

基于视觉显著计算的视频流媒体渐进式表达方法

梁永生^{1,2}, 柳 伟¹, 周 莺², 魏泽锋², 张基宏²

(1. 深圳信息职业技术学院可视媒体处理与传输深圳市重点实验室, 广东深圳 518029;

2. 深圳大学信息工程学院, 广东深圳 518060)

摘 要: 为了有效解决视频流媒体传输网络带宽、播出视频质量和用户实时性访问之间的矛盾, 本文提出了一种基于视觉显著计算的视频流媒体渐进式表达方法. 在视频内容分析和理解的基础上, 首先进行场景分类和视觉敏感区域提取; 然后根据编码信息确定视频序列中各帧的重要性, 估计帧内片层数据重要性; 最后基于视觉显著计算的结果提出一种适应网络带宽和质量可伸缩的视频流媒体渐进式表达方法. 采用中粒度质量可伸缩(MGS)编码, 在模拟网络测试平台上分别针对集中式和分散式视觉敏感区域视频序列进行实验研究, 实验结果验证了本文提出的基于视觉显著计算的视频流媒体渐进式表达方法的正确性和有效性.

关键词: 视频流媒体; 视觉显著计算; 渐进式表达; MGS 编码; 码流提取

中图分类号: TN911.73 **文献标识码:** A **文章编号:** 0372-2112 (2017)07-1567-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2017.07.004

An Approach to Progressive Description of Video Streaming Based on Visual Saliency Computation

LIANG Yong-sheng^{1,2}, LIU Wei¹, ZHOU Ying², WEI Ze-feng², ZHANG Ji-hong²

(1. Shenzhen Key Lab of Visual Media Processing and Transmission, Shenzhen Institute of Information Technology, Shenzhen, Guangdong 518029, China;

2. School of Information Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China)

Abstract: In order to balance the contradiction among network bandwidth, video quality and subscriber's real-time access, a new approach to progressive description of video streaming based on visual saliency computation is proposed in this paper. On the basis of video content analysis and comprehension, scene classification and Visual Sensitive Region (VSR) extraction are performed firstly. Secondly, frame importance according to coding information and slice data importance are determined. Finally, based on visual saliency computation, a new approach to progressive description of video streaming adapted to network bandwidth and quality scalability is proposed in this paper. Applying MGS Coding, experimental study is performed on video sequence with salient region and cluttered regions in network simulation platform, the experimental results show that the progressive description based on visual saliency computation proposed in this paper is accurate and effective.

Key words: video streaming; visual saliency computation; progressive description; MGS coding; bitstream extraction

1 引言

视觉是人类获取信息的重要手段. 人类通过视觉接受的信息包括文本、图像、音视频、三维动画和数字几何等, 约占获取信息总量的 83%. 互联网用户希望在较低的网络带宽条件下, 能够接收到清晰生动的音视频信息. 传统的下载方式很难满足需求, 因此流媒体应运而生, 其中视频流媒体最具代表性, 已成为近年来的研

究热点. 视频流媒体的数据流随时传送随时播放, 用户不必等到视频文件全部下载即可观看, 缩短了等待时间, 降低了客户端缓存的需求, 能够满足实时传播要求^[1].

目前, 互联网上的信息呈现爆炸性增长趋势, 仅视频流媒体资源就已达海量级, 为了有效地表达视频流媒体内容, 需要对视频流媒体内容进行分析和理解. 在当前网络带宽有限的前提下, 传统的视频流媒体表

收稿日期: 2014-06-19; 修回日期: 2014-09-30; 责任编辑: 郭游

基金项目: 国家自然科学基金(No. 61172165); 广东省自然科学基金(No. S2011010000697); 广东省高等职业院校珠江学者岗位计划(2011); 深圳市基础研究计划-杰出青年基金(No. JC201005280613A); 深圳市基础研究(学科布局)计划(No. 基 20160225)

达方法在非结构化、高维海量数据和语义多样性的处理上存在局限. 新的视频流媒体表达方法要求最大限度地利用网络资源, 有效解决视频流媒体质量与用户实时性访问之间的矛盾, 这就需要融合主观感知, 对视频流媒体内容进行分析 and 理解, 以利于接下来的视频编解码和传输处理^[2].

为了有效解决网络带宽、视频质量和用户实时性访问之间的矛盾, 在视频内容分析和码率分配方面开展了相关研究. 文献[3]提出了一种基于 H. 264/SVC (Scalable Video Coding) 的可伸缩 ROI 算法, 根据视觉向心性将 ROI 安排在画面中间区域, 并通过帧间运动来估计的 ROI 大小, 进一步对不同 ROI 区域大小制定不同 SRL (Scalable ROI Layer) 方案. 文献[4]通过调整预滤波器的参数保护不同 ROI 区域, 以较低的复杂度提高视觉效果. 另一种传输可分级视频的思路是联合信源/信道编码方法^[5], 采用一种新型的缩减格栅算法与拉格朗日优化技术对 SVC 数据域差错控制保护级别进行最优分配; 文献[6]针对人眼不同时间段内进行眼动策略分析, 然后结合显著计算得到图像注视区域结果得到 SVC 增强层码率分配算法; 根据可伸缩码流的结构可以进一步分析 H. 264/SVC 各编码层间的相关性, 将空域-质量域两维度码率分配问题描述为一个两阶段最优化问题进行求解^[7]. 文献[8]提出一种聚簇的运动估计方法, 使 SVC 适合硬件实现. 文献[9]总结了 H. 264 率失真优化和码率控制技术的研究进展, 指出主观失真建模和针对高清视频的码率控制是研究的重点和难点. 相关研究都面临一个共性的问题, 即如何采用渐进式表达方式使视觉体验最大化.

