

基于多变换的细粒度视频编码算法

蒲菊华 熊璋 韩军

(北京航空航天大学计算机学院 601 教研室, 北京 100083)

摘要 针对现有细粒度视频编码算法计算复杂度大或视频恢复质量有各种效应的缺点,提出了一种基于联合小波变换和MP变换的细粒度编码算法。该算法在运动估计与补偿的基础上,用小波变换来消除帧间冗余,然后对变换结果根据不同帧的数据特征分别进行二维小波变换或MP变换。算法还提出了新的运动估计和像素调整策略、基于人眼视觉特性的MP原子分配策略和基于能量查找的原子搜索机制。实验表明,该算法可同时兼顾视频恢复质量、计算复杂度和控制粒度。

关键词 视频 细粒度 小波 匹配追踪 运动补偿

中图分类号: TN919.81 **文献标识码**: A **文章编号**: 1006-8961(2004)11-1342-06

A Fine Granularity Video Coding Algorithm Based on Multi-transformation

PU Ju-hua, XIONG Zhang, Han Jun

(Beijing University of Aeronautics and Astronautics, School of Computer Science and Engineer, lab 601, Beijing 100083)

Abstract With the widespread development of video applications, it is very urgent to develop fine granular scalable video compression algorithms, among which discrete wavelet transformation and matching pursuit (MP) transformation are the most popular. However, the complexity of MP video coding and the ringing and rippling of discrete wavelet transform video coding are annoying obstacles to many video applications. In this context, an algorithm based on both discrete wavelet and MP transform is presented in this paper for fine granularity video coding. Firstly, using every eight frames as a unit, this algorithm applies motion prediction and eliminates the inter frame redundancy by 1-dimensional wavelet transformation at motion vector direction. Thus, 1 low-frequency and 7 high-frequency frames are obtained. The transformed low-frequency frame is further processed by 2-dimensional wavelet transformation to eliminate the intra frame redundancy, with the other 7 high-frequency frames coded by Matching Pursuit. A new motion prediction and pixels regulation strategy is also presented in this paper. The MP atom assignment is based on human eye's vision characteristics and the energy of residual after motion estimation. The experiments on its performance and the analysis of computing complexity indicate that this algorithm balances the recovery video quality, computation complexity, and control granularity.

Keywords video, fine granularity, wavelet, matching pursuit (MP), motion compensation

1 引言

随着多媒体技术和网络技术的迅速发展,因特网上的视频应用有着越来越广泛的前景,同时也面临着诸多挑战。视频(尤其是实时视频)对带宽、延迟和丢失率等都有一定要求,由于现有的因特网提供的是一种“尽力而为”的服务,因而不能为视频应用

提供服务质量保证,并且由于网络和接收端的异构性,使得很难充分利用现有网络带宽和资源、以实现服务的灵活性。欲解决视频在现有“尽力而为”因特网上应用传输所面临的这些问题,公认的经济有效的方式是应用层的服务质量控制策略^[1],其中尤以可伸缩性视频编码较突出。由于不同的可伸缩性视频编码算法提供的码流可控粒度是不同的,其解码恢复的视频质量也有很大区别,因此具有不同的适

用背景和应用前景^[2]。由此可见,为了保证一定网络条件下传输恢复的视频质量尽可能高,需要进一步研究如何提供细粒度的视频编码方式。

目前,研究人员在视频细粒度编码方面做了很多工作,如位面编码、小波变换编码和用匹配追踪(matching pursuit, MP)变换实现的视频压缩编码等等,这些方法虽已在视频码流粒度方面获得一定成功,但在提高视频质量或者降低复杂度方面仍存在很多需要不断完善的地方^[2~4]。

针对上述研究背景,本文提出了一种联合的细粒度视频编码算法。该算法充分利用了现有小波变换和 MP 变换理论的特点,并将其分别应用到帧间或帧内不同编码场合,以折衷考虑编码效率和视频恢复质量。

2 细粒度视频编码相关的研究工作

目前的细粒度视频编码公认的有位面编码、小波变换编码和 MP 变换编码 3 种方式。由于位面编码一般仅仅是对传统离散余弦变换(discrete cosine transformation, DCT)的数据采用按位编码,所以 DCT 变换所带来的弊端都不能避免。为此国内外很多研究人员都比较关注小波变换和 MP 变换编码。

