

一种适用于 MPEG-4 形状编码的快速运动估计算法^{*})

倪 伟 郭宝龙

(西安电子科技大学智能控制与图像工程研究所 西安710071)

摘 要 运动估计是 MPEG-4 形状编码的一项关键技术,本文提出了一种适用于形状编码的快速运动估计算法。算法首先在参考帧中进行扫描,得出视频对象的二位边界掩模;在匹配运算时使用 1bit 的异或运算代替原有的加法运算;设定有效的中止准则,对于静止点直接中止搜索;在搜索过程中采用了渐进消除算法,能够在不影响搜索精度的前提下减少搜索点。实验结果表明使用本文的快速搜索算法,运动估计中的运算量比 MPEG-4 VM 原有搜索算法有较大幅度的降低,且编码后的码字长度与原算法基本一致。

关键词 视频编码,运动估计,MPEG-4,形状编码

A Fast Motion Estimation Algorithm for MPEG-4 Shape Coding

NI Wei GUO Bao-Long

(Institute of Intelligent Control and Image Engineering, Xidian University, Xi'an 710071)

Abstract Motion estimation is a key technology for MPEG-4 shape coding. In this paper, a fast motion estimation algorithm is proposed. By examining the content of Alpha plane, a binary mask for the effective search area can be generated. In the searching procedure, the traditional add and plus operation are replaced by 1bit XOR operation, the successive elimination algorithm is used which can lower computational complexity while maintaining search accuracy. Experiment results show that the proposed algorithm can reduce the computational complexity of shape coding significantly and be suitable for real-time software and hardware applications.

Keywords Video coding, Motion estimation, MPEG-4, Shape coding

1 引言

基于视频对象(Video Object, VO)的编码是 MPEG-4^[1] 的一项关键技术。传统的视频编码标准大都采用基于 DCT 变换的混和编码算法,而 MPEG-4 打破了这种矩形帧分块编码的限制,通过 VO 的引入提供了基于对象的编码功能。在编码时能够针对帧图像中任意形状的对象和背景进行编码;解码时可以根据需要对各对象进行不同的组合和编辑。基于对象的编码方案大大提高了视频图像的压缩率,同时还使得 MPEG-4 具备了强大的交互性和灵活性。

VO 由独立的视频对象面(Video Object Plane, VOP)构成,而每个 VOP 又包括纹理信息(YUV)和 Alpha 平面。Alpha 平面的作用是刻画 VO 的形状信息和空间属性,因而必须进行形状编码。在编码过程中使用运动估计以消除时间冗余,使用基于上下文的算术编码算法(Context based Arithmetic Encoding, CAE)编码。然而在 MPEG-4 VM 的编码过程中,运动估计部分占据了形状编码运算量的 90% 以上,不利于实时应用。而近年来所提出的一些快速运动估计算法,如 TSS、DS、HEXBS^[2]、MVFAST^[3] 等,大多只适用于普通的纹理编码。为了实现快速高效的形状编码运动估计,本文提出了一种用于形状编码的快速运动估计算法,在编码后 VOP 码字长度相当的条件下,运动估计的运算量得到了大幅度下降。

2 MPEG-4 中的运动估计算法

VOP 形状编码基于 16×16 的二值 Alpha 块(Binary Alpha Block, BAB),当 VOP 的编码模式为 P-VOP(前向预测模式)或 B-VOP(双向预测模式),此时必须对待编码 BAB 进

行运动估计,得到形状运动矢量(Motion Vector for Shape, MVS)^[4]。MPEG-4 中定义阈值 AlphaTH 以区分有损和无损两种模式的形状编码。形状编码的运动估计分为两步:

(1) 首先计算出 MVS 的预测点(MVPS)。MVPS 由与当前 BAB 相邻 3 个 BAB 的运动矢量和对应纹理宏块的 3 个相邻运动矢量所决定,按照 {MVS1, MVS2, MVS3, MV1, MV2, MV3} 的顺序进行扫描,第一个有效的运动矢量被置为当前 BAB 的运动矢量预测点 MVPS,如果在该集合没有运动矢量满足条件,此时将 MVPS 置为原点(0, 0)。