为了解决这个问题, 需要应用数学模型来研究类似于人类视觉处理方式的信息处理方法, 即视觉显著计算, 通过视觉显著计算确定输入视觉信号的重要程度. 按照人类视觉系统处理信息由粗到精, 从低到高的顺序, 视频流媒体的视觉重要性分为帧重要性, 片层重要性和视觉敏感区域重要性. 基于视觉显著计算的渐进式表达方式的关键在于视觉敏感区域提取和编码, 难点在于渐进式表达方式的编解码一致性和效率.

为了以较低的复杂度有效地保护视觉敏感内容, 本文基于视觉显著计算提取视觉敏感区域, 进而提出一种适应网络带宽和质量可伸缩的视频流媒体渐进式表达方法. 采用 MGS 编码, 在模拟测试平台上进行实验研究, 验证本文提出的渐进式表达方法的正确性和有效性.

2 相关工作

2.1 视频场景分类和视觉敏感区域提取

神经计算学的研究表明, 对于同类视频场景而言,

提取视觉敏感区域的方法是相似的, 即同一类场景具有近似的视觉特征和特征通道线性组合权值, 根据场景类别确定视觉敏感区域及其重要性. 在实际应用中, 场景分类器的构造存在已标注训练数据不足的问题, 而未标注的数据却大量存在. 因此, 本文采用线性邻域扩散 (Linear Neighborhood Propagation, LNP) 算法^[10]. 提高分类器的准确性. LNP 算法是一种新颖的基于图的半监督学习方法, 依据多维分类特征搜寻每个未标注数据的已标注近邻样本数据, 只需要少量的标注数据就可以训练场景分类器, 其核函数如式(1)所示.

$$x_i = \sum_{x_j \in N(x_i)} \alpha_j x_j \Rightarrow f_i = \sum_{x_j \in N(x_i)} \alpha_j f_j \quad (1)$$

其中, $N(x_i)$ 表示样本数据 x_i 的邻域, f_i 为数据 x_i 的标注, α_j 为加权系数.

在场景分类的基础上, 采用两层视觉敏感特征动态提取视频的 VSR^[11].

- (1) 底层特征, 如亮度、颜色、纹理等;
- (2) 中层特征, 如目标运动和摄像机的运动.

由于同类场景具有近似的视觉敏感分布和线性组合权值. 计算综合视觉敏感区域分布图的公式如式(2)所示.

$$S = \alpha_{k,1} \times \text{FFM}_1 + \alpha_{k,2} \times \text{FFM}_2 + \dots + \alpha_{k,n} \times \text{FFM}_n \quad (2)$$

其中, S 为综合视觉敏感区域分布图, FFM_i 为第 i 个特征通道的视觉敏感区域分布图 ($1 \leq i \leq n$), n 是视觉敏感特征的个数, $\alpha_{k,i}$ 是第 k 类视频镜头的第 i 个视觉敏感区域分布图的权值.

要确定其中各个视觉敏感区域的中心点和大小仅靠视觉敏感区域分布图还不够, 因此用正则矩的方法计算每个视觉敏感区域的中心点和大小, 如式(3)所示.

$$m_{pq} = \sum_1^M \sum_1^N x^p y^q s(x, y), \quad p, q = 0, 1, 2, \dots,$$

$$\eta_{pq} = \frac{\sum_1^M \sum_1^N (x - \bar{x})^p (y - \bar{y})^q s(x, y)}{m_{00}^{(p+q+2)/2}}, \quad \bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \quad (3)$$

其中, M, N 是视觉敏感区域分布图的大小, $s(x, y)$ 是对应像素 (x, y) 的视觉敏感值函数. (\bar{x}, \bar{y}) 是视觉敏感区域的中心点, $(\bar{x}, \bar{y}) = (m_{10}/m_{00}, m_{01}/m_{00})$, 视觉敏感区域的大小为 $S_o = \sqrt{\eta_{20}} \times \sqrt{\eta_{02}}$. 根据视觉敏感区域的中心点和大小可以在视觉敏感区域分布图中计算出主观重要性因子 $P_o = S_o / (M(N))$, P_o 表示该区域对人眼视觉系统的激励程度.

视频帧中的 VSR 的中心点和大小会随着视频内容的不同而变化. 为简化起见, 将视频帧中的 VSR 分为分散式和集中式两种^[12]. 其中, 分散式有多个 VSR, 按照视觉敏感程度的大小分别记为 VSR1, VSR2, ..., VSRi.

2.2 MGS 编码

视频编码的可分级性分为空间可分级、时间可分级和质量可分级,其中质量可分级主要分为粗粒度质量可伸缩(Coarse Grain Scalability, CGS)、中粒度质量可伸缩(Medium Grain Scalability, MGS)以及细粒度(Fine Grain Scalability, FGS)质量可伸缩^[13]. 由于 MGS 编码能够在显著降低计算复杂度的情况下,提供与 FGS 相似的率失真结果^[14],因此成为目前主要的视频编码技术.

采用 MGS 编码方式,宏块数据以 4×4 块为单位进行整数变换量化后的系数分配可构建多个增强层^[15]. 配置文件中可以设置多个增强层,例如 $(\text{MGSVector0}, \text{MGSVector1}, \text{MGSVector2}) = (3, 3, 10)$, 表示将对应的系数按照相应的分配比例写入不同的增强层. 为了达到编码的复杂度与精细程度的折中,选择不同的 MGS 增强层个数和系数分配方案. 片层的数据以包为单位传输,可以在任意截断点提取 MGS 子码流,从而实现 MGS 质量可伸缩.

3 基于视觉显著计算的渐进式表达方法

视觉显著计算模拟人类视觉系统对输入视觉信号的滤波操作,即保留/增强重要的信号并滤除/抑制次要的信号^[16]. 视觉显著计算的核心任务即为对视频场景中的各个视觉子集进行排序,以表明其重要程度和处理优先级. 因此,基于视觉显著计算的渐进式表达方法按照“帧-片层-VSR”的顺序逐步分析数据的重要性.

