

约束条件帧间差值运动估值匹配准则*

肖自美 肖舟 梁凡 郑伟国

(中山大学电子系, 广州 510275)

摘要 运动估值是视频压缩编码的核心。针对块匹配算法中广泛采用的 MAD 准则的不足, 提出了一种约束条件帧间差值运动估值匹配准则, 通过理论分析和实验比较, 证实了新匹配准则的优越性。

关键词 视频压缩编码, 运动估值, 匹配准则

1 引言

运动估值是活动图象压缩编码的关键环节, 对消除时间轴(temporal)冗余度, 提高编码效率有极其重要的影响。在相继的图象序列中, 运动估值应能够估计或预测出要编码的下一幅图象的象素值, 再将需要压缩的当前图象的各象素值减去其相应的估计值(预测值), 如果这个估计值良好, 需传送的仅是一个极小的残留差值。这些差值代表了需要传输的信息。如果该估计值是完美的, 那么帧间差值图象中所有象素差值都是零, 而无需传送任何新的信息。估值(预测)的目标是象素的帧间差值为零或接近于零。

如果图象序列中出现某区域的运动, 帧间差值通常不为零, 从统计上来说可能非常大。造成大的帧间预测差值的图象序列的类型主要包括剧烈的运动和对比度尖锐的细节。

运动估值一般假定场景中的物体运动仅出现平移, 平移导致了物体在帧与帧之间的位移。也即图象序列的帧与帧之间的变化可近似看作构成物体的那些象素在相继帧之间产生了位移。实际上, 在绝大多数的自然景物中, 物体的运动不是杂乱无序的, 在大多数情况下可以局部地近似地用平移来表示。基于这个理由, 视频帧间编码可采用“有运动补偿的”

(motion compensated)预测, 用相对的垂直和水平方向的平移, 以补偿局部的运动, 这就是运动估值/运动补偿(ME/MC)的基本思想。

传统的运动估值方法分为两类: 象素递归法(Pixel Recursive Algorithm, PRA)和块匹配法(Block Matching Algorithm, BMA)。前者以单象素为基础进行递归运动估值, 其准确性较好, 但计算极为复杂, 且对噪声敏感, 实际中很少采用。块匹配法将图象分割为固定大小的象素块, 通过比较被匹配块与候选匹配块的相似(匹配)程度来确定运动估值。

运动补偿的实现和优劣关键在于运动估值。运动补偿的效率取决于运动估值的准确性。估值准确, 当前帧绝大多数象素可由前一帧预测而得到, 传送信息量很少, 码率大为降低; 反之, 运动估值不准确, 甚至出现误匹配(mismatch), 将导致码率上升。因此运动估值的准确性是运动估值算法的主要关键。成为运动估值研究的一个热点。

一个视频编码器最主要计算复杂度在运动估值(ME)上。由于 ME 的计算量庞大, 实时视频编码或者难于实现或者成本昂贵。为了解决 ME 计算量庞大这一问题, 许多快速搜索块匹配算法应运而生, 如二维对数搜索算法、三步搜索法、共轭方向搜索法、正交搜索法、交叉搜索法、分层搜索法等^[1]。它们只适用于少量图象类型而不具有普遍性; 或者估值精

* 广东省重点科技攻关项目资助(96220202)。

收稿日期: 1996年10月; 收到修改稿日期: 1997年1月。

准确性欠佳,应用范围还不广泛。仍是运动估值研究的另一个热点。

各种运动估值算法欲达到实用,必须在运动估值的精确性和计算复杂度两者之间权衡。虽然多种改进的算法仍在深入研究和发展中,相对而言,块匹配法获得了广泛应用。目前的国际标准如 H. 26x 和 MPEG 都推荐使用块匹配法作运动估值。

本文针对块匹配法中广泛应用的 MAD 准则的精确性不足,提出一种准确性高的运动估值算法。

2 块匹配运动估值算法的基本原理

块匹配法最基本的方法是全搜索匹配法(Full Search Matching, FSM)。其原理是:将图象的当前帧(第 K 帧)划分为固定大小 $M \times N$ 像素的图象子块,并假定位于同一图象子块内的所有像素具有相同的位移,对于每一个子块只需计算一个运动矢量。

设 $S_K(m, n)$ 为图象序列中第 K 帧位于坐标 (m, n) 位置的像素值。为了寻找第 K 帧中大小为 $M \times N$ 像素的子块(CB)在前一帧(参考帧,第 $K-1$ 帧)中相对应的位移子块,预先确定一个搜索区域(SW),其几何关系如图 1 所示。

