

一种适用于 H. 264 的基于自适应搜索范围的快速运动估计算法

李翔 吴国威

(清华大学电子工程系, 北京 100084)

摘要 块匹配运动估计在视频编码中有着举足轻重的地位, 其性能的优劣在很大程度上影响着输出码流的质量。全搜索是效果最好的运动估计算法, 但其巨大的运算量是实际应用, 特别是实时应用无法承受的。为了解决这一问题, 各种快速算法不断涌现。本文提出了一种适用于最新视频编码标准——H. 264/MPEG 4-AVC 的快速运动估计算法。该算法基于自适应搜索范围, 利用视频图像序列的帧间统计特性以及运动向量时域、空域的相关性, 在保证 PSNR 性能的同时, 使运动估计部分的运算复杂度大为降低。实验仿真表明, 该算法适用面广, 对大运动与小运动序列都有很强的自适应能力。在保持与全搜索相同 PSNR 的同时, 平均速度超过全搜索 280 倍有余, 超过三步法 1.8 倍, 性能明显优于新三步法、四步法等经典快速运动估计算法。

关键词 H. 264/MPEG 4-AVC 运动估计 自适应搜索范围

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2004)04-0471-06

A Fast Motion Estimation Algorithm Based on Adaptive Search Range for H. 264

LI Xiang, WU Guo-wei

(Dept. of Electronic Engineering, Tsinghua University, Beijing 100084)

Abstract Block matching motion estimation (BMME) plays a very important role in video coding. The performance of BMME greatly affects the quality of the encoded bit stream. Full Search (FS) is the optimal algorithm for BMME. But the tremendous computation payload of it is unaffordable for practical applications, especially for real-time applications. To overcome this problem, a lot of fast algorithms were developed. In this paper, a novel fast motion estimation algorithm designed for the new video coding standard H. 264/MPEG 4-AVC is proposed. Based on an adaptive search range, the proposed algorithm makes use of the statistical characteristics between successive frames and the spatial and temporal correlations among the motion vectors of adjacent blocks to speed the process of motion estimation while maintaining the PSNR with no loss. The simulation results show that the algorithm is able to fit for all types of video sequences adaptively in spite of the degree of the motions. Compared with other algorithms, it is more than 280 times faster than FS (Full Search) and more than 1.8 times faster than 3SS (3 Steps Search) in average with no PSNR loss. And it is a better one than the traditional fast motion estimation algorithms, such as N3SS (New 3 Steps Search) and 4SS (4 Steps Search).

Keywords H. 264/MPEG 4-AVC, motion estimation, adaptive range search

1 引言

块匹配运动估计 (block matching motion estimation, BMME) 在目前已有的各个视频编码标

准中都有着举足轻重的作用。其原理是先将当前帧分割成相互不重叠的矩形块 (如大小为 16×16 像素), 然后在参考帧中寻找最佳匹配位置作为预测, 最后仅对位移向量 (运动向量) 与块残差进行编码传输。由于大多数视频图像序列变化缓慢, 于是仅传输

运动向量与相对较小的块残差能够很好的消除图像序列帧间冗余度,从而降低码率。

在块匹配运动估计中,运动估计的精度会对码率产生很大的影响。而全搜索以其极高的精度受到广泛的关注。但由于其要求对搜索窗内所有点进行匹配,所需运算量十分巨大,很难适应实际应用,特别是实时应用的要求。为了追求较高运动估计精度下的较低运算量,人们不断的提出各种快速算法。概览这些快速算法,大致可分为以下几类:

(1)通过在搜索窗内按一定规则,仅搜索较少的几个点来降低运算复杂度,经典算法有 3 步法(3SS)、新 3 步法(N3SS)^[1]、4 步法(4SS)^[2]等;

(2)通过使用相对较为简单的匹配准则来降低运算复杂度,如最小最大误差等^[3];

