

基于运动复杂度的快速运动估计算法

王 洁, 张太镒, 乐培玉

(西安交通大学 电子与信息工程系, 陕西 西安 710049)

(janewj006@163.com)

摘 要:提出了一种基于运动复杂度的快速运动搜索算法(Motion Complexity-based Searching, MCS)。根据运动复杂度来决定是否采用空间相关性来判定初始搜索点的位置,在确定了初始搜索点后采用与其相对应的十字搜索模式。仿真实验结果验证了此搜索算法的有效性。

关键词:运动估计;运动复杂度;空间相关;十字搜索

中图分类号:TP391.41 **文献标识码:**A

Fast motion searching algorithm based on motion complexity

WANG Jie, ZHANG Tai-yi, YUE Pei-yu

(Department of Electronics & Information Engineering, Xi'an Jiaotong University, Xi'an Shaanxi 710049, China)

Abstract: A fast motion searching algorithm based on motion complexity(MCS) was proposed. It was decided whether to adopt spatial correlation to determine the position of the first searching point according to motion complexity. And after the position of the first searching point was determined a comparative cross searching pattern was adopted. The results of imitation experiments prove the validity of the motion estimation algorithm.

Key words: motion estimation; motion complexity; spatial correlation; cross searching

0 引言

视频编码技术在电视会议、视频电话、数字电视、高清晰度电视、固定卫星服务、多媒体邮件等视频通信服务中起着越来越重要的作用。视频序列在时间上有很强的相关性,因此图像帧间的冗余度很高,而运动估计是降低此冗余度,提高图像压缩比的关键技术。^[1]

运动估计通常采用的是块匹配算法(Block Matching Algorithm, BMA)。块匹配算法的实现效果取决于三个因素:搜索范围、匹配准则和搜索算法。搜索范围可以根据图像的剧烈程度加以调整。常用的块匹配准则有:绝对平均误差函数(SAD)、均方误差函数(MSE)、互相关函数(CCF)等,为了减少运算量,便于硬件的实现,本文将采取 SAD 准则:

$$SAD(x, y) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |F_n(i, j) - F_{n-1}(i+x, j+y)| \quad (1)$$

其中, $-W \leq x, y \leq W$, F_n, F_{n-1} 分别为当前块和前一帧中的对应宏块, $N \times N$ 为宏块大小,在本文中 $N = 16$ 。 W 为运动矢量在水平和垂直方向上的最大位移,搜索窗的大小为 $(2W+1)^2$ 。

搜索算法决定了运动估计的运动搜索的复杂度。目前运动估计的搜索算法有很多种,其中全搜索算法(FS)^[2],由于精度很高,应用也比较广泛,但是它的搜索复杂度很高,运算量很大,不易于硬件实现。为了减少 FS 算法的运动搜索复杂度和数据读取复杂度,近年来出现了许多改进的搜索算法,如新三步搜索算法(N3SS)^[3]、四步搜索算法(4SS)^[4]、交叉搜索算法(CS)^[5]、钻石搜索算法(DS)^[6]等。但是它们都是以

原点作为初始搜索中心的,没有充分利用相邻块之间的运动相关性。因此有人提出了采用相邻块的运动矢量,来预测当前块运动矢量的方法。但是如果当前块与相邻块的相关性不大,或是并不存在相关性,那盲目地采用相邻块的运动矢量来预测,只会使搜索陷入误区。

为此,本文提出一种基于运动复杂度的方法来进行初始搜索点的判定,并根据中心偏置分布特性,对双十字搜索算法进行改进。仿真结果表明,本文提出的基于运动复杂度的搜索算法(MCS)与 FS 算法和综合性能较其他算法优越,与被 MPEG4 国际标准采纳并收入验证模型的 DS 算法相比较,其精度相当,复杂度大幅度降低,对于运动相对剧烈的序列, MCS 的准确性略高。