(2) 在得到 MVPS 后,将当前 BAB 分为 4×4 像素的子块,对于所有 4×4 子块,如果计算的 MC 误差小于或等于阈值 16×AlphaTH,则匹配过程结束,该 MVPS 就是所需的运动矢量 MVS,即 MVS=MVPS。如果上述条件不成立,则以该 MVPS 为中心,在 [-16, 16]×[-16, 16] 的搜索窗内进行搜索。在 MPEG-4 VM 中检测匹配窗中的所有候选点,得出匹配窗内和原始 BAB 相似度最高的搜索点,即 MVS。

3 形状编码的快速运动估计算法

MPEG-4 VM 中所采用的 FS 算法对匹配窗中的所有候选点都进行计算,搜索精度高,但也不可避免带来了计算量过大的缺陷^[5-6]。事实上,基于 BAB 的运动估计具有不同于纹理编码的固有特性,如果能够充分利用这些有利特性,便能够有效提高算法效率。据此,本文提出了一种适用于 MPEG-4 形状编码的快速运动估计算法。

3.1 二值边界掩模

为提高编码效率,MPEG-4 对形状编码设定了如下限制:如果 BAB 完全位于 VOP 内部或完全位于 VOP 区域外部,此

^{*}) 基金项目:国家自然科学基金资助项目(69975015),教育部优秀青年教师资助计划项目(2003)。倪 伟 博士生,研究方向为多媒体通信和图像处理。郭宝龙 教授,博士生导师,研究方向为电路与系统、图像通信等。

时直接编码,无需进行运动估计。因此在搜索窗内进行穷尽的完全搜索是不必要的,只有含有边界的参考帧 BAB 才是有效搜索区域;如果参考帧的 BAB 只含有背景、或完全属于 VOP 内部,那么该 BAB 一定不属于最优匹配块,搜索时可以跳过这部分 BAB。我们将参考帧中所有有效区域的集合定义为边界掩模(boundary mask, BM)。构造 BM 的过程如下:

将参考帧分为 16×16 像素的 BAB;对所有的 BAB 进行扫描,如果当前 BAB 中的像素全部一致,那么说明该 BAB 是单一的背景或位于 VOP 内部,属无效搜索区域,此时将这些

BAB 宏块全部置0;当 BAB 宏块具的点具有不同的数值,说明当前 BAB 宏块中必然含有边界边界,是有效搜索区域,此时将该 BAB 的元素全部置1。当扫描完所有的 BAB 宏块后,参考帧中标记为1的区域即为二值边界掩模。在搜索 MVS 时,所有位于 BM 外的搜索点都无需进行搜索,从而能有效降低运算量。

图1(a)中给出了 Clarie 序列中的一帧图像,(b)是(a)中 VOP 的 Alpha 平面,(c)中的白色区域便是该 VOP 的二值边界掩模 BM。

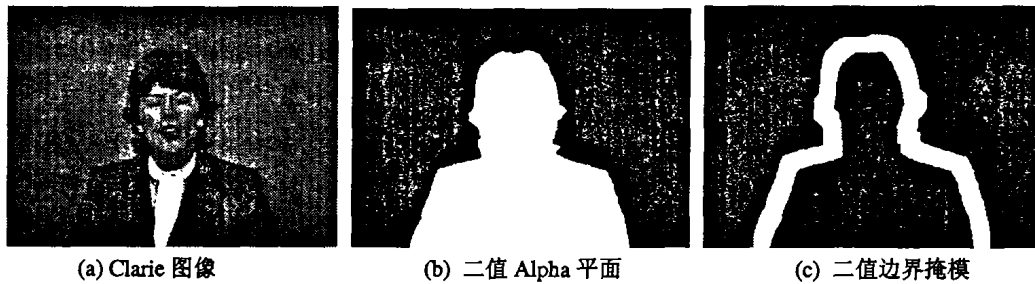


图1 边界掩模的构造示例

3.2 匹配准则

在运动估计过程中需要按照匹配准则对各搜索点进行相似度计算,具有最小匹配准则函数值的搜索点就是 MVS,所以匹配准则直接影响运动估计的速度和效率。MPEG-4 VM 中使用了形状绝对误差和准则(Sum of Absolute Difference for Shape, SADS),定义如下:

$$SADS(i, j) = \sum_{x=1}^M \sum_{y=1}^N |f_k(x, y) - f_{k-1}(x+i, y+j)| \quad (1)$$

