

一种基于时空相关性的运动估计算法¹

施建良 潘 健 余松煜 张文军

(上海交通大学图像处理与模式识别研究所 上海 200030)

摘 要 研究了一种基于时空相关性的运动估计方法。提出在块匹配法中,对于运动相关性好的块根据相邻块或前一帧图像的运动预测出它的初始运动矢量,然后在小范围内搜索,做运动矢量的优化。对传统的块匹配运动估计,给出了提高运动矢量一致性的修正准则。实验结果表明,本文算法所得的运动矢量一致性远高于传统块匹配法,同时能保持和全局搜索法相媲美的预测质量,而运算时间则有较大幅度缩短。

关键词 运动估计, 块匹配法, 时空相关性

中图分类号 TN941

1 引 言

运动估计是序列图像编码的关键技术,其实现的方法很多,如块匹配法、像素递归法、基于频域的方法、基于特征匹配的方法等^[1]。块匹配法由于简单和易于用硬件实现而被 H.261/263、MPEG1/2 等国际标准和建议广泛采用^[2]。其缺点是将图像固定分块并假设整个块做平移运动,没有考虑图像景物内容,也没有充分考虑运动的时间相关性,故运动矢量一致性差,尤其是静态背景的矢量场很杂乱^[3]。为了满足块内各像素运动一致这一基本假设,往往要求将块划分得很小,从而估计结果易受噪声的影响,且造成运动信息的大量重复。为此,人们研究了各种改进方法。除了扩大搜索范围^[4]外,还提出了更复杂的运动模型^[5,6];基于模型和对象的运动估计也是解决块匹配法问题的很好的思想^[7]。

上述改进方法,往往由于运算量大或算法不成熟而难以实用。本文提出的基于时空运动相关性的运动估计方法,从实用角度出发,既保持了高的准确性,又提高了运动矢量的一致性,减少了所需传输的运动信息,运动矢量抗噪能力也得到较大提高。

2 基于时空相关性的运动估计算法

算法的基本思想是:对于运动相关性好的块根据相邻块或前一帧图像的运动预测它的初始矢量,然后在小范围内全搜索,做运动矢量的优化;对于运动相关性差的块,用传统的块匹配法做全搜索,所得的运动矢量根据本文第 3 节所给的一致性匹配准则进行修正。

算法的工作框图如图 1 所示。具体过程如下:

(1) 对序列图像中的第一个需做运动估计的图像帧,用传统块匹配法作运动估计,同时记录该帧图像运动估计时平均块匹配误差 (Average Sum of Absolute Difference per Block: ASAD): $ASAD_{previous}$ 。

(2) 后续帧的运动估计:

(a) 由前一帧同一位置块及其邻块的运动预测当前帧当前块 k 的初始矢量 V_{p1} (相应误差 $Error_{p1}$)。块 k 同前帧块的位置关系如图 2 所示。设 $V_i (i = 1, \dots, 9)$ 为前帧各块运动矢量,则:

¹ 1997-10-16 收到, 1998-08-23 定稿
国家自然科学基金资助项目 (No.69572024)

