

# 一种高效的基于 H.264/AVC 压缩域信息的全局运动估计方法

王 喆, 刘贵忠, 钱学明

(西安交通大学电信学院信通系, 陕西西安 710049)

**摘 要:** 本文提出了一种基于 H.264/AVC 压缩域的高效全局运动估计算法. 由于 H.264 采用了多种新的视频压缩编码技术, 使得其压缩码流的运动矢量(MV)场中包含大量噪声运动矢量, 可参与全局运动估计的运动矢量相对较少. 噪声运动矢量这里指的是与全局运动不相符的运动矢量. 为了降低噪声运动矢量的影响, 提高全局运动估计的精度和效率, 在全局运动估计过程中就必须对噪声运动矢量进行有效的去除. 本文利用 H.264 码流中的运动矢量预测残差(MVD)和运动矢量预测参考块的信息, 对噪声运动矢量做了有效的去除. 实验的结果验证了本文提出的全局运动估计算法的计算效率和精度.

**关键词:** 全局运动估计; H. 264/AVC 压缩域; 运动矢量预测; 运动矢量预测残差

**中图分类号:** TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2011) 3A-019-05

## An Efficient Global Motion Estimation Algorithm on H.264/AVC Compression Domain

WANG Zhe, LIU Gui-zhong, QIAN Xue-ming

(Department of Information and Communication Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

**Abstract:** An efficient global motion estimation algorithm on H.264/AVC compression domain is proposed. Due to many technologies of video compression coding used in H.264/AVC, on H.264/AVC compression domain, the motion vector field contains a lot of noise blocks for global motion estimation. The valid motion vector number for global motion estimation is limited. In order to reduce the influence of noise motion vectors and improve accuracy and computational efficiency of global motion estimation, it is necessary to eliminate noise motion vectors during global motion estimation. In this paper, the noise blocks were eliminated efficaciously by using information of MVD (the difference between the current vector and the predicted vector) and reference block for MV prediction in H.264 bits stream. The extensive experiments show that the proposed method is efficient and robust in terms of both computational complexity and accuracy.

**Key words:** global motion estimation; H.264/AVC compression domain; motion vector prediction; motion vector difference

### 1 引言

全局运动是视频的一种非常重要的特征, 它广泛应用于视频编码, 视频检索, 对象分割及视频拼接等领域. 视频的运动可分解为全局运动和由运动物体引起的局部运动. 全局运动是由摄像机的运动引起的, 这种运动通常可以用一个二维图像的参数变换来表达. 估计这些参数的方法称之为全局运动估计.

常用全局运动估计的方法可以分为基于像素域的方法和基于压缩域的方法. 前者大多采用的是基于最小二乘的迭代方法以及快速的金字塔算法实现全局运动

参数估计, 如 MPEG-4 中所使用的由粗到精的估计方法<sup>[1,2]</sup>, 其中采用一个三层的高斯金字塔模型对原始分辨率下的图像进行下采样处理, 在金字塔的最高层进行全局运动的平移分量的估计, 然后在其它层中使用一个基于迭代的方法来逐渐更新参数. 在文献[3]中运用梯度幅值较大的像素点子集的方法进行全局运动估计. 相对于基于像素域的方法而言, 基于压缩域的全局运动估计方法则是利用压缩码流中的相关信息实现快速的全局运动信息估计.

视频媒体融合语音、图像、文字等多种信息于一体. 相对于一般的文本信息而言, 视频媒体具有数据量大的

特点.为了提高传输和储存效率,大部分视频都以压缩码流的形式存在.因此,直接利用压缩码流的信息进行全局运动估计在视频分析和检索研究领域受到广泛的关注.其中 Su 等人提出了一种基于运动矢量场的全局运动参数估计方法<sup>[4]</sup>,并且在文献[5]中他们提出了一种采用固定空间采样模式的运动矢量组来进行全局运动估计.在我们前期工作中我们将全局运动参数的估计过程看成是一个参数优化过程并且用 GA 进行压缩域中全局运动参数的优化<sup>[6]</sup>,为了加快全局运动估计的速度,初始种群的参数是按照运动矢量场的约束来生产的.另外,我们对文献[7]中的方法的局限性进行了改进,进一步提高了全局运动参数的估计精度.