搜索窗大小为 $SW = (M+2d) \times (N+2d)$,其中 d 为垂直和水平方向的最大位移。

$$NCCF(i, j) = \frac{1}{MN} \frac{\sum_{m=1}^M \sum_{n=1}^N S_K(m, n) \cdot S_{K-1}(m+i, n+j)}{\left[\sum_{m=1}^M \sum_{n=1}^N S_K^2(m, n) \right]^{\frac{1}{2}} \left[\sum_{m=1}^M \sum_{n=1}^N S_{K-1}^2(m+i, n+j) \right]^{\frac{1}{2}}} \quad (2)$$

均方误差准则:

$$MSE(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N [S_K(m, n) - S_{K-1}(m+i, n+j)]^2 \quad (3)$$

平均绝对帧间差值准则 MAD(Mean of the Absolute Frame Difference):

$$MAD(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N |S_K(m, n) - S_{K-1}(m+i, n+j)| \quad (4)$$

上述几种主要的匹配准则,所给出的运动估值准确度相差不大^[2]。(2)式的相关峰值匹配方法和(3)式的最小均方误差匹配方法均需乘法运算,较为复杂。由于 MAD 没有乘法运算,计算量较小,得到了普遍应用。

3 现有块匹配准则存在的问题

采用 MAD 匹配准则进行最简单可靠的穷尽搜索法(全搜索),对搜索范围内的每一点都计算

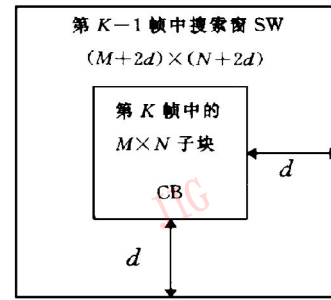


图 1 块匹配算法搜索区域几何关系示意图

Fig 1. The search area of BMA.

定义一个匹配误差函数 $D(i, j)$

$$D(i, j) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N G\{S_K(m, n), S_{K-1}(m+i, n+j)\} \quad (1)$$

对于每一位移 (i, j) , $-d \leq i, j \leq d$ 。

FSM 算法在寻找运动矢量过程中,将第 K 帧中大小为 $M \times N$ 像素的子块在搜索区域 SW 内滑动,共有 $Q = (2d+1)^2$ 种可能匹配的选择方块,也就是进行搜索的次数。在一定的准则下,使匹配误差函数 $D(i, j)$ 值为最小(或最大)的位置 (I, J) 即为最佳匹配位置。得到位移矢量 (I, J) 。

最常用的匹配准则有如下几种:

归一化二维互相关函数(NCCF)准则:

MAD 值,总共进行 $Q = (2d+1)^2$ 次匹配运算,在所有 Q 个 $MAD(i, j)$ 中,求出最小 $MAD(i, j)$,即:

$$MAD(I, J) = \min_{i, j \in [-d, d]} [MAD(i, j)] \quad (5)$$

$MAD(I, J)$ 即对应于最佳匹配。

块匹配准则是将图象分割成块,孤立地逐块进行匹配运算。这里涉及一个基本假设, MAD(或 NCCF, MSE)最小是否能准确地反映真实的运动和图象的细节。我们举一简单例子说明 MAD 块匹配准则的不足。图 2 是一个 4×4 待匹配像素块和两个参考匹配像素块。

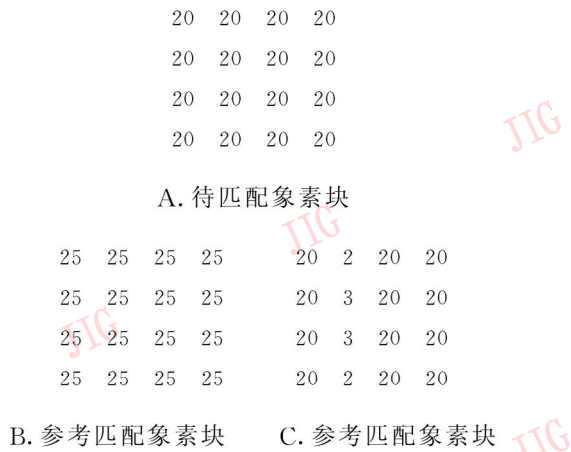


图 2 MAD 匹配的简单例子

Fig. 2 An example for shortage of MAD criterion.

