

◎数据库、信号与信息处理◎

感兴趣区域分割的非平衡多描述视频编码

余海瑞,倪林

YU Hai-rui, NI Lin

中国科学技术大学 电子工程与信息科学系,合肥 230027

Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China

E-mail: yuhairui@mail.ustc.edu.cn

YU Hai-rui, NI Lin. Novel unbalanced multiple description coding scheme for video using ROI segmentation. *Computer Engineering and Applications*, 2009, 45(4): 137-139.

Abstract: The lost of information in the attraction regions will decrease the quality of the video. This paper detects the attraction regions using some method and puts this data of inter-macroblocks into B-kind data partition through modifying the approach defined in H.264. Then it encodes the newly-formed three separate data using unbalanced multiple description method. The simulations show that the proposed method can effectively protect the data in the regions of interest and gain better subjective video quality than other unbalanced multiple coding methods.

Key words: Region of Interest(ROI); unbalanced multiple description coding; data partition; H.264

摘要: 视频序列中感兴趣区域信息的丢失将导致视频质量的严重下降,通过一定的算法检测出感兴趣区域,并通过修改 H.264 编码标准中的数据分割算法,把这部分信息的 Inter 宏块放入 B 类数据分割中,然后对新形成的三类数据运用非平衡多描述编码方法进行编码,实验结果表明,该方法在网络出错时可以有效地保证视频中感兴趣区域的数据,很好地改善接收到的视频质量。

关键词: 感兴趣区域;非平衡多描述编码;数据分割;H.264

DOI: 10.3778/j.issn.1002-8331.2009.04.038 文章编号: 1002-8331(2009)04-0137-03 文献标识码: A 中图分类号: TP391

1 引言

在无线网络中,网络拥塞或信道的突发错误常常使视频流数据包出错或者丢失,引起接收质量的下降。视频流本身具有很强的实时性,超出有效时间范围的数据包即使被正确接收到,对接收方来说也是无用的,因此针对数据业务设计的出错重传机制通常不能用于视频流业务。为了解决这一问题,研究人员提出许多容错技术:帧内编码,多参考帧,数据分割,以及可分级编码和多描述编码等。容错技术使视频流在发生传输错误时,接收端利用收到的部分数据尽可能恢复出较高的视频质量,减小数据丢失带来的影响。

多描述编码是近年来兴起的一种新的面向不可靠信道传输的编码方法,它主要解决网络传输可靠性问题,具有无须重传,延时小,易于实时视频传输的特性^[1]。它假定在发送端和接收端之间有多条相互独立的信道,且同时发生错误,数据包丢失的概率很小。原信号被分解成多个相关描述,每个描述单独编码并通过独立的信道传输,任何一个描述都能由边沿解码器解码得到质量可接受的信号,而多个描述结合则可由中央通道合成更高质量的信号。多描述编码可分为平衡多描述编码(BMDC)和非平衡多描述编码(UMDC),在平衡多描述编码方

法^[2]中产生的两个描述具有同等重要性,而非平衡多描述编码方法^[3-4]将产生具有高分辨率的描述(HR)和具有低分辨率的描述(LR),其中 LR 描述主要作为冗余实现对 HR 描述中的错误进行隐藏和恢复。相对于 BMDC 来说,UMDC 可以更灵活地控制冗余的大小。

以往的 UMDC 方法^[5-6]两个描述的产生是由两个编码器采用不同的编码参数独立形成,HR 描述由小的量化参数编码产生可以解码得到好的视频序列,LR 描述相对 HR 描述是采用大的量化参数编码而成,其重构出的视频质量相对来说比较粗糙,并且它对视频帧中各部分同等对待,没有区分各部分信息的重要性。在很多实际应用中,传输图像和视频的各部分的重要性不同。根据人眼的视觉特性,为了获得最大信息量,用户在浏览图像和视频时,往往只关注画面中的部分区域,即感兴趣区域(Region of Interest, ROI),而对图像中的背景等非感兴趣区域并不关心。因此可以对图像中的感兴趣区域采用不同的编码方法,以牺牲一定的编码效率为代价,确保图像或视频在数据丢失的情况下能够重构出好的质量。

本文结合 H.264 编码标准中的数据分割方法^[7],提出了一种基于感兴趣区域分割的非平衡多描述编码。数据分割是视频

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60672055)。

作者简介:余海瑞(1984-),硕士生,主要研究方向:信号与信息处理、多描述视频编码;倪林(1967-),副教授,主要研究方向:图像检索、生物特征识别、视觉信息处理。

