

分布式视频编码的关键帧提取算法

宋晓丽, 刘冀伟, 张晓星

(北京科技大学 信息工程学院, 北京 100083)

摘要: 分布式视频编码方案中, 目前常用固定周期的方法选取关键帧. 该方法忽略了视频序列的帧间相关性、运动变化情况. 针对这些缺陷, 研究了基于聚类的自适应关键帧提取算法, 在此基础上, 提出基于互信息量的改进算法. 最后, 针对以上2种算法中的时延问题给出了解决方案. 实验证明, 对于不同的测试序列, 基于互信息量改进算法相比固定选取关键帧算法, 边信息 PSNR 均值有 0.67 ~ 1.4dB 的提高. 此外, 解决时延的算法比改进算法在效率上有很大提高.

关键词: 分布式视频编码; 关键帧; 互信息

中图分类号: TP18; TN919.81 **文献标识码:** A **文章编号:** 1673-4785(2011)06-0539-05

A key frame selection algorithm for distributed video coding

SONG Xiaoli, LIU Jiwei, ZHANG Xiaoxing

(School of Information Engineering, Beijing University of Science and Technology, Beijing 100083, China)

Abstract: In most of the existing distributed video coding schemes, the fixed period method is usually applied for selecting a key frame. This strategy ignores the correlation between video frames and the changes of the motion activity along the video sequence. To avoid these flaws, this paper studied the adaptive key frame selection method based on hierarchical clustering; on this basis, an improved algorithm based on mutual information was proposed. Finally, a solution was given to overcome the delay in the above-mentioned methods. Experimental results show that for various video sequences, a 0.67-1.4dB gain in the quality of side information has been achieved. In addition to being delay-efficient, the key selection algorithm requires the lowest performance time.

Keywords: distributed video coding; key frame; section algorithm; mutual information

传统视频编码方案在编码端隐含一个解码器, 使编码端的运算复杂度是解码端的 5 ~ 10 倍以上^[1]. 这种编码方案适用于编码端复杂的领域. 而近年来, 一些新的移动视频设备如: 移动视频相机、移动视频电话、无线 PC 机等需要低复杂度编码. 此时, 传统的视频编码方案难以胜任, 迫切需要一种编码端复杂度低的编码方案. 在此背景下, 一种新的视频编码框架—分布式视频编码 (distributed video coding, DVC) 应运而生. 该编码方案是基于 20 世纪 70 年代 Slepian 和 Wolf^[2] 的分布式无损编码理论以及 Wyner 和 Ziv^[3] 的使用解码端边信息 (side information) 的有损编码理论而建立. 从 2002 年起 DVC

的实现算法逐渐引起关注, 成为视频编码领域关注较多的前沿课题之一.

Wyner-Ziv 视频编码 (Wyner-Ziv video coding, WZVC) 是 DVC 编码的一种主流框架. 在 WZVC 中, 待编码的帧分为 K 帧和 Wyner-Ziv (WZ) 帧, K 帧是通过传统的视频编码方案进行帧内编解码, 而 WZ 帧则是通过信道编码, 仅传输校验位给解码端. 大多数 WZVC 方案采用周期性选择关键帧的方法, 该方法有许多弊端^[4]: 如果视频序列运动缓慢, 选择过多的关键帧会造成冗余, 不利于压缩比的提高; 如果视频序列运动剧烈, 选择较少的关键帧则难以生成高质量的边信息. 为了解决这些问题, 需要一种有效的选择关键帧的方法, 使得关键帧会随着视频序列的运动情况而灵活变化.

文献[5]提出一种基于聚类算法的自适应关键

收稿日期: 2011-07-11.

基金项目: 国家自然科学基金资助项目 (60903067).

通信作者: 宋晓丽. E-mail: xysxl02@163.com.

帧提取算法;文献[4]利用感兴趣点来描述图像的信息,提出一种基于感兴趣点匹配的关键帧选择算法;文献[6]利用 SURF 算法得到的特征点信息作为对帧间相关性的近似估计,提出一种自适应选取关键帧的方法.上述几种方法从不同角度进行了关键帧提取算法的研究,系统性能比周期性选取关键帧提取算法均有所提高.