(3)通过利用代数不等式排除不可能成为最佳匹配的待搜索点,从而减少匹配点数,降低运算复杂度,如近两年提出的一些快速全搜索算法^[4]等。

在这些算法中,第 1 类算法受到最为广泛的关注,这主要是因为与其他两类相比,该类算法能够提供较大的速度提升。但与此同时,这类算法在 PSNR 方面的性能普遍欠佳,尤其是对大运动的序列,这主要是由于大部分该类算法所基于的两条假设:“现实世界大部分运动向量都很小”与“残差空间的单调特性”(即所谓的 Unimodal 模型)。这两条假设虽然使运动向量的模型得以简化,但同时也限制了快速算法的性能,即第 1 条假设会使快速算法在大运动的情况下性能迅速退化,而第 2 条假设则会使搜索容易陷入局部极小值。

针对上述块匹配运动估计快速算法中存在的问题,同时考虑到 JVT(joint video team, ITU-T 与 MPEG 的联合专家组)视频编码标准 H. 264/MPEG 4-AVC^[5]的新特性,本文提出了一种新的混合快速算法:扩展自适应范围搜索 EARS(extended adaptive range search)。该算法利用参考帧统计特性确定当前帧搜索范围,结合预测相关思想,可依据视频图像序列的自身特性,自适应地调整搜索复杂度。该算法适应面广,效率高。与 3 步法相比,平均 PSNR 高出近 0.2dB,平均速度提高近 1.8 倍。

2 算法介绍

为解决经典快速算法中存在的问题,使算法具有对不同序列更强的自适应性以及更强的鲁棒性,

分别采用自适应范围搜索(ARS, adaptive range search)与预测搜索(PS, predictive search)两种技术。

2.1 自适应范围搜索

在 H. 264/MPEG 4-AVC 中,为适应不同运动程度的视频图像序列,提高运动估计的性能,采用了 ± 32 点的搜索窗。这使得单块的全搜索点数上升到 4225 个点。扩大的搜索窗对于大运动序列,无疑可以进一步提高运动估计的精度,但对于较小运动序列,却是不必要的浪费。显然,如果能够根据序列本身的运动特性,自适应地确定一个搜索范围,使绝大部分最佳匹配点都落于该范围内,则可在保证 PSNR 方面性能的同时,降低运算量,提高效率。

考虑到运动向量可以认为是当前块从当前帧到参考帧的偏移量,则可以用运动向量的二阶矩来评价当前帧相对参考帧的运动程度。同时,研究表明视频图像序列是一个相对缓变的过程,从整体上看,其各帧运动向量的二阶矩变化也较为缓慢。图 1 给出了 table tennis 与 grandmother 序列各帧运动向量二阶矩的统计。

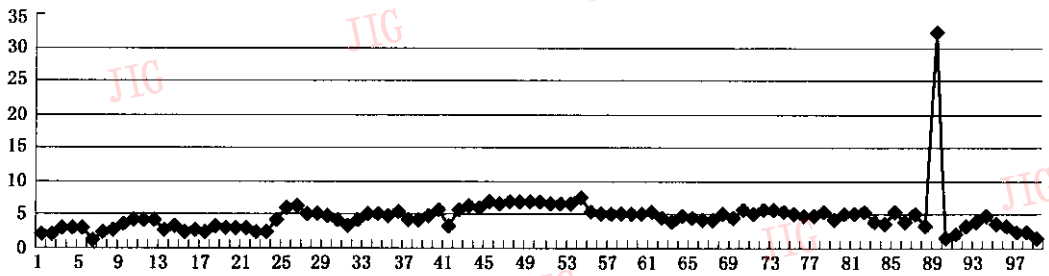
从图 1 可以看出尽管这两个序列运动特性明显不同,一个运动很大,一个运动很小,但其相邻帧运动向量二阶矩的变化幅度都相对较小(应指出的是,图 1 table tennis 中第 89 帧处的尖点对应镜头场景的变换,因此在考虑序列统计特性时,将该点特征予以忽略)。基于此项观察,提出建立在视频图像序列统计特性之上的自适应范围搜索,即根据前一帧运动向量的二阶矩来为后一帧确定一个较为合适的搜索范围,然后在该范围内进行搜索。定义

$$\begin{cases} D = \sqrt{\frac{\sum_{i=0}^{N_{MV}} MV_i^2}{N_{MV}}} \\ AR = 1.5 \times D \end{cases} \quad (1)$$