1 预测初始搜索点的判定

初始搜索点通常直接选择参考帧的(0,0)位置,这种方法简单,但容易陷入局部最优点。由于在运动图像中,一个对象的运动在时间和空间上有很强的相关性。因此有人提出用空间上的相邻块,或是前一帧相对应块的运动矢量来预测当前块的运动矢量。但这是建立在相邻块与当前块运动相关的前提下的。相反,如果图像的运动复杂度非常高,也就是说块与块之间的相关性很小,那么盲目地用相邻块的运动矢量来预测当前块,会使搜索陷入误区,以致加大运算量,降低精度。本文用运动复杂度高低来进行预测初始搜索点的判定。

如图 1^[8],可见块 A、B、C 与当前块相关性较大,而图像的时间相关性相对于空间相关性则较小,而且利用时间相关将需记录整个前一帧的运动矢量域,复杂度较高,实际执行中

收稿日期:2005-09-16 收稿日期:2005-11-14

作者简介:王洁(1980-),女,浙江杭州人,硕士研究生,主要研究方向:MPEG2 图像编解码; 张太镒(1944-),男,陕西西安人,教授,博士生导师,主要研究方向:新一代移动通信技术、软件无线电、无线带宽接入技术、数字图像和数字音频压缩编码; 乐培玉(1981-),男,江苏南京人,硕士研究生,主要研究方向:基于 TI DSP 的 MPEG 视频编解码。

不理想。所以本文采用空间上的相邻块 A, B, C 的运动向量 MV_1, MV_2, MV_3 来计算运动的复杂度 C 。其中 $C(x), C(y)$ 分别代表水平和垂直方向的运动复杂度, $MV_i(x), MV_i(y)$ 分别代表运动向量 MV_i 的水平和垂直分量:

$$C(x) = \sum_{i=1}^3 \left| MV_i(x) - \frac{1}{3}(MV_1(x) + MV_2(x) + MV_3(x)) \right| \quad (2)$$

$$C(y) = \sum_{i=1}^3 \left| MV_i(y) - \frac{1}{3}(MV_1(y) + MV_2(y) + MV_3(y)) \right| \quad (3)$$

设阈值 T , 如果 $C(x), C(y)$ 中有一个大于阈值 T , 我们认为这部分的运动复杂度较高, 不易用相邻块的运动向量来估计当前块, 则搜索点直接从参考帧的原始位置开始; 如果 $C(x), C(y)$ 都小于 T , 则认为运动复杂度较低, 相关性较大, 可以用相邻块的运动向量来确定当前块搜索的起始点。 $E(x), E(y)$ 分别为当前宏块运动向量的预测值 E 的水平和垂直分量:

$$E(x) = \frac{1}{3}(MV_1(x) + MV_2(x) + MV_3(x)) \quad (4)$$

$$E(y) = \frac{1}{3}(MV_1(y) + MV_2(y) + MV_3(y)) \quad (5)$$

当前宏块的中心点为 (i_0, j_0) , 预测点为 (i_1, j_1) , 则 $i_1 = i_0 + E(x), j_1 = j_0 + E(y)$ 。

时间相关性 (第 $t-1$ 帧)

0.140 0.168 M	0.161 0.186 I	0.141 0.168 N
0.151 0.186 H	0.166 0.218 G	0.160 0.201 J
0.148 0.146 L	0.179 0.166 K	0.163 0.159 O

空间相关性 (第 t 帧)

0.385 0.395 D	0.520 0.521 B	0.426 0.377 C
0.488 0.487 A	Current Block	

图1 时空相关性(包括水平、垂直两个分量)

2 基于运动复杂度的搜索算法(MCS)

表1 不同序列的运动矢量在不同范围内的概率

格式	图名	3×3 范围内的概率(%)		1×1 范围内的概率(%)
		起始点为 (i_0, j_0)	起始点为 (i_1, j_1)	起始点为(i_1, j_1)
QCIF	Foreman	86.51	90.17	84.03
	Miss-American	97.50	99.90	99.02
	silent	97.84	98.54	97.55
CIF	stefan	82.10	86.47	82.28
	football	36.82	56.07	42.11
	moblie	98.93	99.72	99.28

