

运动估计算法匹配准则研究

向友君 雷娜 余卫宇 谢胜利

(华南理工大学电子与信息学院 广州 510640)

摘要 运动估计算法是视频编码技术的研究重点,高精度、高效率的匹配和补偿可以减少预测误差,提高视频压缩效果,因此块匹配的准确性是核心问题。对最小绝对误差MAD、最小均方误差MSE、归一化互相关函数NCCF、像素差值分类、子采样、VOD等匹配准则进行了理论介绍和实验分析,并提出了一种利用图像差值分布情况作为匹配准则的方法,即图像差值均方差DVAR匹配准则。实验结果表明,图像差值均方差匹配准则获得了比较高的编解码质量。最后展望了更符合人眼视觉系统(HVS)特性的匹配准则发展方向。

关键词 运动估计,匹配准则,块匹配,均方差,结构相似度

中图法分类号 TN919.81 **文献标识码** A

Research of Block Matching Criterion for Motion Estimation

XIANG You-jun LEI Na YU Wei-yu XIE Sheng-li

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510640, China)

Abstract Motion estimation algorithm is the research priority of video encoding technology. High precision and high efficiency of the match and compensation can reduce prediction error and improve video compression effect, so the accuracy of block matching is the core issue. The Minimum Absolute Difference (MAD), the Minimum mean Square Error (MSE), Normalized Cross-Correlation Function (NCCF), the classification of pixel difference, sub-sampling, Variance Of Difference (VOD), and other matching criterions were theoretically and experimentally analysed. And a method of using the distribution of the image difference as a matching criterion was proposed, that is image Difference Variance (DVAR) matching criterion. Experimental results show that image difference variance matching criterion received relatively high codec quality. Finally the paper pointed out the future development of block matching criterion which is more in line with the human visual system (HVS) characteristics.

Keywords Motion estimation, Matching criterion, Block matching, Mean square deviation, Structural similarity(SSIM)

1 引言

运动估计是视频编码中的一项核心技术,能有效去除序列图像的帧间冗余,从而实现高效编码。研究快速有效的运动估计算法一直是视频编码领域的热点问题。块匹配运动估计算法(Block Matching Algorithm, BMA)由于简单和易于软硬件实现而被广泛应用于包括MPEG-1/2/4和H.26x在内的多项视频编码标准中。其基本思想是将图像序列的每一帧分成 $M \times N$ 的宏块,然后对当前帧中的每一块到参考帧某一给定搜索范围内根据一定的匹配准则找出与当前块最相似的块,即匹配块,由匹配块与当前块的相对位置计算出运动位移,所得运动位移即为当前块的运动矢量。

以往的研究重点都集中在运动估计算法的搜索策略上,而对匹配准则的研究很少,大部分直接采用最小绝对误差和SAD作为算法搜索中的匹配准则。通过运动估计基本思想的分析,可以看出匹配准则对于匹配块的搜索和算法性能也存在着很大的影响。选择一个既能精确描述块之间的匹配程

度,又能快速得到结果的匹配准则,将会极大地提高搜索速度,改进算法性能。本文对最小绝对误差MAD、最小均方误差MSE、归一化互相关函数NCCF、像素差值分类、子采样、VOD等匹配准则进行了理论介绍和实验分析,并提出了一种利用图像差值分布情况作为匹配准则的方法,即图像差值均方差匹配准则。实验结果表明,图像差值均方差对于图像边缘比较明显的视频序列能获得比较高的编解码质量。这些匹配准则大多不能很好地和视觉质量相吻合,是因为它们没有充分地考虑人眼的视觉特性,这对图像这种最终信宿为人眼的信号而言是非常不合理的。文章最后将结构相似度SSIM理论引入运动估计块匹配,对更符合人眼视觉系统特性的匹配准则发展方向进行了展望。

2 匹配准则理论分析

匹配准则实质上是一种误差度量函数,以此为标准衡量两个子块的相似程度。匹配准则的选取,直接关系到运动位移估值的准确性。在块匹配算法中采用不同的匹配准则会获

到稿日期:2008-10-24 返修日期:2008-12-30 本文受国家杰出青年科学基金(60325310),国家自然科学基金重点资助项目(U0635001)资助。

向友君(1976—),女,讲师,主要研究方向为图像处理、视频编码,E-mail: yjxiang@scut.edu.cn;雷娜(1987—),女,主要研究方向为图像处理等;余卫宇(1972—),男,讲师,主要研究方向为图像处理等;谢胜利(1956—),男,教授,博士生导师,主要研究方向为信号与信息处理等。