本文研究了文献[5]中的算法,并且从另一角度,利用互信息量来描述相邻2帧的相关性,提出基于互信息量的自适应关键帧提取改进算法,从系统性能和计算复杂度上对文献[5]中的算法进行了改进,然后对算法中的时延问题给出了解决方案.

1 Wyner-Ziv 视频编码框架

本文采用变换域分布式视频编码系统^[7],编码框架如图1所示,首先将视频序列分为K帧和WZ帧(每个GOP里的首帧为K帧,其余帧为WZ帧).K帧采用传统帧内编解码方法.WZ帧经过基于 8×8 的块DCT变换后,进行量化,然后提取位平面,从高位到低位依次送入LDPC编码器编码,编码后根据解码端的反馈信息,将要求的校验位送到解码端.

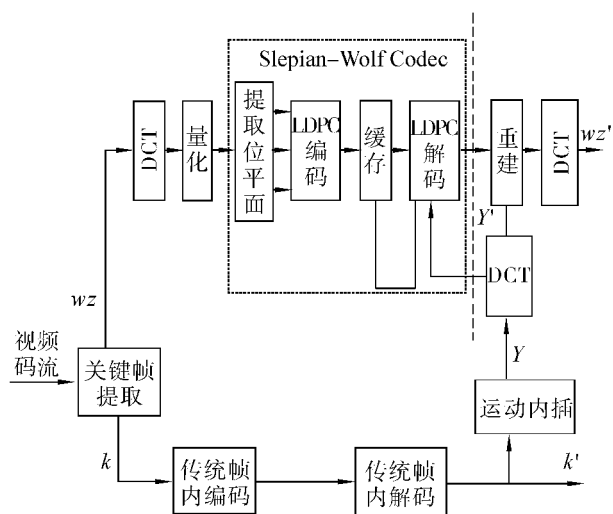


图1 Wyner-Ziv 视频编码框架

Fig.1 Architecture Wyner-Ziv video coding

在 Wyner-Ziv 视频编码系统框架中,解码是最复杂的部分,对于K帧,只需要进行传统的帧内解码就可以得到相应的解码帧;而对于每个WZ帧,解码器都会利用相邻的已经解码的K帧,使用时间域上的插值或者外推的方法形成作为估计WZ帧的辅助信息,对辅助信息做DCT变换,和编码端一样提取位平面,送入LDPC解码器;LDPC解码器,如果不能可靠地解码出符号流,会通过反馈信息从编码器的缓冲区中申请附加的奇偶校验码;利用解码后的

比特流和DCT变换后的边信息重建WZ帧.

2 自适应选取关键帧

2.1 基于聚类的自适应关键帧提取

文中利用以下4个低水平特征指标来评估视频序列的运动剧烈程度^[5]:1)直方图差DH;2)差的直方图HD;3)块的直方图差BHD;4)块的方差的差BVD;

它们的定义如下:

$$DH(i,j) = \frac{1}{D_f} \sum_{k=0}^L |h_i(k) - h_j(k)|, \quad (1)$$

$$HD(i,j) = \frac{1}{D_f} \left(\sum_{k=0}^{1/2-a} h_{i-j}(k) + \sum_{k=1/2+a}^L h_{i-j}(k) \right), \quad (2)$$

$$BHD(i,j) = \sum_{b=0}^{D_f D_B} \sum_{k=0}^L |h_i(b,k) - h_j(b,k)|,$$

$$BVD(i,j) = \sum_{b=0}^{D_f D_B} \sum_{k=0}^L |\sigma_i^2(b,k) - \sigma_j^2(b,k)|.$$

式中: i,j 代表帧的索引, h 代表 L 区间的直方图, D_f 和 D_B 分别代表帧和块的大小, σ 代表方差,对于指标HD, α 代表与原值接近的阈值.前2个指标在帧水平上,检测全部运动的变化,HD指标是非常有效的.BHD和BVD指标对于局部运用更为敏感.