实际中绝大部分图像具有中心偏置的分布特性, 即它们

的运动矢量通常总是高度集中分布在搜索窗的中心位置附近。也就是说绝大部分的块具有较小的运动矢量, 较少的块具有较大的运动矢量。表1给出不同序列的运动矢量的统计结果。可见对于不同的图像序列, 大部分的运动矢量都分布在 3×3 的范围内, 特别是当起始点为预测点的时候, 概率就进一步提高了, 即使对于运动相当剧烈的图像 football 来说这个概率也超过了 50%。因此 3×3 范围内的搜索速度将对整个搜索起决定性作用。另外, 统计了在采用预测点为搜索起点时, 1×1 范围内运动矢量的概率。发现此时大部分图像 80% 以上的运动矢量都在 1×1 的搜索范围内, 甚至有些概率高达 99%。

在分析了图像的中心偏置分布特性后, 我们也发现图像在水平或垂直方向的运动居多数, 而同时在水平和垂直方向运动的较少。也就是说运动矢量出现在 X, Y 方向上的概率很大^[9]。

根据图像的以上两点特性, 本文采用改进的大小十字搜索模型(如图2所示)。

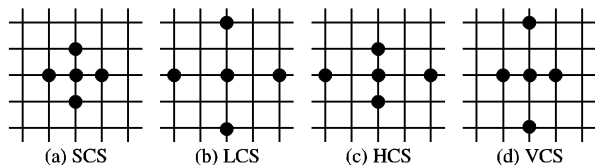


图2 搜索模型

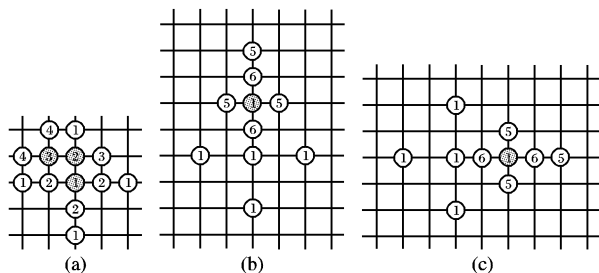


图3 搜索步骤(灰色为每步的 SAD 最小点, 其余的为搜索点)

当运动复杂度 C 大于阈值 T 时, 即搜索起始点为参考帧的原始位置, 我们将用大十字模型 LCS 开始搜索。

1) 搜索 LCS 的中心点和 4 个十字点, 如果中心点的 SAD 值最小, 则认为运动较小, 采用 SCS 继续搜索, 进入 2); 否则, 认为运动幅度较大, 采用变形的十字模型, 至 5)。

2) 如图 3(a), 搜索中心点周围的 SCS 的 4 个顶点, 如果中心点为 SAD 最小值, 则停止搜索, 输出中心点对应的运动矢量; 否则, 到 3)。

3) 以上一步的 SAD 最小点为中心点继续以 SCS 模型进行搜索, 但此时只要搜索中心点上下, 或左右两点。如果中心点的 SAD 值最小, 则停止搜索, 输出此时的运动矢量; 否则, 到 4)。

4) 以上一步的 SAD 最小点为中心点继续以 SCS 模型进行搜索, 此时也只要搜索两点, 左上两点, 或左下两点, 或右下两点, 或右上两点。重复 4), 直到中心点的 SAD 值最小, 输出此时的运动矢量。

5) 以上一步的 SAD 最小点为中心点进行搜索, 如果最小点在水平线上, 我们认为水平运动的概率较大, 则采用 HCS 模型搜索; 如果最小点在垂直线上, 则垂直运动的概率较大, 则采用 VCS 模型进行搜索。如果中心点为最小点, 则 6); 否则, 重复 5)。如图 3(b)~(c)。

6) 搜索中心点周围两点, SAD 最小点为最终点, 输出此时的运动矢量。

当运动复杂度 C 小于阈值 T 时, 即搜索起始点为预测点时, 根据上面统计的结果, 由于运动矢量在 1×1 范围内的概率很高, 所以我们采用小十字模型 SCS 进行搜索, 直到中心点为最小点为止, 输出其运动矢量。

