

基于可见损伤持续时间贝叶斯网络的视频 QoE 评估方法

耿杨, 邵苏杰, 郭少勇, 喻鹏

(北京邮电大学网络与交换技术国家重点实验室, 北京 100876)

摘要: 目前, 对视频业务体验质量 (QoE) 的客观评估方法中, 视频图像质量下降造成的可见显示损伤持续时间对用户体验质量的影响往往被忽略。针对该问题, 讨论了可见显示损伤持续时间对用户体验质量的影响。同时, 分析了造成可见显示持续损伤的客观原因, 构建了包含分层客观指标和主观评估结果的贝叶斯网络。在此基础上, 提出了针对视频业务体验质量的无参考评估方法, 并通过主观测试验证了该方法的准确性。

关键词: 视频体验质量; 贝叶斯网络; 显示持续损伤

中图分类号: TP393

文献标识码: A

Bayesian network-based video QoE assessment method using image sustained damage analysis

GENG Yang, SHAO Su-jie, GUO Shao-yong, YU Peng

(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: At present, the influence of the image sustained damage is often neglected in the objective method of the quality of user experience in video service. In view of this problem, the effect of the visible duration of visual impairment on the quality of the user experience was discussed. At the same time, the objective reason of continuous damage was analyzed, and a Bayesian network was constructed which included hierarchical objective index and subjective evaluation result. On this basis, a no-reference evaluation method for the quality of video service experience was proposed, and the accuracy of the method is verified by subjective test.

Key words: video QoE, Bayesian network, image sustained damage

1 引言

随着网络视频业务的逐渐流行, 2017 年, 视频业务流量将占整个网络资源的 74%。相对于传统的以客观质量指标为基础的 QoS, 以用户体验质量为中心的 QoE (quality of experience) 能够更加直接和准确地反映业务质量的实际情况。QoE 通过主观测试来评估业务的实际体验质量, 在这个过程中需要耗费很大的成本。因此, 通过视频业务的客观质量进行 QoE 的评估和预测成为了现阶段研究的焦点。

现阶段已存在的 QoE 客观评估方法是以视频图像的损伤程度作为评估指标, 如 PSNR^[1]和

SSIM^[2], 其评估过程基于原始视频序列和对应的用户终端接收到的图像之间的比较结果。在评估过程中, 视频图像损伤出现的持续时间被忽视了。事实上, 当一个 24 frame/s 的视频流在播放过程中, 孤立的一帧图像出现损伤时, 用户是感知不到的, 这是由于人类视觉延迟^[3]和标志性记忆^[4]导致的。但是当一段视频图像中持续性地出现显示损伤, 即使这个损伤是很小的, 也依然能够影响用户体验质量。

此外, 通过目前已有的视频 QoE 评估方法获取的确定性评估结果存在 2 个问题: 1) 在视频业务进行的过程中, 客观质量指标的变化对业务体验质量的影响是具有随机性的; 2) 不同用户对体验质量的

收稿日期: 2017-03-29; 修回日期: 2017-05-18

基金项目: 国家科技支撑计划基金资助项目 (No.2015BAG10B00)

Foundation Item: The National Key Technology R&D Program (No.2015BAG10B00)

感知也是呈概率分布的。因此，在本文提出的评估方法中，使用了贝叶斯网络来构建视频 QoE 评估方法，反映了客观质量指标对体验质量影响的随机性。同时，评估结果为用户评分的概率密度分布，反映了真实的用户体验质量情况。

2 相关工作

现阶段，根据在评估过程中对原始视频信息的参考程度，可以将 QoE 客观评估方法分为 3 类：全参考评估 (FR, full-reference) 方法、半参考评估 (RR, reduced-reference) 方法和无参考评估 (NR, no-reference) 方法，它们使用不同的客观质量指标对视频 QoE 进行测量和评分。