其中 $f_k(x, y)$ 和 $f_{k-1}(x, y)$ 分别表示当前 BAB 和参考帧 BAB 的像素点, (i, j) 为搜索点的偏移量,搜索窗尺寸为 $M \times N$ 。

进行形状编码的 BAB 以二维矩阵的形式表示形状信息。如果我们用0表示该点位于 VOP 区域之外,而1表示该点位于 VOP 区域内,则定义二值 Alpha 块误差准则(BABD)为:

$$BABD(i, j) = \sum_{x=1}^M \sum_{y=1}^N f_k(x, y) \oplus f_{k-1}(x+i, y+j) \quad (2)$$

在式(2)中, $\oplus$ 表示异或运算(XOR),具有最小 BABD 准则函数值的搜索点就是 MVS。

使用 SADS 准则进行运算时需要进行 $2 \times M \times N$ 次加减运算和 $M \times N$ 次取绝对值运算;而 BABD 准则使用 XOR 运算代替加法运算,只需要 $M \times N$ 次异或运算和 $M \times N$ 次加法运算。相比较于 SADS 的加减运算,逻辑函数 XOR 具有更快的软硬件执行速度。实验证明使用 BABD 准则进行相似度匹配,其计算速度可以达到 SADS 准则的1/4左右。

3.3 搜索中止准则

MPEG-4 VM 中采用的运动估计算法只考虑了残差 BAB 对编码的影响,如何快速找到相似度最高的 BAB 是其唯一目标。事实上,运动估计服务于形状编码,具有最小 SADS 值并不意味着最好的性能,如何寻找到最有利于编码的 MVDS 才是形状编码运动估计的目的。由于在 MPEG-4 形状编码中采用了 CAE 算法,对 MVS 和 MVPS 的运动矢量差值 MVDS 进行编码,在相邻的运动矢量差别较大的情况下,这种编码算法反而会增加码长。而在保证搜索精度的前提下,一致的运动矢量在进行差分编码时利于提高压缩比。

为此设定中止阈值 TH_S ,在搜索时使用 MVPS 作为初始 MVS 候选点。设当前 BAB 在 MVPS 处的 MC 误差为 BAB_{DMVPS} ,若当前 BAB 在该点的 MC 误差小于阈值 TH_S ,即:

$$BABD_{MVPS} < TH_S \quad (3)$$

则认为 MVPS 就是最优 MVS,此时终止运动估计运算。通过式(3)的搜索中止准则,搜索过程能够在保证足够精度的前提下适时中止,避免出现与周围运动矢量差别较大的运动矢量,从而使 VOP 的运动矢量场具有较好的均匀性,提高形状编码效率。

3.4 搜索算法

搜索方式是决定运动估计效率的另一个重要因素。MPEG-4 VM 采用全搜索法,搜索匹配窗内的所有候选点,运算量大,实时性较差。SEA 算法(Successive Eliminating Algorithm)^[7,8]是由 W. Li 和 E. Salari 提出的一种无损运动估计算法,其特点是能够通过判定不等式判别出匹配窗中的部分无效搜索点,这些搜索点无需做进一步运算,从而达到降低运算量的目的。与其它快速搜索算法不同,SEA 是一种无损预测算法,其搜索精度与 FS 完全相同。在本文中,我们将 SEA 算法与 BAB 的二值特性以及 BABD 准则相结合,提出了一种基于 BABD 准则的二值 SEA 搜索算法,能够对 MPEG-4 形状编码运动估计中的无效搜索点进行判定。

