

一种基于 MPEG-4 的感兴趣区域视频编码新方法

张 方,吴成柯,肖 嵩,王养利

(西安电子科技大学 ISN 国家重点实验室,陕西西安 710071)

摘 要: 本文提出了一种基于 MPEG-4 的感兴趣区域视频编码新方法.通过对感兴趣区边沿宏块采用强制帧内编码及宏块内预测编码时参考位置的自适应选择,该方法有效地抑制了数据噪声的扩散和传播.通过采用自适应宏块大小的运动估计/补偿算法,提高了感兴趣区特别是图像运动复杂的感兴趣区的编码效率及质量.在码率分配部分,本文方法通过计算不同区域的图像复杂性和能量,依据用户可设定的感兴趣权重因子不等重地分配可用码率资源.实验证明,本文方法较大程度改善了感兴趣区视频编码的压缩效率,提高了码率分配地灵活性和有效性.

关键词: 视频编码;感兴趣区;MPEG-4;自适应块大小;可分级编码;码率分配

中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2005) 04-0634-04

A New Regions of Interest Video Coding Method Based on MPEG-4

ZHANG Fang, WU Cheng-ke, XIAO Song, WANG Yang-li

(National Key Lab. on ISN, Xi'dian University, Xi'an, Shaanxi 710071, China)

Abstract: A new ROI video coding method is presented based on MPEG-4. In order to improve the video coding gain and the compression quality, the intra macroblock prediction, adaptive block-size transformation (ABT) and adaptive block-size motion estimation (ABME) are applied to ROI coding, then error propagation from other regions to ROI is prevented by encoding border macroblocks of the ROI in intra mode. Furthermore, content-based bit allocation strategy is proposed according to MPEG-4 rate-distortion model. Software simulation results show that the method can improve the image quality both objectively and subjectively compared to MPEG-4 and other ROI coding methods.

Key words: video coding; ROI; MPEG-4; adaptive block size; scalable encoding; bit allocation

1 引言

网络视频流业务的迅速发展无可避免的受到有限带宽资源的制约,而在用甚低比特率进行视频压缩时,往往会丢失一些细节信息,甚至其中部分信息是人们特别感兴趣的.比如视频会议中的人物头肩区域,航空图像中携带重要信息的小目标区域,干涉多光谱图像中的干涉条纹区域,以及医学图像中病灶区域等等,这些可以统称为感兴趣区域(ROI: regions of interest).从主观视觉角度来说,这些区域的恢复质量直接影响到重建图像的整体主观感受质量.感兴趣区域编码将有限的码率优先分配给视频序列中人们感兴趣的区域,可以明显提高恢复图像质量,同时便于对码流实现分层不等重保护功能,提高视频网络通信系统的鲁棒性,因此在网络视频流业务的不断发展中起着至关重要的作用.

许多文献在感兴趣区域视频编码方法上做了有益的研究^[1-6].文献[3]将感兴趣区域分割为子图像进行独立编码,并且提出了一种渐变的码率分配方案,文献[4]在 MPEG-4 FGS (Fine Granular Scalable)的基础上,通过位平面提升的方法提高

感兴趣区域图像增强层的压缩质量.这些方法取得了一定的效果,但都不能从压缩的角度对感兴趣区域特别是图像运动复杂的感兴趣区的视频质量加以改善,也不能基于实际的图像内容更有效的利用码率资源.本文提出了一种基于 MPEG-4 的感兴趣区视频编码新方法,为了有效抑制数据噪声的扩散和传播,本文方法对感兴趣区边沿宏块采用强制帧内编码,在宏块内预测编码时自适应选择参考宏块的位置.通过采用自适应宏块大小的运动估计/补偿算法,提高了感兴趣区特别是图像运动复杂的感兴趣区的编码效率及质量.在码率分配部分,本文方法通过计算不同区域的图像复杂性和能量,依据用户可设定的感兴趣权重因子不等重地分配可用码率资源.实验证明,本文方法较大程度改善了感兴趣区视频编码的压缩效率,提高了码率分配地灵活性和有效性,其性能在不同参数上明显优于其他方法.