基于上述4个指标的聚类关键帧提取算法描述如下:

- 1) 计算所有帧中相邻帧间的4个指标,建立四维矢量,并进行归一化;
- 2) 累积相邻2帧帧间运动矢量,找到相邻2帧帧间运动矢量规范化式最小值对应的下标值;
- 3) 将该下标值对应的帧与后一帧或后一类进行聚类;
- 4) 重复步骤2)、3),直到相邻2帧间运动积累大于设定的阈值 φ 时,停止聚类.

φ 控制聚类的类数,同时也控制了关键帧帧数(每类里的首帧为关键帧).

2.2 基于互信息量的关键帧提取

2.2.1 互信息量

互信息量源于信息论,不仅是2个随机变量相似性的量度,同时也决定了每一个独立的随机变量在交迭处各自表示的信息量的大小.利用互信息量作为图像的相似性测度是Collignon^[9]于1995年提出的,它在应用上取得了很大成功.

2幅图像的互信息量定义如下:

$$I(X,Y) = \sum_{xy} P_{XY}(x,y) \log P_{XY}(x,y) / P_X(x) P_Y(y). \quad (3)$$

式中: $P_X(x)$ 和 $P_Y(y)$ 分别表示前一帧 X 和当前帧 Y 的概率密度函数,由图像的直方图除以图像总的像素个数得到; $P_{XY}(xy)$ 表示相邻 2 帧的联合概率密,由图像 X 、 Y 的联合直方图除以图像总的像素个数得到. 由定义可知,互信息量并不依赖于灰度值本身,而是信赖于这些灰度值出现的概率. 当互信息量为 0 时,说明视频序列相邻帧发生剧烈运动,该相邻帧相互独立;当互信息量值较大时,意味着相邻帧间的相似性程度较高,所以互信息量能很好地反映视频序列中相邻 2 帧间的相似性程度.

2.2.2 关键帧提取的改进算法

互信息量从全局运动来反映视频相邻 2 帧间的变化,2.1 节中的 DH 和 HD 这 2 个指标也是从全局角度出发来衡量 2 帧间的相似性. 但互信息量表示的是 2 幅图像相互包含对方的信息量,即帧 X 包含帧 Y 的信息量或帧 Y 包含帧 X 的信息量,相比于 DH 和 HD,互信息量更加直接,全面描述了 2 帧的相似程度. 因此在改进算法中用互信息量替换 DH 和 HD 作为衡量视频序列全局变化的指标,同时保留原有的局部指标: BHD 和 BVD.

改进的关键帧提取算法步骤如下:

1) 计算所有视频序列的相邻 2 帧间的 I、BHD 和 BVD,建立三维矢量,并进行归一化;

2)、3)、4) 同 2.1 节中基于聚类的关键帧提取算法中步骤 2)、3)、4)。

由式(1)、(2)和(3)得知,改进算法的计算复杂度明显减小. 首先,互信息量的计算量小于 DH 和 HD 的计算量和;其次,相比于原算法中的基于四维矢量聚类,改进算法中基于三维矢量进行聚类较为简单.

3 解决时延的关键帧提取算法

基于聚类的自适应关键帧提取算法和基于互信息量的关键帧提取算法,在编码端的聚类过程引进了时延,这不满足实际应用的实时性. 因而有必要对上述 2 种方法的时延问题进行研究,本文给出了合理的解决方案.

假设视频序列比较平稳,根据系统已经选择好的 GOP 可以预测后续视频序列的 GOP,减少编码端等待时间,同时选择几个不同 GOP 中率失真性能最优(用每个 GOP 中 WZ 帧的率失真性能代替整个 GOP 的率失真性能)的 GOP 作为编码组,进而保证了选择的关键帧的有效性. 其中速率的估计可以直