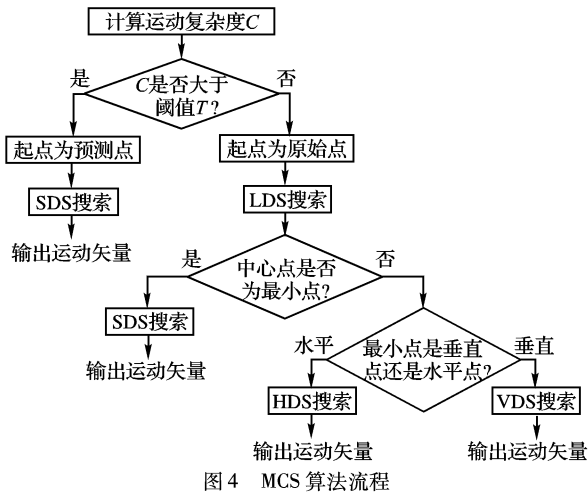


图4 MCS算法流程

3 实验结果与分析

下面用全搜索算法(FS)、钻石搜索算法(DS)与MCS比较, 从以下三方面来分析MCS算法的有效性: 1) 复杂度, 平均每个宏块的搜索点数; 2) 峰值信噪比(PSNR); 3) 准确性, 相对于FS算法的平均运动矢量偏移距离。

实验使用 Visual C++ 6.0 编程实现, 在 MPEG2 的 MP@LL 下进行, 搜索范围为: 15×15 , SAD 匹配准则, 色度格式为 4:2:0, 帧频为 25 frame/s, 压缩码率为 1.152Mbps。

采用以下不同性质序列:

1) CIF (Common Intermediate Format) 格式, 图像大小 352×288 ;

- a) mobile (300 帧): 物体的缓慢平移;
- b) football (90 帧): 剧烈运动序列;
- c) Stefan (90 帧): 背景几乎静止, 主体运动剧烈。

2) QCIF (Quarter Common Intermediate Format) 格式, 图像大小 176×144 。

- a) Foreman (400 帧): 物体移动以及镜头一般性旋转和平移;
- b) Container (300 帧): 物体缓慢移动;
- c) Miss-American (150 帧): 背景静止, 主体运动微小。

3.1 复杂度比较

用平均每个宏块的搜索点数来衡量, 搜索点数越少, 计算的复杂度就越小, 花费的时间也越少。

表2 平均每个宏块搜索点数

算法	CIF			QCIF		
	mobile	football	Stefan	Foreman	Container	Miss-American
FS	478.11	478.11	478.11	430.67	430.67	430.67
DS	14.39	28.42	17.16	14.41	10.49	11.62
MCS	5.13	16.73	7.03	6.91	4.41	5.08

由表2可以看出, FS 平均搜索的点数最多, DS 算法次之, 但已经大大减少了平均搜索点数, MCS 算法平均搜索的

点数最少, 比 DS 算法几乎减少了一半多。这充分体现了 MCS 算法的优越性。

3.2 峰值信噪比比较

算法的性能, 采用峰值信噪比 PSNR 来衡量:

$$PSNR = 10 \times \lg \left(\frac{x_{\max}^2}{MSE} \right) \quad (6)$$

其中 $MSE = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (f(m, n) - f'(m, n))^2$; x_{\max} 为 255; M, N 分别为图像的长宽; $f(m, n), f'(m, n)$ 分别为原图和重建图像对应的像素灰度值。

表3 峰值信噪比(PSNR)

算法	CIF			QCIF		
	mobile	football	Stefan	Foreman	Container	Miss-American
FS	28.437	38.633	33.362	42.399	46.980	44.966
DS	28.436	37.666	32.776	42.272	46.985	44.968
MCS	28.447	38.394	33.112	42.331	46.984	44.971

由表3可见, 基本上 FS 算法得到的 PSNR 最优, 但是个别序列也可能是 DS 算法或是 MCS 算法得到最优的 PSNR。总的来说, MCS 算法的性能与 FS 不相上下, 略优于 DS 算法。

3.3 准确性比较

准确性比较将采用相对于 FS 算法的平均运动矢量偏移距离来衡量。FS 得到的是最优运动矢量, 记为 (F_x, F_y) ; 新算法得到的运动矢量记为 (M_x, M_y) 。则偏移距离 $Dist = \sqrt{(M_x - F_x)^2 + (M_y - F_y)^2}$, $Dist$ 越小说明新的运动矢量越接近全局最优值。