2 系统结构及感兴趣区域编码控制

本文提出的基于 MPEG-4 的感兴趣区域视频编码方法的系统结构及信号处理流程如图 1 所示.视频预处理及 VO

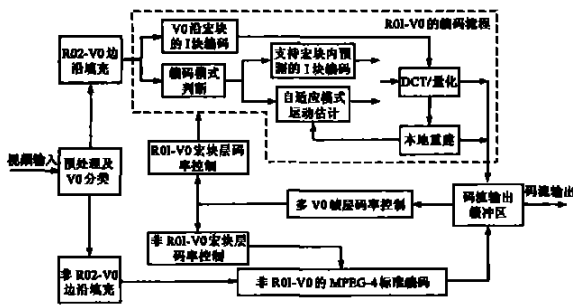


图 1 本文方法的系统结构图



图 2 任意形状 ROI-VO 的边沿填充方法

(Video Object) 分类模块将输入图像分割成任意形状的感兴趣 VO(ROI-VO) 及非感兴趣 VO(非 ROI-VO), 然后由边沿填充模块判断单个 VO 边沿是否与 16 × 16 的宏块边沿重合, 若不重合就在 VO 边沿处填充透明像素点, 使得每个 VO 均包含整数个宏块, 如图 2 所示。对于非 ROI-VO 的编码采用 MPEG-4 标准算法, 考虑到不是本文的研究重点, 这里不再赘述, 具体流程可参见相关文献 [8]。在误码信道传输时, 感兴趣区域视频编码要优先保证 ROI-VO 的图像压缩、传输质量, 因此防止非感兴趣区域数据噪声向感兴趣区域的扩散非常重要。为了避免由于运动估计、运动补偿及传输引起非 ROI-VO 的误差扩散到 ROI-VO, 我们采用了如下两种特殊的处理方法:

(1) 对 ROI-VO 的边沿宏块采用强制帧内编码模式 I-VOP, 从而在一定的运动估计搜索范围内, ROI-VO 内部宏块的搜索窗口不会超出 VO 边界, 也就不会由于非 ROI-VO 的误差而造成运动估计、运动补偿的失准。

(2) 对于 ROI-VO 内做 I-VOP 编码的宏块, 在经过 DCT、量化处理后做变换块(频域)的宏块级帧内 AC、DC 预测时, 我们规定其不能选择非 ROI-VO 的宏块作为参考。对于 VO 交接处或由于 P-VOP 编码导致的不可得到的参考宏块, 用 DC 值 128, AC 值 0 来构建虚拟参考宏块^[8]。

对于感兴趣区采用帧内编码的 16 × 16 宏块, 系统采用自适应模式的运动估计方法。这里引入一个宏块图像运动复杂度 v_o 及宏块内图像运动复杂度 MB 的概念。定义如下: 设观察对象包括 N 个作 P-VOP 编码且运动矢量不为零的宏块, 每个都作 16 × 16 结构的运动估计, 其各自运动矢量表示为 $MV_i = [MV_i(x), MV_i(y)] (i = 1, 2, \dots, N)$, 则该 VO 的图像运动复杂度计算如下:

$$v_o = \frac{1}{N} \sum_{i=1}^N \{ [MV_i(x) - MV_{aver}(x)]^2 + [MV_i(y) - MV_{aver}(y)]^2 \} \quad (1)$$

$$MV_{aver} = [MV_{aver}(x), MV_{aver}(y)] \\ = \left[\frac{1}{N} \sum_{i=1}^N MV_i(x), \frac{1}{N} \sum_{i=1}^N MV_i(y) \right]$$

当一个宏块具有 M 个运动矢量时(比如采用 8 × 8 结构做运动估计时, 宏块会有 4 个运动矢量), 分别表示为 $MV_i = [MV_i(x), MV_i(y)] (i = 1, 2, \dots, M)$, 则宏块图像运动复杂度计算如下:

$$MB = \frac{1}{M} \sum_{i=1}^M \{ [mv_i(x) - mv_{aver}(x)]^2 + [mv_i(y) - mv_{aver}(y)]^2 \} \quad (2)$$

其中,

$$mv_{aver} = [mv_{aver}(x), mv_{aver}(y)] \\ = \left[\frac{1}{M} \sum_{i=1}^M mv_i(x), \frac{1}{M} \sum_{i=1}^M mv_i(y) \right]$$

