

多解析度自适应零块判决算法*

栗强, 崔慧娟, 唐昆, 杜文

(清华大学 电子工程系 微波与数字通信国家重点实验室, 北京 100084)

E-mail: liq@tsinghua.org.cn; {cuihj,tangk,duw}@ee.tsinghua.edu.cn

http://www.ee.tsinghua.edu.cn

摘要: 在视频会议等应用中, 视频编码的速度是保证系统实时性的关键. 为了减小视频编码的运算量并提高系统的整体性能, 提出了多解析度的自适应零块发现判决算法. 多解析度通过对图像的子采样从而大幅度降低了搜索最佳匹配块的次数; 而自适应的零块判决算法使得运动搜索速度收敛的更快, 在不影响图像质量的前提下提高了编码的效率, 并且减少了后续模块(DCT、量化等)所占用的时间. 利用 H.263 算法进行了测试, 结果显示在图像主观质量基本不变的前提下, 编码时间最多减少了近 40%.

关键词: 视频编码; 运动搜索; 多解析度; 离散余弦变换; 自适应零块判决算法

中图法分类号: TP391 文献标识码: A

目前现有的视频标准(MPEG 1,2,4 以及 ITU 的 H.261 和 H.263^[1]等)都是以运动估计、DCT 变换和熵编码作为其核心算法. 这种编码框架由于其简单和高效性而得到了广泛的应用. 由于所面向的应用是中低码率的视频会议等交互式通信系统, 因此视频编解码的实时性就显得格外重要. 对于典型的 H.263 算法而言, 运动搜索和 DCT 模块就占用了近一半的编码时间, 如何有效地减少它们的运算量成为实时实现编码器的关键. 对于中低码率的应用而言, 由于量化系数的选择一般都比较大, 这样经运动预测和补偿后的残差信号在量化后大多变成了全零块, 根据这一特点, 可以仅在时域范围内提前预测全零块的存在, 从而节省大量不必要的后续处理时间. 文献[2,3]中都提出了相应的处理方法, 但是文献[3]中给出的门限太小, 文献[2]中虽然给出了合适的门限, 但在图像变化细节较小的情况下效果并不是很理想, 其 PSNR 下降可达 1 dB ~ 2dB. 而本文提出的自适应的零块判决算法不但给出了合适的门限, 而且还可以通过自适应调节门限大小来控制零块的发现比例, 从而在保证编码效率的前提下使误判达到最小.

在应用零块判决算法的同时, 本文还提出了多解析度^[4]的运动搜索策略, 通过对图像进行子采样从而大幅度减小了运动搜索的运算量, 同时还保持了运动搜索的精度. 本文将这两种方法有机地结合在一起形成了多解析度的自适应零块判决算法. 仿真结果显示, 该方法大幅度减少了编码时间, 并在一定程度上提高了编码效率.

1 自适应的零块判决算法

在帧内编码方式下, 对运动预测和补偿后得到的残差信号进行提前判断, 以判决该宏块中的 6 个块是否量化为全零块. 这里将残差信号定义为 $R^k = \sum_{x=0}^7 \sum_{y=0}^7 |f^k(x, y) - \hat{f}^k(x, y)| = \sum_{x=0}^7 \sum_{y=0}^7 |r^k(x, y)|, k=1,2,\dots,6$. f^k 和 \hat{f}^k 分别为当前宏块和预测宏块的第 k 个块的像素值, R^k 为相应的残差信号. 残差信号经过式(1)的 DCT 变换后生成 $C(u, v)$.

* 收稿日期: 2001-06-15; 修改日期: 2002-05-21

基金项目: 国家自然科学基金资助项目(69972020); 国家重点基础研究发展规划 973 资助项目(G1998030406)

作者简介: 栗强(1976 -), 男, 山西太原人, 硕士, 主要研究领域为图像编码, 多媒体通信; 崔慧娟(1945 -), 女, 辽宁沈阳人, 副教授, 主要研究领域为信源编码, 多媒体通信系统; 唐昆(1945 -), 男, 辽宁人, 教授, 博士生导师, 主要研究领域为数字通信, 语音编码, 自适应信号处理; 杜文(1973 -), 男, 湖南长沙人, 讲师, 主要研究领域为多媒体通信系统, 信源编码.