FR 评估方法通过比较原始视频序列和用户终端接收到的视频图像信息对 QoE 进行评估。在评估过程中，对完整的图像数据进行比较，检测用户观看的图像出现的质量损伤，根据损伤程度对 QoE 进行评价^[2,3]。RR 在评估过程中，同样需要原始视频序列作为参考，但是在比较的过程中，使用的是基于图像提取的特征信息，而不是完整的图像信息^[5]。在使用 FR 和 RR 评估方法进行评估的过程中，需要将用户端接收到的视频数据和作为参考的原始视频数据通过网络传输到评估服务器，才能进行比较评估，因此，额外消耗了网络资源，同时，无法满足实时性的需求。在 NR 评估方法中，整个评估过程不需要任何用于比较的参考信息，利用能够通过客观测量的业务质量指标直接对视频 QoE 进行评估。NR 评估方法能够满足实时性的需求，因此受到了广泛的关注。

现阶段在针对 NR 的一些研究中，通过传统 QoS 的网络层客观质量指标，建立其与主观体验评分 MOS (mean of score)^[6]之间的映射关系^[7,8]。此外，在对视频流本身属性的研究中，文献[9]研究了图像组 (GOP, group of picture) 结构对视频 QoE 的影响。在文献[10]的研究中建立了从网络层性能指标到视频播放质量的映射关系。文献[11]利用 PSNR 评估结果和网络 QoS 指标建立了用于视频 QoE 评估的神经网络。

上述对 NR 评估方法的研究中，能够实现对视频 QoE 的实时评估，但是通过这些评估方法获得的确定性评估结果不能真实地反映用户对视频体验质量的真实情况。因此，本文讨论了可见显示损伤持续时间与视频 QoE 之间的关系，并建立了能够用于视频 QoE

评估的贝叶斯网络，其评估过程体现了客观质量指标对视频 QoE 影响的随机性，评估结果反映了整体用户对业务质量接受程度的分布情况。

3 视频 QoE 评估贝叶斯网络

本文建立了一个贝叶斯网络用于评估视频流用户体验质量 (QoE)。该贝叶斯网络由以下 5 层组成：媒体层、网络层、帧层、终端层和用户层，体现了视频流服务端到端的特性。

媒体层反映了视频内容属性和视频流数据格式。网络层是表示视频流数据在网络传输过程中的网络性能指标。帧层反映了用户终端接收到的视频帧数据质量情况。终端层包含了视频流显示过程中的图像质量。用户层反映了视频业务的用户体验质量，视频 QoE 贝叶斯网络如图 1 所示。

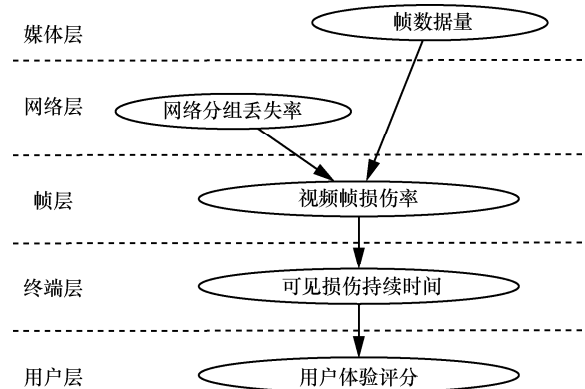


图 1 视频 QoE 贝叶斯网络

3.1 媒体层

在本文的贝叶斯网络中，媒体层节点反映了视频帧数据量的大小。定义 $S(f)$ 为一个视频帧数据在 UDP 网络中完整的传输所需要的网络数据分组数量，其中， f 表示视频帧类型。在本文中， $S(f)$ 数值未直接反映视频帧数据量。通过需要负载的网络数据分组的数量，能够反映网络传输错误对视频帧数据完整性的影响。表 1 为本文中使用的样本视频流 Coastguard 在网络中传输时，3 种类型的视频帧需要使用的网络数据分组的平均值。