设当前 BAB 中像素为 $f_k(x, y)$,其参考 BAB 为 $f_{k-1}(x, y)$,匹配窗尺寸为 $M \times N$ 。由于 $f_k(x, y)$ 和 $f_{k-1}(x, y)$ 均为1bit的二值变量,即 $f_k(x, y), f_{k-1}(x, y) \in \{0, 1\}$,则必有:

$$f_k(x, y) - f_{k-1}(x+m, y+n) \leq f_k(x, y) \oplus f_{k-1}(x+m, y+n) \quad (4)$$

$$f_{k-1}(x+m, y+n) - f_k(x, y) \leq f_k(x, y) \oplus f_{k-1}(x+m, y+n) \quad (5)$$

对于匹配窗中的所有点使用式(4)和式(5),并在等式两边分别进行累加运算,即:

$$\sum_{x=1}^M \sum_{y=1}^N f_k(x, y) - \sum_{x=1}^M \sum_{y=1}^N f_{k-1}(x+m, y+n) \leq \sum_{x=1}^M \sum_{y=1}^N (f_k(x, y) \oplus f_{k-1}(x+m, y+n)) \quad (6)$$

$$\sum_{x=1}^M \sum_{y=1}^N f_{k-1}(x+m, y+n) - \sum_{x=1}^M \sum_{y=1}^N f_k(x, y) \leq \sum_{x=1}^M \sum_{y=1}^N (f_k(x, y) \oplus f_{k-1}(x+m, y+n)) \quad (7)$$

在式(8)和式(9)中,不等式右方就是就 BABD 函数 $BABD(m, n)$;若令不等式左方

$$F_1 = \sum_{x=1}^M \sum_{y=1}^N f_1(x, y),$$

$$F_2(m, n) = \sum_{x=1}^M \sum_{y=1}^N f_{1-1}(x+m, y+n) \quad (8)$$

其中 F_1 表示当前 BAB 中的所有元素之和; F_2 表示 MVS = (m, n) 处 BAB 的所有元素之和, 则式(6)和式(7)可进一步简化为:

$$\begin{aligned} F_1 - F_2(m, n) &\leqslant BABD(m, n) \\ F_2(m, n) - F_1 &\leqslant BABD(m, n) \end{aligned} \quad (9)$$

假设当前已经计算得出的最小 BABD 函数值为 $BABD_{\min}$, 在后续搜索过程中, 只有当 (m, n) 点处的 $BABD(m, n)$ 函数小于现有的 $BABD_{\min}$ 时才说明 (m, n) 是具有更高匹配度的 BAB, 需要进行匹配计算, 此时必然有

$$BABD(m, n) < BABD_{\min} \quad (10)$$

因此, 式(9)也满足如下不等式:

$$F_1 - F_2(m, n) \leqslant BABD_{\min} \quad F_2(m, n) - F_1 \leqslant BABD_{\min} \quad (11)$$

最后我们得到如下判定不等式:

$$F_1 - BABD_{\min} < F_2(m, n) < F_1 + BABD_{\min} \quad (12)$$

也就是说, 如果在 (m, n) 点处的 MVS 匹配度更高, 则必然满足式(12); 反之, 若参考点像素和 $F_2(m, n)$ 不满足式(12)的条件, 则 (m, n) 一定不是最优运动矢量。因此通过式(12)的判定不等式, 我们便可以通过所在位置的像素点之和判断该点是否有可能为最优匹配点, 只有满足此式的候选块才有必要进行下一步的计算, 否则不予考虑; 对于不满足式(12)的点则可以直接排除, 而无需对每个搜索点都进行全部

匹配计算, 以有效降低运动估计的运算量。在另一方面, 不等式(12)是最优点的必要条件, 因此使用式(12)的运动估计精度与 FS 算法完全一致, 不会造成任何误匹配。

