

多分辨率运动估值的改进算法^{*}

刘红梅, 黄金科, 肖自美, 梁 凡
(中山大学电子与通信工程系, 广东 广州 510275)

摘 要: 多分辨率运动估值和补偿算法是一种广泛应用于小波视频编码器的消除时间冗余的算法, 针对其运动矢量编码效率低及采用全搜索方法、运算复杂度较高等问题, 提出了一种基于区域分割的多分辨率运动估值算法. 该算法具有如下特点: ①充分利用了小波分解后各子带所表征的运动结构的相关性, 有效地提高了运动矢量的编码效率; ②通过对视频帧小波分解后的低通子图像进行静止和运动区域分割, 以分割结果指导运动估值过程, 在有效地减少运动估值时间消耗的同时, 进一步提高了运动矢量的编码效率; ③在码率受限的情况下, 重建图像的质量得到提高. 实验结果表明文章所提出算法是有效的.

关键词: 小波分解; 视频编码; 多分辨率运动估值; 运动检测

中图分类号: TP391 **文献标识码:** A **文章编号:** 0529-6579 (2001) 02-0034-05

近年来, 小波变换由于其卓越性能, 已在图像压缩领域得到广泛应用^[1]. 在静止图像压缩领域, 小波变换压缩编码的性能明显优于基于 DCT 的 JPEG 方案. 能否将小波变换成功地推广应用于序列图像的视频压缩编码中自然成为图像压缩编码的研究热点.

在视频编码系统中, 基于块匹配法的运动估值/补偿算法是消除时间冗余的最基本和最重要的方法之一^[2-5]. 传统的运动估值/补偿算法一直在空间域进行, Zhang 等^[6]将小波理论应用于视频压缩时, 提出了一种基于小波分解域的可变块大小多分辨率运动估值 (multiresolution motion estimation, MRME) 算法, 它利用小波金字塔结构每层间的相关性, 来降低运动估值过程的运算复杂性. 但是 MRME 算法中运动矢量编码占用码率较高, 针对这一问题, 文献 [7] 中提出了两阶段法以改进 MRME 算法, 这种方法通过合并相邻块以提高运动矢量的编码效率, 但运算复杂度较高. 文献 [8, 9] 针对 MRME 算法中使用全搜索进行运动估值所耗费时间多的问题, 提取潜在运动区, 使运动估值运算只在运动区域进行, 使编码速度得到提高, 但没有提高运动矢量的编码效率.

为了同时解决 MRME 算法中运动估值运算复杂度和运动矢量的编码效率问题, 本文提出了

基于区域分割的 MRME 算法. 该算法在基本的 MRME 算法的基础上, 充分应用了小波分解后子带间的相关性, 以较少的时间代价, 提高了运动矢量的编码效率. 提出对小波分解后的低通子图像进行区域分割, 以分割结果指导运动估值过程, 在有效地降低了运动估值的时间消耗的同时, 进一步提高了运动矢量的编码效率.

1 利用各子带相关性的可变块 MRME 算法

一个图像 $f(x, y)$ 经过 3 级小波分解, 可以得到 10 个子带 $\{P_3f, D_1^1f, D_2^1f, D_3^1f, j=1\sim 3\}$. 其中, P_3f 是 $f(x, y)$ 在最低分辨率 2^{-3} 上的低通近似图像, 而 D_1^1f, D_2^1f 和 D_3^1f 则分别代表分辨率 2^{-j} 上的水平、垂直和对角线方向高频部分. 这种多分辨率表示法提供了一个等级结构来说明图像的信息. 在金字塔不同层的运动行为虽然不同却具有很高的相关性, 因为他们事实上是在不同的尺度空间和频率范围内, 刻画相同的运动结构. 因此, 可以充分地利用这一点减小运动估值的计算复杂度.