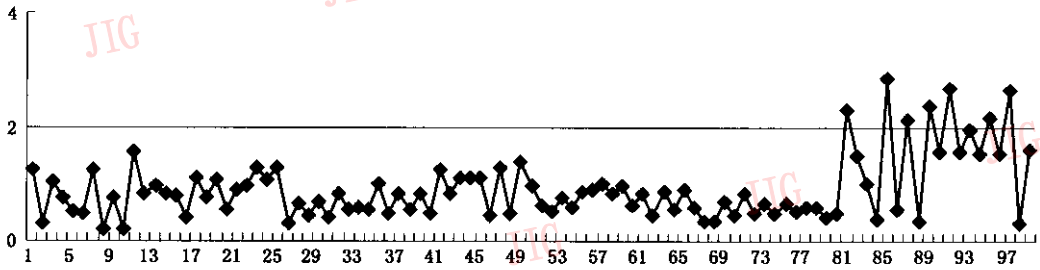
其中, N_{MV} 是前一帧运动向量的个数, D 为运动向量的二阶矩,而 AR 则为用于当前帧的自适应搜索范围。

得到 AR 后,按照如图 2 所示的方式进行自适应范围搜索。其中(0,0)点为静止点(运动向量为 0 的点),搜索范围为 AR ,即不搜索超出 AR 所限矩形范围的点。当最优优点出现在搜索点间距大于 1 的范围内时(如图 2 中的三角形点),采用类似三步法的方式进行再搜索,直至搜索间距为 1,停止搜索。

由于二阶矩体现了帧平均意义的运动速度,因



(a) table tennis 序列



(b) grandmother 序列

图 1 两序列各帧运动向量的二阶矩(前 100 帧)

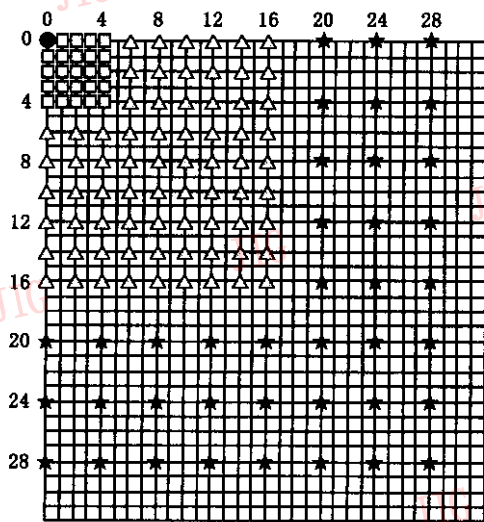


图 2 ARS 搜索方法

致的匹配误差。

由于 ARS 利用了视频图像序列自身的帧间统计特性,对不同运动特性的序列都有很好的自适应性,因此能够克服以往快速算法在大运动情况下的性能退化。即通过 ARS,能够较好地解决前面提到的第 1 条假设所带来的问题。

2.2 预测搜索

以经典的三步法为例,分析一下当 Unimodal 假设与实际不符时,导致的快速算法性能退化的原因。3 步法在进行搜索时,首先在匹配误差曲面上平均选取 9 个点,然后以其中误差最小点为中心,减小搜索半径,再采用相同方法继续搜索,直到半径为 1。当误差曲面不是单调的,即若存在多个极小点,在每一轮选取新的搜索中心时,便很容易陷入局部极小值,从而导致性能退化。

总结起来,在不一定单调的误差曲面上采用过大的搜索半径,将导致快速算法性能退化。于是,类比微积分的思想,人们很自然地提出,如果起始搜索点距离全局最优很近,那么在该小范围内,Unimodal 的假设将仍然保持正确。考虑到实际物体运动的整体性,无论是空域相邻的块还是时域相邻的块,其运动向量间都有很强的相关性。而利用这些相关性,则可得到十分接近全局最优的搜索起始点,于是既可使 Unimodal 模型更好的符合实际,又可提高搜索速度。这就是预测算法的主要思想^[6]。

此如果场景中存在多目标,且运动速度差异较大,那么得到的 AR 也会较大,即运动复杂的帧会有较大的自适应搜索范围。这虽然对帧内小运动目标造成一定的计算量浪费,但保证了运动估计的性能。而采用 1.5D 作为 AR 则是针对一帧内有且仅有小区域的快速运动,而其他目标相对静止的情况,这在一定程度上防止了因对该区域的搜索范围会过小而产生的匹配误差。同时,由于 D 是一个统计量,因此如果小区域的快速运动不足以引起 D 的明显变化,可以认为该区域为“小概率事件”。当综合考虑速度和匹配误差时,就有理由较少考虑由“小概率事件”所导