$$C(u,v) = \frac{1}{4}a(u)a(v) \times \sum_{x=0}^7 \sum_{y=0}^7 r(x,y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right], u,v = 0,1,\dots,7, a(u) = \begin{cases} 1/\sqrt{2}, & u=0 \\ 1, & u=1,2,\dots,7 \end{cases} \quad (1)$$

根据余弦函数值小于 1 以及“和的绝对值小于绝对值的和”的特性可以得出 $C(u,v) < \frac{1}{4} \sum_{x=0}^7 \sum_{y=0}^7 |r(x,y)|$, 而帧间编码的量化公式为 $C_q(u,v) = (|C(u,v)| - Q/2) / (2 \times Q)$, 式中 $C_q(u,v)$ 为量化后的 DCT 系数, Q 为量化参数. 如果 $C(u,v)$ 被量化为零值(即 $C_q(u,v) < 1$), 则由上式可知, $\max_{u,v \in (0,\dots,7)} |C(u,v)| < 5/2 \times Q$, 因此 $5/2 \times Q$ 为 $C(u,v)$ 被量化为零值的一个下界.

如果 $C(u,v)$ 量化为零值, 只需使 $\frac{1}{4} \sum_{x=0}^7 \sum_{y=0}^7 |r(x,y)| < 5/2 \times Q$ 即可, 因此可以得到

$$\sum_{x=0}^7 \sum_{y=0}^7 |r(x,y)| < 10Q. \quad (2)$$

利用上式就可以在 DCT 变换和量化处理之前对残差信号进行判断, 从而避免了大量不必要的计算. 由于上述公式只是建立在简单的余弦函数小于 1 的基础上的, 是判决的一种弱势表示. 文献[2]利用 DCT 变换后直流信号最大的特性只对直流信号进行估计, 从而得出了以下结论:

$$\sum_{x=0}^7 \sum_{y=0}^7 |r(x,y)| < 20Q. \quad (3)$$

与式(2)比较, 式(3)得出的门限提高了一倍, 但由于只是简单地考虑直流系数的影响, 从而会不可避免地导致误判的发生. 在图像运动变化较小的情况下, 这种判决算法比正常情况下 PSNR 要下降 1~2 个 dB, 从而影响了恢复图像的质量.

针对这一矛盾, 本文提出了自适应的零块判决算法, 它不但给出了判决的最佳门限, 同时还可以自适应地改变门限的大小以反映图像本身的变化. 由柯西-许瓦兹公式可知:

$$C(u,v) = \frac{1}{4}a(u)a(v) \sum_{x=0}^7 \sum_{y=0}^7 r(x,y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right] \leq \frac{1}{4} \left[\sum_{x=0}^7 \sum_{y=0}^7 r^2(x,y) \right]^{1/2} \times \left[\sum_{x=0}^7 \sum_{y=0}^7 \left(a(u)a(v) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right] \right)^2 \right]^{1/2} = \frac{1}{4} \left[\sum_{x=0}^7 \sum_{y=0}^7 r^2(x,y) \right]^{1/2} \times 4 = \left[\sum_{x=0}^7 \sum_{y=0}^7 r^2(x,y) \right]^{1/2}. \quad (4)$$

为了使 $C(u,v)$ 量化为零值, 只需使 $\left[\sum_{x=0}^7 \sum_{y=0}^7 r^2(x,y) \right]^{1/2} < \frac{5}{2}Q$ 即可, 该式给出了 MSE 算法下的结果. 由于在 H.263 中规定, 像素亮度的最小值 $Black=16$, 最大值 $White=235$, 于是在通常意义下有公式