3.1 帧的重要性

根据帧的编码信息如帧类型、镜头代表帧、场景切换帧确定视频流中各帧的重要性. 帧类型 k 属于客观参

数,镜头代表帧 w 与场景切换帧 p 属于主观参数. $H(\cdot)$ 是帧级重要性函数,各参数分量的 *Minkowski* 之和,如式(4)所示:

$$I = H(k, w, p) = \sqrt{\frac{1}{3}(k^2 + w^2 + p^2)} \quad (4)$$

其中,帧类型的重要性主要体现在其被参考的次数. 具体计算方法如式(5)所示:

$$k = \sum_m \frac{1}{2^m} p_m n_m \quad (5)$$

$$D = \sum_m n_m$$

其中 m 为时间级, $1/2^m$ 为参考系数, n_m 为 GOP 中时间级 m 对应的帧数, D 为 GOP 大小, p_m 为时间级 m 的加权系数. 镜头代表帧 w 计算方法如式(6)所示:

$$w = \frac{2}{S(P, O) + S(O, P)} \quad (6)$$

其中, $S(P, O)$ 和 $S(O, P)$ 是根据包含度理论计算的视频帧 P 和视频帧 O 之间的关联度. 两者相差越小且越接近 1, 即为镜头代表帧可能性越小. 场景切换帧 p 非 0 即 1.

3.2 MGS 片层数据重要性

3.2.1 片层数据自重要性

在 MGS 编码中,每个子块数据单元中的 16 个系数是 4×4 子块数据通过整数变换后得到的,因此各系数的重要性不相同^[17]. 根据片层所包括的系数个数,定义片层自重要性权重矢量 $K = [k_0, k_1, \dots, k_i]$, $0 \leq i < 16$. 各片层数据自重要性分量 k_0, k_1, \dots, k_i 计算过程如图 1 所示.

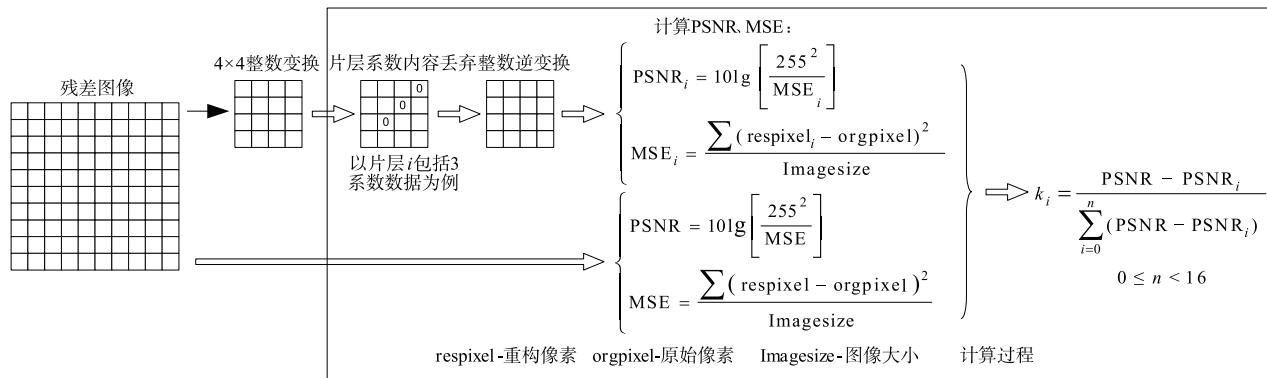


图1 各片层数据自重要性分量计算

通过图 1 可计算出片层自重要性权重矢量 K 中各片层重要性分量的大小. 其中 PSNR 为片层数据未丢失的重构图像质量表征, PSNR_i 为丢弃了片层 i 数据后重构图像质量表征.

3.2.2 片层数据相关重要性

除考虑各片层的自重要性系数以外,还要考虑片

层数据的相关重要性. 定义片层参考度矢量 $C = (c_0, c_1, \dots, c_{N-1})$, 如图 2 所示, 设 $F = \{f_0, f_1, \dots, f_{N-1}\}$ 为由 N 帧所构成的编码图像组, 当 N 取 4 时, f_2 帧图像由于通过双向预测所得因此被认为是 f_0 和 f_4 帧图像的子帧. 类似的, f_1 和 f_3 分别是 f_0 和 f_2 以及 f_2 和 f_4 的子帧. 由于帧间预测作用, 父帧数据丢失会造成对子帧的失真影

响. 片层参考度矢量 $\mathbf{C} = (c_0, c_1, \dots, c_{N-1})$ 中各分量 c_j 的计算如式(7)所示.

$$c_j = \frac{\omega}{T+1} \left[1 + \sum_{i=1}^{N-1} (f_i^F + f_i^B) \right] \quad (7)$$

其中 $0 \leq j \leq N-1$, f_i^F, f_i^B 为前后预测参考权重, 取 $0 \leq f_i^F, f_i^B \leq 1$, ω 为空间层因子, 由于数据上采样原因使得低空间层数据丢失将会对高空间层数据造成影响, 因此 $0 < \omega < 1$, 第一空间层数据 ω 取 1. T 为帧时间层等级. 在图 2 中, 当双向预测参考权重相同时可视 $f_i^F = f_i^B = 0.5$. 可求得 $0 \leq f_i^F, f_i^B \leq 1$. 图像组第 j 帧空间层 n 中堆栈 m 第 s 片层 I_j^s 重要度如式(8)所示:

$$I_j^s = k_s c_j = \frac{\omega_n k_s}{T+1} \left[1 + \sum_{i=1}^{N-1} (f_i^F + f_i^B) \right] \quad (8)$$

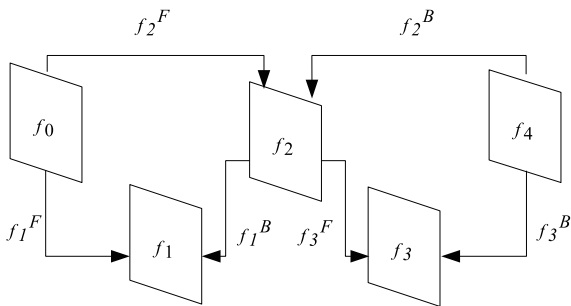


图2 图像组MGS编码示意图

3.3 渐进式表达方法

渐进式表达方法的优势在于充分利用网络带宽, 解决视频质量和用户实时性请求之间的矛盾. 在实际视频编码和传输过程中, 首先传送、接收并显示基本的和重要性高的数据, 然后在进一步传送、接收并显示更精细的和重要性低的数据.