结合上述思想,选取当前块的左上方、上方、左边3个块,以及前一帧相同位置以及其周围8个块的相关运动向量作为预测(起始搜索点候选点)。同时,考虑到H.264/MPEG 4-AVC引入的“可以使用7种不同大小块进行匹配”的新特性,还将一些空域相邻,但类型不同的块的运动向量加入了起始搜索候选点。

具体进行预测搜索PS的步骤是:首先找到这些候选点中的最佳匹配点,然后按照文献[7]所述方式进行再搜索,直至该算法停止。

由于PS充分利用了运动向量在空域与时域上的相关信息,调整了搜索起始点,使Unimodal模型能够较好地符合实际,因此提高了搜索效率,能够较好地解决前面所述第2条假设所带来的问题。同时,它还可以使整帧的运动向量场较为平滑,从而产生十分好的效果。

2.3 初始搜索(IS)

初始搜索(initial search, IS)与ARS、PS相比,是不需要其他辅助信息的搜索方式,仅用于第1个P帧或场景突变(可通过每帧I块的数量来估计场景突变)。它针对尚无空域、时域以及自适应搜索范围等信息可以利用的情况。其搜索方式与图2所示类似,也可用全搜索代替。由于它不涉及算法主要思想,且对算法整体性能影响不大(一般从视频序列整体上看,场景突变的概率很小),因此不加以详细介绍。

2.4 本文算法

将以上算法功能模块结合,就构成了本文所提算法(EARS)。考虑到前面两个主要算法模块的不同特性,即PS功能模块搜索效率高而ARS功能模块自适应性强、性能稳定,首先进行PS中候选点的筛选。如果候选点中有优于静止点的,说明PS能够较为准确地确定最优点,于是进一步完成PS;反之,则进行ARS。

以上算法的执行步骤如下:

(1)计算静止点的匹配误差;

(2)进行PS候选点的筛选,如果候选点的最小匹配误差小于静止点,则采用文献[7]中的方式再搜索,然后转第5步;

(3)如果尚无法得到相应的自适应范围AR(如第1个P帧或场景突变),则进行IS,然后转第5步;

(4)进行ARS;

(5)保存最佳运动向量,停止。

3 仿真结果与分析

3.1 运算量分析

众所周知,块匹配的运算复杂度远远高于运动估计中的其他部分,因此它成为衡量快速算法优劣的主要标准。

若以SAD(sum of absolute difference)为匹配准则

$$SAD = \sum_{n=1}^N |currPel_n - predPel_n| \quad (2)$$

其中, $currPel_n$ 为当前点像素值, $predPel_n$ 为预测点像素值。

则每帧中块匹配所需运算量可用下式表示

$$\begin{cases} OP_{Add} = (B_W \times B_H - 1) \times MT \times \frac{F_W \times F_H}{B_W \times B_H} \\ \approx F_W \times F_H \times MT \\ OP_{Sub} = (B_W \times B_H) \times MT \times \frac{F_W \times F_H}{B_W \times B_H} \\ = F_W \times F_H \times MT \\ OP_{Abs} = (B_W \times B_H) \times MT \times \frac{F_W \times F_H}{B_W \times B_H} \\ = F_W \times F_H \times MT \end{cases} \quad (3)$$

其中, OP_{Add} 、 OP_{Sub} 、 OP_{Abs} 分别代表所需的加法、减法以及绝对值运算的次数, B_W 与 B_H 代表匹配块的宽与高, F_W 、 F_H 为每帧的宽与高, MT 则为每块平均进行块匹配的次数。

这里引入额外的运算为运动向量的预测以及每帧运动向量二阶矩的计算。其中预测除块匹配(计入块匹配的运算量)外,只涉及每块一次的运动向量保存(赋值运算),与块匹配相比,其可以忽略不计。对于运动向量的二阶矩,由式(1),其引入的每帧额外运算可表示为

$$\begin{cases} OP_{Add} = 2 \times \frac{F_W \times F_H}{B_W \times B_H} - 1 \\ OP_{Mul} = 2 \times \frac{F_W \times F_H}{B_W \times B_H} \end{cases} \quad (4)$$