小波视频编码的研究重点在于小波视频编码算法、小波基的选取和小波变换系数的有效编码方法等,如文献[5,6]就分别提出了基于小波变换的多适应性算法和三维小波视频编码算法。这些算法虽然弥补了 DCT 变换的诸多缺点,而且其频率划分可满足人眼的视觉特性,但是用小波变换来消除视频帧内冗余,不能避免其所引起的 ringing 效应和 rippling 效应对视频质量的影响(相关实验见文献[3])。

针对小波变换的不足,MP 变换越来越得到人们广泛关注^[7,8]。大家知道,MP 变换是一种信号变换方法,其是用某定义了内积的信息空间中的一组超完备的信号簇(由若干信号函数构成,通常称为字典集合)来表示该空间中的其他任何信号^[7]。如果用一维信号来解释,则 MP 变换的过程为:对信号 $f(t)$,首先从字典集合 D 中选择一个函数 $g_{r_0}(t)$,使得下式成立:

$$\forall g(t) \in D, \|f(t) - \langle f(t), g(t) \rangle g(t)\| \geq \|R_1(t)\| \quad (1)$$

其中, $R_1(t) = f(t) - \langle f(t), g_{r_0}(t) \rangle g_{r_0}(t)$,表示信号 $f(t)$ 经过一次变换以后的差信号。式(1)表明, $g_{r_0}(t)$ 的选择应该使得残差信号最小,这样就完成了对

$f(t)$ 的一次分解,接着对差信号 $R_1(t)$ 进行分解,然后循环上述操作,直到所得的差信号 $R_n(t)$ 小于某个给定的阈值为止。若定义 $R_0(t) = f(t)$,则信号 $f(t)$ 就可以用如下序列来近似表示: $\{\langle R_i(t), g_{r_i}(t) \rangle, i = 0, 1, \dots, n\}$ 。

由于这种视频编码不但可以克服小波变换的 ringing 和 rippling 效应或者传统混合编码的方块效应,而且能保证视频的最重要特征优先编码,而相对不重要的特征稍后编码,并可以通过有效的选择字典集合来控制哪些特征需要更有效的编码,因此,它在同等码率下使视频的失真少,但其面临一个主要问题就是内积计算的迭代运算复杂度太大,而且算法的实现复杂度与基本信号的选取有很大关系。Neff 和 Zakhor 就提出了需要对基本信号近似、内积搜索计算采用快速算法的方法^[9]。

上述分析表明,由于单纯用小波或 MP 变换进行视频编码,都不能很好地兼顾编码效率、视频的恢复质量和算法复杂度,因此,文献[3]提出了一种结合小波变换和 MP 变换的细粒度视频编码算法,该算法虽然充分考虑了两种编码算法的特点,取得一定成功,但是,由于其时间维小波变换仅仅利用了帧间像素的线性相关性,其对于运动较剧烈的序列并不太理想,因此,本文将运动估计引入到帧间变换中,并改进了文献[3]中算法对 MP 原子的搜索和编码策略,这样就进一步提高了算法性能。

3 多变换细粒度视频编码算法的基本思路

该算法首先将输入的连续 8 帧视频进行缓存,并对缓存的视频帧进行运动估计与补偿;然后在时间维对这 8 帧进行运动矢量方向的三级一维的小波变换,以消除视频帧间的冗余信息。对于小波变换后的图像,第 1 帧代表了时间维的低频信息,其与原始的第 1 帧图像最相近;而后面的几帧则代表了不同分辨率下的时间维高频信息。对这些变换后的视频帧,第 1 帧低频图像的数据量较大,应进一步采用二维的小波变换,而对于后续的代表高频变换信息的视频帧,则采用 MP 变换编码的方式。经过上述变换以后,再通过将小波系数、MP 原子码流和运动矢量信息进行熵编码来形成输出码流,当接收端收到这些码流以后,再采用相应的反变换就可以解码恢复视频。该算法的编码框架原理如图 1 所示。

