适合于小波视频编码的尺寸自适应图像组结构

赵凡，张娜，任斐斐，刘贵忠
（西安交通大学电子与信息工程学院，710049，西安）

摘要：为了减小采用固定大小的图像组(GOP)结构对存在快速或者复杂运动的视频序列进行运动补偿时域滤波(MCTF)时造成的编码性能损失，提出了一种基于运动模式的适合于小波视频编码的尺寸自适应GOP结构(MDA-GOP)．其主要思想是：利用均匀抽取的4个运动向量场的方向直方图来决定一个初设尺寸为16帧的临时GOP的运动模式，再根据得到的运动模式进一步决定最终的GOP结构，进而在选定的GOP结构上进行相应的MCTF．仿真结果表明：与固定大小的GOP结构相比，采用MDA-GOP结构的峰值信噪比提高了0.05～0.50 dB；与均方误差性能自适应决定GOP结构的方法相比，可以节省至少40%的编码时间．

关键词：视频编码；运动补偿时域滤波；图像组结构；方向直方图
中图分类号：TN919.8 **文献标志码**：A **文章编号**：0253-987X(2009)04-0052-05

An Adaptive Group of Pictures Structure for Wavelet-Based Video Coding

ZHAO Fan，ZHANG Na，REN Feifei，LIU Guizhong
(School of Electronics and Information Engineering，Xi′an Jiaotong University，Xi′an 710049，China)

Abstract：A motion mode driven adaptive group of pictures (MDA-GOP) is proposed to reduce the coding performance loss that is caused by motion compensation temporal filtering (MCTF) using fixed group of pictures (GOP) structure when the video sequence contains fast motion or complex motion. The MDA-GOP adapts to wavelet-based video coding. The main idea is as follows: every four motion fields are extracted to decide the motion mode of the GOP with initial size 16; The final GOP size is then determined by the motion mode; Finally, MCTF and coding are performed in the final GOP structure. Experimental results compared show that the proposed MDA-GOP structure achieves 0.05-0.50 dB PSNR gain in comparison with the fixed size GOP structure, and that the coding time is saved at least 40% compared with the adaptive GOP structure that is decided by mean square error.

Keywords：video coding; motion compensation temporal filtering; group of pictures structure; direction histogram

目前，在基于小波变换的视频编码器算法中，引入运动补偿时域滤波(MCTF)[1-5]方法是为了消除帧间冗余．在进行MCTF时通常都采用固定尺寸(8帧或者16帧)的图像组(GOP)结构，对于那些运动平缓的序列，采用较长的GOP(比如16帧)会获得较好的编码性能，而对于存在快速或者复杂运动的序列，编码性能可能会因选择较长的GOP而有所下降．文献[6]提出了一种用于可扩展编码的自适应GOP结构，根据各个结构MCTF后的均方误差(MSE)性能来决定GOP的最终尺寸，虽然取得了理想的效果，但在选择模式时很大程度上增加了计算复杂度．

收稿日期：2008-06-03． 作者简介：赵凡，(1969—)，女，博士生；刘贵忠(联系人)，男，教授，博士生导师． 基金项目：国家重点基础研究发展规划资助项目(2007CB1002)；国家自然科学基金资助项目(60572045)；教育部高等学校博士学科点专项科研基金资助项目(20050698033)．

本文提出了一种由运动模式自适应决定 GOP 尺寸的方法. 利用事先均匀抽取的 4 个运动向量场的方向直方图[7-8]来决定一个初设尺寸为 16 帧的临时 GOP 的运动模式，再根据得到的运动模式进一步决定 GOP 的最终尺寸，最后在新的 GOP 上进行 MCTF 和编码. 仿真结果表明，与固定大小的 GOP 结构相比，采用基于运动模式的尺寸自适应的 GOP 结构(MDA-GOP)的峰值信噪比提高了 0.05～0.50 dB，与均方误差性能自适应决定 GOP 结构的方法相比，可以节省至少 40%的编码时间.

1　运动模式决定的尺寸自适应的 GOP 结构

1.1　编码结构

图 1 显示了编码器的各个结构模块. 根据临时均匀抽取的 4 个运动向量场建立运动场方向直方图，再根据直方图表示的运动特性决定 GOP 的最终尺寸，进而在选定的 GOP 单元上进行 5/3 小波的 MCTF 和空域的 9/7 小波变换，最后对运动矢量、运动模式和变换后的三维小波系数分别进行熵编码，打包形成最后的整体位流.