得不尽相同的运动估计结果。

2.1 最小绝对误差 MAD

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

式中, (i, j) 为位移矢量, f_c 和 f_r 分别为当前帧和参考帧的灰度值, $M \times N$ 为宏块的大小, 若在某一个点 (i_0, j_0) 处 $MAD(i_0, j_0)$ 达到最小, 则该点为要找的最优匹配点。

MAD 准则不需作乘法运算, 实现简单、方便, 因此使用最多, 通常使用 SAD (Sum of Absolute Difference) 代替 MAD。SAD 即求和绝对误差, 定义如下:

$$SAD(i, j) = \sum_{m=1}^M \sum_{n=1}^N |f_c(m, n) - f_r(m+i, n+j)| \quad (2)$$

2.2 最小均方误差 MSE

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

MSE 值最小时即为最优匹配点。

2.3 归一化互相关函数 NCCF

$$NCCF(i, j) = \frac{\sum_{m=1}^M \sum_{n=1}^N f_c(m, n) f_r(m+i, n+j)}{[\sum_{m=1}^M \sum_{n=1}^N f_c^2(m, n)]^{1/2} [\sum_{m=1}^M \sum_{n=1}^N f_r^2(m+i, n+j)]^{1/2}} \quad (4)$$

这里, 最优匹配即为找 NCCF 的最大点。

2.4 像素差值分类准则

H. Gharavi 和 M. Mills 提出的像素差值分类准则^[1]不同于以上几种准则。它基于最小风险准则, 在参考图像的搜索窗中选取含匹配像素最多的子块作为最佳匹配块。

该准则将参考块中的每一个像素分类为匹配像素和失配像素进行统计。对选定的阈值 T_0 , 进行如下分类:

$$T(m, n, i, j) = \begin{cases} 1, & \text{if } |f_c(m, n) - f_r(m+i, n+j)| \leq T_0 \\ 0, & \text{others} \end{cases} \quad (5)$$

式中, “0”表示该像素对位移为 (i, j) 失配, “1”表示 $f_r(m+i, n+j)$ 是 $f_c(m, n)$ 的匹配像素。在整个块内对 T 值进行求和运算:

$$S_M(i, j) = \sum_{m=1}^M \sum_{n=1}^N T(m, n, i, j) \quad (6)$$

在运动搜索中选取参考块中 $S_M(i, j)$ 最大, 也即匹配像素最多的块作为匹配块。

可以看到采用这种匹配准则的运动估计体现了最小风险准则, 在运动搜索中选取参考块中失配像素最少的块作为匹配块, 对保持当前块与匹配块形状的一致性有较强的约束。但由于各种噪声的干扰, 在亮度变化较为缓慢的区域, 该匹配准则下会具有相同的匹配像素数, 即有相同的风险。对于此种情况有些人采用基于最小相对位移约束的判决方法进行二次匹配^[2], 也有些人利用人眼对物体边缘运动比较敏感的特性^[3], 提出了边缘像素点分类准则。其基本思路是对每个子块中物体边缘位置的像素点进行分类, 根据像素亮度值的差异分为匹配和不匹配两类, 统计参考帧中各候选子块所含边缘匹配像素点的数目, 最多的那个子块为最优匹配块。

2.5 子采样匹配准则

$$SAD(i, j) = \sum_{m=1}^M \sum_{n=1}^N |f_c(m, n) - f_r(m+i, n+j)| p(m, n) \quad (7)$$

式中, $p(m, n) = \begin{cases} 1 & m, n \text{ 都是偶数} \\ 0 & \text{其它情况} \end{cases}$

子采样匹配方法大大降低了计算复杂度, 其计算量只为

原来的 1/4。有的不是 4:1 模式, 而是水平方向或竖直方向子采样, 如

$$SAD(i, j) = \sum_{m=1}^{M/2} \sum_{n=1}^N |f_c(2m, n) - f_r(2m+i, n+j)| \quad (8)$$

还可以采取偶数行偶数列、奇数行奇数列等子采样方式。

2.6 VOD (Variance Of Difference) 匹配准则^[4]

$$VOD(i, j) = \sum_{m=1}^M \sum_{n=1}^N f^2(m, n) - \frac{1}{MN} (\sum_{m=1}^M \sum_{n=1}^N f(m, n))^2 \quad (9)$$

$$f(m, n) = f_c(m, n) - f_r(m+i, n+j) \quad (10)$$

$$\text{令 } a_c = \frac{1}{MN} (\sum_{m=1}^M \sum_{n=1}^N f_c(m, n)), a_r = \frac{1}{MN} (\sum_{m=1}^M \sum_{n=1}^N f_r(m, n))$$