收稿日期:2008-08-05 修回日期:2008-10-31

传输在易错环境下抗误码的有效途径,特别是在反馈信道不可用的场合,它将编码后的码流按照一定关系组织成多个有重要性区别的子流(称为 partition),而不是像传统的做法那样组织成一个流。在传输过程中对重要的子流采用较强的保护以使其被正确接收的概率更大。H.264 编码标准中给出的数据分割方法做法简单,但灵活性差,不能适应图像内容和网络条件的变化。将其进行改进提升感兴趣区域数据的重要性,即将视频序列中检测出的感兴趣区域的 Inter 宏块信息放入 B 类数据分割中,这样 C 类数据分割中将只包含背景等不感兴趣区域宏块的 Inter 信息。HR 描述将按照标准的编码方法包含三类所有数据,而 LR 描述将作为 HR 描述的冗余只包含修正的 A 和 B 类数据,以提高重要数据的保护力度。相比以前的非平衡视频多描述编码而言,本文提出的方法在网络出错或拥塞时可以很好地提高视频质量。

2 H.264 的数据分割算法

传统编码器将宏块头信息与宏块系数交织在一起形成单一码流,这样当其中某个信息位出现传输错误时,由于 VLC 编码,出错位后面的数据,包括运动矢量等重要信息,都变得不可用,这可能使视频质量严重下降。在 H.264 标准中,使用三种不同类型的数据分割,其中, A 类分割是头信息划分,包括宏块类型、量化参数和运动矢量; B 类分割是帧内信息划分,包括帧内 CBP 和帧内系数,帧内信息可以阻止错误的传播; C 类分割是帧间信息划分,包括帧间 CBP 和帧间系数,一般情况下它是编码分片的最大分区。在这种组织方式下, A 类数据包含了正确解码图像所需的最重要的信息; 其次 B 类数据重要性强于 C 类数据; C 类数据最要性最小,一些帧的 C 类数据即使不被正确接收到,解码器也能重建出可以接受的图像质量。三类数据有这样的依赖关系: B 类和 C 类数据各自独立,但是它们的解码都必须依赖于 A 类数据。基于这样的分割方式,信道编码器可以有选择地对不同类数据进行不等重要错误保护,以使在信道条件不好时,首先确保那些重要的信息能被接收端正确接收。

H.264 的数据分割方法虽然较好地改善了丢包环境下的视频传输质量,但是其还有不足之处。例如: B 类和 C 类数据的分割过于僵化,编码器实际上不能控制这两类的数据量。这两类数据的比例很不均衡: B 类包含的数据太少,对很多不含 Intra 宏块的帧而言 B 类数据量为 0,而 C 类数据包含的宏块数据过多,几乎为一帧图像全部宏块的信息。在信道条件较差的环境中,如果保护力度最弱的 C 类数据全部被丢弃,则依然会使一些帧的重建质量明显下降,有时甚至不可容忍。而实际上, C 类中包含的 Inter 宏块对重建图像质量的影响是各不相同的,不宜将它们等视之。根据人眼视觉特性,用户在浏览图像和视频时,往往关注画面中的部分区域,即感兴趣区域。就是说感兴趣区域的宏块相对来说在重构视频信息时具有更重要的特性。因此,本文提出一种改进的数据分割方法,该方法保留原有的 A 类和 B 类数据,但是将 C 类数据中的感兴趣区域宏块划分入 B 类,剩余保留在 C 类中。这种方法在数据包丢失时可以更好地改善重建视频的主观质量。

3 基于图像内容分割的非平衡多描述编码

3.1 感兴趣区域的检测

视频序列中感兴趣区域由一系列的感兴趣对象组成。将采

用^[8]定义的模型来表示感兴趣对象,每个感兴趣对象有两个属性: SR(Spatial Region)和 AV(Attention Value)。SR 表示感兴趣对象在图像中的具体位置, AV 反映了感兴趣对象的重要性,即在整个图像所包含的信息中的贡献。为了精确地检测图像中的感兴趣信息,基于上述模型,本文采用三种感兴趣对象: 运动对象(motion object),人脸对象(face object)和显著图对象(saliency object)。由于视频压缩编码大都基于块对序列进行编码,将以块的形式给出感兴趣区域检测的结果。