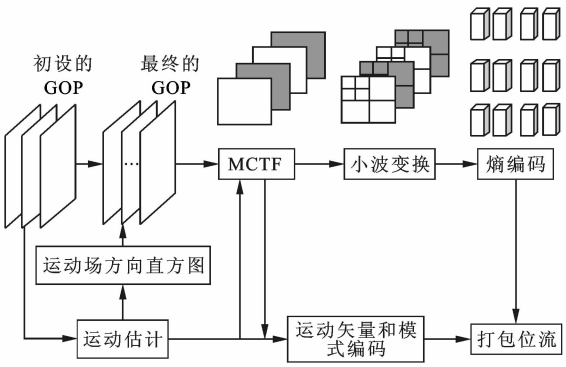


图 1　编码器的结构模块

1.2　运动向量的方向直方图

运动向量的方向直方图表示图像在各个运动方向上的宏块数，它反映了运动的特性，直方条越高表示该运动方向上的宏块数越多，即图像在此方向上的活动程度越强. 本文利用直方图来选择初设的 GOP 结构的运动模式，从而最终确定 GOP 的尺寸.

运动向量的方向直方图的建立步骤如下：

(1)用($\boldsymbol{x}$, $\boldsymbol{y}$)表示运动矢量对，用 $\theta=\tan^{-1}(\boldsymbol{y}/\boldsymbol{x})$表示运动矢量的角度或者方向，取值范围为 0～$2\pi$；

(2)把所有的角度量化到 8 个方向，分别为 $\pi/4$ 的不同倍数，在图中用直方图阶数 2～10 来表示；

(3)每个阶数的量值表示对应方向上的宏块数，阶数 1 表示 1 帧内静止的宏块数，也就是$|(\boldsymbol{x},\boldsymbol{y})|\leqslant 1$ 的宏块.

运动向量的方向直方图只能测试 1 帧图像的运动特性. 要判定 GOP 的运动特性，就需要建立方向直方图组. 本文就是利用初设的 GOP 单元均匀抽取的 4 个运动向量场，分别画出它们的方向直方图，组成运动直方图组，从而判定 GOP 的模式.

1.3　GOP 的模式决定

本文通过以下对运动矢量的方向直方图的峰值位置和活动程度的定义，说明 GOP 运动模式的决定原则.

对每一个方向直方图 j，它的峰值位置定义为 ρ_j，宏块数定义为 h_j，活动程度定义为

$$\gamma_j=\sum_{i=2}^{9}h_i/\sum_{i=1}^{9}h_i \quad j=1,2,3,4 \qquad (1)$$

GOP 运动模式决定算法如图 2 所示，也就是：

(1)如果 4 个直方图的峰值距离 $d(\rho_j,\rho_{j+1})$都小于等于 1，并且活动程度不小于 0.75，这时认为存在 1 个快速的全局活动，模式被设定为 1；

(2)如果 4 个直方图的峰值距离彼此也小于等于 1，但活动程度小于 0.75，这时认为虽然存在一致性的运动，但活动程度不剧烈，模式被设定为 0；

(3)如果任意 2 个直方图的峰值距离都大于 1 且活动程度不小于 0.35，这时认为存在剧烈的局部运动，模式也被设定为 1，否则模式被设定为 0.

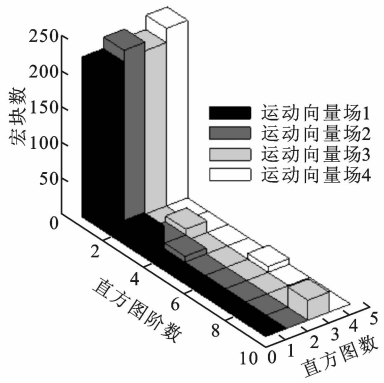
```
输入:j, h_j, γ_j, ρ_j
输出:运动模式M
if d(ρ_j, ρ_{j+1})=|ρ_j−ρ_{j+1}|≤1 ∀j=1,2,3
  if γ_j≥0.75 ∀j=1,2,3,4
    M=1
  else
    M=0
else
  if γ_j≥0.35 ∀j=1,2,3,4
    M=1
  else
    M=0
```