根据帧的重要性确定最优码率分配, 依据帧中各片层的重要性排序, 确定片层重要数据(编码重要数据的范围可以通过带宽及实际需求进行具体确定). 视频的渐进式表达方法如图 3 所示. 在增强层中首先划分视觉敏感区域(VSR)数据和非视觉敏感区域(N-VSR)数据, 然后按照 MGS 片层重要性计算式(8)计算 MGS 各片层数据对视频序列质量的整体重要性, 将对应的数据分别填入相应区域的 MGS 片层数据. 针对前文所述的集中式 VSR 和分散式 VSR, 采用不同的渐进式表达方法. 在增强层各个片层的数据分配上, 如果是集中式的 VSR, 将 VSR 的数据直接进入重要性最高的片层; 如果是分散式的 VSR, 将各个 VSR 的编码数据按照重要性高低依次进入不同重要性的片层(图中用不同的灰度表示不同的重要性). 在网络带宽有限的条件下, 按照基于视觉显著计算的渐进式表达方法, 确定数据的丢弃排序策略, 能够使视频码流在当前码率下获得良好的播出质量.

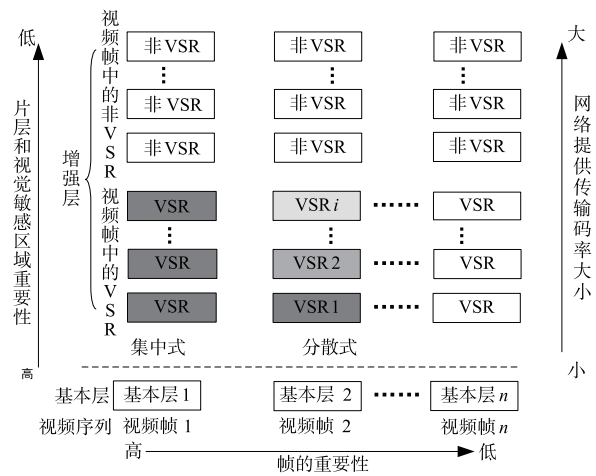


图3 渐进式表达方法示意图

由 MGS 编解码结构可知, 提取了中间增强层数据后, 由于层间预测的原因将会导致中间增强层以上的增强层数据无法利用. 为了保持层间预测的一致性, 可采用文献[18]中通过改变 MGS 层间依赖性结构的方法来解决.

4 实验研究与结果分析

本文采用 ITU 的视频质量专家组(VQEG)提供的标准测试序列作为数据源, 在 ITU-T 标准提案 ITU-T VCEG P206d0 中定义了通用的 SVC 测试环境^[19], 在模拟测试平台上进行实验研究, 实验平台及流程如图 4 所示. H.264 码流的编码和结构分析采用 Tektronix 公司的压缩视频测试系统 MTS4EA, 可以实时对图像数据进行逐帧逐块分析, 给出全面的语义跟踪文件, 确定编码器的逐块判决, 并生成统计分析数据.

视频质量评价可以归结为视频每帧图像质量的评价, 将源视频图像作为参考来度量解码后视频图像与失真程度. 目前常用的客观评价方法主要有峰值信噪比(Peak Signal to Noise Ratio, PSNR)和均方误差(Mean Square Error, MSE)等^[20-22]. 在客观评价指标方面, 本文采用 PSNR 和注意力加权的峰值信噪比(Attention Weighted PSNR, AW-PSNR)评价视频质量; 在主观评价指标方面, 利用泰克公司的视频质量分析仪 PQA600 测试视频质量. 采用的评价参数包括差异平均主观评分(DMOS)和注意加权的差异平均主观评分(ADMOS), DMOS 和 ADMOS 值越低越好.

基于前述的视觉显著计算, 本文为验证渐进式表达方法的一致性和有效性, 设计了 3 组实验: 实验 1 给出 MGS 片层重要性的客观数据, 实验 2 采用码流提取的方式验证渐进式表达的有效性, 实验 3 通过模拟信道验证渐进式表达提升了视觉感受.

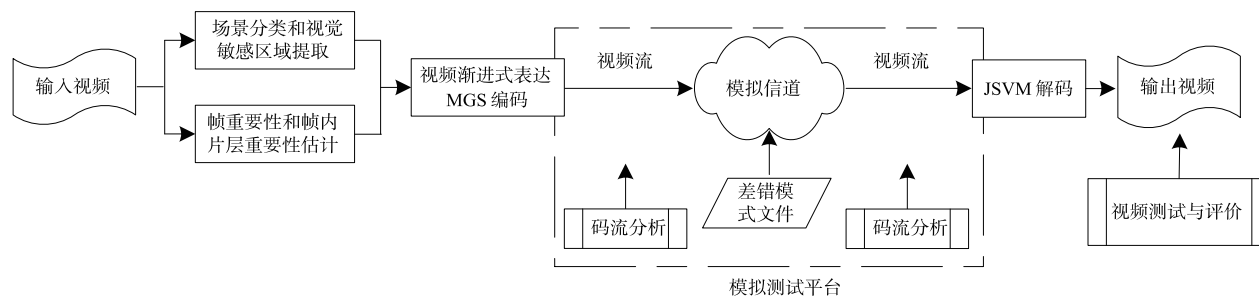


图4 基于视觉显著计算的渐进式表达方法的实验研究平台

实验 1 数据重要性渐进排序测试

实验条件:采用 Foreman 序列,序列采用 QCIF 视频格式,帧率为 15f/s,图像组为 8, MGS 堆栈包括 5 个片层,系数矢量为[0 1 3 8 4]. 基本层量化参数及质量增强层量化参数分别为 42 和 32,输出 32 帧编码数据流. 由于增强层运动矢量信息存放在 MGS 堆栈的首片层中,本文不考虑增强层运动矢量因素对视频流的影响,因此设置首片层(第 0 层)系数为 0.