分别计算像素块 A 与像素块 B、C 之间的 MAD 值

$$MAD(A, B) = 80/16,$$

$$MAD(A, C) = 70/16$$

显然, $MAD(A, C) < MAD(A, B)$, 按 MAD 匹配准则, C 像素块应为最佳匹配块。但从主观质量评价看, 最佳匹配块不是 C 而是 B。

实际上, 像素块 A 和 B 都是图象的平坦区域, 它们对应像素值之差均在 5 以内, 人眼主观感觉不出它们之间的明显差别, 反之像素块 C 却存在一条明显的黑色竖线, 与 A 的主观感觉相差甚大。

从上述简单例子可以看到 MAD(或 NCCF、MSE) 准则只是简单地将像素块之间的差异进行累加平均, 并以此平均效果作为匹配程度的判断准则, 没有考虑两匹配像素块中各像素之间的差值大小程度, 忽视了像素块内的细节特征, 因而造成了主、客观评价的差异, 带来了运动估值的不精确。

4 基于视觉特征的约束条件帧间差值运动估值新匹配准则

4.1 新匹配准则的基本思想

景物中物体的剧烈运动和图象的丰富细节(边缘、纹理等)是形成帧间预测差值大误差的主要原因。大的帧间预测误差易被人眼视觉感知, 造成的图象损伤也最为敏感。任何大误差形成的图象损伤都是局域性的(localized), 将局域性的少量的大误差在整个匹配块内进行空间域统计平均, 显然降低了匹配准则的精确性和可靠性, 匹配块尺寸愈大, 这种不精确性和不可靠性愈严重。最小 MAD 或最小 MSE 准则从主观视觉角度并非最佳准则, 我们可将

MAD 或 MSE 的误差分解为小误差与大误差之和, 凡大误差的数值或数量超过一定大小, 即使 MAD 或 MSE 最小, 也将排除在最佳匹配之外。应最大限度充分利用人们的视觉特性以获得更为精确、可靠的运动估值匹配准则。

4.2 新的运动估值匹配准则

大小为 $M \times N$ 的像素块之间对应像素之差值的绝对值为 $K(0 \leq K \leq 255)$ 的像素数目用 n_K 表示(图 1)。如果搜索区域 SW 中的一像素块 PB 是被匹配像素块 CB 的候选匹配块, 计算不同的帧间对应之差值绝对值 K 和相应的像素数目 n_K , 作出 $n_K \sim K$ 分布曲线(图 3)。

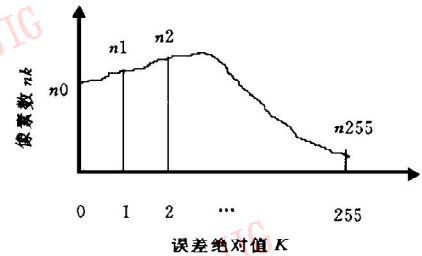


图 3 预测误差绝对值与像素数关系曲线

Fig. 3. Distribution of the absolute value of predictive error and the number of pixels.

像素块内总像素数为 $M \times N$, 有

$$\sum_{K=0}^{255} n_K = M \times N \quad (6)$$

两个像素块匹配愈好, 图 3 中的曲线将越靠近原点呈狭窄脉冲状特性。当两个块达到完美匹配, 即完全一致时, $n_K \sim K$ 曲线退化成为一条位于原点、幅度为 $M \times N$ 的竖线, 即 $K = 0, n_0 = M \times N$ 的竖线。

如果在横坐标轴上取某一阈值 T (T 为正整数), 计算如下的数值:

$$N_T = \sum_{i=0}^T n_i \quad (7)$$

N_T 表示像素块 PB 与像素块 CB 中对应像素的差值的绝对值小于或等于 T 的所有像素的数目。如果像素块 PB 与像素块 CB 匹配愈好, 在适当的 T 值下, N_T 值就会越大; 也就是说在一定的 T 值下, 如果 N_T 值越大, 表明两个像素块间彼此相近的像素数目越多, 因此两个像素块匹配越好。

根据以上分析可以看出: N_T 值的大小可以用来度量两个像素块之间的匹配程度。我们将这一思想应用到运动估值中, 可以得到以下算法:

在搜索区域 SW 中, 用全搜索的方法进行匹配运算, 选取一个门限 T , 在每一次匹配运算时符合约

束条件:

$$|S_k(x, y) - S_{k-1}(x + i, y + j)| \leq T, x \in [0, M - 1], y \in [0, N - 1] \quad (8)$$

对于每对 (i, j) , $i, j \in [-d, d]$ 可得到符合约束条件(8)的按(7)式累加而得的 $N_T(i, j)$ 值。

进行匹配搜索的运算次数:

$$Q' \leq Q \leq (2d + 1)^2 \quad (9)$$

相应地得到 Q' 个 $N_T(i, j)$ 值。求出最大的 $N_T(i, j)$, 即:

$$N(I, J) = \max_{i, j \in [-d, d]} [N_T(i, j)] \quad (10)$$

与 $N(I, J)$ 对应的象素块就是所寻找的最佳匹配块, (I, J) 便是运动矢量。

新算法中将帧间象素绝对差值大于某个阈值的那些象素排除在匹配准则 $N_T(i, j)$ 的计数之外, 在比较匹配程度过程中不予计算在内。因为运动估值的基本假设是物体的平移, 运动估值精确程度愈高, 出现帧间预测大误差的概率愈小。新算法将大误差摒弃, 从而避免了由于统计平均将大预测误差掩盖以致造成可能出现的误匹配, 提高了运动(位移)估值的准确度。我们称这一新匹配准则为约束条件帧间差值运动估值匹配准则 (Restrictive Conditional Interframe Difference, RCID)。