表 1 3 种类型的视频帧需要使用的网络数据分组平均值

视频帧类型	负载网络数据分组的平均值/个
I 帧	36.52
P 帧	26.47
B 帧	16.69

3.2 网络层

本文使用网络分组丢失率 (PLR, packet loss ratio) 作为网络层元素。PLR 直接反映了 UDP 网络的传输性能。在视频流数据经由 UDP 网络传输时, 无重传机制来解决数据分组的丢失问题, 因此, 当 UDP 网络数据分组在传输过程中出现分组丢失情况时, 会对用户终端接收到的视频帧数据完整性产生影响。在传输中, 每个 UDP 网络数据分组发生分组丢失事件是具有相同概率的独立事件。一个 UDP 网络的 PLR 可以通过统计方法获取。利用终端接收到的网络数据分组数量和服务器端发送的网络数据分组总数量, PLR 可以表示为

$$PLR = 1 - \frac{\text{接收到的网络数据分组数量}}{\text{网络数据分组总数量}} \quad (1)$$

3.3 帧层

视频流帧数据通过网络数据分组负载并在 UDP 网络中传输。在网络数据分组出现丢失的情况下, 用户终端接收到的视频帧数据会出现损伤。视频帧数据通过 $S(f)$ 个网络数据分组负载经过 UDP 网络传输时, 用户终端接收到完整的视频帧数据等同于一个发生概率为 $1-PLR$ 的 $S(f)$ 重伯努利实验。因此, 定义 $D(f)$ 为经过网络传输后, 接收到的视频帧数据出现损伤的概率。

$$D(f) = 1 - (1 - PLR)^{S(f)}, f \in (I, P, B) \quad (2)$$

3.4 终端层

终端层元素反映在视频播放过程中图像显示状态。在视频播放过程中, 受损帧的帧类型不同, 影响到出现显示损伤的帧的数量不同。在 H.264 标准中, 通过一组帧之间的参考关系实现了帧间相关性。通过帧间相关性, 能够降低视频流数据量, 节省在网络传输过程中的资源消耗; 同时, 帧间相关性能够降低网络传输错误对视频播放质量产生的影响。H.264 格式的视频流帧有 3 种格式: I 帧 (intra coded frame)、P 帧 (predicted frame) 和 B 帧 (bi-directional predicted frame)。在一个以开放图像组 (open-GOP) 形式组成的视频流中, 使用变量 $GOP(N, M)$ 描述其结构, 其中, N 为整个 GOP 中包含的视频帧数量, M 表示在相邻的 2 个参考帧 (I 帧或 P 帧) 之间 B 帧的数量。定义变量 V 为视频播放过程中出现显示损伤的帧的数量, 当一个受损的视频帧出现时, 对 V 产生的影响分析如下。

1) I 帧: I 帧在编码过程中包含了完整的图像信

息, 在用户终端对 I 帧解码的过程中, 不需要其他参考帧。在视频流数据中, 通常每间隔 0.5 s 或 1 s 就会出现一个 I 帧。由于包含了完整的图像信息, I 帧通常具有更大的数据量。因此, 频繁地出现 I 帧会影响视频流数据的压缩效率, 同时增大在网络传输中的负担。通常在视频会议等需要实时传输的业务中, 视频流中只包含有少量的 I 帧。I 帧在视频流传输过程中, 通常被设置成随机接入点。在经过网络传输后, 终端解码器将 I 帧作为另外 2 类帧的参考帧。

I 帧作为它所在的整个 GOP 组的参考帧, 同时, 也是它之前的 M 个 B 帧的参考帧, 当出现帧数据损伤时, 造成的可见显示损伤持续时间会出现以下 2 种情况。

当这个 I 帧之前的 M 个 B 帧没有受到其他参考帧的影响时

$$V_1 = N + M \quad (3)$$

在视频播放过程中, 出现这种情况的概率为

$$P(V_1) = D(I) \cdot \overline{D(I)} \cdot \overline{D(P)}^{\binom{N}{M+1}} \quad (4)$$