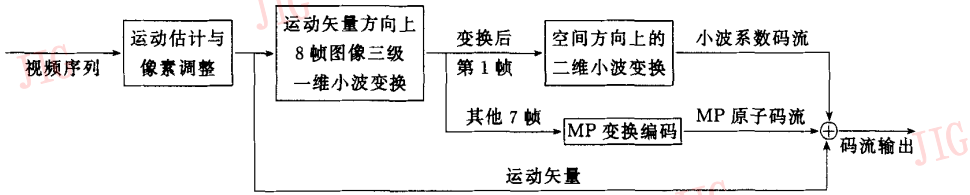


图 1 编码器流程框图

4 算法关键技术

(1) 运动估计及像素块调整

文献[3]在进行时间维小波变换时,是将视频帧的对应位置的值看成一个一维信号,而直接对这样的一维信号进行变换。由于该方法实质上是利用了帧与帧之间相同位置像素的线性相关性来消除帧间冗余,因此其有很大局限,即对那些帧间的空间线性相关性较弱、存在运动相关性的情况,这种方式就很难消除帧间的冗余信息。针对这些问题,本文算法采用了基于运动估计的时间维小波变换。

本算法按从左到右、从上到下的顺序对视频帧的宏块进行编号。对于要编码的视频帧,采用二维对数搜索法进行运动估计寻找各块的运动矢量,该运动估计以 SAD(Sum of Absolute Difference)作为匹配准则函数^[5]。为了便于后续的像素调整,在运动估计过程中,记录下各块与其匹配块的 SAD 值,并按宏块编号顺序形成一个 SAD 序列。

经运动估计后,下一步的工作就是要进行运动矢量方向上的小波变换,即,如果将第 1 帧的各个块和后续各帧中处于其运动矢量方向的块组成一个序列,则应对这个序列上各块的像素进行一维小波变换。由于经过运动估计以后,相邻两帧的块之间并不一定是——对应的,因此,为了保证每一帧的各个 16×16 的块都处于某个序列中,且只处于一个序列中,就需要进行像素块的调整。像素块的调整采用如下原则:

- ① 对于——对应的区域不予调整,并将两帧中的这些块都标记为“映射区域”;
- ② 对于 B 帧中多个像素块对应 A 帧中同一像素块的块,按块编号顺序扫描,并根据 SAD 值最小原则进行匹配。如果图 2 中 B 帧的第 4、6 块都映射到 A 帧的第 5 块,则根据 SAD 值最小原则调整后,仅将 B 帧的第 6 块对应 A 帧的第 5 块,并标记为“映射区域”,而将 B 帧的第 4 块标记为“非映射区域”;
- ③ 因 A、B 两帧像素块数目相等,故经上述规则

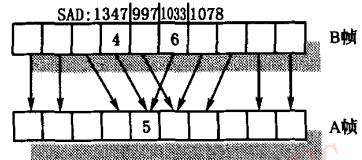


图 2 运动矢量和帧间块的对应

处理后,A 帧的未标记为“映射区域”的块必然和 B 帧的“非映射区域”数目相等。为此,按从左到右、从上到下的顺序扫描视频帧,然后将这些区域强制进行一一映射。采用该原则映射的 B 帧的块,其运动矢量需要调整为相对于修改后的映射块的矢量。如图 3 所示,调整以后,B 帧的第 4 块对应 A 帧的第 3 块,相应的运动矢量也要修改为相对于块 3 的矢量。

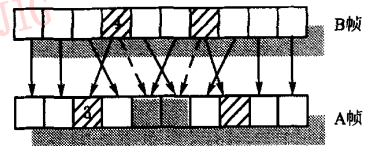


图 3 强制映射未一一对应的区域

经过上述调整以后,相邻 8 帧的像素块就在运动矢量方向上构成了一个长度为 8 的序列,然后对这些序列进行 3 级的一维小波变换,以消除视频帧间的冗余。

(2) 小波滤波器的选取

本文算法包含以下两种小波变换:一种是时间维 8 帧图像在运动矢量方向上的一维变换;另一种是每 8 帧中第 1 帧低频变换图像的空间二维变换。这两种变换有很大区别,由于其中的时间一维变换的数据序列短(只有 8),因此时间轴小波分解要求滤波器阶数不能太高,本文算法实验采用了如下的滤波器^[6]:

$$h_n = \{0.6830127, 1.1830127, 0.3169872, -0.1830127\}$$

$$g_n = \{0.1830127, 0.3169872, -1.1830127, 0.6830127\}$$

二维空间变换,本文采用 Daubechies 的 9/7 小波函数^[5],并采用行列分离变换的方式来实现。

(3) 信号延拓与处理

由于视频图像数据的长度有限,因此进行小波分解时,一般都需对信号的边界进行延拓,本文算法对时间维的一维小波变换和低频帧的二维小波变换采用了不同的延拓方式。其中,时间一维变换采用了周期延拓,而低频帧的二维小波变换则采用了对称延拓,以保持边界的相对光滑。

(4) MP 变换函数的选取与原子分配

本文算法的 MP 变换所采用的字典集合是基于可分离 Gabor 函数的,即按照某种准则查找需要编码的残差信号,并以如下五元组的形式表示: $\{X, Y, x, y, p\}$, 其中 X, Y 表示与当前原子对应的字典集合中的映射元素; x, y 表示其所对应的高频图像中的匹配块的中心坐标; p 为内积映射值,即 (x, y) 处的图像块在字典集合中的映射元素 (X, Y) 上的投影。经过变换编码后,高频帧信号就可以用上述定义的若干原子来表示,而 MP 原子参数采用算术编码。

采用这种 MP 变换方式编码的视频图像,其质量的好坏与原子的个数及其在视频帧中的相对位置(即原子的编码顺序)密切相关。由于现有的 MP 视频编码算法大都仅仅重视了原子编码位置的合理性^[8]或个数的合理设计^[3],并没有兼顾二者,因此本文算法将从这方面进行改进。

本文算法采用了基于能量的机制来搜索需编码的高频信号,即先查找视频帧中能量最大的块,再从该块开始进行 MP 变换编码,然后继续搜索视频帧中未编码,且能量最大的块,并对其进行 MP 变换编码,如此循环操作,直到未编码块的能量都小于某个最小阈值或已经编码的 MP 原子个数达到所设定的原子个数为止。可见,由于该方法总是编码能量最大的信号,因此是同等数目原子情况下对高频帧信号的最佳表示。

本文算法对每帧 MP 原子个数,采用了动态设计编码原子个数和上述能量阈值的方式,且不同高频帧的 MP 原子个数和能量阈值还借助人眼的视觉特性和小波变换后的数据分布特点来确定^[6]。实验中,高频帧编码原子个数满足如下关系:对于第 1 个高频帧,编码原子为 N_1 个,第 2 和第 3 两个高频帧的编码原子个数满足 $N_{2,3} = \lceil \alpha_1 \times N_1 \rceil$,余下的 4 帧高频帧编码 MP 原子个数为 $N_{4..7} = \lceil \alpha_2 \times N_1 \rceil$,其中 $0 < \alpha_2 < \alpha_1 < 1$,能量阈值也采用了类似比例规则。为了保证算法的灵活性,在实现过程中, N_1, α_1 和 α_2 都可以动态调整。

(5) 低频帧小波变换系数编码

对时间维小波变换以后的高频数据,可采用上面

设计的原子个数进行 MP 变换,并对 MP 原子参数采用算术编码;而其中的低频数据经过二维小波变换以后,则利用基于空间方向的等级树方法(SPIHT)来编码^[9]。另外,运动矢量也采用算术编码。

5 算法性能评价

本文算法在时间维充分利用了小波变换的诸多优点来消除了帧间的冗余,同时为了减少小波变换可能引入的 ringing 效应和 rippling 效应,高频帧的帧内编码采用 MP 变换方式来实现,这不但使得编码压缩率较高,而且其编码的视频有较高的恢复质量。由于算法采用了基于运动估计的小波变换,因此适合于多种应用场合。本文还对算法的时间复杂度、视频恢复质量进行了分析和实验,并和传统 DCT 变换编码、小波变换编码和文献[4]中的算法进行了比较。下面是相关实验及结果分析。