实验内容:对编码数据图像组中各片层进行丢弃解码影响测试.

表 1 是 Foreman 序列图像组中各帧各片层丢失对视频流质量影响程度的测试统计表,通过实际的测试统计可知各片层对视频质量的影响反映了各片层的重要性不同. 利用本文提出的 MGS 片层数据重要性排序计算方法得到的渐进式排序如图 5 所示,两者进行对比验证了 MGS 编码数据重要性渐进排列计算的准确性.

表 1 Foreman 序列 MGS 各片层影响测试表

视频流 PSNR 均值(dB)	丢弃第 1 片层	丢弃第 2 片层	丢弃第 3 片层	丢弃第 4 片层
第 1 帧	34.3769	34.1053	33.9756	36.2764
第 2 帧	36.3186	36.3409	36.3554	36.3597
第 3 帧	36.2	36.2422	36.3165	36.36
第 4 帧	36.3165	36.3396	36.3513	36.36
第 5 帧	35.8865	35.94	36.1794	36.3587
第 6 帧	36.3281	36.3467	36.355	36.36
第 7 帧	36.2206	36.251	36.3069	36.36
第 8 帧	36.3327	36.3382	36.3492	36.36

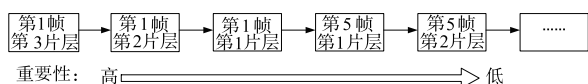


图5 Foreman序列数据渐进式排序

实验 2 渐进式表达方法码流提取测试

实验条件:选取 Foreman 序列和 Bus 序列分别代表集中式 VSR 序列和分散式 VSR 序列. 每个序列采用 CIF 视频格式,帧率为 30f/s,图像组为 8, MGS 堆栈包括 4 个片层,系数矢量为[2 5 6 3]. 基本层量化参数及质

量增强层量化参数分别为 40 和 25,输出 32 帧编码数据流. 性能测试结果如图 6 所示.

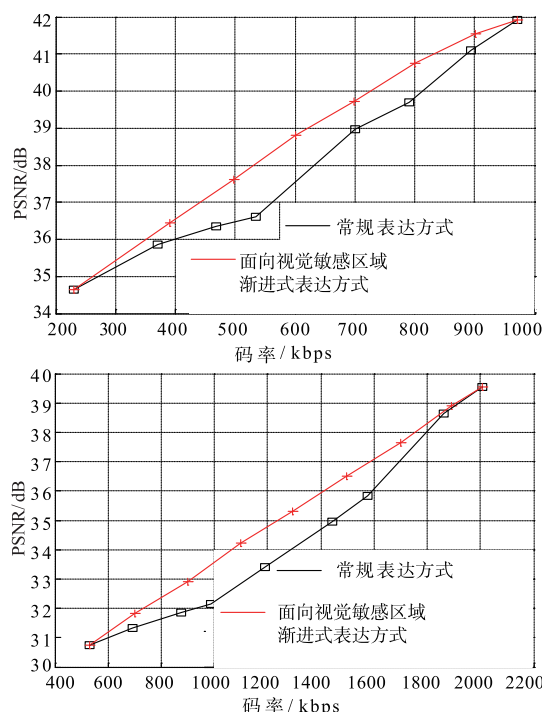
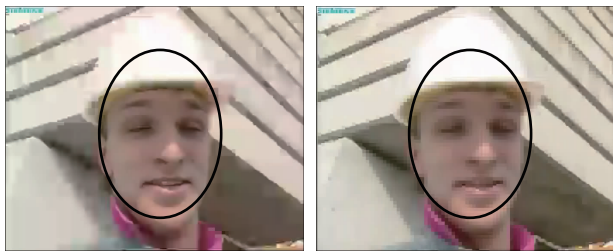


图6 视频流媒体不同表达方式性能测试

在 Foreman 序列中,当基本层量化参数及质量增强层量化参数为 50 和 25,码率限制为 500kbps,可以发现对视觉敏感区域(人脸)的保护效果比较明显,如图 7 所示. 在 Bus 序列中,当基本层量化参数及质量增强层量化参数为 50 和 25,码率限制为 1000kbps,可以发现提高了 3 个视觉敏感区域(绿化带、台标和站牌)的视觉清晰度,如图 8 所示.

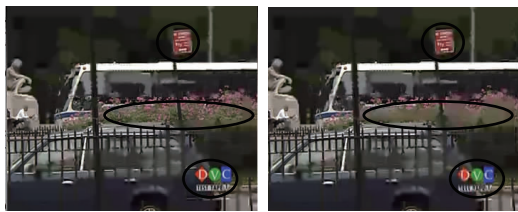
实验 3 通过模拟信道接收端性能测试

在码流数据通过模拟信道发生丢包过程中,对包含视觉敏感区域编码数据的最重要 MGS 片层在可用带宽资源范围内优先传输. 在测试过程中,进行了基本层数据和增强层数据分离,分别通过低差错率和高差错率信道进行模拟. 由于基本层数据包的丢失或误码会对视频解码带来严重影响,本文在不讨论差错隐藏、恢



(a) 渐进式表达方式 (b) 常规表达方式

图7 Foreman序列渐进式表达方式码流提取测试



(a) 渐进式表达方式 (b) 常规表达方式

图8 Bus 序列渐进式表达方式码流提取测试

复的前提下,主要针对增强层进行测试.