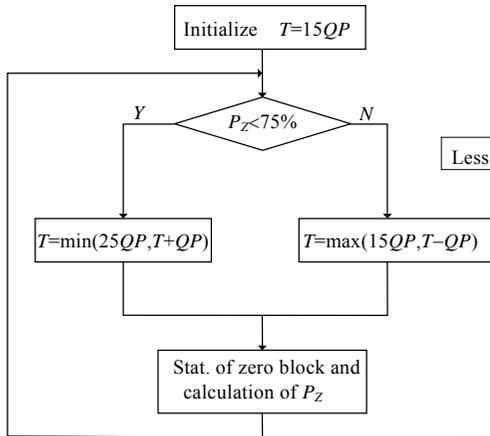
$\left[\sum_{x=0}^7 \sum_{y=0}^7 r^2(x,y) \right]^{1/2} < \frac{1}{4\sqrt{2}} \times \left[\sum_{x=0}^7 \sum_{y=0}^7 |r(x,y)| \right]$ 成立. 考虑到式(4)中等号成立的条件可以推出

$$\sum_{x=0}^7 \sum_{y=0}^7 |r(x,y)| < 15Q. \quad (5)$$

式(5)给出了零块判决的最低门限, 虽然它比式(3)的门限减小了 1/4, 但是式(3)的结论不具有一般性, 而式(5)才给出了最佳门限. 由于这里给出的是最坏情况下的判决结果, 而在实际应用中这种情况并不多见, 为进一步有效地判决零块, 本文在式(5)的基础上提出了自适应的零块判决算法. 具体判决过程如图 1 所示, 其中判决门限 T 最小为 $15Q$, 最大为 $25Q$, P_z 代表判决得到的零块数目和实际结果中零块数目的百分比.

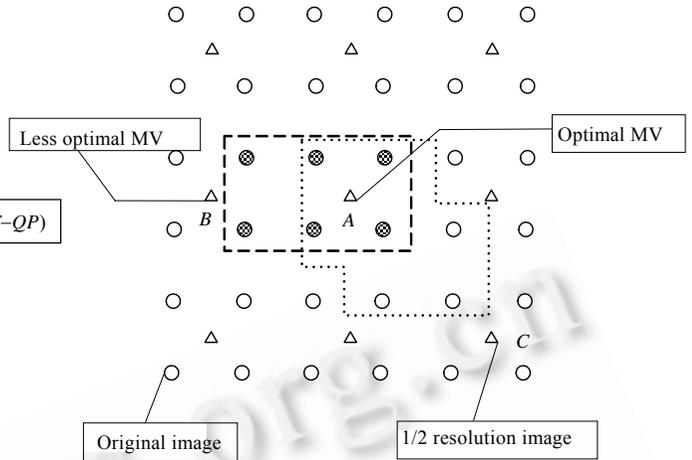
对于解析度为 $2m \times 2n$ 的原始图像而言, 首先采用降解析度的方法把原始图像缩小至以前的 1/2, 我们称生成的图像为解析图像. 它的像素值可由式(6)得出, 其中 $P(x,y), R(x,y)$ 分别代表原始图像和解析图像的像素值.

$$R(x,y) = \frac{1}{4} [P(2x,2y) + P(2x+1,2y) + P(2x,2y+1) + P(2x+1,2y+1) + 2]. \quad (6)$$



初始化, 统计零块并计算 P_z .

Fig.1 Adaptive zero block detection algorithm
图1 自适应零块判决算法



最佳运动矢量, 准最佳运动矢量, 原始图像, 1/2 解析图像.

Fig.2 Multi-Resolution motion search algorithm
图2 多解析度运动搜索算法

由于解析度下降为原始图像的 1/2, 这样, 以前建立在 16×16 宏块基础上的运动搜索就可以转化为 8×8 的块搜索. 本文采用的搜索算法为降解析度空间的菱形搜索法, 其搜索过程如下:

- (1) 对解析图像的每个块在上一帧的重建图像中搜索最佳的匹配块, 搜索范围为原图像的一半. 在搜索到最佳的运动矢量之后, 记录下最佳运动矢量(图 2 中 A 点)和准最佳运动矢量(图 2 中 B 点).
- (2) 将所得到的解析图像的最佳运动矢量和准最佳运动矢量乘以 2.
- (3) 在图 2 中, 根据所得到的最佳运动矢量和准最佳运动矢量的位置和方向关系, 在这两个位置之间选择若干点作为候选矢量. 如果 B 点为准最佳运动矢量, 那么在短折线框内包含的 6 个圆点就作为候选的整像素运动矢量; 如果斜方向的 C 点为准最佳运动矢量, 那么在不规则虚线中包含的 7 个圆点则为候选的整像素运动矢量.
- (4) 对所选择的候选运动矢量进行 16×16 大小的宏块运动搜索, 找到最佳的整数像素运动矢量.
- (5) 在半像素插值图像中进行半像素搜索, 找出最佳的半像素运动矢量.

