multimedia. Computers in Human Behavior, 2008, 24: 1317−1329.

11 Apple Inc. http://dss.macosforge.org/. [2013-07-18].

12 NIST Internetworking Technology Group (ITG). http://snad.ncsl.nist.gov/nistnet/index.html. [2013-06-21].

13 Wei Y, Bhandarkar SM, Li K. Client-centered multimedia content adaptation. Transactions on Multimedia Computing, Communications, and Applications, 2009, 5(3): 22−29.

14 ITU-T Recommendation P.910. Subjective Video Quality Assessment Methods for Multimedia Applications. 1999, 9

15 Önür ÖD, Alatan AA. Video adaptation for transmission channel by utility modeling. IEEE ICME 2005, Amsterdam. IEEE. 2005. 1400−1403.