通常情况下,视频中的运动对象比其他类型的对象更加吸引用户的注意,也就是说具有更多的感兴趣信息。因此在本文系统中,运动对象的检测是最重要的。在视频压缩编码中采用了运动估计对视频序列进行预测编码,将根据视频编码中的运动矢量来确定每个运动对象的 SR。首先由运动估计方法^[9]给出图像中每个宏块的运动矢量 $(dx_{i,j}, dy_{i,j})$,计算出运动场 $I(i,j) = \sqrt{dx_{i,j}^2 + dy_{i,j}^2}$,然后采用中值滤波器滤除噪声,设置一阈值 T 和运动场比较,大于 T 的宏块将是运动强度比较强的运动对象的宏块,也就是所要检测的运动对象的位置。AV 值则根据区域的尺寸和运动强度来估计, AV 值和运动对象的面积和运动强度的强弱成正比,也就是拥有较大面积、较高运动强度的运动对象具有更高的 AV 值。具体计算公式为:

$$AV_{motion} = Area_{motion} * W_{motion}^{intensity} \quad (1)$$

显著对象的检测^[10]将同时采用多个 low-level 特征即对比度、方向和色度生成显著图,进而确定显著对象所在的区域。显著对象的 AV 值根据下式计算:

$$AV_{saliency} = \sum_{(i,j) \in R} B_{ij} * W_{saliency}^{i,j} \quad (2)$$

其中 B_{ij} 表示显著图中宏块 (i,j) 处的像素值。通常情况下,用户对图像中心更加关注,所以在上式中加入一个以图像中心为分布中心的标准化的高斯模板,给图像中不同位置加入不同权值。

在本文的系统中,人脸对象的检测采用其他学者已提出的成熟的检测算法^[11], AV 值将根据人脸对象的尺寸计算。

为了将三种感兴趣对象结合到一起,将他们的 AV 值归一化到(0,1)范围内,最终的 AV 值计算公式为:

$$AV_i = w_k * AV_i^k / \sum_i AV_i^k \quad (3)$$

其中 AV_i^k 表示第 i 个感兴趣对象的 AV 值,它是采用模型 k 检测的, w_k 是模型 k 的权值。利用该权值可以调整不同的感兴趣对象对整个感兴趣信息的贡献程度,从而可以根据接收端用户的具体需求调整感兴趣检测算法。如果运动对象更加吸引用户的关注,则运动对象的权值比其他对象要高。

图 1 给出了 Foreman 第 3 帧的运动对象、显著图对象和人脸对象,以及三者结合在一起的感兴趣对象的检测结果,以宏块形式给出。

3.2 非平衡多描述编码

本文给出的基于感兴趣区域分割的非平衡多描述编码整体框架如图 2 所示,编解码过程均采用 H.264 标准,对编码产生的码流将采用数据分割技术进行保护。首先输入视频序列经编码器编码产生 HR 描述和 LR 描述,HR 包括完整的码流即 A、B、C 三类数据分割,LR 将只包括编码的一些重要参数,运动矢量和感兴趣区域的信息即 A 和修正后的 B 类数据分割。接

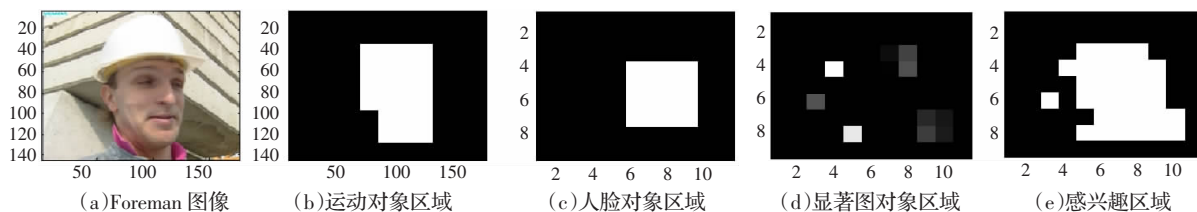


图1 感兴趣对象的检测结果

着两套数据流以 RTP 的格式封装放入信道中传输,在接收端先对数据包进行解封封装,然后解码。在解码端将视具体情况而决定如何解码,如果传输过程中没有发生错误,解码器将采用 HR 描述的信息重构视频序列,如果传输过程中发生错误,具体的处理过程将在下面给出。

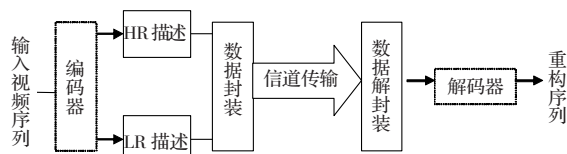


图2 本文给出的非平衡多描述编码结构图

如果 HR 中的某一 P 帧丢失,将采用 LR 中的重构帧作为输出,首先它既可以保证感兴趣区域的正确解码,其次由于一幅图像中的非感兴趣部分大都只包含运动幅度较小的背景等区域信息,可以根据 A 类数据中的运动矢量和之前正确解码的视频帧进行错误恢复。为了减弱 HR 描述中 P 帧的丢失对后续帧的影响,在编码过程中将周期性地插入 I 帧。

