

基于兴趣区域的 H. 264/AVC 会话视频传输算法

李 凡

(西安交通大学 电子与信息工程学院, 陕西 西安, 710049)

摘 要: 针对会话视频的严格的实时性要求, 提出一种基于 H. 264/AVC 会话视频的误差抵抗编方法; 建立一种低复杂度的感兴趣区域分割算法; 基于所提分割算法, 对感兴趣区域进行长期参考帧预测编码, 从而达到提高感兴趣区域的视频传输质量的目标. 所提方法具有低复杂度的特点, 能够满足会话视频的实时性要求, 并且易于移植入 H. 264/AVC 编码器. 仿真结果证明了算法的鲁棒性和编码有效性.

关键词: 视频传输; 感兴趣区域; 误差抵抗; 视觉系统

中图分类号: TN919. 8

文献标识码: A

文章编号: 1000—7180(2012)08—0028—05

ROI-Based Transmission for H. 264/AVC Conversational Video

LI Fan

(School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: It is also a challenge for transmission for conversational video due to the strict real-time requirements. In this paper, we propose a low-complexity error resilient coding scheme for H. 264/AVC conversational video communications. For the conversational video, the attention-attracted areas make more contributions to the perceptual quality for end users. Therefore, the proposed scheme combines the fast region-of-interest (ROI) segmentation and the long-term reference (LTR) frame coding in order to improve the perceptual quality of ROI areas. The scheme can be easily integrated into the H. 264/AVC encoder. Simulation results verify the robustness and coding efficiency as well as improved subjective visual quality.

Key words: video transmission; region-of-interest (ROI); error resilience; human visual system

1 引言

会话视频是网络承载的一种典型的视频应用, 会话视频在内容上具有显著的特点. 其中一个最大的特点就是会话视频对观众来讲总具有感兴趣区域 (Region of Interest, ROI). 感兴趣区域的质量对整个视频的质量起到至关重要的作用.

当今的视频在编码中均采用预测编码, 视频帧之间具有非常强的相关性. 如果一个视频帧中的某个宏块在传输中丢失, 不但会影响所丢失宏块位置的视频质量, 还会对后续的视频帧中直接或间接与丢失宏块相关的位置的视频质量带来影响, 此现象叫做错误传播.

错误抵抗是解决错误传播问题的一个有效的方法^[1]. 常用的方法是将具有重要内容的视频宏块采用帧内编码的模式, 而不采用帧间预测编码, 从而杜绝前一视频帧的错误传播. 但是, 与帧间预测编码相比, 帧内编码的编码效率低, 会造成视频的压缩比降低. 另一种错误抵抗的方法是采用前向纠错编码 (Forward error correction, FEC) 对具有重要内容的视频宏块进行编码, 从而降低重要宏块的丢失概率^[2]. 但是, FEC 编码会增加冗余的数据包, 从而增大传输数据量. H. 264/AVC 编码标准中, 引入了长期参考帧来达到编码效率和鲁棒性的折衷^[3-4]. 采用长期参考帧模式, 可以对具有重要内容的视频宏块采用长期参考帧进行预测. 由于长期参考帧的零错

收稿日期: 2012—03—12; **修回日期:** 2012—05—09

基金项目: 国家自然科学基金(61001095); 国家科技重大专项(2012ZX03002023)

误特性,可以达到重要内容的有效的保护。

近年来,基于感兴趣区域的视频传输越来越引起学术界的关注^[5-9]。在此基础上,本文提出一种基于会话视频感兴趣区域传输的错误抵抗算法。

2 系统架构

图1中给出本文所提算法的系统架构。首先,采用快速 ROI 分割方法将会话视频中的每一视频帧内容进行分割,划分出 ROI 区域和非 ROI 区域。然后,采用错误抵抗编码方法对视频进行编码。对于 ROI 区域和非 ROI 区域,选择长期参考帧、短期参考帧进行帧间编码,或直接进行帧内编码。