实验条件:选取 Foreman 序列代表集中式 VSR 序列,采用 CIF 视频格式,帧率为 30f/s,图像组为 8, MGS

堆栈包括 3 个片层,系数矢量为 [1 6 9]. 基本层量化参数及质量增强层量化参数分别为 12 和 60. 通过模拟信道时使其每一 GOP 中的 I 帧从尾部开始丢弃四个数据包. 原始码流从最后一个增强层开始丢弃,其中包括 VSR 区域的数据包;而通过渐进式表达方式排序的码流则从每层非 VSR 区域的数据包开始丢弃. 在带宽资源有限时,VSR 区域的数据包优先传输,保护了对视频播出质量贡献大的数据包,提高了人眼主观感受. 选取 Soccer 序列代表分散式 VSR 序列,编码参数同 Foreman 序列.

实验结果如图 9 ~ 图 12 所示. 可以看出对于集中式 VSR 序列,在客观评价指标 PSNR 上没有改进,平均值非常相近,这是因为保护的数据虽然对主观视觉感知贡献很大,却不能提升客观评价指标. 渐进式表达在客观评价指标 AWPSNR 上的改进比较明显,Foreman 序列平均值增加了 2.3dB, Soccer 序列平均值增加了 3.9dB,这是因为 AWPSNR 在视觉敏感区域上分配了较高的权重. 对于集中式 VSR 和分散式 VSR 序列,渐进式表达在主观评价指标 DMOS 和 ADMOS 上的改进比较明显,平均降低了约 11%,说明渐进式表达能够在有限资源条件下提高主观视觉感受. 图 13 和图 14 是 Foreman 序列和 Soccer 序列采用不同的表达方式的主

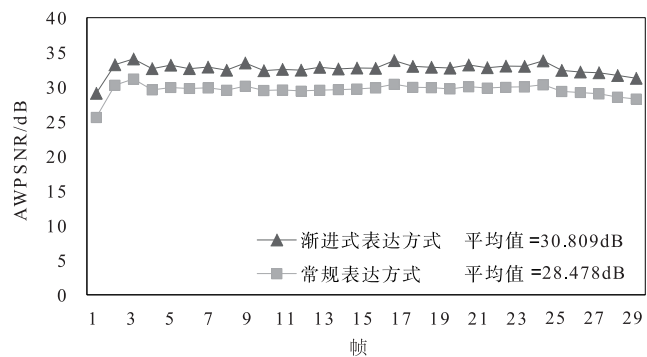
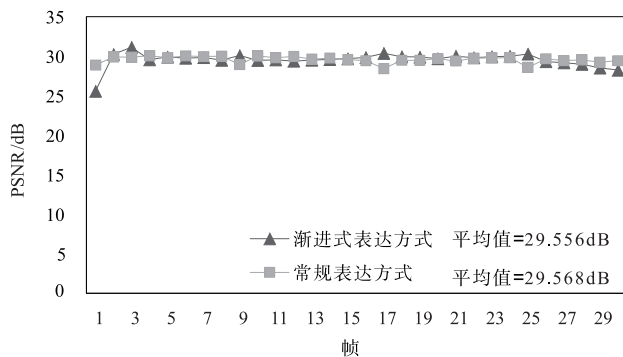


图9 Foreman序列采用客观评价指标实验结果比较

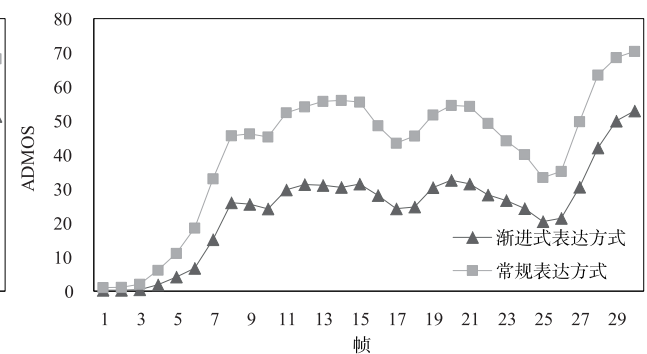
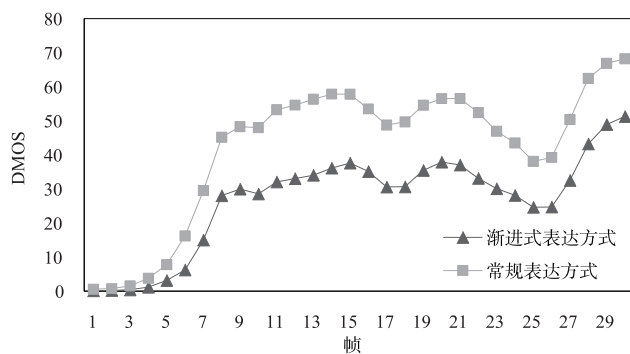