上述准则中 T 值的选定十分关键。如果 T 值得太大, 将会使所有象素匹配块间的对应象素的差值都小于 T 值, 从而反映不出到底哪些象素是匹配的; 如果 T 值得太小, 则可能使小于 T 的象素的数目太少而反映不出象素块间其它象素之间的差异, 因而也反映不出匹配程度。在通常情况下, 一般取 5 到 10 之间的正整数为宜。

5 实验结果和讨论

为了比较约束条件帧间差值匹配准则 (RCID) 与 MAD 准则下块匹配算法的性能, 我们分别运用两种准则进行了实验。

采用 Miss America 和 Claire 两种标准的序列图象, 其大小为 CIF 352(288, 象素子块大小为 16×16 宏块, 采用全搜索。Miss America 的搜索范围 $d = (15 \text{ pel})$, 以第 30 帧预测第 31 帧。Claire 的搜索范围 $d = \pm 7 \text{ pel}$, 以第 10 帧预测第 11 帧。

在两种匹配准则下, 经过 ME/MC, 得到恢复图象的 PSNR 值(表 1)。

表 1. 两种匹配准则的实验结果

Table 1 Experimental results of performance of RCID and MAD matching criteria (dB)

	Miss America		Claire		
	RCID	MAD	RCID	MAD	
T	PSNR	PSNR	T	PSNR	PSNR
5	37.6		5	38.2	
7	36.9	35.5	7	37.3	36.1
9	37.9		9	38.1	

通过以上理论分析和实验结果比较, 我们得到如下结论:

(1) RCID 匹配准则与 MAD 匹配准则相比, 所预测的图象的主客观质量得到了提高, 客观质量评价 PSNR 提高了 1—2 dB, 主观质量评价也有改善提高。说明新的匹配准则能较好的搜索匹配象素块, 运动估值的准确度明显提高, 所求得运动矢量更为精确。

(2) 约束条件中 T 值的选取直接影响运动估值的准确度, 选择合适的 T 值有利于找到最佳匹配块, T 值取 5—10 为宜。

(3) 算法复杂度取决于搜索和运算的方法。RCID 匹配准则没有采用乘除法复杂运算, 仅运用了减法, 求绝对值, 比较, 累加等简单运算, 与 MAD 算法复杂度相当。

RCID 匹配准则是一种高性能的运动估值匹配准则, 它提供了比 MAD 准则更高的估值准确度。

运动估值的估计精确度和实时性是视频压缩中的两个基本问题和一对矛盾。在目前技术条件下, 虽然块匹配法已达到实用程度, 但随着应用的深入、广泛, 对运动估值的精度和速度提出了更高要求。本文仅从提高运动估值精度角度提出了一种新的准则, 性能有所提高。

参考文献

- Ghanbari M. The cross-search algorithm for motion estimation. IEEE, trans. Commun. 1990, 38(7):950~953.
- Musmann H G, Pirsch P, Grallert H J. Advance in picture coding. IEEE 1985, 73(4):523~548.



肖自美 1961年毕业于武汉大学无线电专业,1981—1984年留学德国。现任中山大学电子系教授,中山大学信息与通信技术研究中心主任,中国图象图形学会常务理事。曾获1993年度国家教委科技进步二等奖。主要从事图象压缩编码与处理,语音压缩与处理,多媒体技术与通信系统,数字传输与调制技术等领域的研究和教学工作。

A Novel Motion Estimation Match Criterion of Restrictive Conditional Interframe Difference

Xiao Zimei Xiao Zhou Liang Fan Zheng Weiguo

(*Electronics Department Zhongshan University, Guangzhou 510275*)

Abstract Motion estimation and motion compensation (ME/MC) is the core technology of motion picture coding. A new motion estimation scheme based on a novel matching criterion, Restrictive Conditional Interframe Difference (RCID), is proposed in this paper. The new matching criterion overcomes the shortages of MAD criterion and has the equivalent computing complexity with MAD. The theory analysis and experimental results show the effectiveness and superiority of RCID criterion.

Keywords Video coding, Motion estimation, Match criterion