图 2　GOP 运动模式决定算法

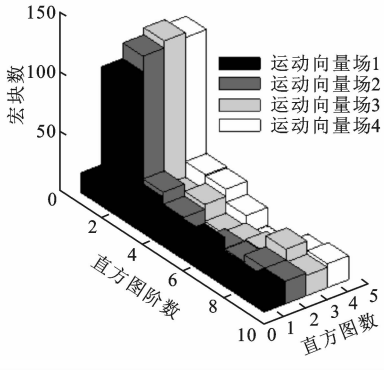
模式 1 被认为存在快速或者复杂运动，而模式 0 被认为运动相对缓慢或均匀. 峰值位置相差可以为 1 是为了允许像机有一个轻微的移动或偏向. 这里，运动活动程度参数 γ_j($j=1,2,3,4$)是活动的宏块数占整个宏块数的比例，取值范围为 0～1. r_j 值越大，表明活动程度越强，值越小，表明活动程度

越弱. 根据对大量的具有不同运动特性的视频序列进行测试发现，运动活动程度参数取值在 0.75～1.0 之间，而且直方图峰值位置相差在 1 阶之内，就可以表明像机运动的一致性. 活动程度参数取值在 0.35～1.0 之间，直方图峰值位置相差在 1 阶之内，就可以表明局部运动的复杂性，所以把全局运动的运动活动程度参数的下界设定为 0.75，把复杂的局部运动的活动程度参数的下界设定为 0.35.

从 Foreman 序列 0～15 帧等间隔抽取 4 场运动向量，从 Football 序列 112～127 帧等间隔抽取 4 场运动向量，分别建立如图 3a、图 3b 所示的方向直方图组. 由图 3 可以看到：Foreman 序列的 4 组直方图的峰值位置都在直方图的阶数 1 上，表示大部分宏块都是静止的；Football 序列的 4 组直方图的峰值位置都在直方图的阶数 2 上，表示该序列组的主要运动是一个全局的左移运动，而且其活动程度较 Foreman 序列要强得多. 从这 2 组的方向直方图也可以看到，运动不仅在某一方向上占绝对的主导地位，而且运动的强度也得到了不同的体现. 对于 Foreman 序列 0～15 帧，计算结果为：$d(\rho_j, \rho_{j+1})=0, \forall j=1, 2, 3$；$\gamma_1=0.125$，$\gamma_2=0.063$，$\gamma_3=0.375$，$\gamma_4=0.03$，$M=0$. 对于 Football 序列 112～127 帧计算结果为：$d(\rho_j, \rho_{j+1})=0, \forall j=1, 2, 3$；$\gamma_1=0.92$，$\gamma_2=0.95$，$\gamma_3=0.94$，$\gamma_4=0.91$，$M=1$.



(a)Foreman 序列



(b)Football 序列

图 3 运动向量场的方向直方图组

运算结果进一步验证了算法的正确性.

2 GOP 结构的重新划分

针对初设的尺寸为 16 帧的 GOP，利用运动估计得到均匀间隔的 4 个运动向量场，在对这些运动向量场建立的方向直方图组进一步分析的基础上，得到初设的 GOP 运动模式，从而最终判定要选择的 GOP 结构. 如果 $M=0$，则认为运动相对缓慢，如图 4a 所示的初设尺寸为 16 帧的 GOP 结构将不再变化. 如果 $M=1$，则认为存在快速或复杂的运动，把初设的尺寸为 16 的 GOP 进一步划分为如图 4b 所示的(8，8)结构，然后在最终选定的 GOP 上进行如图 4 所示的 MCTF. 图 4a 所示的 GOP 尺寸为 16 帧，进行 4 级 MCTF；图 4b 所示的 GOP 结构为(8,8)，在每个(8,8)结构的 GOP 上进行 3 级 MCTF.

在基于 MCTF 的小波视频编码算法中，GOP 的尺寸可以是 16,8,4 和 2 帧. 通常情况下选择固定大小是 16 帧或者 8 帧的 GOP 结构，所以本文根据运动特性自适应地设定 GOP 的尺寸为 8 帧或者 16 帧，具有一定的代表性.