由于式(6)综合考虑了像素间的相关性, 因此本文提出的搜索算法保持了和以往算法相同的精确性, 同时降解析度搜索算法的使用使得搜索次数相对于原钻石搜索算法有了大幅度的减少. 表 1 给出了几种搜索算法对单个宏块平均搜索次数(搜索范围为 $-16 \sim 15$)的比较结果.

Table 1 Search count comparison of single macroblock

表 1 单个宏块的搜索次数比较

Algorithm	Full search algorithm	Three-Step search algorithm	Diamond search algorithm	Multi-Resolution search algorithm
Average search count	1 089	33	16	11

算法, 全搜索法, 三步搜索法, 菱形搜索法, 多解析度搜索法 平均搜索次数.

可以看出, 改进的搜索算法在保证搜索精度的前提下, 在运算量上有了极大的改进. 由于多解析度算法需要将图像大小缩小到原始图像的 1/2, 因此对于格式较大(CIF 以上)的图像效果会好一些, 而对于 QCIF 格式而言, 效果并不是很明显.

前面提到零块判决算法可以提前判决全零块的存在, 它可以应用在运动搜索过程中以减少运动搜索的次数. 由于原始图像的一个宏块被映射为解析图像的一个块, 因此解析图像的不规则性远大于原始图像. 如果解析图像 8×8 的块经过自适应零块判决算法被判决为全零块, 那么根据统计的观点可以认为, 原始图像的 4 个亮度块也被量化为全零块. 这样, 在运动搜索过程中就可以简单地利用相应的判决方法来判断全零宏块的存在, 一旦满足结论立刻中止搜索过程并记录下当前的运动矢量. 由于运动矢量离中心点越近, 其 Huffman 码字越短, 因此

在节省运算量的同时,输出码流的长度也相应地减小了,在一定程度上提高了编码的效率。

对宏块中部分非零块的处理可以在后续的整像素搜索中进行,这样,对 16×16 宏块的判断就可以简化为相应 8×8 块的判断,记录下来相应全零块的位置以避免相应的后处理.上面讨论的仅仅是亮度块,对于色度块,本文给出一个简化的结论:如果一个宏块中的亮度块为全零块,则色度块也为全零块.这里利用的是亮度信号的分辨率比色度信号分辨率高的特点,如果亮度信号满足一定的关系,则色度信号也基本满足相应的关系。

由此可见,两种算法的结合使运动搜索和 DCT 以及相应的后处理得到了很大的简化.其实质是在降采样空间的搜索过程中加入判决停止条件,即是否满足图 1 给出的判决条件.一旦满足就将当前的运动矢量标记为本宏块的最佳运动矢量,并把残差信号标记为全零值,同时后续的 DCT、量化、反量化、IDCT 以及运动补偿等过程都被省略掉。

2 仿真结果及其分析

仿真采用了符合 H.263 的 TMN8.0^[6]模型的标准软件(后面简称为常规算法)作为评测的编解码器平台,对大量图像序列进行了测试.这里给出了具有代表性的标准测试序列 Miss America (Miss.cif), Tennis Player (Player.cif)和 Mobile(Mobile.cif)的仿真结果,其中 Miss America 只是简单的头肩部, Tennis Player 和 Mobile 为背景和前景都有运动的复杂图像.实验选取了每个序列 50 帧逐行的 CIF 格式的图像作为数据源。

下面主要从运算复杂度、编码效率和重建图像的质量等几个方面来说明改进算法的优势.其中算法的运算复杂度是以每个宏块的运动矢量搜索次数以及采用自适应零块判决算法统计出的全零块和实际的全零块的比例关系来衡量的.图 3 给出了不同图像序列 PSNR 和比特率的对比关系.从图中可以看出,对于较为复杂的图像而言,在相同码率条件下,图像的客观质量有所提高;而对于头肩部这种背景占主体的简单运动图像,在 PSNR 相同的条件下,在较低码率时,码流有了明显的减小。