表4 平均运动矢量偏移距离(Dist)

算法	CIF			QCIF		
	mobile	football	Stefan	Foreman	Container	Miss-American
FS	0	0	0	0	0	0
DS	0.107	4.812	1.026	0.532	0.020	0.316
MCS	0.112	4.677	0.743	0.502	0.021	0.342

由表4可见, 对于 football, Stefan, Foreman 这三个序列来说, MCS 得到的 $Dist$ 优于 DS; 而其余三个序列则是 DS 的 $Dist$ 优于 MCS。我们可以说两种算法的准确性相当, 对于运动相对剧烈的序列, MCS 的准确性略高。

通过以上的比较分析, 说明 MCS 算法与 FS 和 DS 相比不但大大减少了平均搜索点数, 而且在性能和准确性上都与 FS 以及 DS 不相上下, 甚至在某些序列中优于 DS 算法。

4 结语

本文提出了一种基于运动复杂度的搜索算法 MCS, 根据运动复杂度的大小来决定初始搜索点的位置。同时, 充分利用中心偏置分布, 以及图像的运动在水平或是垂直方向居多的特点, 在不同的运动复杂度下采用不同的十字搜索模式进行搜索。实验结果表明, MCS 算法在保证性能和准确性的前提下, 减少了搜索点数, 加快了运动估计的速度。

参考文献:

- [1] 钟玉琢, 王琪, 赵黎, 等. MPEG-2 运动图像压缩编码国际标准及 MPEG 的新进展[M]. 北京: 清华大学出版社, 2002.

虑到嵌入式的特点,算法应当满足计算量小,所占用存储空间也小的要求。

本文算法在背景建模时,采取了使用动态权值矩阵判决背景像素的方法来降低建立背景模型的计算量,在计算判决动态背景与运动目标所使用的状态权值矩阵 $M_{i,j}(t)$ 和状态变化频率矩阵 $P_{i,j}(t)$ 时,只需要根据现时刻和前一时刻的状态权值和状态变化频率进行简单统计,所以计算量较小。在存储方面,只需在内存中存储三个 $N \times N$ 的矩阵(背景图像信息、状态权值矩阵、状态变化频率矩阵),与 LOTS^[6] 和 KDE^[11] 等算法相比,本文只需要较小的存储空间。

但是,本算法也有其不足之处。由于需要统计一段时间内的状态权值,所以在程序刚开始运行或者全局光照突变时,会经历 λ 时长的不稳定阶段,在此阶段中会将所有运动的像素(包括运动前景和运动背景)都当作前景分割出来,但经过 λ 后,就会进入稳定阶段,此时这些动态背景像素将被融合到背景中,所以在检测动态背景像素时会表现出一定的延迟。 λ 的大小与算法中 T 的取值成正比。此外,该算法没有考虑摄像机全局运动的情况。

3 实验结果

为了验证算法的有效性,我们在室外场景中对一个运动的人进行跟踪。场景中包含了运动目标、静止的背景、动态背景像素。在该实验中我们成功地将静止的背景和动态的背景像素融合到背景像素中,而把运动目标从视频序列中抽取出来。为了更好地显示跟踪效果,我们将抽取出来的物体用矩形标识出来。