随着信息技术的发展以及高效实施媒体传输的需求推动,H.264/AVC (MPEG-4 part 10)应运而生<sup>[8,9]</sup>,其中采用了多参考帧补偿、更多的预测模式以及可变块的运动估计方法等新的编码技术,使得编码效率和性能大大提升.由于 H.264/AVC 压缩域和 MPEG(包括 MPEG-1, MPEG-2 以及 MPEG-4 part2)压缩域的码流结构具有很大的不同,其相应的全局运动估计方法也应随之改变.然而,现有的基于压缩域信息的全局运动估计方法大都是基于 MPEG 压缩标准来进行的<sup>[4-7]</sup>.如何利用 H.264/AVC 压缩域中的信息进行快速的全局运动估计是一个具有挑战性的课题,并且迄今还少有相关的研究,因此文本针对 H.264/AVC 压缩视频的特点提出了一种全局运动估计方法.

## 2 基于 H.264/AVC 压缩域的全局运动估计

H.264/AVC(也即 MPEG-4 part 10)是由 ISO 和 ITU 的共同推出的一个全新的视频编码标准,在该标准中使用了一系列高效的视频压缩编码技术,以在保证视频质量的同时进一步降低视频的码率.其中包括帧内预测编码、多参考帧补偿技术、可变块的运动估计等.这些技术的引用增加了编码的复杂度,同时也使得基于压缩域码流的分析变得复杂.这里我们针对 H.264/AVC 压缩域的特点提出了一种新颖的全局运动估计方法.

### 2.1 H.264/AVC 运动估计简介

不同于以前的编码标准(MPEG-1/2 以及 MPEG-4 Part 2),为了降低 B, P 帧编码的码率,H.264/AVC 中使用了基于多参考帧的运动估计,以最大限度的消除帧间冗余,在节省码率的同时保证了视频质量.其中,参考帧的数目最大可以设置为 15,这样同一帧中不同块的参考帧可能不相同.而且 H.264/AVC 编码标准还采用了自适应块大小的运动估计模式.对运动矢量采用预测的方式以减小编码运动矢量的码率.如图 1 所示,当前块的运动矢量由和它邻近的  $4 \times 4$  块( $a, b, c$  或  $d$ )

的运动矢量做预测,根据宏块分割类型选择当前块的运动矢量预测模式,预测模式分为:左预测,上预测,右上预测和中值预测四种.预测得到的运动矢量记为 MVP,它与当前块真实的运动矢量 MV 的差即为 MVD.由于 MVD 可以在一定程度上反映当前块与其邻近块运动矢量的差异,因此我们可以使用 MVD 值去除与全局运动估计不相符的运动矢量,可以使估计的结果更为精确.

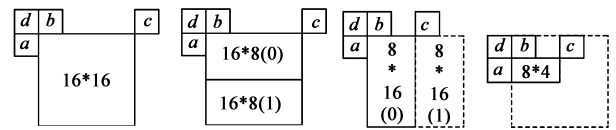


图1 各种宏块分割及其邻近预测块

### 2.2 H.264 压缩域中的全局运动估计噪声块

视频中的运动可分解为由镜头运动引起的全局运动和由运动物体引起的局部运动,从而,我们把视频分为全局运动区域(GMR)和局部运动区域(LMR).全局运动估计噪声块指的是对全局运动估计带来误差的块,它可以具体的分为全局运动区域噪声块和局部运动区域噪声块,我们分别称之为内噪块和外噪块.外噪块就是由物体的局部运动引起的,而引起内噪块的原因主要有以下几点:

- (1)多参考帧的引入使得 GMR 中存在不是以前一参考帧做预测的块,其运动矢量与全局运动不相符.
- (2)由于块匹配误差产生的与全局运动不相符的运动矢量.
- (3)平滑区域为 0 或者近似为 0 的运动矢量.
- (4)GMR 中的帧内模式编码宏块.

为了提高全局运动参数估计的精度,我们必须对上述类型的块所对应的运动矢量进行剔除.下面我们具体介绍每种情况下的噪声块剔除的具体方法.

### 2.3 全局运动估计噪声块的去除

在 H.264/AVC 中能够根据运动补偿的结果自动确定块的划分模式.一般而言,对于帧间预测模式的 MB 可以划分  $16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8, 8 \times 4, 4 \times 8, 4 \times 4$  等多种模式.这种灵活的块预测模式,使得从码流里直接提取得各种信息所指向的块并不是同样尺寸的,为了后续分析的方便起见,我们首先对块划分的大小进行统一,在这里我们以  $4 \times 4$  块为单位,给每个  $4 \times 4$  块分配运动矢量等信息.具体的去除噪声块的步骤如下:

- (1)去除不是以前一参考帧做运动估计的块,以及帧内编码模式的块.我们用  $I_1(x, y)$  标记第一步执行后的结果, $(x, y)$  表示当前块的坐标, $I_1(x, y)$  等于 1 时,表示该块可用于全局运动估计, $I_1(x, y)$  等于 0 时,表示该块为噪声块.