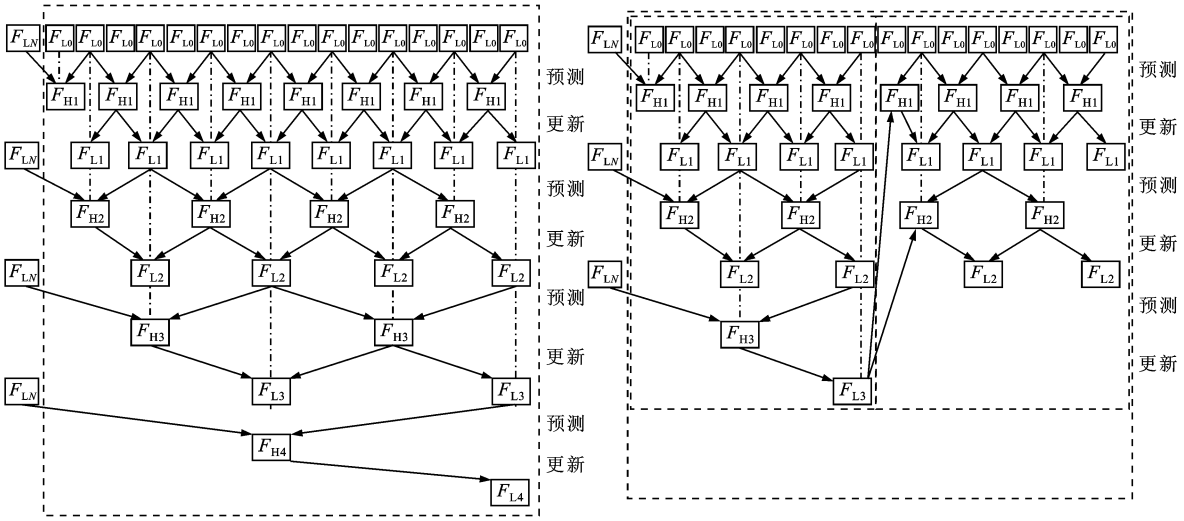
虽然更小尺寸的 GOP 结构在理论上也是可行的，但是它在大幅度增加复杂度的同时并不能确保一定会取得比尺寸为 8 帧的 GOP 结构更好的编码性能. 如果要进一步划分，不仅要抽取更多的运动矢量场，而且要对划分的机制做进一步研究，这将是我们下一步要做的工作之一.

3 实验结果

本文利用 MC-EZBC 软件平台[9]对本文方法进行了实验测试. 实验采用 3 个标准 CIF 格式的测试序列 Foreman、Soccer 和 Football，帧率均为 30 帧/s. 先对尺寸为 16 帧的 GOP 进行 4 级 5/3 MCTF，对(8,8)的 GOP 进行 3 级 5/3 MCTF，再进行 5 级空间 9/7 小波变换，然后对 GOP 模式、运动向量和变换系数分别编码，最后经过打包形成最终的位流.

3.1 峰值信噪比性能比较

为了验证本文方法在各种位率下均有更好的编码性能，对 Football 和 Soccer 序列进行了 2 种不同的 GOP 结构的峰值信噪比的比较. 从图 5 可以看出，相对固定尺寸为 16 帧的 GOP 结构，采用本文 MDA-GOP结构对编码性能有一定的改进，尤其是



F_{LN}表示前一个 GOP 进行最后一级 MCTF 后得到的低通帧；F_{L0}表示读入的原始视频帧；
F_{Li}和F_{Hi}(i=1～4)分别表示进行了 i 级 MCTF 后得到的低通帧和高通帧

(a)尺寸为 16 帧的 GOP 结构　　(b)(8,8)的 GOP 结构

图 4　在 GOP 结构上进行的 MCTF

对运动相对剧烈的视频序列，峰值信噪比的提高可以达到 0.5 dB.

3.2　编码时间比较

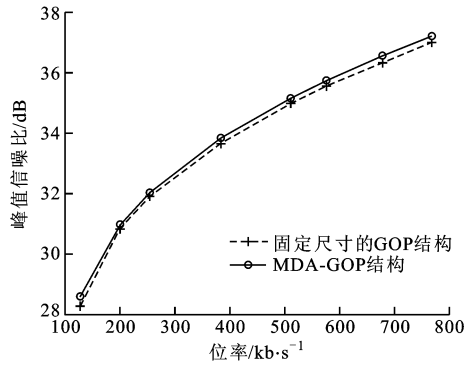
表 1 比较了采用 2 种不同的自适应决定 GOP 尺寸的编码时间，可以看出用本文方法可节省至少 40%的时间. 文献[6]的最佳模式的选择是根据采用几种模式 MCTF 后对相应的 MSE 性能进行比较得到的，不仅需要用每种模式进行 MCTF，而且要对各自 MCTF 后的 MSE 进行比较，很耗时. 本文方法仅需对抽取的 4 个运动向量场进行方向直方图分析，从而得到初设尺寸为 16 的 GOP 的运动模式，进而判断是否对 GOP 进行细分. 不管初设的 GOP 是否要进一步划分，抽取的4个运动向量场将被保留下来而无需重新估计，这就降低了计算复杂度，节省了时间. 因此，在保持性能不变的前提下，本文方法较之文献[6]方法在计算复杂度上有所降低.