(1) 算法时间复杂度分析

由于 DCT 变换和小波变换在时间复杂度方面远低于 MP 变换,因此如果对同一帧图像进行编码,则前者所需的时间是后者所需时间的 β 倍, $\beta \ll 1$ 。此外,由于 MP 变换编码的时间主要依赖于原子数,因此可以用 MP 原子数量来估算编码所需时间。下面分析编码 N 帧视频序列,不同算法的时间复杂度。

文献[8]的算法只对视频序列的第 1 帧进行 DCT 变换,其余都进行 MP 变换,其时间复杂度为 $T_1 = N_1 \times (N-1) + \beta \times N_1$ 。

本文算法对视频序列的每 8 帧中的一帧进行小波变换,其余 7 帧进行不同原子个数的 MP 变换,其时间复杂度为 $T_2 = \frac{1+2\alpha_1+4\alpha_2}{8} N_1 \times N + \frac{1}{8} \beta \times N \times N_1$ 。

一般,若 $T_2 < T_1$,则

$$\frac{T_1 - T_2}{T_1} = \left[N_1 \times (N-1) + \beta \times N_1 - \left(\frac{1+2\alpha_1+4\alpha_2}{8} N_1 \times N + \frac{1}{8} \beta \times N \times N_1 \right) \right] / [N_1 \times (N-1) + \beta \times N_1]$$

可用下式求其极限:

$$\lim_{N \rightarrow \infty} \frac{T_1 - T_2}{T_1} = 1 - \frac{1 + \beta + 2\alpha_1 + 4\alpha_2}{8}$$

因为 $0 < \alpha_2 < \alpha_1 < 1$ 以及 $\beta \ll 1$, 一般有 $\beta + 2\alpha_1 + 4\alpha_2 < 6$, 所以

$$\lim_{N \rightarrow \infty} \frac{T_1 - T_2}{T_1} > \frac{1}{8}$$

可见,本文算法所编码的 MP 原子较少,即其时间复

杂度较低。

(2) 算法图像恢复质量评价

为了解本文算法的效果,对几种算法的视频恢复质量进行了对比实验。测试序列包括:Mother、Foreman Missbm 和 Coast 等标准视频测试序列,由于主要测试低码速率下各算法的编解码性能,因此码流速率控制在 30Kbps 左右,输入序列都采用了 176×144 的 QCIF 格式的视频序列。实验中, N_1 、 α_1 和 α_2 分别为 120、0.8 和 0.65,基本能量阈值为 120。H. 263 的码率控制则通过调节量化器参数和 I 帧间隔来实现^[10]。表 1 是实验过程中采用的参数或实际测量值,其中 PSNR 表示亮度信号的峰值信噪比,反映了视频恢复质量。实验表明,与文献[10]算法相比,本文算法有较高的视频恢复质量。

表 1 算法质量实验结果对比

序列名称	平均每帧 MP 原子数	本文算法 PSNR(dB)	H. 263 PSNR(dB)	PSNR 差量(dB) 本文算法-H. 263
Forman	86	29.774	29.543	+0.231
Mother	61	35.691	35.449	+0.242
Coast	87	27.572	27.291	+0.281
Missbm	56	38.977	38.796	+0.181
Hall	70	33.527	33.359	+0.168

(3) 与传统编码算法的比较

为验证本文算法效果,对本文算法和传统 DCT 变换编码算法^[10]在低码速率下的视频恢复质量进行了对比实验,码速率控制在 30Kbps。图 4 从左到右依次是 Carphone 序列的第 48 帧到 51 帧,即一次时间维小波变换的前 4 帧图像用不同算法恢复的结果对比。



(a) 原始视频图像



(b) 本文算法恢复结果



(c) 传统 DCT 变换恢复结果(存在方块效应)



(d) 存在 ringing 和 rippling 效应的小波变换恢复结果

图 4 不同算法恢复视频质量对比

其中,传统 DCT 变换的量化参数是 18,本文算法中,3 帧高频图像共编码 329 个 MP 原子。图 4(d) 仅是示意存在 ringing 和 rippling 效应的小波变换恢复帧,由图 4 对比可见,本文算法缓解了传统编码方式各种效应的影响。