- (2)在(1)的基础上,对  $I_1(x, y)$  等于 1 的块进行处

理. 根据这些块的 MVD 值设定阈值, 根据此阈值标记块. 前面已经提到 MVD 为运动矢量残差, 它是 MV 与 MVP 的差值, MVP 是由邻近已编码分割的 MV 预测得到的. H.264 在 MV 预测上做了很大的改进, 不仅有 MPEG 的左预测模式, 还增加了上预测, 右上预测和中值预测三种模式, 这就使得 MVD 可以在一定程度上反映当前块与邻近块的运动差异, 若当前块的 MVD 值大于一定的阈值  $\alpha$ , 我们标记  $I_2(x, y)$  等于 0, 否则,  $I_2(x, y)$  等于 1.  $I_2(x, y)$  同  $I_1(x, y)$ , 表示第二步后的标记结果. 阈值  $\alpha$  由下式确定,

$$\alpha = \frac{1}{n} \sum_{i=1}^n |P_{MVD}(x_i, y_i)| + \frac{1}{n} \sqrt{\sum_{i=1}^n P_{MVD}^2(x_i, y_i)} \quad (1)$$

式中,  $n$  表示所有  $I_1(x, y)$  等于 1 的块的数目,  $P_{MVD}(x_i, y_i)$  表示块  $(x_i, y_i)$  的 MVD 值.

(3) 由于局部运动区域内部的块大部分是由其它局部运动块做参考来预测 MVP 的, 所以这些块的 MVD 值有时也会很小; 同时, 处于全局运动区域边界的某些块可能是由局部运动块做预测的, 这些块的 MVD 值也可能较大. 针对以上两种情况, 我们必须根据参考块的坐标信息对已标记块进行更新. 参考块的坐标信息在 H.264 码流中不是直接传输的, 而是隐藏在码流中需要通过一定的计算得到. 在得到参考块的坐标后, 对  $I_2(x, y)$  等于 1 的块集合按照从左到右、从上到下的顺序的做一次再标记. 具体过程如下:

$$\begin{cases} I_2(x, y) = 0, I_2(x', y') = 0 \cap |P_{MVD}(x, y)| < \alpha \\ I_2(x, y) = 1, I_2(x', y') = 0 \cap |P_{MVD}(x, y)| > \alpha \end{cases} \quad (2)$$

式(2)中,  $(x', y')$  为当前块做 MV 预测的参考块坐标.

然后, 令  $I_3 = I_1 \cap I_2$ ,  $I_3$  为第三步标记后的结果.

(4) 去除平滑区域的运动矢量为零的块. 这一步工作我们利用 H.264/AVC 中的 SKIP 模式进行标记. 在第三步的基础上, 我们把  $I_3(x, y)$  等于 1 中编码模式为 SKIP 且运动矢量为零或近似为零的块标记  $I_4(x, y)$  等于 0.



图2 foreman测试序列去除结果

## 2.4 全局运动参数的估计

本文采用 6 参数的摄像机运动模型, 如式(3)所示

$$\begin{cases} x'_i = a_0 x_i + a_1 y_i + a_2 \\ y'_i = b_0 x_i + b_1 y_i + b_2 \end{cases} \quad (3)$$

点  $(x'_i, y'_i)$  表示当前帧中块的中心点坐标, 点  $(x_i, y_i)$  表示参考帧中块的中心点坐标. 全局运动估计即是

对式(3)中的  $(a_0, a_1, a_2, b_0, b_1, b_2)$  进行估计. 此处我们可运用二元线性回归算法对全局运动参数进行估计<sup>[10]</sup>. 首先, 我们采用最小二乘法求  $a_0, a_1, a_2$  的估计量. 做离差平方和

$$Q = \sum_{i=1}^n (x'_i - a_0 x_i - a_1 y_i - a_2)^2 \quad (4)$$

这里,  $n$  表示去除噪声块后的运动矢量块的数目.