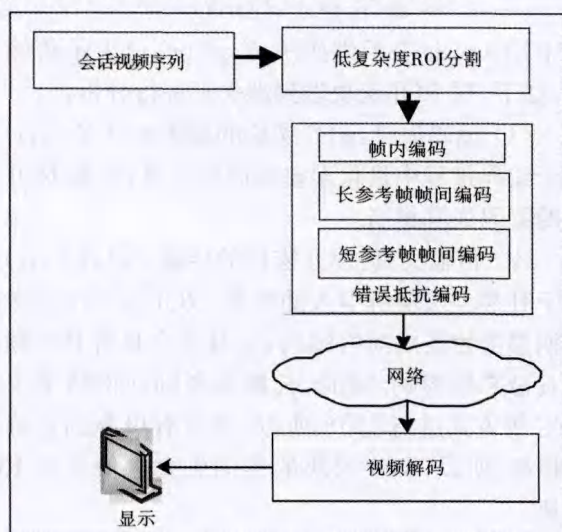


图1 系统架构

需要指出的是,错误抵抗编码方法直接移植进 H. 264/AVC 编码器。经过网络传输,视频在解码器中解码后显示在接收端。

3 基于 ROI 的错误抵抗编码算法

3.1 低复杂度 ROI 分割算法

ROI 分割算法是要在视频帧中确定感兴趣区域部分。需要注意的是,ROI 分割算法需要具有较低的算法复杂度,从而保证实时性。

对于会话类视频应用,人脸区域是图像中最显著、最能引起人们注意的区域^[8],所以,本文提出一种基于人脸检测的 ROI 分割算法。目前,人脸检测的算法成果较多,例如,支持向量机算法,基于隐马尔科夫模型的高阶统计算法等。尽管这些方法能够得到较为精确的人脸检测结果,但是这类算法均具有很高的算法复杂度,不能在实时传输应用中使用。

本文提出的基于人脸检测的 ROI 分割算法的

核心思想是利用肤色信息和人脸形状信息进行人脸定位。

由于目前的视频压缩算法主要是基于视频帧的宏块,我们将 ROI 分割算法应用于在宏块级。算法的步骤包含如下四步:

(1) 肤色检测。首先,采用肤色检测算法检测出具有人脸肤色的区域 p_1 ,图2(a)表示了视频 foreman 第三帧采用文献[8]中提到的模型进行检测的结果。可以看到,简单的肤色检测会得到许多噪声点。这些噪声点是对背景区域的误检。

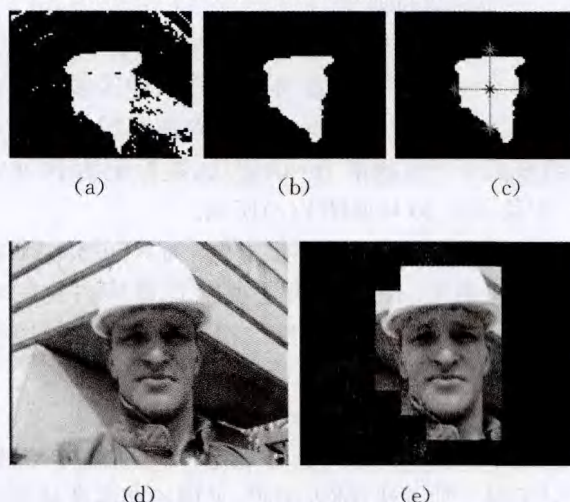


图2 视频 foreman 第三帧的 ROI 分割

用两种肤色模型检测人脸区域。设 p_1 和 p_2 分别表示两种模型的检测结果,都是二值图。图2(a)表示了视频 foreman 第三帧采用文献[8]中提到的模型进行检测的结果,用 p_2 表示,在图中,背景区域中有很多检测错误的像素点。原始图像的人脸区域掩膜 p 是 p_1 和 p_2 的交集,表示为

(2) 人脸形状限制。人脸的形状具有椭圆形的特性。这里,我们采用文献[9]提出的椭圆肤色模型,对肤色检测得到的区域进行人脸形状限制,从而将背景区域的噪声点去除,如图2(b)所示。去除的方法采用交集运算,设椭圆肤色模型得到的区域为 p_2 ,则去除噪声后的人脸区域为

$$p = p_1 \& p_2 \quad (1)$$

(3) 区域膨胀。仅仅采用人脸检测算法,得到的感兴趣区域只包括人脸部分。这与真正的人眼注意力是不相符的。对于观察者来讲,不止人的皮肤能引起注意,人头部的其他部分,如头发和帽子等,均是观察者感兴趣的区域。但是,人脸检测无法得到这些区域。