接获得^[9], WZ 帧 PSNR 值估计^[10] 如式(6):

$$\tilde{d}_{iWZ} = \begin{cases} L_{Bi}, d_{iSI} < L_{Bi}; \\ U_{Bi}, d_{iSI} > U_{Bi}; \\ d_{iSI}, U_{Bi} < d_{iSI} < L_{Bi}. \end{cases} \quad (4)$$

$$L_{Bi} = (C_{iWZ} - 1/2) \times qs_j,$$

$$U_{Bi} = (C_{iWZ} + 1/2) \times qs_j,$$

$$MSE_{WZ} = \sum_{i=0}^{N-1} (d_{iWZ} - \tilde{d}_{iWZ})^2 / N. \quad (5)$$

式中: \tilde{d}_{iWZ} 为重建 WZ 帧的 DCT 系数, d_{iSI} 为边信息的 DCT 系数, L_{Bi} 、 U_{Bi} 分别表示原始 WZ 帧量化的上、下限, C_{iWZ} 为原始 WZ 帧的量化系数, qs_j 为 DCT 系数第 j 个系数带的量化步长.

解决时延的关键帧提取算法的具体过程如下:

1) 以当前帧为起始帧,在编码端分别计算 GOP 为 2、3、4 时的率失真性能;

2) 选择最优性能的 GOP 作为特定窗口内的第一个 GOP,窗口的初始大小为 1,当且仅当当前获得的 GOP 与先前的 GOP 一样时,窗口尺寸才会增加 ($W = 2 \times Wold, W < 20$),否则 $W = 1$;

3) $Wold = W$;

4) W 为 1 时,使用选择好的 GOP 进行编码,从下一帧算起,回到 1);

5) 当 W 不为 1 时,连续进行 W 个 GOP(从当前 GOP 的前一个 GOP 作为初始值算起)编码,未完成时重复进行 4),否则重新开始测试 GOP,回到 1)。

根据上述分析及算法步骤可知,该算法选择性能最优 GOP 作为编码 GOP,能保证系统的性能,且在满足一定条件时,允许连续编码,可以用当前的 GOP 作为后续窗口内的 GOP,明显减少测试 GOP 所用的时间,因此该算法在保证系统性能的前提下,能有效减少时延.

4 实验结果

本部分首先给出传统周期性关键帧提取算法(periodical key frame selection, PKFS)、基于聚类的自适应关键帧提取算法(clustering adaptive key frame selection, CAKFS)和基于互信息量关键帧提取改进算法(mutual information key frame selection improving algorithm, MIKFS)3 种方案的性能对比,结果如表 1 所示. 实验中,以 Foreman、Coastguard 序列作为测试序列,以边信息和原始 WZ 帧的平均 PSNR 作为评判标准,通过调整阈值使 3 种方案中的关键帧数(key frame number, KFN)相同.

表1 PKFS、CAKFS 与 MIKFS 的性能比较

Table 1 Comparison results of the performance of PKFS, CAKFS and MIKFS dB

算法	KFN 为 25		KFN 为 33		KFN 为 50	
	Foreman	Coastguard	Foreman	Coastguard	Foreman	Coastguard
PKFS	30.7	29.9	31.6	29.4	34.3	31.9
CAKFS	30.1	28.5	31.8	31.1	35.1	34.5
MIKFS	30.7	29.2	32.0	31.4	35.9	34.9

分析表1的数据可知,在测试序列为 Foreman、Coastguard 时,相比于 PKFS 方案,CAKFS 方案中边信息质量分别平均提高了 0.13 dB、0.97 dB. 本文给出的 MIKFS 改进方案中边信息质量相比于 CAKFS 方案提高了 0.53 dB、0.47 dB; 相比于 PKFS 方案提高了 0.67 dB、1.4 dB. 可见说明了本文 MIKFS 方案的优越性.

关于解决时延的关键帧提取算法的实验,用 Foreman 序列的前 100 帧来测试,测试架构是本文第2部分给出的 DVC 框架,1)量化部分采用 JPEG 标准里量化矩阵,量化因子为 $QF = (0.5, 2, 4)$,取不同的量化因子,可以获得不同的输出码率和不同质量的重建帧. 量化因子为 0.5 时,DEKFS 算法获得的 GOP 情况如图2;2)调节 CAKFS 和 MIKFS 中的阈值,使3类算法的关键帧数相等;3)边信息使用传统的运动补偿内插方案来插出. 首先给出以 PSNR 值作为评判准则,DEKFS 算法与 PKFS 算法比较的结果如表2所示.