文 [6] 中提出 MRME 算法先估值出 P_3f 子带的运动矢量场, 利用此信息作为其他子带运动矢量场的初始估值, 在此基础上修正这些子带的运动矢量场. 这种运动估值仍然是基于块的, 但

^{*} 基金项目: 国家自然科学基金资助项目 (69975011); 广东省自然科学基金资助项目 (980442); 广东省自然科学基金重点资助项目 (963037)

收稿日期: 2000-08-22; 作者简介: 刘红梅 (1969—), 女, 讲师.

各层所对应的块大小不同，第 j 层上的子带采用的块大小为 $p \circ 2^{3-j} \times q \circ 2^{3-j}$ ， $p \times q$ 为最低分辨率子带中所取块的大小。

设时刻 t 在 P_3f 中位于 $s = (s_x, s_y)$ 的像素值为 $I_{3,t}^0(s)$ ，其中下标 3 和上标 0 分别表示该像素所在分解层为第 3 层以及在该层中的 P_3f 子带。于是，时刻 t 在 D_j^kf 中位于 $s = (s_x, s_y)$ 的像素值可表示为 $I_{j,t}^k(s)$ 。将 P_3f 中块的运动矢量定义为 v_3^0 ，它由绝对误差和准则 SAD (sum of absolute difference) 确定

$$v_3^0 = \underset{v \in W}{\operatorname{argmin}} \sum_{s \in W} \| I_{3,t}^0(s) - I_{3,t-1}^0(s+v) \|$$

其中， W 为 2×2 的块。

在基本的 MRME 算法中，在得到了低通子带 P_3f 的运动矢量后，所有子带以该运动矢量为基础进行各子带寻优。位于第 j 层和频带 k 的 D_j^kf 中的运动矢量由 $v_j^k = v_3^0 \times 2^{3-j} + \Delta_j^k$ 进行估值，其中 $\Delta_j^k = (\delta x, \delta y)$ 是按下面准则决定的修正值：

$$\Delta_j^k = \underset{v \in W}{\operatorname{argmin}} \sum_{s \in W} \| I_{j,t}^k(s) - I_{j,t-1}^k(s + v_3^0 \times 2^{3-j} + v) \|$$

其中， D_j^kf 的块大小 W 为 $2 \circ 2^{3-j} \times 2 \circ 2^{3-j}$ 。

在本文实现各子带寻优的过程中，考虑到子带表征的运动结构之间除了相关性之外，还具有空间方向选择性，一个某个特定方向的运动，将主要反映在 3 个子图像中的 1 个或 2 个，因此采用文献 [8] 中的处理方法：位于第 j 层和频带 k 的 D_j^kf 中的运动矢量由 $v_j^k = v_j^{/k} + \Delta_j^k$ 进行估值。其中 $v_j^{/k}$ 代表估值矢量的基础值，对于顶层其等于 v_3^0 ，其它层则为上一层对应方向的子带寻优后得到的运动矢量 v_{j+1}^k 的 2 倍，而不再是 MRME 算法中低通子图像运动矢量 v_3^0 的 2^{3-j} 倍；其中 $\Delta_j^k = (\delta x, \delta y)$ 是按下面的准则决定的修正值：

$$\Delta_j^k = \underset{v \in W}{\operatorname{argmin}} \sum_{s \in W} \| I_{j,t}^k(s) - I_{j,t-1}^k(s + v_j^{/k} + v) \|$$

其中， D_j^kf 的块大小 W 为 $2 \circ 2^{3-j} \times 2 \circ 2^{3-j}$ 。这样做，使下一分解层的基础运动矢量比基本的 MRME 中的更加精确可靠。

在 MRME 算法中，由于低通子带的运动矢量是所有其他层进行运动估值的基础，它们的失真将对所有其他层的预测产生较大影响，从而导致在恢复图像中出现不可忽视的质量损伤，为了