4 实验结果

在仿真实验中视频编码器采用修改的 H.264JM86,图像序列采用 Foreman(100 帧)和 Mother(100 帧),两个序列均为 QCIF(176×144)格式,编码帧率设为 30 fps,编码模式为 IP P...P,每 6 帧插入一次 I 帧。将和文献[6]中的编码方案 UMD 进行比较,在文献[6]中两个描述分别采用不同的量化参数编码而成,LR 描述相对 HR 描述是采用大的量化参数编码而成,其重构出的视频质量相对来说比较粗糙,并且它对视频帧中各部分同等对待,没有区分各部分信息的重要性。通过比较将看到本文所采用的方法(ROI-UMD)在数据发生丢失的情况下,能有效地保护图像中的感兴趣区域,能够重构出更好的视频序列,极大地提高了主观视频质量。

图 3 给出了传输过程中 HR 描述中 P 帧丢失由不同方法中相应 LR 描述的数据进行解码恢复的结果图。图(a)为无错情况下的重建图,(b)为 ROI-UMD 方法中的重建图,(c)为 I-UMD 方法中的重建图。由图中可以看出,本文方法恢复的图像获得较好的质量,感兴趣区域信息没有丢失,其余部分由于运动强度较小采用帧间补偿可以很好地恢复,和 I-UMD 方法恢复出的效果相当,而 I-UMD 方法恢复出的图像整体效果不错,但是人脸部分效果不好。

下面将比较 ROI-UMD 和 I-UMD 方法在丢包情况下的性能,其中两种方法中都将以周期为 6 的方式插入 I 帧,在实验中将模拟 6 种丢包率即 5%、10%、15%、20%、25%、30%。从图 4 中可以看出两种方法随着丢包率的增加,接收端重构视频质量均下降,但本文方法所获得的均值 PSNR 较 I-UMD 方法高 0.1~0.7 dB。

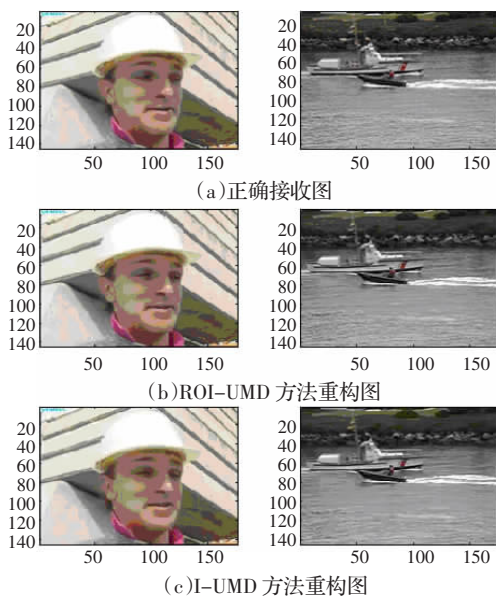


图3 重建效果比较图

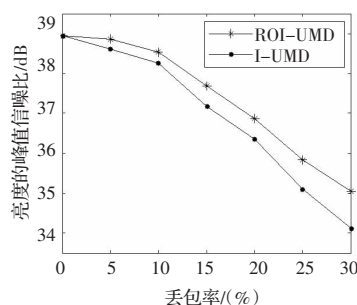


图4 本文方法与 I-UMD 方法均值 PSNR 性能比较图

5 结论

本文提出了一种基于感兴趣区域分割的非平衡多描述编码方法,这种方案提高了视频序列中感兴趣区域信息正确接收的概率,接收到的图像具有很好的视觉效果。相比一般的非平衡编码方法而言,该方法在发送端和接收端只需要一个编码器和一个解码器,简化了编解过程。它充分结合 H.264 编码标准的数据分割方法,进一步保证了数据的正确传输。

参考文献:

- [1] Goyal V K. Multiple description coding: Compression meets the network[J]. IEEE Signal Processing Magazine, 2001, 18(5): 74-93.
- [2] Matty K R, Kondi L P. Balanced multiple description video coding using optimal partitioning of the DCT coefficients[J]. IEEE Transactions on CSVT, 2005, 15(7): 928-934.