

# 适用于 H.264 的高性能快速运动估计算法

喻占武<sup>1</sup> 黄晓杰<sup>1</sup> 贾佳<sup>2</sup> 范励<sup>2</sup>

(1 武汉大学多媒体通信工程中心, 武汉市珞喻路 129 号, 430079)

(2 武汉第二炮兵指挥学院基础部, 武汉市工农兵路 46 号, 430012)

**摘要:** 提出了一种适用于 H.264 的快速运动估计算法: 使用一维映射作为块排除准则, 并充分利用 H.264 中的运动估计算法间的匹配冗余, 使得 7 种运动估计算法并行执行。试验结果表明, 在估计精度几乎无损的情况下, 本文算法比传统的全局搜索速度快 6 倍以上。

**关键词:** FS; PFS; 一维映射; 运动估计; H.264

**中图法分类号:** TN919.81; TP311.12; TP393

在自然视频编码中, 运动补偿预测编码的方法可以有效地减少连续帧之间的相关性, 所以成为决定编码性能的关键技术。但是全局搜索运动估计需要耗费编码器极大的运算时间。如在 MPEG-4 中, 运动估计占据了 66% 的编码器运算量<sup>[1]</sup>, 而在 H.264<sup>[2]</sup> 语法中, 支持多重参考帧选择和多达 7 种的运动估计算法, 所以运动估计在 H.264 编码器中占用运算量更大。各种用来减小运动估计的计算量的算法一般可以分为两大类: ① 全局搜索(FS), 通过辅助的策略减少搜索的计算量。如可以动态选择搜索中心或者提前跳出匹配搜索等来减少块匹配所需的计算量。这类算法可以在搜索窗口内找到最匹配的块, 所以具有了最优的搜索性能, 但是需要较多的计算量。② 只对部分位置进行搜索, 通过减少搜索空间中的候选位置或使用特定的匹配准则来减少计算量。如新三步搜索法(TSS)<sup>[3]</sup>、新四步搜索法<sup>[4]</sup> 和钻石搜索法<sup>[5]</sup> 等。这类算法可以取得较小的运算复杂度, 但是很容易陷入局部最优, 所以使得搜索性能下降。

本文结合以上两种方法的优点, 并针对 H.264 中的 7 种运动估计算法, 提出了一种新的快速算法(PFS)。它的主要特点如下: ① 采用全局搜索, 这样可以在搜索窗口内找到全局最优的运动矢量, 提高编码率失真性能。② 首先进行粗略的搜索, 把块的二维信号映射为一维信号进行

匹配, 这样可以快速有效地排除大部分不匹配的位置。然后对剩余的位置再使用二维匹配准则进行精细的搜索。上面的过程虽然把二维信号转换为一维信号进行匹配, 但是不会引起匹配失真, 这是因为, 如果两个块的一维累加信号不能很好地匹配, 则这两个块的二维信号也不会很好地匹配。③ 充分利用了 H.264 中的多种运动估计算法间的匹配冗余, 使得 7 种运动估计算法可以并行执行, 极大地减少了 H.264 中运动估计的计算量。

## 1 H.264 中的运动估计

H.264 中支持 7 种分块模式, 包括  $16 \times 16$ 、 $16 \times 8$ 、 $8 \times 16$ 、 $8 \times 8$ 、 $8 \times 4$ 、 $4 \times 8$  和  $4 \times 4$ (图 1)。使用这些新的分块模式可以为编码器带来以下优点。

1) 较小的分块可以对精细的运动细节进行更好的模拟, 减少大块斑的产生, 能够取得更好的主观视觉效果。

2) 虽然较小的分块需要传送较多的运动矢量, 增加了系统开销, 但是由于较小的分块提供了更准确的预测, 所以整体的压缩比有很大的提高。与只使用  $16 \times 16$  模式的运动估计相比, 使用全部 7 种模式的运动估计可以节省多达 15% 的比特率。但这个新特性使得编码器的计算量急剧增加, 所以对运动估计的性能提出了更高的要求。