当这个受损 I 帧之前相邻的 M 个 B 帧已经受到其他参考帧损伤的影响时

$$V'_1 = N \quad (5)$$

这种情况出现的概率为

$$P(V'_1) = D(I) \left(1 - \overline{D(I)} \cdot \overline{D(P)}^{\binom{N}{M+1}} \right) \quad (6)$$

2) P 帧: P 帧在编码过程中包含了图像信息和运动向量。在解码过程中, P 帧需要参考帧才能够正确的解码。P 帧的参考帧是 I 帧或相邻的上一个 P 帧。P 帧包含的数据量是少于 I 帧的。

在一个 $GOP(N, M)$ 中包含 P 帧的数量为 $\frac{N}{M+1} - 1$ 。

当 P 帧出现损伤时, 对可见显示损伤持续时间的影响与它在这组 GOP 中的位置有关。当这个 GOP 中第 i 个 P 帧出现损伤时, 造成的可见显示损伤持续时间为

$$V_p = N - (M + 1)(i - 1), i \in \left[1, \frac{N}{M + 1} - 1 \right] \quad (7)$$

在视频流播放的过程中, 这种情况出现的概率为

$$P(V_p) = D(P) \cdot \overline{D(I)} \cdot \overline{D(P)}^{i-1} \quad (8)$$

3) B 帧: B 帧在解码过程中, 同时需要多个 I

帧或 P 帧作为参考帧，因此，B 帧所包含的数据量是 3 类帧中最少的。在 MPEG-2 中，B 帧的参考帧被规定需要前后相邻的 2 个 I 帧或 P 帧作为参考帧。在 H.264 中，B 帧可以具有任意数量的参考帧。在视频流数据中，B 帧不会作为其他帧的参考帧。因为 H.264 视频编码中的帧间相关性，在经过 UDP 网络传输时，单一参考帧数据出现的损伤能够影响一段时间内多帧显示图像损伤的出现。

在一个视频流播放过程中，受到损伤的 B 帧只能影响它对应的显示图像，不会对其他帧的显示质量产生干扰，因此

$$P(V_B = 1) = D(B) \cdot \overline{D(I)} \cdot \overline{D(P)}^{\binom{N}{M+1}} \quad (9)$$

综上所述，一段时长为 T 的视频流在播放过程中，视频可见显示损伤持续时间为

$$t = \frac{T \sum V \cdot P(V)}{N} \quad (10)$$

3.5 用户层

本文通过主观测试对在视频播放过程中可见显示损伤持续时间对用户质量的影响进行了评估。利用 6 个具有不同内容的原始视频序列生成测试样本视频流。每个原始视频序列都对应一组包含 10 个不同的可见显示损伤持续时间的测试样本视频流，如表 2 所示。在本文研究中，主观数据分析处理有如下 2 个目标：1) 对不同测试样本的主观评估结果进行聚类分析；2) 主观测试结果的统计特征分析。通过概率密度拟合方法获得主观体验质量的分布。

表 2 样本视频流信息

样本名称	可见显示损伤持续时间/s				
Cosatguard	0.08	0.20	1.20	2.80	3.20
	4.40	6.40	6.80	8.00	10.00
Container	0.16	0.24	0.80	2.40	4.00
	4.80	5.60	7.20	8.40	9.60
Foreman	0.08	0.20	1.60	2.00	3.20
	5.20	6.40	7.60	8.40	10.00
Hall	0.12	0.24	1.60	2.80	3.60
	4.40	6.00	6.80	8.80	9.20
Mobile	0.16	0.40	1.20	2.40	4.00
	5.20	5.60	7.20	8.80	9.20
News	0.12	0.40	0.80	2.00	3.60
	4.80	6.00	7.60	8.00	9.60

1) 样本视频聚类分析

通过 Wilcoxon 秩和检验将主观测试结果具有相似概率分布的测试样本视频进行聚类。Wilcoxon 秩和检验是一种非参数假设检验，用于测试 2 组数值是否具有相似的概率分布情况。在本文中，对全部样本视频流的主观测试结果两两进行 Wilcoxon 秩和检验，将主观测试结果具有相似概率分布的样本视频流进行聚类。