表 2 和表 3 给出了不同图像编码复杂度的对比结果.从中可以看出运动搜索的次数相对于常规算法有显著的减少,同时全零块的发现使得 DCT 等后续运算可以忽略.统计结果显示,新方法的使用在保证 PSNR、比特率基本不变的情况下,整个编码过程的时间减少了 20%~40%。

Table 2 Search count of single macroblock

表 2 单个宏块搜索次数对比表

Parameter of quantization	Average search count					
	Miss.cif		Mobile.cif		Player.cif	
	TMN8.0 method	Proposed method	TMN8.0 method	Proposed method	TMN8.0 method	Proposed method
5~15	18.2	12.1	15.6	13.1	28.0	14.8
15~25	18.7	11.6	15.5	13.9	27.0	14.3

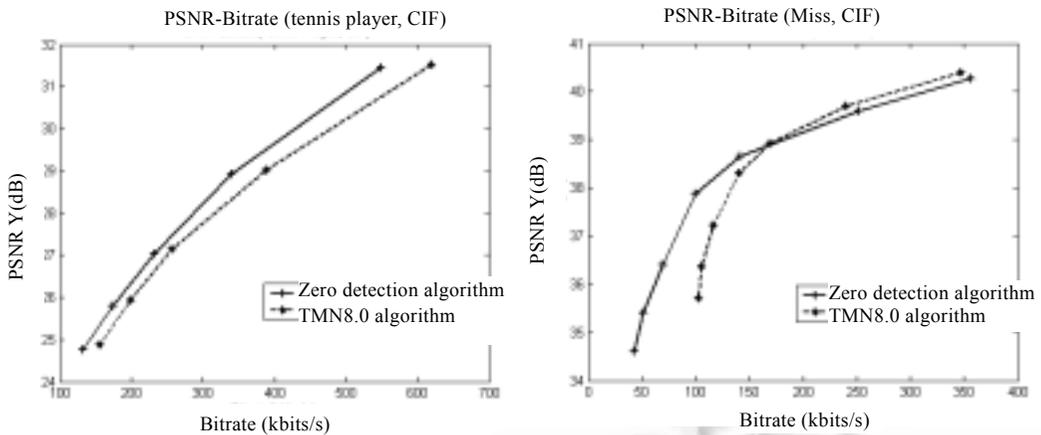
平均搜索次数, 量化参数, TMN8.0 算法, 本文提出的算法.

Table 3 Percentage of predication zero block versus actual zero block

表 3 预测零块与实际零块比例对比

Parameter of quantization	Percentage of prediction zero block versus actual zero block (%)					
	Miss.cif		Mobile.cif		Player.cif	
	Threshold =15Q	20Q	15Q	20Q	15Q	20Q
5	68	72	25	30	56	74
10	77	82	38	47	63	73
15	88	88	37	43	56	66
20	89	93	36	38	52	61
25	92	95	35	37	52	61

测零块和实际零块的比例, 量化参数, 门限.



自适应零块判决算法, TMN8.0 标准算法.
 Fig.3 Comparison of video coding efficiency
 图 3 视频编码性能比较

3 结 论

本文提出了一种多解析度的自适应零块判决算法,它通过降低图像的解析度以及利用零块判决算法来减少运动搜索的次数以及其后的 DCT 等模块的编码时间.实验结果显示,该方法在客观图像质量和比特率基本不变的条件下,编码时间减少了 20%~40%,从而提高了系统的整体性能.

References:

- [1] ITU-T Recommendation H.263 -Video coding for low bit rate communication, 1997.
- [2] ITU-T Study Group 16. Intelligent pre-quantization in motion compensated video coding. Q15-D-35. Tampere, Finland ITU-T, 1998.
- [3] Xie, bo, Zhu, Xue-long. Implementation method of global decision for H.263 very low bitrate image coding. Acta Electronical Sinica, 1999,27(4):66~70 (in Chinese).
- [4] Byung Cheol Song*, Jong Beom Ra. A fast multiresolution block matching algorithm for motion estimation. Signal Processing: Image Communication, 2000,15:799~810.
- [5] Tabatabai, Ali j, Jasinschi, *et al.* Motion estimation methods for video compression—a review. Journal of the Franklin Institute, 1998,335(8):1411~1441.
- [6] ITU-T Study Group 16. TMN8 video codec test model near-term version8. Q15-B-46. Sunriver ITU- T, 1997.