图10 Foreman序列采用主观评价指标实验结果比较

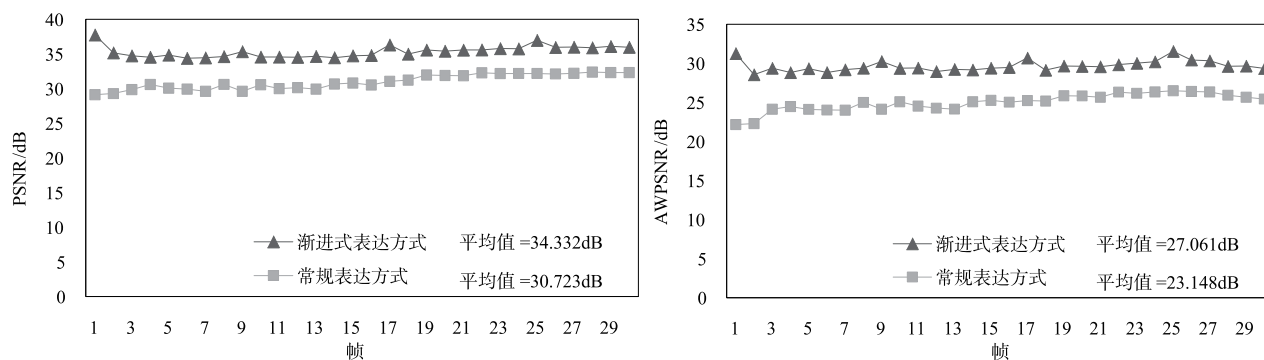


图11 Soccer序列采用客观评价指标实验结果比较

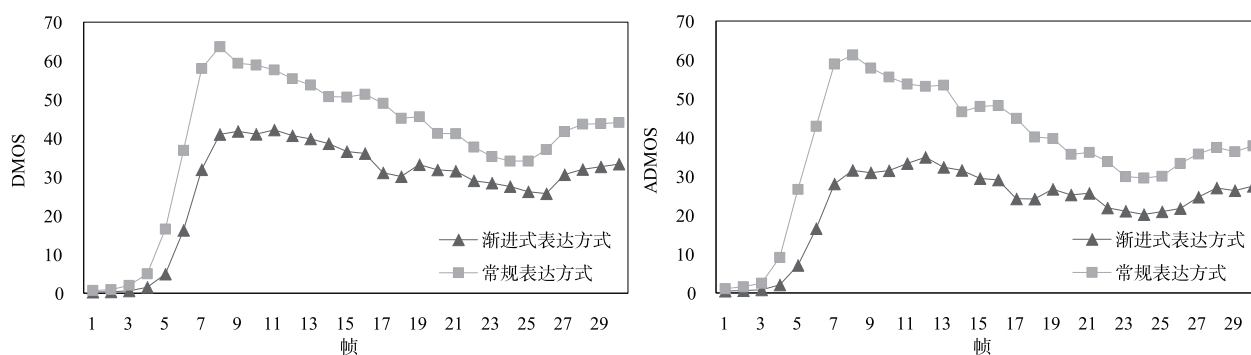


图12 Soccer序列采用主观评价指标实验结果比较



(a) 渐进式表达方式

(b) 常规表达方式

图13 Forman序列渐进式表达方法接收端性能测试



(a) 渐进式表达方式

(b) 常规表达方式

图14 Soccer序列渐进式表达方法接收端性能测试

观视觉对比. 实验结果与人眼的视觉习惯一致: 在视觉敏感区域集中的情况下, 人眼关注的内容集中于视觉敏感区域; 在视觉敏感区域分散的情况下, 人眼更注重整体的视觉感受. 由于视频评价指标侧重点不一样, 指

标上的增益会随着视频内容的变化而有所不同.

5 总结

本文提出了一种基于视觉显著计算的视频流媒体渐进式表达方法, 能够有效解决网络带宽、视频质量和用户实时性访问之间的矛盾. 在视频内容分析的基础上提取视觉敏感区域; 然后根据帧的重要度和帧内片层数据重要性, 结合集中式和分散式视觉敏感区域提出一种适应网络带宽和质量可伸缩的视频流媒体渐进式表达方法. 采用 MGS 编码, 在模拟测试平台上分别进行数据重要性渐进排序测试、渐进式表达方法码流提取测试和通过模拟信道接收端性能测试. 实验结果验证了本文提出的渐进式表达方法的正确性和有效性, 具有极强的应用价值.

下一步工作将研究用简洁有效的数学模型对视觉注意机制中的融合机理进行建模, 实现低复杂度高精度的视觉显著计算. 同时研究针对下一代高性能视频编码 (High Efficiency Video Coding, HEVC), 即 H. 265 的渐进式表达方法. 借鉴提案中的视觉注意机制方面的相关内容, 例如 AHG19 (Verification Test Preparation and Testing, JCTVC-P1011) 中的主观测试验证部分, 以低复杂度提高 HEVC 码率控制精度, 在异构网络不同的目标码率条件下, 结合视觉显著计算和码率分配确