经过 Wilcoxon 秩和检验后的样本视频聚类结果如表 3 所示。

表 3 样本视频聚类结果

聚类	可见显示损伤持续时间/s							
C ₁	0.08	0.12	0.16	0.2	0.24	0.4	0.8	1.2
C ₂	1.6	2.0	2.4	2.8	—	—	—	—
C ₃	3.2	3.6	4.0	4.4	4.8	5.2	5.6	6.0
C ₄	6.4	6.8	7.2	7.6	8.0	—	—	—
C ₅	8.4	8.8	9.2	9.6	10.0	—	—	—

经过 Wilcoxon 秩和检测后生成的 5 个聚类 (C_i)，每个聚类所包含的样本视频流的主观测试结果是符合统一概率分布假设的。各个聚类中包含的样本视频流主观测试结果组成的集合的统计信息如图 2 所示。图 2 中，每个聚类分别对应一个统计箱型图，矩形上沿和下沿分别对应数据集中上四分位数和下四分位数，矩形中的线表示集合数据的中位数，虚线上下沿表示根据上下四分位数确定的上下边界超出边界值的统计数据，被视为异常值，用“+”标记。

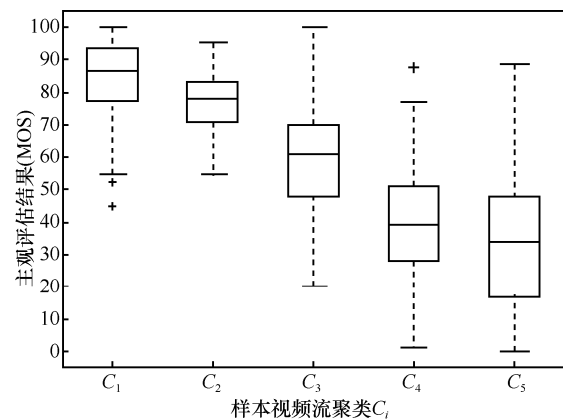


图 2 主观测试结果

2) 概率分布集合

通过正态概率分布拟合获得主观测试结果的

概率密度函数。

$$f(m|C_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(m-\mu)^2}{2\sigma^2}} \quad (11)$$

其中, m 代表用户体验质量评分, $f(m|C_i)$ 为用户整体评分在聚类 C_i 条件下的 m 取值的概率密度分布。通过拟合后的概率密度函数能够实现视频 QoE 的评估和预测。

拟合后的概率密度函数参数如表 4 所示。

表 4 拟合后的概率密度函数参数

聚类	μ	σ
C_1	87.23	14.19
C_2	77.38	10.30
C_3	60.78	17.63
C_4	44.15	17.59
C_5	31.79	21.50

4 QoE 评估方法及验证

在本节中, 提出了基于贝叶斯网络的视频流媒体业务 QoE 实时评估方法, 并对评估结果进行了验证。

4.1 QoE 评估方法

视频流媒体 QoE 能够通过第 3 节中提出的由客观质量指标构成的贝叶斯网络进行评估。评估算法如算法 1 所示。

算法 1 视频 QoE 评估算法

- 1) 分析视频流属性, 计算 $S(f)$;
- 2) 分析视频流帧结构, 初始化参数 N, M ;
- 3) 统计网络分组丢失率 PLR ;
- 4) 计算 I 帧出现损伤时: $V_I, P(V_I), V'_I, P(V'_I)$;
- 5) 计算 P 帧出现损伤时: $V_P, P(V_P)$;
- 6) 计算 B 帧出现损伤时: $P(V_B = 1)$;
- 7) 计算可见显示损伤持续时间 t ;
- 8) 通过 t 将评估视频流对应到聚类中, 得到 MOS 的分布情况。