考虑到人的头部接近于一个椭圆,因此,我们将

感兴趣区域膨胀成为一个椭圆. 用 a 和 b 分别表示椭圆的长轴和短轴, θ 是长轴和横轴的夹角. 三个参数分别表示为

$$\begin{cases} a = \sqrt{\varphi_1 + \sqrt{\varphi_2}} \\ b = \sqrt{\varphi_1 - \sqrt{\varphi_2}} \\ \theta = \frac{\arctan[2\eta_{11}/(\eta_{20} - \eta_{02})]}{2} \end{cases} \quad (2)$$

式中, φ_i 是 i 阶不变矩, η_{ij} 是 ij 阶归一化中心距, 基于轴线和方向, 我们将椭圆限制在一个外接矩形内, 图 2(c) 为椭圆的轴线.

(4) ROI 宏块标注. 一个宏块是否被标注为 ROI 的宏块, 这由其内部具有的感兴趣区域的像素点个数决定. 宏块中被标记为感兴趣区域的像素点的数量高于一定阈值 thr 的宏块, 被标记为 ROI 宏块. 宏块级的 ROI 如图 2(e) 所示.

经过以上四步, 视频帧被分割为 ROI 宏块和非 ROI 宏块两类. ROI 宏块是观察者最感兴趣的区域, 需要在传输中采取特殊的措施保证其在接收端高质量的恢复.

3.2 基于 ROI 的长期参考帧编码方法

H. 264/AVC 编码标准规定, 对宏块进行编码可以采用三种方法: 帧内编码, 采用长期参考帧的预测编码以及采用短期参考帧的预测编码. 本文的算法中, 基于 ROI 的编码方法充分利用长期参考帧与短期参考帧的不同编码方式, 从而保证 ROI 宏块的传输质量.

本算法中, 我们将视频中的 I 帧设置为长期参考帧, 置于长期参考帧缓冲器中. 由于 I 帧仅采用帧内编码的模式, 不采用帧间预测的模式, 所以, 在编码时不会因为其它帧中的错误而引入失真. 并且, 在网络传输中, 对长期参考帧采用前向纠错编码等方式进行保护, 保证其无错误的传输. 同时, 当前编码帧的前一帧设置为短期参考帧, 也就是说每帧都可以担任短期参考帧.

如图 3 所示, 对视频帧进行编码时, 如果当前宏块是 ROI 宏块, 则采用长期参考帧进行编码; 如果当前宏块是非 ROI 宏块, 则采用短期参考帧进行编码. 即:

$$\text{编码模式} = \begin{cases} \text{LTR 预测} & \text{ROI 宏块} \\ \text{STR 预测} & \text{非 ROI 宏块} \end{cases}$$

令 p 表示丢包率, $\hat{f}_n^i$ 和 $\tilde{f}_n^i$ 分别表示第 n 个视频帧中第 i 个像素在编码器端和解码器端相应的重建值. 因此, 端到端失真可以表示为:

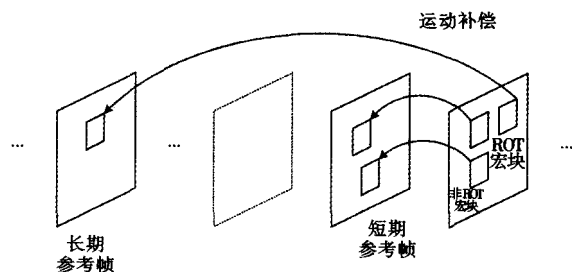


图3 基于 ROI 的参考帧选择示意图

$$\begin{aligned} d(n, i) &= E\{(f_n^i - \tilde{f}_n^i)^2\} \\ &= E\{(f_n^i - \hat{f}_n^i + \hat{f}_n^i - \tilde{f}_n^i)^2\} \\ &= E\{(f_n^i - \hat{f}_n^i)^2\} + E\{(\hat{f}_n^i - \tilde{f}_n^i)^2\} \\ &= d_e(n, i) + d_c(n, i) \end{aligned} \quad (3)$$

式中, $d_e(n, i)$ 表示编码失真, $d_c(n, i)$ 表示传输失真. 如下, 对 ROI 宏块端到端失真进行分析:

(1) 编码失真: ROI 宏块的编码失真 $d_e(n, i)$ 来源于编码过程中量化误差引起的失真, 与非 ROI 宏块的编码失真相当.