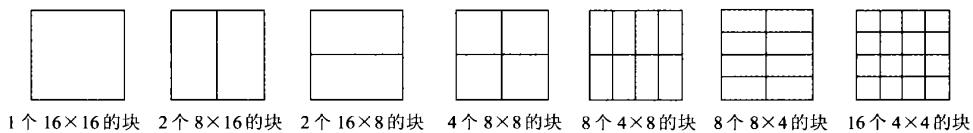


图 1 H. 264 采用的 7 种运动估计模式

Fig. 1 Block Modes of Motion Estimation in H. 264

## 2 基于一维映射的块匹配准则

### 2.1 一维映射<sup>[6]</sup>

假设帧的大小为  $W \times H$ , 块的大小为  $B_h \times B_v$ , 搜索窗口的大小为  $W_h \times H_v$ 。并设  $B^{x, y}$  表示左上角坐标为  $(x, y)$  的二维块,  $B_{i, j}^{x, y}$  表示此块中第  $j$  行、第  $i$  列的像素值。定义  $B^{x, y}$  的垂直一维映射为  $P_{B_i}^{x, y}$ , 它的第  $i$  个分量为  $B^{x, y}$  的第  $i$  列像素值的和:

$$P_{B_i}^{x, y} = \sum_{j=0}^{B_v-1} B_{i, j}^{x, y} \quad (1)$$

式中,  $0 \leq i \leq B_h$ , 并且  $0 \leq j \leq B_v$ 。

通过映射, 一个  $B_h \times B_v$  的二维信号就变成了只有  $B_h \times 1$  的一维信号, 如图 2 所示。

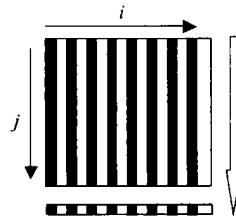
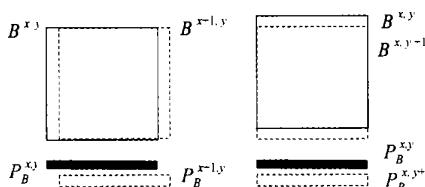


图 2 二维块的一维垂直映射

Fig. 2 Projection of a 2D Block

### 2.2 快速一维映射

如果进行全局运动估计, 则会有  $W \times H$  个不同的候选块要进行块匹配。 $B^{x, y}$  和其相邻的块  $B^{x+1, y}$  只有两列不同, 类似地,  $B^{x, y}$  和它相邻的块  $B^{x, y+1}$  只有两行不同, 所以根据  $P_{B_i}^{x, y}$  只需要计算不同的部分就可以有效地获得  $P_{B_i}^{x+1, y}$  和  $P_{B_i}^{x, y+1}$  的值, 如图 3。



### 3 $B^{x, y}$ 和 $B^{x+1, y}$ 、 $B^{x, y+1}$ 的位置关系和相应的映射关系

Fig. 3 Relation Between  $B^{x, y}$ ,  $B^{x+1, y}$  and  $B^{x, y+1}$ 

由  $P_{B_i}^{x, y}$  快速得到  $P_{B_i}^{x+1, y}$  和  $P_{B_i}^{x, y+1}$  的数学表达式如下:

$$P_{B_i}^{x, y+1} = P_{B_i}^{x, y} - B_{i, 0}^{x, y} + B_{i, B_v-1}^{x, y+1} \quad (2)$$

$$P_{B_i}^{x+1, y} = \begin{cases} P_{B_i}^{x, y}, & i < B_h - 1 \\ \sum_{j=0}^{B_v-1} B_{i, j}^{x, y}, & i = B_h - 1 \end{cases} \quad (3)$$

这样获得一个块的一维映射平均只需要两行或者两列的操作, 效率可提高 3~4 倍。

### 2.3 快速一维映射在 H. 264 中的应用

以往的运动估计算法中普遍采用 SAD 值 (sum of absolute difference, SAD) 来作为匹配准则。定义下面的公式为新的一维匹配准则:

$$P_{SAD}(dx, dy) = \sum_{i=0}^{B_h-1} |P_{C_i}^{x, y} - P_{R_i}^{x+dx, y+dy}| \quad (4)$$

式中,  $P_{SAD}$  为投影后的 SAD 值;  $P_C$  为当前帧中  $B_h \times B_v$  大小块的一维垂直投影;  $P_R$  为参考帧中待匹配块的一维垂直投影。如果将 SAD 放缩, 可以发现  $P_{SAD}$  和  $S_{SAD}$  有以下关系:

$$\begin{aligned} S_{SAD}(dx, dy) &= \sum_{i=0}^{B_h-1} \sum_{j=0}^{B_v-1} |C_{i, j}^{x, y} - R_{i, j}^{x+dx, y+dy}| = \\ &\sum_{i=0}^{B_h-1} \sum_{j=0}^{B_v-1} |C_{i, j}^{x, y} - R_{i, j}^{x+dx, y+dy}| \geq \\ &\sum_{i=0}^{B_h-1} |P_{C_i}^{x, y} - P_{R_i}^{x+dx, y+dy}| = P_{SAD}(dx, dy) \end{aligned} \quad (5)$$