定优化方案,使得在目标码率下视频序列质量最大化。

参考文献

- [1] Abdelhamid Nafaa, Manon Gaucher. Implementation and analysis of a peer-to-peer retransmissions system for live video services [J]. *IEEE Multimedia*, 2011, 18 (2) : 60 - 71.
- [2] Yan Li, Athina Markopoulou, John Apostolopoulos, Nicholas Bambos. Content-aware playout and packet scheduling for video streaming over wireless links [J]. *IEEE Transactions on Multimedia*, 2008, 10(5) : 885 - 895.
- [3] Jung-Hwan Lee, Chuck Yoo. Scalable ROI algorithm for H. 264/SVC-based video streaming [J], *IEEE Transactions on Consumer Electronics*, 2011, 57(2) : 882 - 887.
- [4] Dan Grois, Ofer Hadar. Complexity-aware adaptive preprocessing scheme for region-of-interest spatial scalable video coding [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2014, 24(6) : 1025 - 1039.
- [5] 李晓峰, 周宁, 刘洪盛, 张敏. 一种基于缩减栅格算法的 SVC 联合信源/信道编码方法 [J]. *电子学报*, 2011, 39 (4) : 859 - 864.
LI Xiao-feng, ZHOU Ning, LIU Hong-sheng, ZHANG Min. A joint source/channel coding with reduced trellis algorithm for the scalable extension of H. 264/AVC [J]. *Acta Electronica Sinica*, 2011, 39 (4) : 859 - 864. (in Chinese)
- [6] 陈旭, 张基宏, 柳伟, 梁永生. 基于视觉注意的的视频可伸缩 ROI 算法 [J]. *山东大学学报 (工学版)*, 2013, 43 (1) : 15 - 21.
Chen Xu, Zhang Jihong, Liu Wei, Liang Yongsheng. New scalable ROI algorithm based on visual attention [J]. *Journal of Shandong University (Engineering and Technology Edition)*, 2013, 43(1) : 15 - 31. (in Chinese)
- [7] 刘家瑛, 郭宗明, Yongjin CHO. 面向 H. 264/SVC 空域-质量域可伸缩编码的码率分配算法 [J]. *电子学报*, 2010, 28(9) : 2112 - 2117.
LIU Jia-ying, GUO Zong-ming, Yongjin CHO. Bit allocation algorithm in H. 264/SVC spatial-quality with dependent R-D modeling [J]. *Acta Electronica Sinica*, 2010, 28 (9) : 2112 - 2117. (in Chinese)
- [8] Ching-Lung Su, Tse-Min Chen, Chih-Yang Huang. Cluster-based motion estimation algorithm with low memory and bandwidth requirements for H. 264/AVC scalable extension. *IEEE transactions on circuits and systems for video technology*, 2014, 24(6) : 1016 - 1024.
- [9] 崔子冠, 朱昌秀, 干宗良, 唐贵进, 刘峰. H. 264 视频编码率失真优化和码率控制技术的研究进展 [J]. *电子学报*, 2013, 41(12) : 2443 - 2450.
CUI Zi-guan, ZHU Xiu-chang, GAN Zong-liang, TANG Gui-jin, LIU Feng. Advance in rate distortion optimization and rate control techniques for H. 264 video coding [J]. *Acta Electronica Sinica*, 2013, 41 (12) : 2443 - 2450. (in Chinese)
- [10] Wang J, Wang F, Zhang C, Shen HC, Quan L. Linear neighborhood propagation and its applications [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(9) : 1600 - 1615.
- [11] Christian Siagian, Laurent Itti. Rapid biologically-inspired scene classification using features shared with visual attention [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(2) : 300 - 312.
- [12] Jun Huang, Xiao Yang, Xiang-zhong Fang, Wei-yao Lin, Rui Zhang. Integrating visual saliency and consistency for re-ranking image search results [J]. *IEEE Transactions on Multimedia*, 2011, 13(4) : 653 - 661.
- [13] Heiko Schwarz, Thomas Wiegand. Implementation and performance of FGS, MGS, and CGS [S]. *Joint Video Team, Doc. JVT-V126*, 2007. 2 - 3.
- [14] Rohan Gupta, Akshay Pulipaka, Patrick Seeling, etc. H. 264 coarse grain scalable (CGS) and medium grain scalable (MGS) encoded video: A trace based traffic and quality evaluation [J]. *IEEE Transactions on Broadcasting*, 2012, 58(3) : 428 - 439.
- [15] 汪大勇, 孙世新. 可伸缩视频编码研究现状综述 [J]. *电子测量与仪器学报*, 2009, (8) : 78 - 84.
WANG Da-yong, SUN Shi-xin. Summary of research on scalable video coding [J]. *Journal of Electronic Measurement and Instrument*, 2009, (8) : 78 - 84. (in Chinese)
- [16] 纪超, 刘慧英, 孙景峰, 贺胜, 黄民主. 基于空域和频域的图像显著区域检测 [J]. *吉林大学学报 (工学版)*, 2014, 44(1) : 177 - 183.
JI Chao, LIU Hui-ying, SUN Jing-feng, HE Sheng, HUANG Min-zhu. Image salient region detection based on spatial and frequency domains. *Journal of Jilin University (Engineering and Technology Edition)*, 2014, 44 (1) : 177 - 183. (in Chinese)
- [17] Edmund Y. Lam, Joseph W. Goodman. A mathematical analysis of the DCT coefficient distributions for images [J]. *IEEE Transactions on Image Processing*, 2000, 9 (10) : 1661 - 1666.
- [18] Truong Cong Thang, Jung Won Kang, Jeong-Ju Yoo, Jae-Gon Kim. Multilayer adaptation for MGS-based SVC bitstream [C]. *Proceeding of the 16th ACM International Conference on Multimedia*, Vancouver, British Columbia, Canada, 2008. 689 - 692.
- [19] Wenger S, Wang Y K, Schierl T, Eleftheriadis A. RTP payload format for SVC video [S]. *IETF Internet Draft draft-ietf-avt-rtp-svc-06*, 2010. 47 - 95.

- [20] Chan-Won Seo, Jong-Ki Han. Rate control scheme for consistent video quality in scalable video codec [J]. IEEE Transactions on Image Processing, 2011, 20 (8) : 2166 - 2176.
- [21] Michal Ries, Olivia Nemethova, Markus Rupp. Video quality estimation for mobile H. 264/AVC video streaming [J]. Journal of Communications, 2008, 3 (1) : 41 - 50.
- [22] Kalpana Seshadrinathan, Rajiv Soundararajan, Aaln Conrad Bovik, Lawrcncc K. Cormack. Study of subjective and objective quality assessment of video [J]. IEEE Transactions on Image Processing, 2010, 19 (16) : 1427 - 1441.



周 莺 女,1984 年生于江苏省,现为深圳大学信息工程学院博士研究生,研究方向为视频编解码和数据通信.

E-mail: zhouying722@163. com



魏泽锋 男,1988 年生于广东省,现为深圳大学信息工程学院硕士研究生,研究方向为视频编码与信号处理.

E-mail: zefwei@ sina. cn



张基宏 男,1964 年生于江苏省,1992 年于东南大学获工学博士学位,现为深圳大学教授,研究方向为信号处理和多媒体信息处理.

E-mail: zhangjh@ sziit. edu. cn

作者简介



梁永生 男,1971 年生于黑龙江省,1999 年于哈尔滨工业大学获工学博士学位,现为深圳信息职业技术学院教授,深圳大学博士生导师. 研究方向为计算机网络与数据通信,信号处理与模式识别.

E-mail: liangys@ sziit. edu. cn



柳 伟 男,1973 年生于湖南省,2005 年于国防科学技术大学获工学博士学位,现为深圳信息职业技术学院教授,研究方向为图像处理,视频编解码,基于内容的多媒体信息处理,多媒体数据库系统.

E-mail: liuwei@ sziit. edu. cn