(2) 传输失真: ROI 宏块的传输失真 $d_c(n, i)$ 来源于其参考宏块所引入的失真. ROI 宏块所参考的长期参考帧采用帧内编码, 自身不会具有其它帧的失真带来的影响. 同时, 长期参考帧在网络中采用 FEC 等方式进行保护, 所以, 其具有极高的正确接收概率. 所以, ROI 宏块的传输失真远低于非 ROI 宏块.

所以, 采用长期参考帧对 ROI 宏块进行编码, 可以降低 ROI 宏块的端到端的传输失真, 对会话视频中的感兴趣区域做到高质量的传输.

4 仿真结果

本文基于 H. 264 参考软件 JM14 进行算法仿真. 测试视频选用标准测试序列 Foreman, Carphone, Claire 和 Suzie, 测试序列格式为 QCIF (352 * 288). 视频帧组 (GOP) 长度选择为 30, 其中包括一个 I 帧和 29 个 P 帧, 帧率为 30 帧/秒.

本文进行两组仿真: 固定视频比特率仿真和固定网络丢包率仿真. 每组仿真中, 比较三种算法的性能: 基准算法, 帧内编码算法和本文算法. 基准算法是不采用错误抵抗的编码算法. 帧内编码算法对 ROI 区域全部采用帧内刷新进行编码. 每次实验进行 30 次, 取平均值. 图 4 中给出了四种行列的仿真结果.