4.2 验证

本文通过利用 NS2 建立的 UDP 网络仿真环境, 对贝叶斯网络中各节点之间的关系及视频 QoE 评估结果进行了验证。在仿真网络中, 设置了 8 个级别的网络分组丢失率 (0.5%、1.0%、1.5%、2.0%、3.0%、5.0%、7.0%、10%), 传输测试视频流属性如表 5 所示。对仿真后接收到的视频流数据进行分析, 各类型帧实际损伤情况和贝叶斯网络计算结果

比较如图 3 所示。在图 3 中, 曲线 $t(\text{Coastguard})$ 和 $t(\text{Highway2000})$ 为经过评估算法 1 获取的随着网络分组丢失率的升高, 可见显示损伤时间 t 的变化趋势。通过 “○” 和 “▽” 标记了 2 种内容的样本视频流经过仿真网络传输后在不同的网络分组丢失率条件下出现的可见显示损伤时间, 可以发现其结果与经过算法 1 获取的变化趋势基本一致。

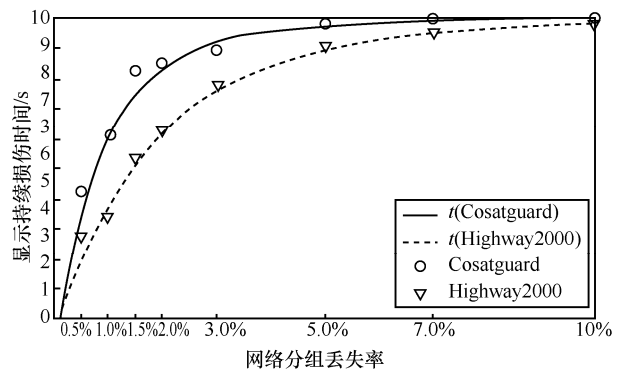


图 3 网络仿真验证结果

对仿真后获取的视频流进行了主观测试。主观测试与通过贝叶斯评估算法结构比较如图 4 所示, 图 4 中显示在经过 30 名测试人员观看后评分结果的分布和经过本文评估方法中各聚类均值之间的比较关系。将具有不同的可见显示损伤时间的样本

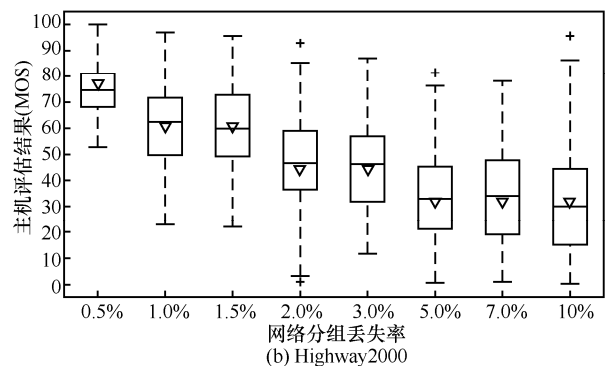
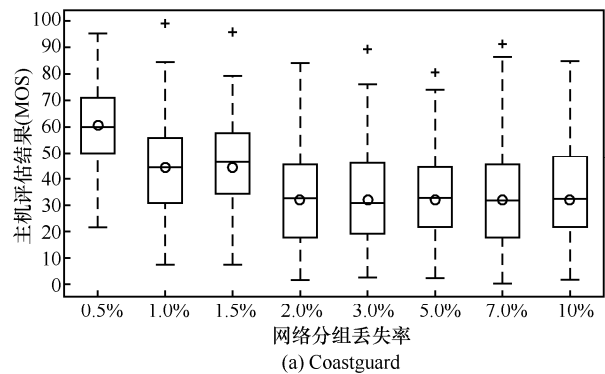


图 4 主观测试结果与评估结果比较

表5 测试样本视频属性

样本名称	帧率/(frame·s ⁻¹)	时长/s	GOP(N, M)结构	S(I)	S(P)	S(B)
Coastguard	24	10	15:2	36.52	26.46	16.69
Highway2000	200	10	12:2	21.26	13.91	12.15