其中, OP_{Mul} 代表平方时所需乘法的次数。计算AR每帧还需1次除法、1次乘法、1次平方根运算,与每帧的块匹配相比(根据式(3)),QCIF每帧所需的加法已不下10万次,其可忽略。考虑到绝对值与平方均可由查表实现,所以 OP_{Abs} 与 OP_{Mul} 可视为同一种运算。对比式(3)、式(4),考虑到一般 MT 至少大于10次,对 4×4 块而言,二阶矩所引入的额外运算量

不足块匹配的 $1/80$, 而对 16×16 块则更是不足 $1/1\,000$, 可以忽略。因此, 在后面的运算复杂度比较时, 将以块匹配作为衡量算法快慢的依据。

3.2 仿真结果与讨论

在 JVT 的 H. 264/MPEG 4-AVC 编码参考模型 JM 5.0 上实现本文所提算法。测试环境是根据 JVT 制定的编码器测试环境^[8]扩展而来。选择具有广泛代表性的 7 个标准测试序列: salesman (352×288 , 150 帧), carphone、Claire、foreman、grandmother (176×144 , 300 帧), flower garden、table-tennis (352×240 , 150 帧)。其中, table-tennis 为大运动序列, carphone、foreman、salesman 为中等运动序列, Claire、grandmother 为小运动或几乎静止序列, flower garden 则包含了较多的细节与镜头的平移。采用的帧率是 30Hz, 搜索范围是 ± 32 点, 单帧参考, 使用全部 7 种块进行匹配。编码后的序列, 除首帧外, 其余各帧均编码为 P 帧, 即采用 IPPP... 的方式。为便于比较, 还在 JM5.0 上实现了 3 个经典算法: 3 步法 (3SS)、新 3 步法 (N3SS) 与 4 步法 (4SS)。

需要特别指出的是, 由于 H. 264/MPEG 4-AVC 视频编码标准本身就已经融入了预测算法的相关思想 (中值预测), 因此在该平台上实现的快速算法也会自然地具有预测算法的特性, 其性能会比快速算法本身有所提高。

表 1 给出了在相同码率下, 4 种快速算法与全搜索 (FS) 相比, 各自亮度分量的峰值信噪比 (PSNR) 损失。其中 PSNR 由 JVT 指定工具 AVSNR^[8] 给出。由表 1 可以看出, 本文所提算法 (EARS) 具有极强的自适应性, 在小运动、大运动或是镜头平移、缩放的情况下, 都能保持优异的性能。与 FS 相比, 7 个不同运动特性序列中最大的 PSNR 损失只不过是 0.03dB, 平均 PSNR 相同, 其损失完全可以忽略不计。而对比 3SS、N3SS、4SS, 不难发现, 它们在小运动情况下, 还可保持良好的性能, 如 Claire 与 grandmother, 而对于中运动和大运动的序列, 则性能退化严重: 对于 carphone 序列, 3SS 与 4SS 的损失都在 0.2dB 以上, 而对于 table tennis, 3SS 的损失更达到了 0.37dB。

表 2 给出了 4 种快速算法与 FS 在速度方面的性能比较。采用 7 种块的平均搜索点数作为算法速度方面的评判依据。由表 2 可以看出 EARS 算法在速度上的自适应性: 对于如 salesman 等中小运动序

表 1 亮度分量 PSNR 比较

单位: dB

测试序列	3SS	N3SS	4SS	EARS
carphone	-0.28	-0.13	-0.20	-0.01
Claire	-0.04	-0.03	-0.06	-0.03
flower garden	-0.33	-0.03	-0.13	0.01
foreman	-0.15	0.00	-0.04	0.04
grandmother	-0.04	0.00	-0.02	0.00
salesman	-0.09	-0.03	-0.05	-0.02
table tennis	-0.37	-0.11	-0.15	-0.11
平均	-0.19	-0.05	-0.09	0.00