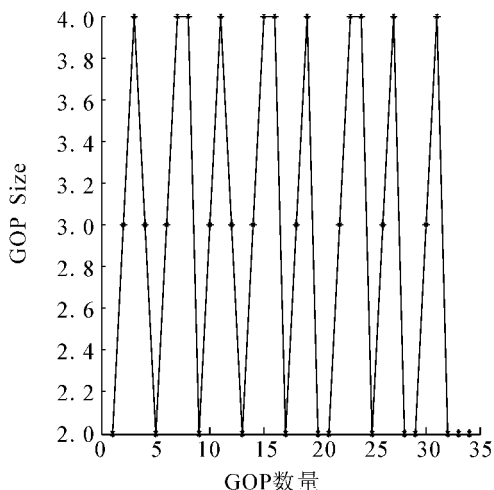


图2 DEKFS 算法的 GOP 情况

Fig. 2 GOP results of DEKFS algorithm

表2 PKFS 与 DEKFS 的性能比较

Table 2 Comparison results of the performance of PKFS and DEKFS dB

序列\方法	KFN 为 33	
	PKFS	DEKFS
Foreman	31.6	31.7

分析表2的数据可知,在测试序列为 100 帧 Foreman 时,相比于 PKFS 方案,DEKFS 方案中边信息质量平均提高了 0.1 dB.

另外给出了 CAKFS 算法、MIKFS 算法与解决时延的关键帧选取算法(delay efficient key frame selection, DEKFS)整体性能比较结果. 分别统计 CAKFS 算法、MIKFS 算法和解决时延的关键帧选取算法中总 WZ 帧的率失真性能,测试结果如图3.

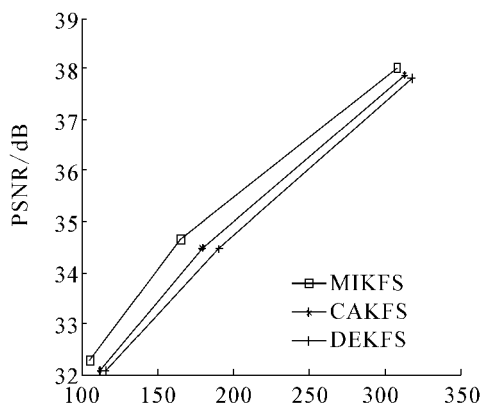


图3 Wyner-Ziv 系统性能对比

Fig. 3 Performance comparison of Wyner-Ziv system

3 类算法的时间统计结果如图4.

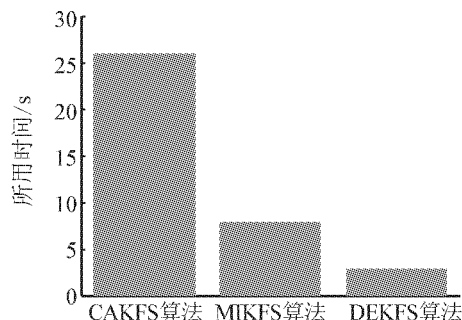


图4 3 种算法的时间对比

Fig. 4 Time comparison of three algorithms

由图3可知,解决时延的关键帧算法与基于聚类的自适应关键帧算法和改进算法相比,率失真性能有略微降低. 主要是因为本文的算法是在假设视频序列比较平稳的前提下进行的,这样可以预测后续视频序列的 GOP,另外,DEKFS 算法限制了每个 GOP 里能取的帧数的范围.

虽然 DEKFS 算法相比于 CAKFS 和 MIKFS, 率失真性能略微降低, 但从图 4 可得, DEKFS 算法相比于 PKFS 在时间对比度上有明显的优越性. CAKFS 和 MIKFS 算法必须把所有的帧先聚类, 才能编码, 这样编码端聚类没有完成时, 编码器闲置起来, 造成时间的浪费. 而 CAKFS 算法在获得一组 GOP 后, 编码器即可开始编码, 编码和后续 GOP 的获得可以同时进行, 允许系统持续编码, 对于平稳的视频序列, 可以通过前面的 GOP 获得即将编码的 GOP 尺寸, 减少计算量, 且更有效地减少了时间. 因此, 从实际角度出发, 能够解决时延的算法更有效.