由上面的不等式可以得到, 如果某个块的  $P_{SAD} > S_{SAD_{min}}$ , 则表明此块的  $S_{SAD} > S_{SAD_{min}}$ , 此时这个块必定不是最匹配的块。也就是说, 如果两个块对应的一维映射不能很好地匹配, 则这两个块也就不可能很好地匹配。这也正是一维映射可以作为块匹配准则排除不匹配位置的理论依据。

需要注意的是, 两个块的一维映射匹配只是两个块匹配的必要条件, 却不是充分条件, 所以即使两个块的一维映射匹配也不能保证两个块匹配。笔者只是使用一维映射匹配的方法来快速排除大部分不匹配的位置, 然后还需要在剩余的位置中使用二维匹配准则来准确寻找最匹配的块。

由于不可能提前得到  $S_{SAD_{min}}$ , 所以本文使用一个线性公式来估计  $S_{SAD_{min}}$ :

$$S_{SAD_{min}} \sim \alpha \cdot P_{SAD_{min}} \quad (6)$$

其中  $\alpha$  是一个可调因子, 由它来控制运动估计的

精度。相应地, 一维匹配排除准则就变为下面的不等式:

$$P_{SAD} > \alpha \cdot P_{SAD_{\min}} \quad (7)$$

也就是说, 如果某个候选位置满足上式, 则表明此块必定不是最匹配的块。而且可以观察到, 如果增加  $\alpha$  的值, 可以排除的候选位置就会减少, 但可以获得更准确的估计; 如果减小  $\alpha$  的值, 就可以排除更多的候选位置, 但同时会使估计的精度下降。

因为快速一维块映射的计算量远远小于二维块匹配, 所以在进行运动估计时, 可以极大地减少计算量。同时在 H.264 的 7 种运动估计模式中, 有大量的块匹配操作都是冗余的。例如, 当进行一个  $16 \times 16$  大小的块匹配时, 事实上已经做了 4 个  $8 \times 8$  大小的块匹配。所以可以并行处理所有 7 种块匹配操作, 这样就使得 7 种模式运动估计的计算量大大减少。

## 2.4 算法流程

本文提出的 PFS 算法流程如下。

1) 把当前宏块和参考宏块都分割成  $4 \times 4$  大小的子块。

2) 首先使用快速一维映射, 计算当前宏块和参考宏块中所有  $4 \times 4$  大小块的一维垂直映射。

3) 从模式 7 到模式 1 依次合并或者累加已经得到的一维垂直映射: 水平方向合并相邻的两个  $4 \times 4$  大小块的一维垂直投影, 就可以得到 8 个  $8 \times 4$  大小块的一维垂直投影; 垂直累加相邻的两个  $4 \times 4$  大小块的一维垂直投影, 就可以得到 8 个  $8 \times 4$  大小块的一维垂直投影。根据  $4 \times 8$  和  $8 \times 4$  大小块的一维垂直投影, 类似地, 可以得到所有 7 种块模式的一维垂直投影。

4) 对于每一种块模式, 如果  $P_{SAD} > \alpha \cdot P_{SAD_{\min}}$ , 则说明当前候选位置不是最匹配的块, 应该排除此位置。如果当前的  $P_{SAD} < P_{SAD_{\min}}$ , 则更新  $P_{SAD_{\min}}$ 。

5) 重复第 4 步, 执行完所有 7 种块模式的一维匹配操作。

6) 重复第 4、5 步, 在搜索窗口内执行完所有位置的一维匹配。

7) 对于一维映射准则没有排除的所有候选位置, 再使用 SAD 或者 MSE 等二维匹配准则从中选取最匹配的块。

## 3 试 验

使用 H.264 的 TML8 模型作为算法的仿真平

台, 测试使用了格式为 QCIF 的经典测试序列 foreman。编码序列的长度选取了 50 帧, 宏块的大小为  $16 \times 16$  点, 搜索窗口的大小为  $32 \times 32$  点。表 1 显示了 FS 和本文的 PFS 计算复杂度的比较。

表 1 对一个宏块进行运动估计时  
FS 和 PFS 的算法复杂度比较

Tab. 1 Speed Comparison Between FS and  
PFS for One Macroblock

算法	平均匹配次数	平均额外操作
FS	41 984 次 SAD	0
PFS ( $\alpha = 4$ )	1 024 次 PSAD, 6 560 次 SAD	15 360 次加操作, 2 048 次合并

可以看到, 表 1 使用 PFS, 计算速度大大提高。这是因为: ① 使用一维快速映射计算  $P_{SAD}$  的计算量远远小于计算  $S_{SAD}$  的计算量; ② 7 种块模式的匹配都是并行操作; ③ 额外只需要少量的加法和数据合并, 这些操作的开销远远小于系统性能的整体增益。