视频流根据表3分类后,在图4中,分别通过“○”和“▽”标记样本所在的聚类对应的 μ 值。由于用户的主观体验感受是因人而异的,在真实情况下,用户对主观体验质量的评价结果是满足正态分布的,变量 μ 代表了正态分布的均值。在图4中,对应聚类的 μ 值与由样本视频流的主观评估结果的组成集合的均值基本一致。

5 结束语

本文分析了视频图像的可见损伤持续时间对视频服务体验质量的影响,建立了可以全面反映业务运行状态的贝叶斯网络。

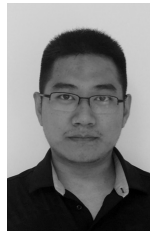
首先分析了分层客观指标之间逐层影响的关系。其次,通过主观测试和分析,得到了可见损伤持续时间和MOS评分之间的关系。通过主观测试的结果,对可见损伤持续时间的影响程度进行聚类。通过统计分析获得的聚类结果得到了生理和心理学的理论支持。由于视觉和标志性记忆的持续性,以及人眼不同的视觉刺激,最小的剩余时间在100~200 ms的范围内。本文的聚类结果中,具有0.08~0.2 s伤害持续时间的视频样本也被聚集成一组。其他群体也受到用户心理能力的影响,这些影响将在未来的工作中进行研究。

另外,本文提出的通过贝叶斯网络对视频业务QoE的NR评估方法。在用户终端可以通过检测网络的传输状态和视频流属性,获取评估过程需要的客观指标,利用基于贝叶斯网络的概率公式来评估MOS评分的概率分布情况。

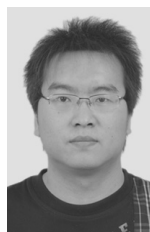
参考文献:

- [1] ITU-T. Methods for subjective determination of transmission quality[S]. T.REC-P.800, 1996.
- [2] HUYNH-THU Q, GHANBARI M. Scope of validity of PSNR in image/video quality assessment[J]. Electronics Letters, 2008, 44(13): 800-801.
- [3] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [4] LONG G M. Iconic memory: a review and critique of the study of short-term visual storage[J]. Psychological Bulletin, 1980, 88(3): 785.
- [5] GOLDSTEIN E B. Cognitive psychology: connecting mind, research and everyday experience[M]. Nelson Education, 2014.
- [6] LI W, LUO Q, YU P, et al. Reduced-reference video QoE assessment method based on image feature information[C]//Network Operation and Management Symposium(APNOMS), IEEE, 2015: 519-522.
- [7] ZHANG F, LIN W, CHEN Z, et al. Additive log-logistic model for networked video quality assessment[J]. IEEE Transactions on Image Processing, 2013, 22(4): 1536-1547.
- [8] ZINNER T, HOHLFELD O, ABBOU D, et al. Impact of frame rate and resolution on objective QoE metrics[C]//Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on. IEEE, 2010: 29-34.
- [9] HUSZÁK Á, IMRE S. Analysing GOP structure and packet loss effects on error propagation in MPEG-4 video streams[C]//Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on. IEEE, 2010: 1-5.
- [10] YU P, LIU F, GENG Y, et al. An objective multi-layer QoE evaluation for TCP video streaming[C]//Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on. IEEE, 2015: 1255-1260.
- [11] GASTALDO P, ROVETTA S, ZUNINO R. Objective quality assessment of MPEG-2 video streams by using CBP neural networks[J]. IEEE Transactions on Neural Networks, 2002, 13(4): 939-947.

作者简介:



耿杨(1987-),男,辽宁盘锦人,北京邮电大学博士生,主要研究方向为视频业务QoE质量管理。



邵苏杰(1985-),男,陕西汉中,北京邮电大学博士后,主要研究方向为网络管理与智能电网。

郭少勇(1985-),男,河北邢台人,北京邮电大学讲师,主要研究方向为网络管理、终端管理与智能电网。

喻鹏(1986-),男,湖北随州人,北京邮电大学副教授,主要研究方向为无线节能管理、5G网络管理与智能电网。