则

式(9)可以改写为

$$VOD(i, j) = \sum_{m=1}^M \sum_{n=1}^N ((f_c(m, n) - a_c) - (f_r(m+i, n+j) - a_r))^2 \quad (11)$$

式中, a_c 表示当前块的平均亮度, a_r 表示参考块的平均亮度。式(11)具有更直观的意义, 与式(3)相比, VOD 匹配准则相当于当前块和参考块都去掉照明度变化影响后的 MSE。

如果当前块与某个位移处的参考块相比只有照明度发生了变化, 则其残差块的每一个差值都是相同的, 因此其 DCT 变换只有 DC 分量, AC 分量为零, 熵编码模块只需要编码 DC 分量, 因此用较少的比特就可以表示该残差块。由式(11)可以知道, 这种情况下, 对应该位移的 VOD 匹配准则的结果为 0, 因此该位移就是 VOD 准则下的最佳运动向量。SAD 和 MSE 匹配准则没有考虑照明度变化的影响, 照明度变化被计算在匹配误差内, 在这种情况下, 会产生较大的匹配误差, 有可能导致没有选择该位移作为最佳运动向量。对照明度的变化不敏感, 这是 VOD 匹配准则的优点之一。与 SAD 相比, VOD 的缺点在于需要乘法运算。对多数 CPU, 乘法运算需要更多的时间和功耗, 因此设计一种快速算法以减少乘法运算是有意义的。

2.7 图像差值均方差匹配准则 DVAR (Difference Variance)

均方差是反映一组数据大小波动的指标, 用当前块与参考块的像素差值的均方差来衡量前后误差分布情况。对于进行匹配的块, 当前块与参考块的差值的均值一般为 0, 差值均方差越小, 说明像素差值越集中在均值附近, 即说明当前块越接近匹配块, 反之则为非匹配块。

定义

$$DVAR(i, j) = \sqrt{\frac{\sum_{m=1}^M \sum_{n=1}^N (f(m, n) - \bar{V})^2}{MN}} \quad (12)$$

$$f(m, n) = f_c(m, n) - f_r(m+i, n+j) \quad (13)$$

式(12)中, $\bar{V} = \frac{\sum_{m=1}^M \sum_{n=1}^N f(m, n)}{MN}$ 为所有差值的均值, 对应 DVAR 最小的运动矢量即为最佳运动矢量。

由于对于诸如 SAD, MSE 的匹配准则来说, 如果在某些像素点上差值很大, 而在其他点上差值很小, 则求出来的 SAD, MSE 值依然比较小, 可能会误判为匹配块, 尤其对于图像边缘部分。而此种方法考虑了所有像素误差的分布, 当所有像素点的误差都比较小时, 才为最佳匹配块。实验结果也证明了这一点, 采用图像边缘比较明显的 Foreman 序列, DVAR 的匹配效果要好于 SAD, MSE。但是此种方法相对于 SAD, MSE 也增加了计算复杂度。

3 实验结果分析

实验基于 H. 264 参考模型 JM10.1^[5] 进行改造, 选取

Foreman. qcif ,Salesman. qcif 标准视频序列进行编解码,编码帧数为 20 帧,编码顺序为 IPPPP,每 19 个 P 帧一个 I 帧。所有实验在个人计算机(赛扬 1G,512M 内存)上进行。

表 1 列出了在相同的输出码率下,对所选视频序列应用本文提到的各匹配准则进行实验的结果,搜索方式为全搜索。表 1 中实验结果的比较,均以 JM10.1 中使用的 SAD 匹配准则为基准。PSNR 反映了匹配精度,时间消耗反映了计算复杂性。

表 1 实验结果

视频序列	匹配准则	PSNR(dB)	Δ PSNR	消耗时间(s)	Δ IME(%)
Foreman	SAD	36.510	0.000	0.610	0.000
	MSE	36.579	0.069	2.393	292.3
	NCCF	36.557	0.047	7.667	1156.9
	像素差值分类	36.513	0.003	0.344	-43.6
	子采样	36.506	-0.004	0.518	-15.1
	VOD	36.580	0.07	7.153	1072.6
	DVAR	36.593	0.083	10.361	1598.5
Salesman	SAD	38.981	0.000	0.814	0.000
	MSE	38.987	0.006	1.905	134.0
	NCCF	38.992	0.011	7.580	831.2
	像素差值分类	38.983	0.003	0.250	-69.29
	子采样	38.980	-0.001	0.439	-46.06
	VOD	38.988	0.007	4.905	502.6
	DVAR	38.992	0.011	10.517	1192.0