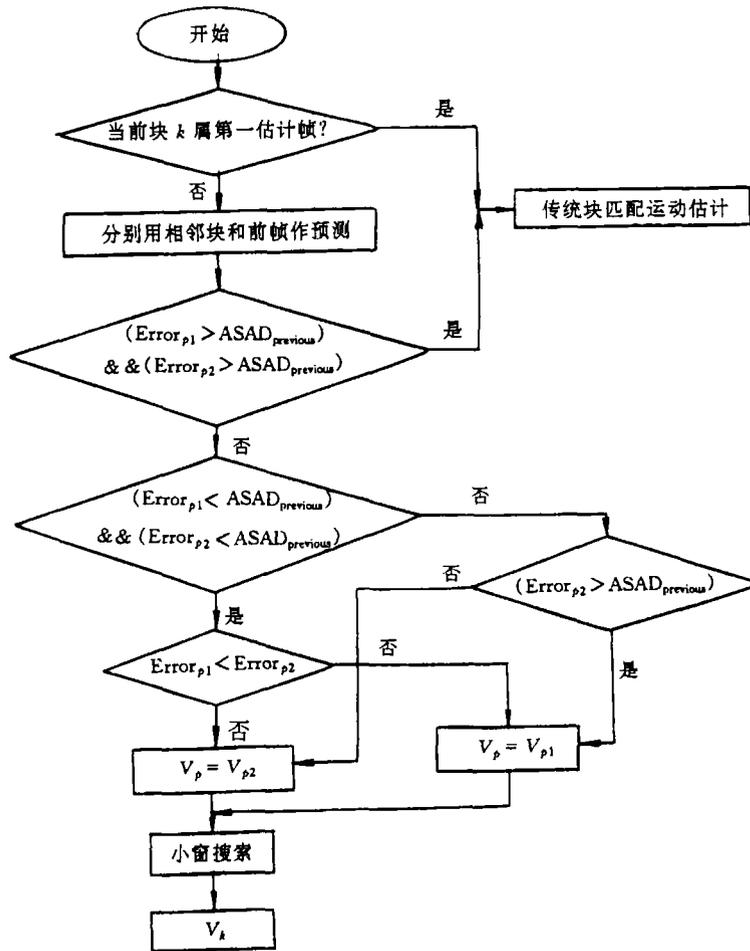


图 1 算法工作流程图

$$V_{p1} = (V_1 + 2V_2 + V_3 + 2V_4 + 2V_5 + V_6 + 2V_7 + V_8 + 12V_9)/24. \quad (1)$$

(b) 对块 k ，再根据它在当前帧中邻块的运动做预测，令预测矢量为 V_{p2} （相应误差 Error_{p2} ）。块 k 同邻块的位置关系如图 3 所示。设 $V_i (i = 1, \dots, 4)$ 为各邻块的运动矢量：

$$V_{p2} = (V_1 + 2V_2 + V_3 + 2V_4)/6. \quad (2)$$

(c) 令 $\text{Error} = \min(\text{Error}_{p1}, \text{Error}_{p2})$ ，若 $\text{Error} < \text{ASAD}_{\text{previous}}$ ，取相应的运动矢量 (V_{p1} 或 V_{p2}) 作为当前块的初始运动矢量 V_p 。否则 Error 太大，我们认为不宜由相邻块或前一帧的运动对当前块做预测，也即当前块运动相关性比较差，需用传统块匹配法做运动估计，并由第 3 节辅以一致性修正。

(d) 以 V_p 为初始矢量，在小窗内做全搜索，得到的最佳运动矢量即为当前块 k 的运动矢量 V_k 。小窗搜索可结合半像素精度情况进行。

(e) 重复 (a)-(d)，直至当前帧的所有块都作了运动估计。

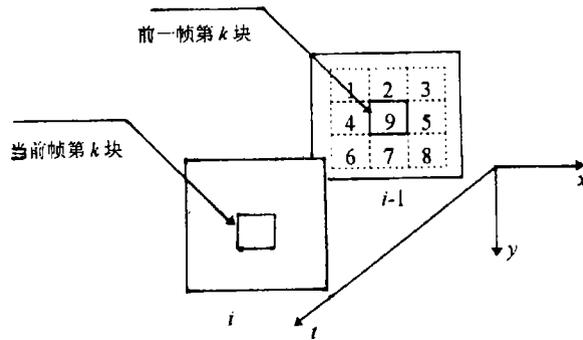


图 2 当前块同前帧同位置块及其邻块关系图

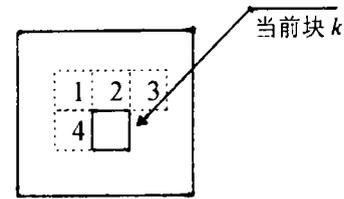


图 3 当前块与其邻块示意图

(f) 计算当前帧的平均块匹配误差 $ASAD_{current}$, 并令

$$ASAD_{previous} = ASAD_{current}. \quad (3)$$

(3) 重复第 (2) 步, 进入下一帧的运动估计。

从上述过程可看出, 由于采用了前一帧的平均块匹配误差来决定是否对当前 k 块进行预测, 从匹配误差方面来说, 该算法具有一定的自适应性。采用前一帧的匹配误差来作为阈值是基于这样一个事实: 图像前后帧 (甚至相邻数帧) 的运动具有较强的相关性。

对于处于当前帧边界的块, 我们只用前一帧同一位置块的运动矢量作为当前块的预测。考虑到运动时间上的差异, 在前一帧进行小窗搜索的窗口大小取为当前帧小窗搜索的两倍。

3 运动矢量的一致性修正准则

对用传统块匹配法作运动估计的块, 为充分利用运动相关性, 文献 [8] 提出了相关性约束的块匹配准则, 即在原 MSE 匹配准则基础上加上预测矢量这一附加项。然而, 由于运动矢量, 特别是物体边界块的运动矢量, 存在较大的误差, 使得附加项本身就存在较大的误差, 不适于做匹配准则。另外, 该附加项中只考虑了帧内相关性, 忽略了帧间相关性。本文提出以下一致性准则, 可对块的运动矢量作较好的校正。