(4) 剧烈运动视频序列的质量评价

对剧烈运动视频序列用本文算法恢复的视频质量进行了测试,并和文献[3]中算法的性能进行了比较。由于 Mobile 序列是公认的最复杂、最不易压缩的序列,故本文选用该序列作为测试序列。该实验的输入序列是 QCIF 格式,码速率控制在 45Kbps 左右,两种算法编码的 MP 原子都是平均每帧 123 个。实验测试了 Mobile 序列的前 104 帧,其对比结果如图 5 所示。无运动估计算法恢复图像,其平均 PSNR 为 19.003dB,而本文算法的平均 PSNR 为 21.298dB,可见,本文算法的视频恢复质量明显高于前者。

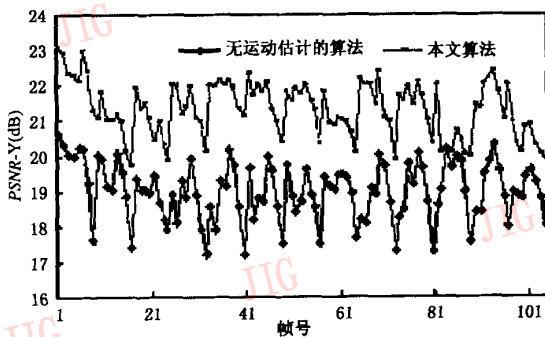


图 5 Mobile 序列的测试结果对比

6 结论及进一步的工作

文章首先介绍了目前细粒度视频编码的相关技术及存在的问题,并在此基础上提出了一种基于小波变换和 MP 变换的带运动估计的细粒度视频编码算法。实验证明,该算法能获得较好的视频恢复质量和具有较好性能。

下一步的工作是在保证一定视频恢复质量的前提下,研究如何进一步降低算法的时间复杂度以及如何更有效地编码所有小波低频帧,并将该算法应用到具体的视频编码标准中。

参考文献

- 蒲菊华,熊璋,史肖燕. 带 QoP 反馈的实时视频服务质量控制框架[J]. 北京航空航天大学学报,2004,30(2): 164~167.
- Li Weiping. Overview of fine granularity scalability in MPEG-4 video standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(2): 301~317.
- 蒲菊华,熊璋,方义. 一种基于小波变换和 MP 变换的细粒度视频编码算法[J]. 计算机工程与设计,2004,25(1): 13~16.
- 胡栋,郑宝玉. 数字视频关键技术的若干新进展[J]. 通信学报, 2003, 24(7): 93~106
- 宋京民. 因特网广播和视频的多适应性编码研究[D]. 北京:北京航空航天大学,2001.
- 沈兰荪. 图像编码与异步传输[M]. 北京:人民邮电出版社, 1998,123~148.
- Mallat Stephane G, Zhang Zhifeng. Matching pursuits with time-frequency dictionaries[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3397~3415.
- Neff Ralph, Zakhor Avidesh. Very low bit-rate video coding based on matching pursuits[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1997, 7(1): 158~171.
- 杨春玲,余英林. 改进的小波域三维等级树集分割视频编码方法[J]. 通信学报,2001, 22(5): 86~92.
- 廖清华. 基于 H. 263+ 的视频编码及实时传输技术的研究与实现[D]. 北京:北京航空航天大学,2002.

蒲菊华 1976 年生,1999 年在北京交通大学计算机与信息科学学院获工学学士学位,现为北京航空航天大学计算机学院博士研究生。主要研究方向为视频编码、传输与服务质量控制以及 H. 264 和 AVS 的优化。

E-mail:scrappy@163.com



熊璋 1956 年生,教授,博士生导师。曾获国家科技进步一等奖、航空部科技进步一等奖各一项。近年发表了学术论文 50 余篇。目前主要研究方向为多媒体技术和分布式信息处理。开发的项目有全国人大、全国政协、香港立法局、澳门立法局电子表决系统。



韩军 1965 年生,高级工程师,2002 年在北京航空航天大学计算机学院获得工学博士学位,曾获得航空部科技进步一等奖一项。目前主要研究方向为视频分割与跟踪等处理技术和视频传输。