使  $Q$  达到最小, 求偏导得:

$$\begin{cases} \frac{\partial Q}{\partial a_2} = -2 \sum_{i=1}^n (x'_i - a_0 x_i - a_1 y_i - a_2) = 0 \\ \frac{\partial Q}{\partial a_0} = -2 \sum_{i=1}^n (x'_i - a_0 x_i - a_1 y_i - a_2) x_i = 0 \\ \frac{\partial Q}{\partial a_1} = -2 \sum_{i=1}^n (x'_i - a_0 x_i - a_1 y_i - a_2) y_i = 0 \end{cases} \quad (5)$$

令,

$$\mathbf{A} = \begin{bmatrix} n & \sum_i x_i & \sum_i y_i \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i y_i \\ \sum_i y_i & \sum_i y_i x_i & \sum_i y_i^2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \sum_i x'_i \\ \sum_i x_i x'_i \\ \sum_i y_i x'_i \end{bmatrix}, \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_0 \end{bmatrix}$$

则式(5)可转化为  $\hat{\mathbf{a}} = \mathbf{A}^+ \mathbf{B}$  的方式求解. 这里系数矩阵  $\mathbf{A}^+$  是矩阵的 Moore-Penrose 广义逆或者称作矩阵  $\mathbf{A}$  的极小最小二乘  $g$  逆. 对式(3)中的  $b_0, b_1, b_2$  参数也可采用同样的方法估计得出.

## 3 实验结果及分析

为了对本文提出的基于 H.264/AVC 压缩域的全局运动估计方法进行测试, 我们选取了四组标准测试序列, coastguard, tabletennis, foreman 和 flower. 其分辨率分别为  $176 \times 144$ ,  $352 \times 240$ ,  $176 \times 144$  和  $176 \times 144$ . 采用 H.264/AVC 的验证程序 JM10.2 对这三组序列进行编码, GOP 格式为 IBBPBBPBBPBB, 长度为 12. 选用 P 帧的信息进行全局运动估计.

(1) 计算复杂度分析. 首先, 由于本文所提出的方法是在压缩域下进行的, 不需要对图像进行完全解码, 而且直接使用了码流中的运动矢量信息, 所以相对于像素域的全局运动估计算法, 节省了大量的运算时间. 其次, 由于噪声块的去除, 使得参与估计的块减少, 不仅使估计结果更为精确, 而且提高了估计的效率. 而由于去除噪声块的方法基本是使用码流的原始数据, 其计算量很小. 所以, 本文所提出的方法在时间上是非常高效的.

(2) 估计精度分析. 由于目前关于 H.264/AVC 压缩域的全局运动算法很少, 为了验证本文估计参数的精度, 我们选用了 MPEG-4 中的全局运动估计算法做参照, 由于其是在像素域进行的, 参数的估计也就相对精

确。

然后,我们用全局运动参数恢复图像,与原始图像做比较.计算不同方法的 PSNR 值,这里我们手动去除了局部运动区域,只比较全局运动区域的 PSNR.

表 1 中,GME\_P 是本文提出的方法,GME\_N 是没有去除全局运动噪声块进行全局运动估计的结果,GME\_MPEG4 是用作对照的 MPEG-4 中所采用的方法.从表中我们可以看出,由于 GME\_MPEG4 是基于像素域的,其估计精度要优于本文的方法,平均 PSNR 要比本文的方法高 1.12dB;同时本文提出的方法要明显好于 GME\_N,其平均 PSNR 比 GME\_N 高 1.25dB,这就证明了本文所提出的去除噪声运动矢量的全局运动估计方法能有效地提高估计的精度.

表 1 各种视频序列全局运动区域的 PSNR 值

	Coastguard	Tabletennis	Foreman	Flower	Average
GME_MPEG4	29.254	32.523	21.015	25.347	27.035
GME_N	27.202	27.392	20.204	23.865	24.666
GME_P	28.269	29.894	20.709	24.804	25.919

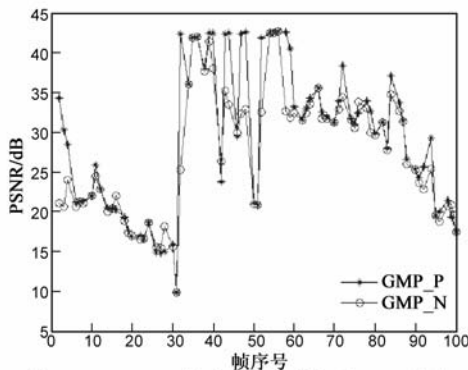


图3 Coastguard测试序列三种方法PSNR比较

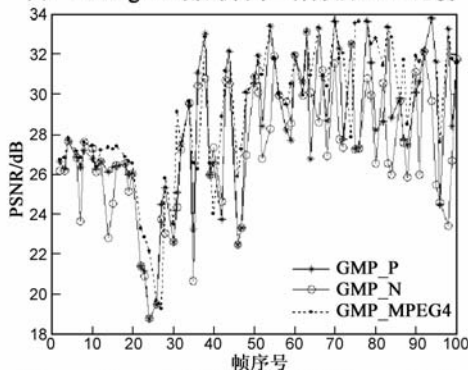


图4 Tabletennis测试序列两种方法PSNR比较

从图 3 中我们看到,本文所提出的方法(GMP\_P)在绝大多数帧 PSNR 值要高于 GMP\_N,而与 MPEG4 中像素域的方法也非常接近.图 4 是利用 Tabletennis 测试序列对 GMP\_P 和 GMP\_N 两种方法进行比较的结果,从图中可以看出,应用本文的方法(GMP\_P)进行全局运动参数的估计更为精确.