令全搜索块匹配法所得的运动矢量为 $(x_{initial}, y_{initial})$ 。对搜索范围内的每一点 i , 若满足:

$$\frac{error_i + 1}{error_{initial} + 1} - 1 < \gamma, \quad (\gamma: \text{实验常数}), \quad (4)$$

则计算并保存 i 点的能量函数 $ET(i)$ 。这里, $error_i$ 为块对应匹配点 i 的误差, $error_{initial}$ 为 $(x_{initial}, y_{initial})$ 点的匹配误差。求出使 $ET(i)$ 最小的点, 假设为 m , 则点 m 相应的矢量即为该块的最终运动矢量。能量函数 $ET(i)$ 的计算充分考虑了在时间和空间上运动的相关性:

$$ET(i) = (x_{mean} + x_i)^2 + (y_{mean} + y_i)^2, \quad (5)$$

其中 (x_i, y_i) 为当前块相应 i 点的运动矢量, x_{mean}, y_{mean} 分别为前一帧同一位置的块及其邻块 (见图 4) 运动矢量 x 、 y 分量的加权平均值。其中阴影块为前帧中与当前块同位置的块, 其它为其邻块。计算公式如下:

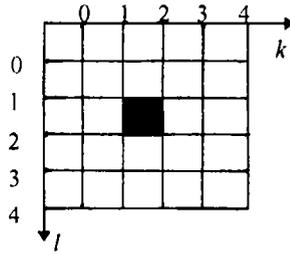


图 4 同位置的前帧块及其邻块示意图

$$x_{\text{mean}} = \frac{1}{\text{BS}^2} \sum_{k,l=0}^4 a[k]b[l]v_x[k][l], \quad y_{\text{mean}} = \frac{1}{\text{BS}^2} \sum_{k,l=0}^4 a[k]b[l]v_y[k][l], \quad (6)$$

其中 BS 为块大小, $v_x[k][l]$ 、 $v_y[k][l]$ 分别为图 4 中块运动矢量的 x 、 y 分量; 权值 $a[k]$, $b[l]$ 与运动矢量 (x_i, y_i) 有关. $a[k]$, $b[l]$ 的取值考虑了运动的空间相关性. 当 $k = \left\lfloor \frac{2 * x_i}{\text{BS}} \right\rfloor + 2$ 时,

$$a[k] = -2 * x_i + (k - 1) * \text{BS}, \quad a[k + 1] = 2 * x_i - (k - 2) * \text{BS}, \quad (7)$$

否则 $a[k] = 0$. 同样, 当 $l = \left\lfloor \frac{2 * y_i}{\text{BS}} \right\rfloor + 2$ 时,

$$b[l] = -2 * y_i + (l - 1) * \text{BS}, \quad b[l + 1] = 2 * y_i - (l - 2) * \text{BS}, \quad (8)$$

否则 $b[l] = 0$.

显然, γ 的取值很大程度上影响了最终预测误差和运算量. 为确保修正后仍有较小的预测误差, γ 不宜取得过大 (本文取 0.1). 能量函数 $\text{ET}(i)$ 则约束着当前块对相邻块的影响.

4 实验结果

我们用标准序列测试图像 Miss American(CIF)、Renata(720*576) 等对本文算法进行了验证, 并将实验结果与全局搜索的块匹配法进行了比较. 实验中取块大小为 16×16 , 传统块匹配法的搜索范围为 $[-16, +16]$, 前一帧的小窗搜索范围为 $[-4, +4]$. 当前帧小窗搜索范围为 $[-2, +2]$.

结果表明, 相关性运动估计所得的运动矢量一致性明显加强, 用于传输运动矢量的比特数有较大幅度的下降 (见表 1、表 3). 从预测质量看, 两者信噪比 SNR 相差很小, 由表 2、表 4 可看出, Miss American 平均比全局搜索法低约 0.12dB, 而 Renata 平均仅低 0.085dB 左右. SNR 计算公式如下:

$$\text{SNR} = -20 \log_{10} \text{SAD} / (255 * \text{pixel}_{\text{col}} * \text{pixel}_{\text{row}}) \quad (9)$$

其中 SAD 为绝对匹配误差之和, $\text{pixel}_{\text{row}}$ 、 $\text{pixel}_{\text{col}}$ 分别为图像的行列像素数.

在运算速度方面, 本文算法的一致性调整准则使得计算复杂度稍有增加, 但平均搜索步数却下降至不足全局搜索法的 1/3. (Miss American 由全局搜索平均每块 1089 步下降至 356 步, Renata 下降至 331 步). 因而总的运算速度仍比全局搜索法要快得多.