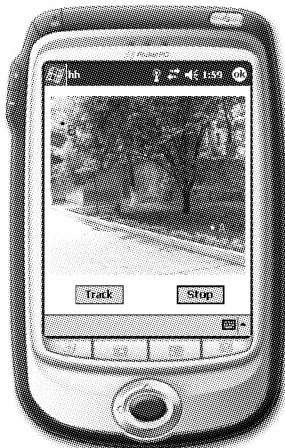


图1 背景场景

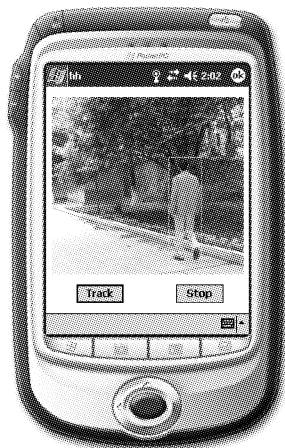


图2 检测到的运动物体

我们将该算法运行在嵌入式操作系统pocketpc2003中,

其 CPU 为 300MHz, RAM 为 64MB。图 1 为初始背景场景, A、B 两点是动态背景, 其中 A 点为场景中有规律运动的小草, B 点为随风摆动的树叶。在图 2 为视频序列中第 382 帧图像, 图像中被矩形标识出来的部分就是运动的对象, 而 A 点、B 点并没有当作前景分割出来, 而是被融合到背景中去了。

参考文献:

- [1] PICCARDI M. Background subtraction techniques: a review[A]. 2004 IEEE International Conference on Systems, Man and Cybernetics[C]. 2004, 4. 3099 - 3104.
- [2] KANG S, KOSCHAN A, ABIDI B, *et al.* Video Surveillance of High Security Facilities[A]. Proceedings of 10th International Conference on Robotics & Remote Systems for Hazardous Environments [C]. Gainesville, FL, 2004. 530 - 536.
- [3] 冯继超. 面向 21 世纪的嵌入式系统及发展方向[J]. 工业控制计算机, 2001, 14(5).
- [4] CHEN SY, MA SY. Efficient Moving Object Segmentation Algorithm Using Background Registration Technique[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(7): 577 - 585.
- [5] TSAIG Y, AVERBUCH A. Automatic Segment of Moving Objects in Video Sequences: A Region Labeling Approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(7): 597 - 612.
- [6] BOULT TE, MICHEALS R, GAO X, *et al.* Framerate omnidirectional surveillance and tracking of camouflaged and occluded targets[A]. Second IEEE Workshop on visual Surveillance[C], 1999. 48 - 55.
- [7] MARVILLE M. A framework for high-level feedback to adaptive, per-pixel mixture-of-gaussian background models[A]. IECCV[C], 2002.
- [8] STAUFFER C, GRIMSON WEL. Adaptive background mixture models for real-time tracking[A]. Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C]. 1999, 2.
- [9] ZIVKOVIC Z, van der HEIJDEN F. Recursive Unsupervised Learning of Finite Mixture Models[J]. IEEE Transactions On PAMI, 2004, 26(5).
- [10] ZIVKOVIC Z. Improved Adaptive Gaussian Mixture Model for Background Subtraction[A]. Proceedings of the 17th International Conference on Pattern Recognition[C], 2004, 2. 28 - 31.
- [11] MITTAL A, PARAGIOS N. Motion-based background subtraction using adaptive kernel density estimation[A]. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C], 2004, 2. II-302 - II-309.
- [12] 朱斌, 何成杰, 叶家骏, 等. 基于嵌入式系统的视频图像传输系统[J]. 电子技术, 2003, 30(1): 22 - 24.

(上接第 597 页)

- [2] 刘滔, 张益贞. VC++ 实现 JPEG/MPEG 编解码技术[M]. 北京: 人民邮电出版社, 2002.
- [3] LI R, ZENG B, LIDU ML. A New Tree-Step Search Algorithm for Block Motion Estimation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1994, 4(4).
- [4] PO L-M, MA W-C. A Novel Four-Step Algorithm for Fast Block Motion Estimation[J]. IEEE Transactions on Circuits Systems for Video Technology, 1996, 6(3).
- [5] GHANBARI M. The Cross-Search Algorithm for Motion Estimation[J]. IEEE Transactions on Circuits, 1990, 38(7).
- [6] ZHU S, MA K-K. A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation[J]. IEEE Transactions on Image Processing, 2000, 9(2).
- [7] ISO/IEC JTCL/SC29/WG11 N2932. MPEG-4 Video Verification Model Version 14.0[S], 1999.
- [8] ANTONIO C, CLSUDIA F, PAU D. A Complexity-Bounded Motion Estimation Algorithm[J]. IEEE Transactions on Image Processing, 2002, 11(4).
- [9] 李炜, 周兵, 李波. 运动矢量场自适应搜索算法[J]. 计算机学报, 2003, 26(2).