本文提出的算法对 ROI 区域的视频质量 PSNR 比基准算法提高了 2.88~5.28 dB, 比帧内编

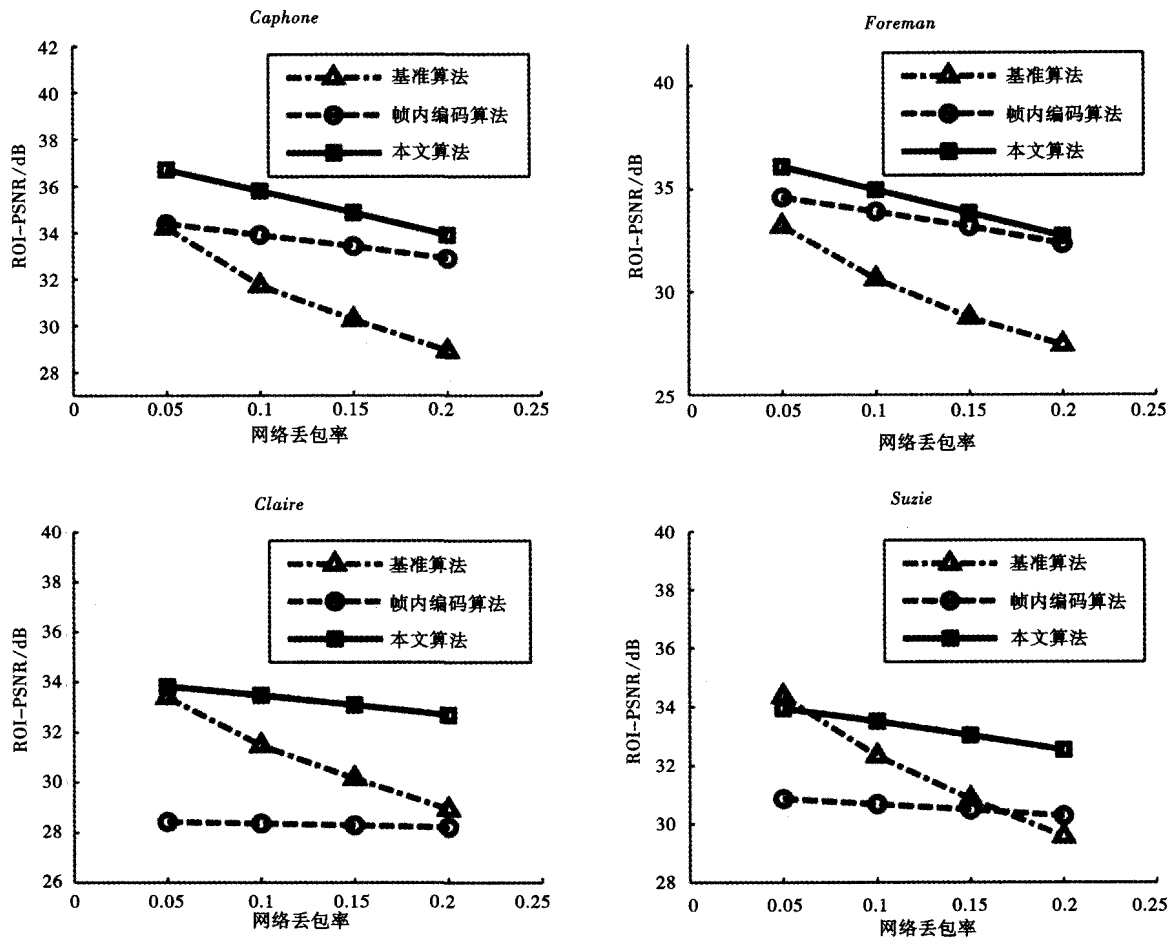


图 4 固定视频比特率仿真结果比较

码算法提高了 1.39~4.90 dB. 对于其他的测试序列,也可以得到类似的结果.

在固定网络丢包率仿真中,显然,本文的算法由于其它两种对比算法.需要说明的是,对于四种测试视频,在编码码率为 100 kbs 的条件下,采用本文算法,ROI 区域的视频质量均高于 32 dB,已经达到人眼良好的接收范围.图 5 中给出了四种测试序列的

仿真结果.

为了更清楚地表示仿真结果,表 1 中给出了 ROI 区域的视频质量和整个视频帧的视频质量.从中可以看中本文算法有效提高了 ROI 区域的视频质量.同时采用本文的算法,整个视频的质量基本与其他两种算法的结果相当.

表 1 算法的视频质量性能 PSNR 比较

	序列	算法	Loss rate			
			5%	10%	15%	20%
ROI	Carphone (300 kb/s)	基准算法	34.217	31.726	30.258	28.907
		帧内编码算法	34.386	33.898	33.407	32.873
		本文算法	36.702	35.794	34.859	33.888
PSNR	Foreman (300 kb/s)	基准算法	33.199	30.614	28.750	27.421
		帧内编码算法	34.589	33.890	33.163	32.327
		本文算法	36.080	34.955	33.838	32.699
整帧	Carphone (300 kb/s)	基准算法	30.629	31.123	31.261	31.685
		帧内编码算法	30.024	31.081	31.779	32.223
		本文算法	30.647	31.428	31.672	32.166
PSNR	Foreman (300 kb/s)	基准算法	29.630	30.043	30.253	30.705
		帧内编码算法	29.004	29.916	30.521	30.930
		本文算法	29.633	30.359	30.744	31.256