5 结束语

DVC 中固定选取关键帧的方法对系统性能的提高有一定的局限性. 基于此, 本文在研究基于聚类的关键帧提取算法的基础上, 提出了基于互信息量的改进算法, 改进算法中用互信息量作为反映视频序列的帧间相关性的全局指标, 优化了率失真性能, 同时提高了计算效率. 但以上 2 种算法都有一定的延迟效应, 解决时延的算法在率失真性能上略微降低, 却有效地减少了编码器的等待时间, 更具有实际应用价值. 后续的工作将考虑如何更精确地度量帧间相关性, 优化 DVC 的率失真性能, 以及在 DVC 的率失真性能和关键帧选取算法的复杂度之间找到平衡点.

参考文献:

- [1] NILSSON M. The advanced video coding standard [C]//IT to HD: Visions of Broadcasting in the 21st Century. London, UK, 2004: 85-100.
- [2] SLEPIAN D, WOLF J. Noiseless coding of correlated information sources[J]. IEEE Transactions on Information Theory, 1973, 19(4): 471-480.
- [3] WYNER A, ZIV J. The rate-distortion function for source coding with side information at the decoder [J]. IEEE Transactions on Information Theory, 1976, 22(1): 1-10.
- [4] YANG Feng, DING Guiguang, DAI Qionghai. Adaptive key frame selection Wyner-Ziv video coding [C]//IEEE Multimedia Signal Processing. Shanghai, China, 2005: 1-4.
- [5] ASCENSO J, BRITES C, PRERIRA F. Content adaptive Wyner-Ziv video coding driven by motion activity [C]//IEEE International Conference on Image Processing. Atlanta, USA, 2006: 605-609.
- [6] 张晓星, 刘冀伟, 张波. 分布视频编码中基于帧间相关性的自适应关键帧选取算法[J]. 光电子·激光, 2010, 21(10): 1536-1541.
ZHANG Xiaoxing, LIU Jiwei, ZHANG Bo. Adaptive key frame selection algorithm based on interframe correlation in distributed video coding [J]. Journal of Optoelectronics · Laser, 2010, 21(10): 1536-1541.
- [7] AARON A, RANE S, SETTON E, et al. Transforms-domain Wyner-Ziv codec for video [C]//Visual Communications and Image Processing. San Jose, USA, 2004: 520-528.
- [8] MAES F, COLLIGNON A, VANDERMEULEN D. Multi-modality image registration by maximization of mutual information [C]//Mathematical Methods in Biomedical Image Analysis. San Francisco, USA, 1996: 14-22.
- [9] YAACOUB C, FARAH J, PESQUET-POPESCU B. Optimal rate allocation in multi-user Wyner-Ziv video coding systems with coded key frames [C]//Personal, Indoor and Mobile Radio Communications. Cannes, France, 2008: 1-5.
- [10] AHMAD I, AHMAD Z, ABOU-FAYCAL I. Delay-efficient GOP size control algorithm in Wyner-Ziv video coding [C]//Signal Processing and Information Technology. Ajman, United Arab Emirates, 2009: 403-408.

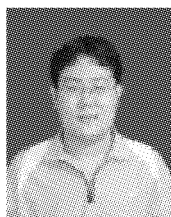
作者简介:



宋晓丽, 女, 1986 年生, 硕士研究生, 主要研究方向为图像处理、视频压缩等.



刘冀伟, 男, 1962 年生, 副教授, 中国人工智能学会、人工心理与情感计算专业委员会理事, 主要研究方向为图像处理、视频压缩、人工智能系统等, 近年来在国内外著名期刊与会议上发表学术论文 40 余篇, 出版著作多部.



张晓星, 男, 1984 年生, 博士研究生, 主要研究方向为图像处理、图像、视频编码等.