提高低通子带运动矢量的准确性，本文算法采用了文献 [7] 中的匹配准则，即低通子带中两组 4 个相同空间位置的块的 SAD 匹配准则：

$$v_3^0 = \underset{v \in W}{\operatorname{argmin}} \sum_{k=0}^3 \sum_{s \in W} | I_{3,t}^k(s) - I_{3,t-1}^k(s+v) |$$

由于低通子带 P_3f 提供了粗糙的信息，而其它子带 D_3^1f 、 D_3^2f 和 D_3^3f 则提供了关于物体的不同方向的边缘信息。上述匹配准则利用最低分辨率层所有的 4 个子带所提供的信息，可以获得更加精确的基础运动矢量。

2 基于区域分割的 MRME 算法

MRME 算法产生的运动矢量编码的编码比特数所占的比例很高，在数据率有限的情况下，由于分配给运动补偿残差图像的码率太低，而影响重建图像的质量。运动矢量编码效率低的原因除了 MRME 所固有的大矢量数目外，还和块匹配算法固有的缺陷有关，如搜索到的运动矢量的准确性不高，而且产生的运动矢量场的一致性不够好，尤其在运动物体的边缘区估值精度不高。

另外，由于采用全搜索方法，进行运动估值时消耗的时间较多。为此，本文提出了基于区域分割的 MRME 算法。首先，针对 P_3f 子带使用运动检测器，将最低分辨率层低通子图像分割成静止区域和运动区域；运动估值只对运动区域进行，这样可减少计算量。同时利用分割指导运动估值过程，获得更真实可靠，一致性更好的运动场。从而有效地降低运动矢量编码所占的比特数。

2.1 用运动检测器进行区域分割

运动检测器主要用来检测当前帧小波分解后低通子图像相对于前一帧小波分解低通子图像有无变化，并将每一帧的小波分解低通子图像分割成静止区域和运动区域。视频编码中所用的运动检测算法通常从被考察的 2 帧图像像素的亮度差入手。它主要包括以下几个部分^[9]：

(1) 计算两帧图像 LL 子带的小波系数差 $FD(i, j) = S_k(i, j) - S_{k-1}(i, j)$

(2) 设定阈值 T_1, T_2 ，判别每个系数 $p(i, j)$ 所对应空间位置的象素的状态

$$p(i, j) = \begin{cases} 0 & \text{静止} & |FD(i, j)| < T_1 \\ X & \text{不确定} & T_1 \leq |FD(i, j)| \leq T_2 \\ 1 & \text{运动} & |FD(i, j)| > T_2 \end{cases}$$

阈值 T_1 和 T_2 的选择必须适应图象序列的噪声幅度. 使用不确定状态可以避免有噪声引起的误判决.

(3) 消除不确定状态. 为了使结果只有静止和运动两种状态, 必须消除不确定状态. 按光栅顺序扫描 $p(i, j)$, 若状态不确定, 则以该系数为中心, 计算测量窗内静止状态的个数 C . 若 C 大于或等于 5, 则该系数 $p(i, j)$ 所对应空间位置的象素的状态是静止的, 否则是运动的.

(4) 采用标号的方法消除由运动区域包围的静止区域及合并小的运动区域. 如果在 2 个相继帧间有较大的平坦区在运动, 则前后 2 帧可能有重叠的区域. 这些区域中的象素值可能变化不大, 这样通过阈值运算, 会误判为静止区域, 为了保持运动物体的整体性, 必须消除这样的由运动区域包围的静止区域.

经过上述步骤的处理, 最后得到由不同状态的连通区域组成的掩模图像.

2.2 提高运动矢量场的一致性和可靠性

根据区域分割的结果, 在算法中为每块匹配块, 也就是每个运动矢量, 引入一个可靠性标志 $reliable$. 如果该块所有的象素都运动或者都静止, 则认为该块是可靠的; 如果, 块内并非所有的象素都具有相同的状态, 则可以判断该块位于运动物体的边缘, 对于运动估值来说, 认为是不可靠的. 算法的基本思想是: 尽可能的利用预测值, 整个估值算法分两次扫描象素, 第一次只针对运动区域中可靠的块, 在较少预测信息的基础上进行估值; 第二次再对运动区域中不可靠的块进行运动估值, 此时由于可靠的块已全部进行了估值, 可以提供较多的预测信息, 在此基础上, 提高预测的可靠性.