从实验结果可以看到因为 Foreman 有比较多的亮度变化,Salesman 有比较多的细节但是运动比较少,亮度变化也比较小,所以对于 Foreman,各匹配准则下的 PSNR 结果差别比较大。除了子采样匹配准则,其余匹配准则的 PSNR 均高于 JM10.1 原来使用的 SAD 匹配准则,其中本文提出的匹配准则 DVAR 所得到的 PSNR 最高。但在计算复杂度方面,子采样的计算复杂度相比 SAD 明显降低了。另外可以看到 VOD 匹配准则相当于当前块和参考块都去掉照明度变化影响后的 MSE,因此与 MSE 的 PSNR 比较相近。DVAR 对于图像边缘比较明显的 Foreman 来说显示了比较好的性能,同时也适用于普通序列。时间上,像素差值分类准则由于其运算简单,因此用了最少的时间,但其 PSNR 有所降低。DVAR

由于计算复杂,因此用时最多,其次是 VOD 与 NCCF。

结束语 本文在对各种匹配准则进行理论及实验分析的基础上提出了一种利用图像差值分布情况作为匹配准则的方法:DVAR 匹配准则。从实验结果来看,与视频标准优先采用的 SAD 进行比较,本文提出的匹配准则不仅适合普通情况,对有亮度变化及边缘明显的序列具有更好的性能。但降低其复杂性是该准则要解决的问题。

文中所研究的这些匹配准则都没有充分地考虑人眼的视觉特性。人眼在观察一幅图像的时候,首先提取图像的结构信息^[6],因为图像像素之间的信息是相关的,人眼判断一幅降质图像质量的好坏,更多的是审视该图像的结构信息与原图像的结构信息是否相近。所以图像结构相似度(SSIM)因其更符合人眼视觉系统(HVS)特性被认为具有客观评判性,已被众多学者研究,该理论作为匹配准则的发展方向将能大大提高匹配性能。该方法包括 3 部分:亮度比较、对比度比较以及结构信息的比较。但算法复杂度较高,所以降低复杂度也是该方向的一个难点。

参 考 文 献

[1] Gharavi H , Mills M . Block matching motion estimation algorithm new result[J]. IEEE Transaction on circuit and systems, 1990,37(5):649-651

[2] 段哲民,周巍,李辉,等. 一种基于 H. 264 的有效运动估计算法[J]. 电子学报,2006,34(12):2406-2409

[3] 贾志科,崔慧娟,唐昆,等. 用于运动位移估值的新匹配准则[J]. 清华大学学报,1999,39(1):43-46

[4] 刘昌进,郭立,李敬文,等. 一种基于 DCT 系数分布特性的块匹配准则[J]. 小型微型计算机系统,2006,27(5):893-895

[5] Joint Video Team (JVT) . Reference Software JM10.1 [EB/OL]. <http://iphome.hhi.de/suehring/tmpl/download/jm10.1.zip>,2006.06

[6] Wang Zhou,Bovik A C,Sheikh H R,et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans on Image Processing,2004,13(4):600-612

(上接第 254 页)

改变 Web 地址栏中的地址,可以调出其他辅助功能的控制窗口(如摄像头的控制、云台控制、视频参数的设置、视频保存等)。实验表明,电路板的功能稳定,音视频质量较好。



图 6 Web 客户端实时浏览 4 路解压后的 CIF 图像

结束语 本文主要阐述一个通用的基于 DSP(TI 公司 DM642 主芯片)的多路实时音视频编码终端的硬件设计方案和实现过程。对硬件整体方案做了论证,对音视频、网络、存储、电源等各个重点模块做了较深入讨论。由于 CPU 主频达到了 600MHz,因此也对硬件 PCB 的布局、布线和相应的仿

真做了全面的分析。最后对硬件做了软件实际运行的测试,表明硬件系统工作性能稳定可靠。由于 DSP 芯片具有通用性和可编程性,因此本系统可以扩展到 MPEG-2,MPEG-4,H. 263和 H. 264 等多种多媒体编码压缩标准以及视频监控、远程医疗、视频会议等绝大多数视音频的应用场合,只需要在软件上做少量的修改即可,因而具有较广泛的应用前景。

参 考 文 献

[1] Texas Instruments Incorporated. TMS320DM642 Video/Imaging Fixed-Point Digital Signal Processor,2003

[2] Li Wei. An Embedded 4-Channel Real-Time Multimedia Coding System[C]// Multi-Media Modelling Conference Proceedings 2006 12th International. 2006,1:364-368

[3] Texas Instruments Incorporated. TVP5150 A PCB Layout Guidelines,2003

[4] Using IBIS Models for Timing Analysis. SPRA839A. Texas Instruments,2003,4:1-26

[5] 曾峰,巩海洪,曾波. PowerPCB 高速电子电路设计与应用[M]. 北京:电子工业出版社,2004