表 1 全局搜索法和基于时空相关性运动估计法 Miss American 矢量编码比特数 (单位 bit)

帧号	25	26	27	28	29	30	31	32	33	34	35	平均
全局法	3288	3155	3122	3182	2996	3056	3268	3226	2962	3244	3146	3150
相关法	1989	2112	2020	2074	1990	1990	2548	2596	2045	2147	2416	2175

表 2 全局搜索法和基于时空相关性运动估计法对 Miss American 的 SNR 计算结果 (单位 dB)

帧号	25	26	27	28	29	30	31	32	33	34	35	平均
全局法	44.025	44.196	44.302	44.497	44.505	44.494	44.508	44.563	44.171	44.145	44.143	44.323
相关法	43.897	44.095	44.177	44.393	44.392	44.371	44.232	44.233	44.223	44.002	44.020	44.185

表 3 全局搜索法和基于时空相关性运动估计法对 Renata 的矢量编码比特数 (单位 bit)

帧号	20	21	22	23	24	25	26	27	28	29	平均
全局法	8617	8459	8906	8757	8536	8425	8270	8259	8229	8214	8467
相关法	7986	7619	7797	8052	7443	7419	7612	7176	6685	6943	7473

表 4 全局搜索法和基于时空相关性运动估计法对 Renata 的 SNR 计算结果 (单位 dB)

帧号	20	21	22	23	24	25	26	27	28	29	平均
全局法	32.310	32.482	32.513	32.744	32.983	32.920	32.906	32.800	32.720	32.816	32.719
相关法	32.222	32.402	32.418	32.675	32.900	32.824	32.814	32.746	32.622	32.719	32.634

5 结 论

本文提出的基于时空相关性的运动估计方法, 用相邻块和前一帧的运动矢量对当前块进行预测, 充分利用了块运动的帧间相关性和帧内相关性. 在预测判决上, 用前一帧的运动估计平均误差来决定当前块的初始矢量是否通过预测得到, 故有一定的自适应能力. 对于传统块匹配法所得的运动矢量, 本文进一步利用了运动的时空相关性进行了一致性调整, 较好地校正了由于噪声的影响而产生的误匹配, 提高了运动矢量的抗干扰性.

参 考 文 献

- [1] Dufaux F, Moscheni F. Motion estimation techniques for digital TV: a review and a new contribution, Proc. IEEE, 1995, 83(6): 858-876
- [2] Haskeu B G, et al. Image and Video Coding—Emerging Standards and beyond, IEEE Trans. on-CAS VT, 1998, 8(7) 814-837.
- [3] Hangu Yeo, Yu Hen Hu. A motion estimation and image segmentation technique based on the variable block size. Proc. ICASSP'97, Munich, Germany: 1997, 3137-3140.
- [4] Bierling M. Displacement estimation by hierarchical block matching SPIE Vol.1001, VCIP'88, 942-951.
- [5] Yuichiro Nakaya, Hiroshi Harashima. Motion compensation based on spatial transformations. IEEE Trans on CASVT, 1994, CASVT-4(3): 339-356.
- [6] Deihl N. Object-oriented motion estimation and segmentation in image sequences. Signal Processing: Image Communication, 1991, 3(1): 23-56.
- [7] Aizawa K, Huang T S. Model-based image coding advanced video coding techniques for very low bit-rate applications. Proc. IEEE, 1995, 83(2): 259-271.
- [8] 薛向阳, 吴立德. 基于运动矢量插值的运动补偿算法. 电子科学学刊, 1997, 19(4) 483-488.

A NEW MOTION ESTIMATION ALGORITHM BASED ON SPATIO-TEMPORAL COHERENCE

Shi Jianliang Pan Jian Yu Songyu Zhang Wenjun

*(Institute of Image Processing and Pattern Recognition,
Shanghai Jiaotong University, Shanghai 200030)*

Abstract A new motion estimation algorithm based on spatio-temporal coherence is presented. In this proposed algorithm, an initial motion vector of the block which has good coherence with its neighbors or that of the previous frame is directly predicated. A more precise motion vector is obtained based on the initial one by full-searching block matching algorithm (FBMA) in a small range. As to blocks with poor coherence, FBMA with ordinary range is used, and their motion vectors are at last adjusted by the coherence criterion given in the paper. Experiments show satisfying results can be reached with the proposed algorithm.

Key words Motion estimation, Block matching algorithm, Coherence

施建良: 男, 1971年生, 博士生, 从事专业为图像压缩编码、图像处理.

潘 健: 男, 1971年生, 博士生, 从事专业为图像压缩编码、ATM传视频.

余松煜: 男, 1941年生, 教授, 博士生导师, 从事专业为图像压缩编码、数字电视和目标识别.

张文军: 男, 1963年生, 教授, 博士生导师, 国家高清晰度电视研究开发工程项目总体组组长从事专业为数字电视系统、数字图像压缩、图像通信工程.