如果在一个 VO 区域内, 有 X 个宏块作 16 × 16 结构的运动估计, 有 Y 个宏块作其他结构(如 8 × 8、16 × 8、8 × 16、4 × 8 等等)的运动估计, 则该 VO 的图像运动复杂度可表示为:

$$= \frac{X}{X+Y} v_o + \frac{Y}{X+Y} MB \quad (3)$$

据此, 我们设计了一种自适应块结构的运动估计策略, 具体算法描述如下:

- (1) 根据式(1)、(2)、(3)分别计算上一帧编码图像的 v_o 、 MB 、;
- (2) 如果 C_{low} , 则对当前帧该 VO 内所有宏块作 16 × 16 结构的运动估计;
- (3) 如果, $C_{low} < C_{high}$, 则对当前帧该 VO 内所有宏块作 16 × 8、8 × 16、8 × 8 结构的运动估计, 然后依据本地重建图像的 MSE 判决来选择具体的块结构;
- (4) 如果 $> C_{high}$, 则对当前帧该 VO 内所有宏块作 4 × 8、8 × 4、4 × 4 结构的运动估计, 然后依据本地重建图像的 MSE 判决来选择具体的块结构。

这里的 C_{low} 、 C_{high} 为用户可设置的图像运动复杂度阈值, 取值越低, 则编码过程中采用较小块结构的概率越大, 从而编码质量越高, 产生的码流越多, 当然计算量也就越大。对于帧间编码的 DCT 模块, 其块变化结构依据运动估计部分的选择为准, 比如对于作 8 × 16 结构运动估计的宏块, 其残差的 DCT 运算也采用 8 × 16 结构。

3 基于内容的感兴趣 VO 码率分配方案

通过分析输出缓冲区中存储空间的剩余量及前一帧图像的实际编码码率, 可以计算出当前帧期望的编码码率 B_{pre} , 同时定义分给感兴趣 VO 及其他 VO 的码率分别为 R_{ROI} 及 R_{Other} , 则对各不同区域的码率分配可以采用如下原则^[1,7]:

$$\begin{cases} R_{pre} = R_{ROI} + R_{Other} \\ \frac{R_{ROI}}{R_{Other}} = \frac{m}{M-m} \times \prod_{j=1}^{j-1} \frac{m}{M-m} \times \prod_{l=1}^{l-1} \frac{m}{M-m} \end{cases} \quad (4)$$

Weight

由于宏块的 MAD 值可以反映出对应位置图像的复杂性及能量,因此方程组(4)第二个等式中引入了 Weight 部分的加权因子,是表示感兴趣 VO 和其他 VO 图像复杂度的统计性比例系数,它的引入可以解决这一问题:当感兴趣 VO 图像信息较为简单时,有限的码率已经可以保证较高的图像质量,此时没必要分配太多的码率资源给它,可以留给或许复杂度远高于它的其他 VO。为用户可设置的感兴趣区重要程度的权重系数,可在区间(1,10)内取值。化简式(4),可得到:

$$R_{ROI} = \frac{\sum_{j=1}^m MAD_{ROI}^j}{M_{-m} + \sum_{j=1}^m MAD_{ROI}^j} R_{pre} \quad (5)$$

MPEG4 中多 VO 编码的率失真模型定义为^[9]:

$$R(i) = \frac{X(i)MAD(i)}{Q(i)} + \frac{X(i)MAD(i)}{Q(i)^2} \quad (6)$$

其中 $R(i)$ 为第 i 个 VO 的码率, $Q(i)$ 为第 i 个 VO 的量化步长, $X(i)$ 、 $X(i)$ 分别为模型的一阶、二阶系数。假定感兴趣 VO 中共包含 m 个宏块,分配给单个宏块的码率可依据下式确定:

$$R_{ROI}^j = \frac{MAD_{ROI}^j}{\sum_{k=1}^m MAD_{ROI}^k} R_{ROI}, (j=1, 2, \dots, m) \quad (7)$$