表 2 7 种块平均搜索点数比较

测试序列	FS	3SS	N3SS	4SS	EARS
carphone	4 225	41.00	19.50	18.27	17.73
Claire	4 225	41.00	17.56	17.24	10.55
flower garden	4 225	41.00	19.16	17.96	18.09
foreman	4 225	41.00	19.36	18.21	15.56
grandmother	4 225	41.00	17.46	17.16	11.65
salesman	4 225	41.00	17.89	17.48	10.67
table tennis	4 225	41.00	19.29	18.16	18.68
平均	4 225	41.00	18.60	17.78	14.70

列, 其速度可接近 FS 的 400 倍, 接近 3SS 的 4 倍, 达到 N3SS 与 4SS 的 1.5 倍以上; 而对于大运动序列, 如 table tennis, EARS 则会自动调节搜索点数到与 N3SS、4SS 相当的水平, 以适应复杂的运动。当从总体上考察速度方面的性能时, 可以看到, EARS 的平均搜索点数为 5 种算法中最少的, 平均速度是 FS 的 287.4 倍, 3SS 的 2.8 倍, 明显快于 N3SS 与 4SS。

综合两表数据, 不难看出, 本文所提算法 EARS 在 PSNR 与速度两方面均全面优于经典快速算法 3SS、N3SS 与 4SS, 具有很强的自适应性及鲁棒性。

4 结 论

本文提出了一种适用于最新视频编码标准——H. 264/MPEG 4-AVC 的快速运动估计算法 EARS。该算法以自适应搜索范围为基础, 利用视频图像序列的帧间统计特性以及运动向量时域、空域的相关性, 在保持 PSNR 方面性能不受损失的同时, 大大降低了运算复杂度。由于自适应范围搜索技术的使用, 使得该算法具有很强的自适应性, 在以往快速算法表现不佳的大运动情况下, 仍能保持良好的性能。在 JVT H. 264/MPEG 4-AVC 的编码参考模型 JM5.0 上实现了算法 EARS。实验仿真表明, EARS 无论在速度还是在 PSNR 方面, 都要优于经

典快速算法 3SS、N3SS 与 4SS。在保持与 FS 相同 PSNR 的同时,平均速度还可达到 FS 的 287 倍, 3SS 的 2.8 倍,明显快于 N3SS 与 4SS。

参考文献

- 1 ITU-T Rec. H. 264 | ISO/IEC 14496-10 AVC-Joint Committee Draft(2002).
- 2 Li Renxiang, Zeng Bing, Liou Ming L. A new three-step search algorithm for block motion estimation[J]. IEEE Transactions on Circuits System Video Technology, 1994, **4**(4): 438~442.
- 3 Po Lai-Man, Ma Wing-Chung. A novel four-step search algorithm for fast block motion estimation [J]. IEEE Transactions on Circuits System Video Technology, 1996, **6**(3): 313~317.
- 4 Chen Mei-Juan, Chen Liang-Gee, Chiueh Tzi-Dar, *et al.* A new block-matching criterion for motion estimation and its implementation [J]. IEEE Transactions on Circuits System Video Technology, 1995, **5**(3): 231~236.
- 5 Kim Jong-Nam, Kang Dae-Kap, Byun Sung-Cheal, *et al.* A fast full search motion estimation algorithm using sequential rejection of candidates from hierarchical decision structure [J]. IEEE Transactions on Broadcasting, 2002, **48**(1): 43~46.
- 6 Tourapis Alexis M, Au Oscar C, Liou Ming L. Highly efficient predictive zonal algorithms for fast block-matching motion estimationM[J]. IEEE Transactions on Circuits System Video Technology, 2002, **12**(10): 934~947.

- 7 Liu Lurng-Kuo, Feig Ephraim. A block-based gradient descent search algorithm for block motion estimation in video coding[J]. IEEE Transactions on Circuits System Video Technology, 1996, **6**(4): 419~422.
- 8 Sullivan G. Recommended simulation common conditions for H. 26L coding efficiency experiments on low-resolution progressive-scan source material[A]. In: ITU-T VCEG[C], Santa Barbara, CA, USA, 2001.



李翔 1978 年生, 2001 年毕业于清华大学电子工程系, 现为该系 2001 级硕士研究生。研究方向为视频编码。



吴国威 1935 年生, 教授, 1958 年毕业于清华大学电子工程系无线电技术专业。研究领域包括信号处理、图像处理、模式识别以及数字视频广播等。已发表论文 80 余篇。