算法的具体过程描述如下:

(1) 检测当前块象素对应的标记是否全部属于静止区域, 如果是则 $reliable = 1$, 运动矢量为零矢量, 不再进行估值. 其他子带相应位置的运动矢量也为零矢量.

(2) 否则, 如当前块全部属于运动区域, 则 $reliable$ 为 1, 根据邻块的运动做预测. 与邻块的位置关系如图 1a 所示. 设 V_i 为各邻块的运动矢量, 则预测值为

$$V_p = \frac{a_1 \times V_1 + 2a_2 \times V_2 + a_3 \times V_3 + 2a_4 \times V_4}{a_1 + 2a_2 + a_3 + 2a_4}$$

其中 a_i 的取值为 0 或 1, 如果相应块与当前块处

于同一连通区域, 则其值为 1, 否则为 0. 实际上表明预测值是场景自适应的. 判断预测值是否越界, 如没有则以预测值为中心进行搜索. 为了提高一致性, 我们给匹配准则增加了一个附加项. 附加项 σ 表征的是和预测值的不符合程度, 中心为 0. 在实验中, 由于我们采用的是螺旋型搜索顺序, 为了简便起见, 我们令 $\sigma = l \times \theta$, 其中 l 为与中心的位移, θ 是预先设定的阈值. 估值出 P_3f 子带的运动矢量, 再对其它子带的对应块进行估值.

(3) 而如果当前块有的象素运动、有的象素静止, 则可以推测它位于运动区域的边缘, 其运动的不可靠性较大, 不能按可靠块同样的方法进行处理. 为了提高其预测值的可靠性, 需要获得更多的信息, 为此先把它滞后处理. 等到第一遍扫描结束, 所有可靠块的运动矢量已经得到, 我们就可以把不可靠块周围相邻块中可靠的运动矢量全部用来预测. 当前 $reliable = 0$ 的块与邻块的位置关系如图 1b 所示. 设 V_i 为各邻块的运动矢量, 则预测值为

$$V_p =$$

$$(a_1 \times V_1 + 2a_2 \times V_2 + a_3 \times V_3 + 2a_4 \times V_4 + 2a_5 \times V_5 + a_6 \times V_6 + 2a_7 \times V_7 + a_8 \times V_8) / (a_1 + 2a_2 + a_3 + 2a_4 + 2a_5 + a_6 + 2a_7 + a_8)$$

其中, a_i 的取值意义同上. 在预测值的准确性和可靠性提高的基础上, 以附加项来提高运动矢量的一致性. 估值出 P_3f 子带的运动矢量, 再对其它子带的对应块进行估值. 这样, 由分割结果指导运动估值过程, 使运动估值的准确性和一致性得到了很大的提高. 再和运动矢量的预测编码相结合, 可以较大地降低运动矢量所需的码率. 在码率受限的情况下, 可以将更多的比特分配给残差信息, 提高重建图像的质量. 同时, 由于它只对运动区域进行估值, 使得运动估值所耗时间从与图像大小成正比变成与运动区域的大小成正比, 能够使估值的时间显著地减少, 对实时编码非常有利.

3 模拟实验

在实验中比较了 3 种方案.

方案 I: 文献 [6] 中提出的基本的 MRME 算法.