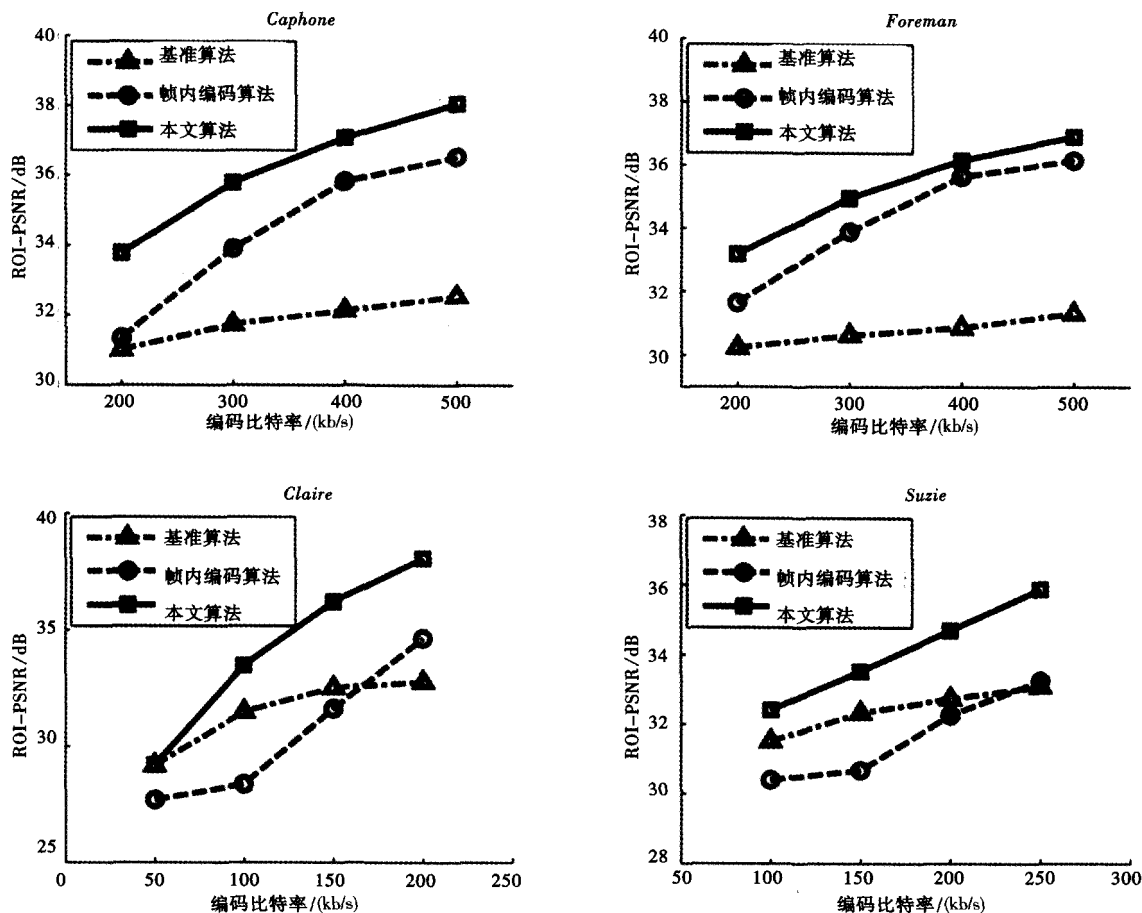


图5 固定网络丢包率仿真结果比较

5 结束语

本文提出一种基于会话视频实时传输的错误抵抗算法,其目标是提高感兴趣区域的视频质量.算法将低复杂度的感兴趣区域分割与 H. 264/AVC 编码器中的长期参考帧编码有机的联合起来,具有低复杂度的特点,能够满足会话视频的实时性要求,并且易于移植到 H. 264/AVC 编码器中.仿真结果验证了算法的有效性和鲁棒性.

参考文献:

- [1] 朱怀宏,姜永兵.一种基于块编码的压缩视频通信差错算法[J].微电子学与计算机,2008(9):69-71.
- [2] Frossard P, Verscheure O. Joint source/fec rate selection for quality-optimal mpeg-2 video delivery[J]. IEEE Trans Image process[J]2001,10(12):1815-1825.
- [3] Wiegand T, Farber N, Stuhlmüller K, et al Error-resilient video transmission using long-term memory motion-compensated prediction[J]. IEEE Journal on Selected Areas in Communications, 2000, 18(6): 1050-1062.
- [4] Zhang Q, Liu G. Error resilient coding of H. 264 using

intact long-term reference frames[C]// in Proc IET 5th International Conference on Visual Information Engineering, Xi'an, China; IET, 2008; 62-66.

- [5] 唐朝霞,章慧,吴海华.基于感兴趣区域多特征加权融合的图像检索算法[J].微电子学与计算机,2011(6):205-208.
- [6] Jerbi A, Wang J, Shirani S. Error-resilient roi coding using preand post-processing for video sequences[C]// Proc IEEE ISCAS'04, 2004(6):57-60.
- [7] Liu Y, Li Z G, Soh Y C. Region-of-interest based resource allocation for conversational video communication of H. 264/AVC[J]. IEEE Trans Circuits Syst. Video Technol, 2008,18(1):134-139.
- [8] Chai D, Ngan K N. Face segmentation using skin-color map in videophone applications[J]. IEEE Trans Circuits Syst Video Technol, 1999,19(4):551-564.
- [9] Hsu R-L, Abdel-Mottaleb M, Jain A. Face detection in color images[J]. IEEE Trans Pattern Anal Mach Intell, 2002,24(5):696-706.

作者简介:

李凡男,(1981—),讲师.博士.研究方向为视频处理与通信.