附中文参考文献:

- [3] 谢波,朱雪龙.H.263 甚低码率图像编码的全局判决实现方法.电子学报,1999,27(4):66~70.

Adaptive Zero-Block Detection Algorithm with Multi-Resolution*

LI Qiang, CUI Hui-juan, TANG Kun, DU Wen

(State Key Laboratory on Microwave and Digital Communications, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

E-mail: liq@tsinghua.org.cn; {cuihj,tangk,duw}@ee.tsinghua.edu.cn

http://www.ee.tsinghua.edu.cn

Abstract: In the current applications such as video conferencing, the video codec speed is the key factor to maintain the real-time performance in the whole system. In order to minimize the computations and improve the performance of video coding, an adaptive zero block detection algorithm combined with multi-resolution motion estimation is proposed in this paper. The multi-resolution technique uses down-sampling method on the whole image, which can dramatically reduce the computations in searching for the best candidate block during the motion search period. At the same time, adaptive zero block detection criteria can make the motion search procedure even faster and improve the coding efficiency without any influence on the image quality. Furthermore, the run-time of DCT(discrete cosine transform) and quantization has been significantly reduced afterwards. Simulation results show that with H.263 codec, the proposed algorithm can obtain 40% savings in computations at most without any losses in the subjective and objective quality of decoded images.

Key words: video coding; motion search; multi-resolution; DCT (discrete cosine transform); adaptive zero-block detection

* Received June 15, 2001; accepted May 21, 2002

Supported by the National Natural Science Foundation of China under Grant No.69972020; the National Grand Fundamental Research 973 Program of China under Grant No.G1998030406

第 3 届 SPIE 多谱图像处理与模式识别国际学术会议

征文通知

继第 1、2 届多谱段图像处理与模式识别国际学术会议顺利召开之后,第 3 届 SPIE 多谱图像处理与模式识别国际学术会议计划于 2003 年 10 月 14 日~16 日在北京举行.大会将为从事多谱图像处理与模式识别领域研究的专家、教授、工程技术人员和研究生提供一个相互学习和交流的国际性论坛,为了提高会议的学术水平,会议将特邀国内外知名专家就本学科的一些前沿作专题报告.大会工作语言为英语,会议论文集将由 SPIE 在美国印刷并正式出版,世界发行,欢迎大家积极投稿,参加会议.

一、征文范围(包括但不限于)

1. 多谱图像获取(红外成像,微波成像,雷达和激光雷达成像,超声波成像,高光谱与超谱、医学成像);
2. 多谱图像处理(红外图像处理,微波图像处理,雷达和激光雷达图像处理,超声波图像处理,高光谱与超谱、医学图像处理);
3. 图像分析技术(图像滤波,小波与分形分析,边缘检测,图像分割,图像特征提取,目标识别与跟踪,图像序列分析,图像检索,数据融合与知识发掘);
4. 模式识别与三维视觉(分类技术,神经网络,定标,立体视觉,Shape from X,3D 建模与重建);
5. 图像的并行处理(算法,结构,工具,系统);
6. 优化技术和迭代算法(顺序和并行算法,优化技术,计算方法,优化技术的接口);
7. 应用(商业应用,工业应用,安全应用,多媒体应用,医学应用,文化应用,环境应用).

二、重要日期

论文摘要截止时间:2003 年 3 月 10 日 正式论文截止时间:2003 年 7 月 30 日

三、论文摘要提交方式

1. 电子邮件: mippr@nlpr.ia.ac.cn

如果您的摘要是通过电子邮件发给我们,可使用以下几种方式:文本格式、PDF 格式、Microsoft Word/Rtf 格式.请您注意:为确保接收到您的邮件,请您在邮件主题栏中注明会议主题.

2. 邮寄:您邮寄 3 份摘要的副本到以下地址: 100080 北京海淀区中关村南一条一号中国科学院自动化研究所模式识别国家重点实验室 MIPPR 2003 秘书处

3. 传真:请传真一份摘要至:+86-10-62551993 大会网址:<http://nlpr-web.ia.ac.cn/MIPPR/>

联系人:尹春燕