其中 $j(j=1, 2, \dots, m)$ 表示感兴趣 VO 内第 j 个宏块的码率分配权重系数,缺省情况下取 $j=1$ 。单个宏块内的率失真模型即期望码率与量化步长的关系与式(6)相似,推导可得如下结果:

$$Q_{ROI}^j = \begin{cases} \frac{MAD_{ROI}^j}{R_{ROI}^j} A, & \kappa < 0 \\ \frac{\sqrt{\kappa + X(i)MAD_{ROI}^j}}{2R_{ROI}^j}, & \kappa \geq 0 \end{cases}$$

其中: $\kappa = [X(i)MAD_{ROI}^j]^2 + 4R_{ROI}^jX(i)MAD_{ROI}^j$, A 、 $X(i)$ 、 $X(i)$ 为模型参数,需要随着已编码宏块实际码率与期望码率之间的关系实时更新,具体可以采用与文献[8]中描述相同的算法。

4 仿真结果

为了将本文提出的方法与其他文献中的方法做比较,我们采用 QCIF 格式 Carphone 和 Akiyo 序列对系统的性能做了测试。编码帧频为 15fps,强制 FVOP 转换间隔为 15,运动估计搜索范围 $[-32, 31]$,半像素精度。为了简化分析,仿真时的感兴趣区域选为一个矩形框, C_{low} 、 C_{high} 分别取值 2.5 和 5,公式(4)中的取值 5,公式(7)中取 $j=$

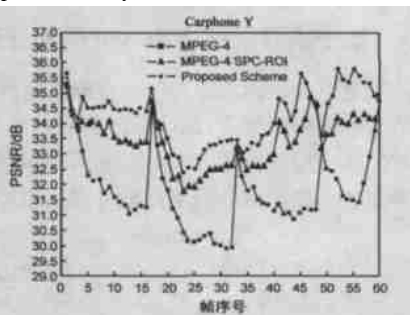


图3 不同编码方法的帧序号-PSNR 曲线

1,采用文献[8]中给定的 MPEG 量化矩阵,目标码率设定为 32Kbps。软件仿真时,噪声信道采用两状态 Markov 链的 Gilbert-Elliott (GE)模型。由于 Carphone 序列感兴趣区(头像像)部分运动较大,具有较强代表性,也鉴于篇幅限制,图3仅绘制出不同编码方法时 Carphone 重建测试序列 ROI 区 Y 分量 PSNR 的时变曲线,显然本文方法的性能要明显优于 MPEG4 及 SPC-ROI^[3]。在相同实验条件下,运动越复杂的图像帧及图像区域,本文方法获得的编

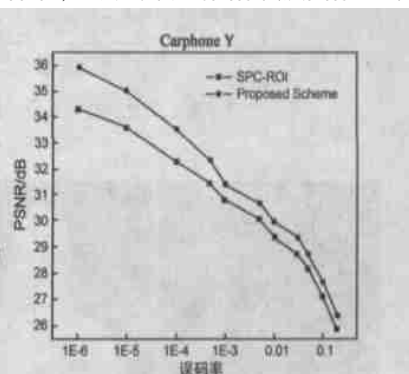


图4 不同误码率下重建图像的 PSNR 曲线

码增益越大,对于相对静止或 FVOP 编码的图像区域,本文方法的编码性能与其他两种方法接近。图4所示为 Carphone 序列第 55 帧图像在不同误码率情况下的解码图像信噪比曲线,这里对本文方法和 SPC-ROI 均采用了与文献[10]相类似的不等重保护 (UEP: unequal protection) 策略。图5为误码率 $Ber = 10^{-3}$ 条件下该帧的重建图像。



图5 Carphone 序列第 55 帧在误码率 10^{-3} 时采用不同方法的重建图像质量

5 结论

本文提出了一种基于 MPEG4 的视频感兴趣区域编码新方法,在引入视频帧多个编码 VO 概念的基础上,将编码对象切割为感兴趣 VO 和非感兴趣 VO,并分别对其采用不同的处理算法。与其他文献相比,本文方法有三个创新点:

- (1) 为了防止编码及传输误差在非感兴趣区域及 ROI-VO 之间的扩散,系统对于 ROI-VO 的边沿宏块均采用强制 I 帧编码,且规定其宏块内压缩域系数的预测不能采用非感兴趣区域的宏块作为参考。
- (2) 对于 ROI-VO 内采用帧间编码的宏块,系统引入了图像运动复杂度的概念。在运动复杂度高时,其运动估计自适应的采用更小的块结构,从而其编码精度也就更高。
- (3) 提出了基于内容的 ROI-VO 码率分配方案,在计算图像不同区域复杂度的基础上,优先给复杂性高的 VO 和宏块划分更多的码率资源。

仿真结果表明,本文方法能较大程度改善感兴趣区域的编码质量,同时实现对感兴趣区域及非感兴趣区域编码质量的精细控制,在相同实验条件下,其编码增益比文献[3]的方法高平均 1.2dB。

参考文献:

- [1] Supavadee Aramvith, Hatairat Kortrakulkij, Datchakorn Tancharoen and Somchai Jitapankul. Joint source-channel coding using simplified block-based segmentation and content-based rate-control for wireless video transport [A]. Proceedings of International Conference on Information Technology: Coding and Computing (ITCC) 2002 [C]. Las Vegas, Nevada, April 08-10, 2002. 71 - 76.
- [2] C W Lin, Y C Chen, M T Sun. Dynamic region of interest transcoding for multipoint video conferencing [A]. Proceedings of Int. Computer Symp. Workshop on Computer Networks, Internet, and Multimedia [C]. Chiayi, Taiwan, 2000. 114 - 121.
- [3] Miska M. Hannuksela, Yè- Kùì Wang, Mòncef Gabbouj. Sub-picture: ROI coding and unequal error protection [A]. IEEE 2002 International Conference on Image Processing (ICIP '2002) [C]. Rochester, New York, USA: Sept. 2002. 537 - 540.
- [4] Changho Shin, Kwang-deok Seo, Jae-kyoon, Kim. Rectangular region-based selective enhancement for MPEG-4 fine granular scalability [A]. Int. Packet Video Workshop [C]. Pittsburgh USA: April, 2002. 101 - 107.
- [5] Nikolaos Doulamis, Anastasios Doulamis, Dimitrios Kalogeras, Stefanos Kollias. Lowbit-rate coding of image sequences using adaptive regions of interest [J]. IEEE Transactions on CSVT, 8(8): 928 - 934.
- [6] D Chai, KN Ngan, A Bouzerdoum. Foreground/ background bit allocation for region of interest coding [A]. IEEE 2000 International Conference on Image Processing (ICIP '2000) [C]. Vancouver, BC, Canada, Sept. 2000, 2: 923 - 926.
- [7] A Vetro, H Sun, Y Wang. MPEG-4 rate control for multiple video objects [J]. IEEE Trans on CSVT, 1999, 9(1): 186 - 199.
- [8] MPEG-4 Video Verification Model version 18. 0, ISO/ IEC/ JTC1/ WG11 [S]. Sep, 1999.
- [9] ITU-T Rec H 264 / ISO/ IEC 11496-10, Advanced Video Coding, Final Committee Draft, Document JVT-F100 [S]. December, 2002.
- [10] 肖嵩, 张方, 吴成柯. 基于小波 SPIHT 的联合信源信道编码新方法 [J]. 计算机学报, 2003, 26(3): 281 - 286.

作者简介:



张 方 男, 1977 年出生于陕西省, 分别在 1998 年、2001 年获得西安交通大学电子工程专业学士、硕士学位, 2004 年 6 月在西安电子科技大学获得信号与信息处理专业博士学位, 主要从事图像压缩编码、网络多媒体业务及视频流的网络传输等方面的研究工作. E-mail: zfang@xidian.edu.cn.

吴成柯 男, 1938 年出生于安徽省, 教授, 博士生导师, 著有《图像通信》、《通信原理》两本书籍, 主要从事图像处理、图像通信、计算机视觉等方面的研究工作.

肖 嵩 女, 1977 年出生于陕西省, 1998 年获西安电子科技大学通信工程专业学士学位, 2001 年获西安电子科技大学通信与信息系统专业硕士学位, 2004 年 6 月在西安电子科技大学获得信号与信息处理专业博士学位, 主要从事图像压缩编码、联合信源信道编码等方面的研究工作.