方案 II: 利用了小波分解后各子带方向选择

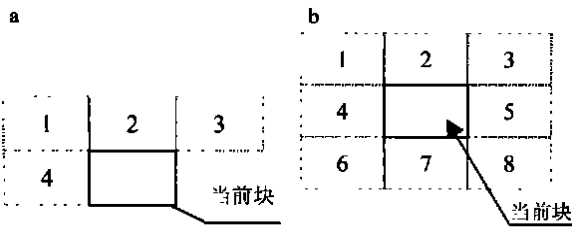


图 1 当前 reliable=1 (a) 和 reliable=0 (b) 的块与相邻块示意图

Fig 1 The relation of the current block (reliable= 1, a) and (reliable= 0, b) with its neighbor blocks

性^[9]，并提高了低通子带基础运动矢量准确性的 MRME 算法^[8]。

方案 III 方案 II 与本文提出的基于区域分割的 MRME 算法相结合。

表 1 给出了在每个方案中运动矢量编码所占用的码率，用 Motion Vector 表示 (Bytes/Frame)、以及所有帧在运动运动估值和运动补偿运算上所耗费的总时间，用 Total Time 表示 (单位为 s，包括运动估值和运动补偿，方案 II 还包括了用运动检测器进行区域分割的时间)。本实验是在内存为 6×16M 的 K6-266 微机上进行的。实验所用的序列是 QCIF 格式序列 Carphone 和 CIF 格式序列 Hall。以 10 帧/s 的帧率，对前 40 帧进行编码。对于 QCIF 格式，本文采用 2 层小波分解，对于 CIF 格式，采用 3 层分解。其它参数如下：

格式	QCIF	CIF
低通子带大小	2×2	2×2
最低分辨率层的搜索范围 Ω_0	8×8	10×10

表 1 运动估值性能实验结果比较

Tab 1 The performances of the motion estimation algorithms

方案	序列	Carphone	Hall
I	Motion Vector	397. 26	269. 41
	Total Time	10. 37	30. 37
II	Motion Vector	348. 56	242. 72
	Total Time	11. 63	31. 48
III	Motion Vector	255. 38	164. 51
	Total Time	8. 42	6. 79

从表 1 可以看出，方案 II 由于利用了小波分解后各子带方向选择性，并提高了低通子带基础运动矢量准确性，减小了每一个子带运动矢量的误差，对运动矢量编码效率的提高较为明显。由

于方案 II 在低通子带运动估值中匹配准则的计算量稍大于方案 I，所以方案 II 中用于运动估值和补偿的时间比方案 I 有所增加。在方案 II 中，40 帧 CIF 格式的序列 Hall 所耗的时间比 40 帧 QCIF 格式的序列 Carphone 所耗的时间还要少。可见，由于应用运动检测器，将视频帧分割为运动区域和静止区域，编码器不再对静止区域进行运动估值和补偿，所耗的时间从由图像大小决定变成主要由运动区域的大小来决定，大大提高了系统效率。同时利用分割结果，进一步提高了运动矢量的准确性和运动场的一致性，使运动矢量编码所需的码率得到了有效的降低。

为了衡量本文提出的算法对编码性能的影响，我们对测试序列 Miss America, Salesman 和 Claire 的前 20 帧亮度分量进行了编码。这些测试序列的分辨率都为 352×288，帧率为 10 帧/s。实验结果如表 2 所示。其中 PSNR 值给出的是限定编码码率时的恢复图像的质量。Motion Vector 与 Total Time 同上，DFD (Bytes/Frame) 表示运动补偿后残差全部传输时所需要的码率，TOTAL (Bytes/Frame) 表示 DFD 与 Motion vector 所需总的码率。

表 2 编码实验结果比较¹⁾

Tab 2 The comparison of the coding results

方案	序列	Miss America	Salesman	Claire
I	PSNR	36. 54	31. 04	38. 48
	Motion Vector	251	377	384
	DFD	4474	14338	3877
	TOTAL	4725	14715	4261
	Total Time	14. 25	17. 95	11. 59
II	PSNR	36. 65	31. 18	38. 70
	Motion Vector	227	336	352
	DFD	4459	14239	3802
	TOTAL	4686	14575	4154
	Total Time	15. 23	18. 67	12. 48
III	PSNR	36. 77	31. 39	38. 97
	Motion Vector	103	192	198
	DFD	4532	14187	3931
	TOTAL	4635	14379	4129
	Total Time	4. 26	5. 35	5. 57