综上所述,本文提出的在去除全局运动噪声块的基础上进行的全局运动估计方法是非常有效的.

## 4 结论

全局运动参数估计在视频分析等领域有着重大的意义,随着 H.264/AVC 压缩编码算法的推广应用,基于 H.264/AVC 压缩域的全局运动估计也就显得尤为重要.本文充分利用 H.264/AVC 的码流特征,制定了一种高效、精确的全局运动估计算法.本文提出的方法的主要特点是:

实现了 H.264/AVC 压缩域的全局运动估计.由于是利用压缩域信息进行估计的,在计算时间上大大优于像素域的全局运动估计.

在充分了解 H.264/AVC 压缩域的码流特征的基础上,利用码流中直接传输的 MVD 值及其 MV 预测模式信息,制定了高效的去除全局运动噪声块的方案.

## 参考文献

- [1] ISO/IEC J TC1/SC29/WG11-2001. MPEG-4 Video Verification Model version 18.0[S].
- [2] Dufaux F, Konrad J. Efficient, robust, and fast global motion estimation for video coding[J]. IEEE Trans Image Processing, 2000, 9(3): 497 - 501.
- [3] Keller Y, Averbuch A. Fast gradient methods based on global motion estimation for video compression[J]. IEEE Trans Circuits Syst. Video Technol., 2003, 13(4): 300 - 309.
- [4] Su Y P, Sun M T, Hsu V. Global motion estimation from coarsely sampled motion vector field and the applications[J]. IEEE Trans Circuits Syst Video Technol, 2005, 15(2): 232 - 242.
- [5] Su Y P, Sun M T. A Non-iterative motion vector based global motion estimation algorithm[A]. IEEE Conference on Multimedia and Expo[C]. Taipei: IEEE, 2004. 703 - 796.
- [6] 钱学明,刘贵忠.基于 GA 的压缩域中全局运动估计及在字幕遮挡区域恢复中的应用[J].电子学报,2006,34(10): 1920 - 1924.
- [7] Qian X M, Liu G Z. Global motion estimation from randomly selected motion vector groups and GM/LM based applications [J]. Signal, Image and Video Processing, 2007, 1(3): 179 - 189.
- [8] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14486-10 AVC [S]. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VECG, JVT-G050, 2003.
- [9] Wiegand T, Sullivan G J, et al. Overview of the H.264/AVC video coding standard [J]. IEEE Trans Circuits Syst Video Technol, 2003, 13(7): 560 - 576.
- [10] Li H L, Liu G Z, et al. A robust, efficient, and fast global mo-

tion estimation method from MPEG compressed video [A].  
IEEE Proc PCM [C]. Taiwan: Springer, 2002. 60 – 65.

### 作者简介



**王 ■** 博士生,男,1981年3月出生于陕西省宝鸡市.现为西安交通大学电信学院信息与信号处理专业博士生.主要研究方向为图像视频分析和检索.

E-mail: redfg@mailst.xjtu.edu.cn



**刘贵忠** 博士生导师,教授,男,1962年9月出生于陕西延安.1982年9月获西安交通大学学士学位;1985年9月获硕士学位;1989年9月获荷兰埃因霍温大学数学与信息科学系博士学位.现为西安交通大学电信学院博士生导师.主要从事小波分析时频分析等非平稳信号分析方法及自适应处理方法、图象视频多媒体压缩编码及传输控制、信号采集与分析系统、遥感图象压缩分析处理及高分辨地震勘探信号处理等方面的研究工作.

E-mail: liugz@mail.xjtu.edu.cn



**钱学明** 博士,副教授,男,1975年1月出生于安徽潜山,2004.9~2008.1在西安交通大学攻读博士学位,于2008年3月获得西安交通大学博士学位,于2008年4月起任教于西安交通大学电信学院信息与通信工程系 (<http://gr.xjtu.edu.cn:8080/web/qianxm>).感兴趣的研究方向是图像视频分析和检索以及图像视频传输中视频质量提升方法. E-mail: qianxm@mail.xjtu.edu.cn