表 1　编码时间比较

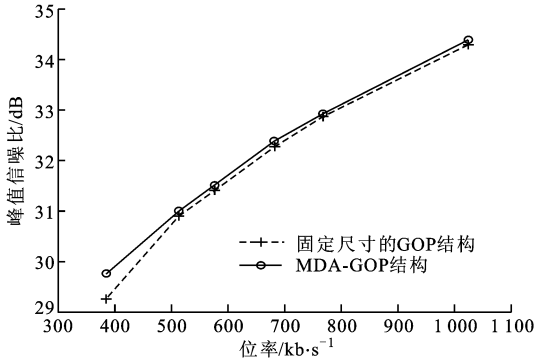
序列	编码时间/s	
	文献[6]方法	本文方法
Foreman	12 754	7 643
Football	47 374	18 522
Soccer	17 344	9 724

3.3　主观效果比较

为了进一步评价 2 种 GOP 结构的主观效果，对位率为 256 kb/s 下的 Foreman 序列第 178 帧的重构图像进行了比较，如图 6 所示. 从视觉效果上也可以明显看到，采用本文方法比采用固定 GOP 的方法效果要好得多.



(a)Football 序列



(b)Soccer 序列

图 5　不同尺寸 GOP 结构的率失真性能比较

(a)采用固定尺寸为16帧的GOP结构

(b)采用MDA-GOP结构

图6 位率为256 kb/s时Foreman序列178帧的重构效果比较

4 结 论

本文提出了一种尺寸自适应的GOP结构方法——MDA-GOP,其GOP的尺寸由运动模式决定.方法的原理是:在对等间隔抽取的4个运动向量场建立的方向直方图进行分析的基础上,自适应地决定初设尺寸为16帧的GOP的运动模式,从而进一步决定最终的GOP结构.MDA-GOP结构非常适合于小波视频编码.实验结果表明,对于存在快速或者复杂运动的序列,采用MDA-GOP结构比采用固定尺寸的GOP结构能取得更好的编码性能,而且相对于文献[6]的方法,在不改变编码性能的基础上,可节省不低于40%的编码时间.

参考文献:

[1] OHM J R. Three-dimensional subband coding with motion compensation [J]. IEEE Transactions on Image Processing, 1994, 3(5): 559-589.

[2] CHOI S J, WOODS J W. Motion compensated 3-D subband coding of video [J]. IEEE Transactions on Image Processing, 1999, 8(2):155-167.

[3] SECKER A, TAUBMAN D. Motion-compensated highly scalable video compression using an adaptive 3-D wavelet transform based on lifting [C]// Proceedings of 2001 IEEE International Conference on Image Processing. Piscataway, NJ, USA: IEEE, 2001: 1029-1032.

[4] PESQUET-POESCU B, BOTTEAU V. Three-dimensional lifting schemes for motion compensated video compression [C]// Proceeding of 2001 IEEE International Conference on Acoust, Speech and Signal Processing. Piscataway, NJ, USA: IEEE, 2001: 1793-1796.

[5] MEHRSERESHT N, TAUBMAN D. An efficient content adaptive motion compensated 3-D dwt with enhanced special and temporal scalability [J]. IEEE Transactions on Image Processing, 2006, 15(6):1397-1412.

[6] GWANG H P. Adaptive GOP structure for SVC [EB/OL]. (2005-01-12)[2008-04-15]. http://wg11.sc29.org/mpeg/docs/71/HongKong/contrib/m11563.zip.

[7] DULAVERAKIS C, VAGIONITIS S, ZERVAKIS M. Adaptive methods for motion characterization and segmentation of MPEG compressed frame sequences [C]// Proceedings of 2004 International Conference on Image Analysis and Recognition. Berlin, Germany: Springer-Verlag, 2004:310-317.

[8] ZHAO Fan, LIU Guizhong, NAN Nan. A high performance motion mode adaptive lifting motion compensation wavelet video codec [C]// Proceedings of 2007 IEEE International Conference on Multimedia and Expo. Piscataway, NJ, USA: IEEE, 2007:743-746.

[9] CHEN P, WOODS J W. Bidirectional MC-EZBC with lifting implementation [J]. IEEE Trans Video Technology, 2004, 14(10):1183-1194.

(编辑 刘杨)