1) BR 代表比特率，单位是 kb/s，CR 代表压缩比；
BR/CR: Miss America 为 58/ 140;
Salesman 为 135/ 60;
Claire 为 108/ 75

4 结 论

实验结果表明本文算法同时在 3 个方面对 MRME 算法进行了有效的改进:

(1) 在基于运动估值/补偿的编码器中, 主要运算开销往往在运动估值上. 基于区域分割的结果, 使运动估值的速度得到有效的提高.

(2) 基于分割结果, 提高了运动矢量场的一致性和准确性, 使运动矢量编码效率得到提高.

(3) 码率有限的情况下, 由于本文算法中运动矢量所需的比特数大大地降低了, 则可以给运动补偿后的残差分配更多的比特, 使重建图像的 PSNR 值得到普遍的提高.

参考文献:

- [1] 肖自美. 图像信息理论与压缩编码技术[J]. 广州: 中山大学出版社, 2000. 282—304.
- [2] ITU-T SGXV. Video codec for audiovisual services at px64kbit/s. ITU-T Recommendation H. 261. July 1990.
- [3] ISO/IEC JTC1/SC29/WG11. MPEG-4 Video Verifica-

tion Model version 13. 3, MPEG99/4960. Melbourne: 1999.

- [4] ISO/IEC 13818-2, Generic coding of moving pictures and associated audio information: video. 1996.
- [5] ITU-T. Video Coding for Low Bitrate Communication, Draft Recommendation H. 263 Version 2. 1998.
- [6] ZHANG Y, ZAFER S. Motion-compensated wavelet transform coding for color video compression[J]. IEEE Trans Circ and Syst for Video Technol, 1992, 2(3): 285—296.
- [7] SEONGMAN K, SEUNGHYEOM R. Interframe coding using two-stage variable block-size multiresolution motion estimation and wavelet decomposition[J]. IEEE Trans Circuits Syst Video Technol, 1998, 8(4): 399—409.
- [8] WEI J, LI Z N. An enhancement to MRMC scheme in video compression[J]. IEEE Trans Circuits and Systems for Video Technology, 1997, 7(3): 564—568.
- [9] 许洁斌, 韦岗, 布礼文. 一种基于形态变换的小波分解运动补偿编码新算法[J]. 中国图象图形学报, 1998, 3(9): 727—733.
- [10] 卢官明, 毕厚杰, 江平. 求取真实运动矢量的一种算法[J]. 数据采集与处理, 1999, 14(3): 267—271.

An Improved Multiresolution Motion Estimation Algorithm

LIU Hong mei, HUANG Jin ke, XIAO Zi mei, LIANG Fan

(Department of Electronic and Communication, Zhongshan University, Guangzhou 510275, China)

Abstract: Multiresolution motion estimation and compensation is a widely used algorithm to reduce the temporal redundancy in the wavelet video encoder. Aimed at resolving the problem of the motion vector coding efficiency and the computation complexity, this paper presents a region segmentation based multiresolution motion estimation (MRME) algorithm. The features of the presented algorithm are as follow: ① by fully using the spacial and orientation correlations between the motion structure represented by the subbands of wavelet decomposition, the coding efficiency of motion vector is improved; ② by segmenting the lowest frequency subimage of wavelet decomposition of the frame into static and motive regions, the computation complexity is reduced effectively and the motion vector coding efficiency is improved farther; ③ when the data rate is limited, the performance of codec using the presented algorithm is improved. The experimental results show the effectiveness of the presented algorithm.

Keywords: wavelet decomposition; video coding; multiresolution motion estimation; motion detection