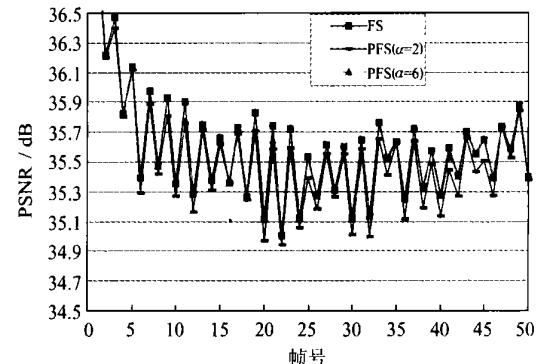


图 4 使用 FS、PFS( $\alpha=2$ ) 和 PFS( $\alpha=6$ ) 峰值信噪比的比较

Fig. 4 Performance Comparison Between  
FS, PFS ( $\alpha = 2$ ) and PFS ( $\alpha = 6$ )

图 4 显示了 FS 和 PFS 算法峰值信噪比的比较。当  $\alpha$  的取值从 6 到 2 逐渐减小时, 重建图像的质量有轻微下降, 但计算复杂度急剧减少。经过测试大量的序列可以发现, 选取  $\alpha=4$  时, 使用 PFS 输出的图像质量基本没有下降, 此时计算复杂度和搜索精度达到较好的平衡。

## 4 结 语

H.264 是 ITU-T 和 MPEG 组织联合制定的新一代视频编码标准, 它提供了丰富的编码算法支持, 所以压缩性能超越了所有现存的视频编码标准。但是由于引入的 7 种运动估计模式极大地增加了编码器的复杂度, 所以如果要将这一标准用于实际系统, 必须选用行之有效的快速算法。本文提出一种可以

并行处理 7 种模式的快速运动估计算法, 极大地减少了编码器耗费在运动估计中的计算量, 使 H. 264 成为一个实用的视频编码技术提供了有力工具。

## 参 考 文 献

- 1 Kuhn P. Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation, Kluwer, 1999
- 2 Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H. 264 | ISO/IEC 14496-10 AVC), 2003
- 3 Li R, Zeng B, Liou M L. A New Three-Step Search Algorithm for Block Motion Estimation. *IEEE Trans. Circuits System*, 1994, 4(4): 438~442

- 4 Po L M, Ma W C. A Novel Four-Step Search Algorithm for Fast Block Motion Estimation. *IEEE Trans. Circuits System*, 1996, 6(3): 313~317
- 5 Zhu S, Ma K K. A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation. *IEEE Transactions on Image Processing* 2000, 9(2): 287~290
- 6 Tu C J, Trac Tran, Topiwala P. A Hybrid Feature/ Image Block Motion-vector Estimation Approach (HyFIVE), VCEG 14th Meeting Santa Barbara CA, USA, 2001

第一作者简介: 喻占武, 教授, 博士生导师。主要从事多媒体视讯技术方面的研究。

E-mail: zw.yu@public.wh.hb.cn

## A High-Quality Motion Estimation Algorithm Applicable to H. 264

YU Zhanwu<sup>1</sup> HUANG Xiaojie<sup>1</sup> JIA Jia<sup>1</sup> FAN Li<sup>2</sup>

(1 Center of Multimedia Communications Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Basic Department, The Second Artillery Command College, 46 Gongnongbing Road, Wuhan 430012, China)

**Abstract:** A fast algorithm (PFS) applicable to H. 264 is proposed: utilizing one dimension projection to eliminate unwanted positions. The algorithm can make 7 modes of motion estimation perform almost simultaneously. The simulation results show that the proposed algorithm is up to 6 times faster than the exhaustive search algorithm, with almost identical performance.

**Key words:** FS; PFS; motion estimation; 1D projection; H. 264

**About the first author:** YU Zhanwu, professor, Ph. D supervisor. His research orientation is multimedia video communication.

E-mail: zw.yu@public.wh.hb.cn

(责任编辑: 涓涓)

(上接第 460 页)

arate noise from signal in difference images, and change information can be effectively concentrated into a few components of the final result. Another change detection method based on PCA is also described briefly for comparison.

Experimental results of a case study using Landsat5 TM imageries are presented to demonstrate the effectiveness of our method. And the characteristics of correlation between results and original imageries are discussed in detail.

**Key words:** orthogonal transformation; change detection; canonical correlation; minimum noise fraction transformation; principle component analysis; remote sensing

**About the first author:** ZHANG Lu, Ph. D candidate, majors in remote sensing data processing and its application.

E-mail: Lu.Zhang@163.com

(责任编辑: 涓涓)