3.5 算法流程

根据上述思想, 可以将本文算法表述如下:

step1: 对参考平面进行扫描, 得出参考平面中的边界掩模 BM;

step2: 依次搜寻当前 BAB 的形状及纹理邻块, 得到预测点 MVPS。令 $MVS = MVPS$, 计算此时的 BABD, 如果满足 $BABD < TH_s$, 则直接中止搜索, 此时最优 MVS 即为 MVPS, 结束搜索; 否则转 step3;

step3: 对当前搜索点进行判断, 如位于 BM 区域内, 则转 step4; 如不在 BM 区域内, 则转向下一搜索点, 继续执行 step3;

step4: 使用基于 BABD 准则的二值 SEA 算法进行搜索, 得出最优 MVS。

4 实验结果与分析

实验采用 4 个 QCIF 格式 (176×144) 的图像序列: News、Akiyo、Football 和 Foreman, 使用本文算法和 MPEG-4 VM 原有算法, 分别在 10 帧/秒和 30 帧/秒的条件下执行无损形状编码, 搜索范围为 $[-16, 16]$ 。实验中算法阈值设定为 $TH_s = 5$ 。表 1 列出了两种情况下本文算法与 MPEG-4 VM 算法的性能比较结果。其中 Bit/VOP 表示对 VOP 编码的比特数, 平均搜索点数代表每个 VOP 运动估计中搜索点数的平均值。

表1 算法性能比较

序列	运动估计算法	10帧/秒			30帧/秒		
		Bit/VOP	平均搜索点数		Bit/VOP	平均搜索点数	
News	MPEG-4 VM	238.3	176038	—	173.6	521649	—
	本文算法	238.6	31933	18.14%	174.1	104069	19.95%
Akiyo	MPEG-4 VM	255.7	322751	—	147.7	732511	—
	本文算法	256.2	80107	24.82%	148.2	161885	22.10%
Football	MPEG-4 VM	1087.6	3981953	—	753.9	8052021	—
	本文算法	1092.5	1658483	41.65%	754.8	3476826	43.18%
Foreman	MPEG-4 VM	283.3	294768	—	193.7	818462	—
	本文算法	285.2	116345	39.47%	194.9	313225	38.27%

分析表1可以发现, 与 MPEG-4 VM 中的算法相比, 本文算法在码字长度基本相同的条件下幅提高了搜索速度。在静止 BAB 占有较大比重的图像序列 Akiyo 中, 本文算法的搜索点数仅相当于 VM 的 24.82% 和 22.10%, 对于 News 序列也仅为 18.14% 和 19.95%; 对于运动较为剧烈的序列 Football 和 Foreman 来说, 本文算法的运算量也有显著改善。

结论 本文分析了 MPEG-4 形状编码运动估计的一些固有特性, 在此基础上提出了一种适用于 MPEG-4 形状编码的快速运动估计算法, 它包括 4 个方面: 构造边界掩模获得有效搜索区域、使用 BABD 匹配准则加速匹配、使用高效的搜索中止准则对静止点及时加以判定和渐进消除的搜索方法。实验结果表明运动估计的运算量得到了大幅度降低, 而编码后的码字长度与 MPEG-4 VM 中的算法基本一致, 在搜索速度和编码效率面具有明显的优势。

参考文献

- 1 MPEG-4 Video Verification Model Version 18.0[S]. ISO/IEC 130.
- 2 JTC1/SC29/WG11 N3908, Jan. 2001
- 3 Zhu C, Lin X, Chau L. Hexagon-based search pattern for fast block motion estimation [J]. IEEE Trans. Circuits System Video Technology, 2002, 12(5):349~355
- 4 Zhu S, Ma K K. A new diamond search algorithm for fast block-matching motion estimation [J]. IEEE Trans. on Image Processing, 2000, 9(2):287~290
- 5 Yu D, Jang S K, Ra J B. Fast motion estimation for shape coding in MPEG-4[J]. IEEE Trans. on Circuits and System for Video Technology, 2003, 13(4):358~363
- 6 Panusopone K, Chen X. A fast motion estimation for MPEG-4 arbitrarily shaped objects [C]. ICIP'2000, (1): 876~879
- 7 Tsai T, Chen C. A fast binary motion estimation algorithm for MPEG-4 shape coding [J]. IEEE Trans. on Circuits and System for Video Technology, 2004, 14(6):908~913
- 8 Li W, Salari E. Successive elimination algorithm for motion estimation [J]. IEEE Trans. on Image Processing, 1995, 4(1):105~107
- 9 Wang Y, Tu G. Successive elimination algorithm for binary block matching motion estimation [J]. Electronic Letter, 2000, 36(24): 2007~2008