

# 极低码率视频编码中块自适应分割运动估值

王辉柏 张春田

(天津大学电子信息工程学院, 天津 300072)

**摘要** 在极低码率视频编码运动估值过程中, 针对视频对象的边缘形状特征, 将宏块按不同形状的模板自适应地进行分割, 同时引入近期、远期 2 个参考帧进行块匹配。为提高搜索速度, 首先判断出不需编码的宏块, 并引用了一种快速搜索算法。实验结果表明, 与 H. 263 先进预测模式相比, 采用所建议的运动补偿方法, 图象的信噪比得到明显的提高。

**关键词** 视频编码, 运动估值, 运动补偿, H. 263, MPEG-4

## 1 引言

运动估值在视频压缩编码中起着关键的作用。H. 261、H. 263、MPEG-1、MPEG-2 都将运动估值和运动补偿预测作为消除时间冗余的最基本的技术, 并且都沿用了块匹配运动估值算法<sup>[1]</sup>。MPEG-4 虽然提出了面向对象的编码策略, 但对各个视频对象 (Video Object) 在编码时, 仍将块匹配作为基本的工具之一<sup>[2]</sup>。在块匹配算法中, 一般以正方形像块作为匹配单元, 在 H. 263 的先进预测模式下, 当一个  $16 \times 16$  的宏块内运动情况比较复杂时, 则将其分解成 4 个  $8 \times 8$  的块分别在参考帧中找相应的匹配块力图获得更好的匹配效果<sup>[1]</sup>。然而,  $8 \times 8$  的块分割方法并没有考虑到图象中各视频对象的形状任意性。所以  $8 \times 8$  的块分割方法就不一定是最佳的, 但每个宏块却为此付出 4 个运动矢量的代码开销。一般来说, 在一个宏块中出现多于 2 个的视频对象的机会不多, 因此可设法将宏块一分为二。只要分割得法, 可用 2 个运动矢量来表示分割边界两侧 2 个视频对象的不同运动, 从而在节省传送运动矢量开销的同时, 有可能更逼近真实运动。

在可视电话中, 大部分宏块长时间采用帧间编码, 随着量化误差的积累, 重建图象的质量在逐渐下

降。相对而言, 前面的帧比后面的帧失真要小。此外, 可视电话序列中各对象的运动速度较慢, 运动范围相对较小, 且具有一定的重复性, 如头部的整体运动、眼睛的运动、嘴的运动等都具有这些特性。所以, MPEG-4 把在运动估值时引入近期、远期 2 个参考帧作为核心实验的内容<sup>[2]</sup>, 文献<sup>[3]</sup>证实了在引入近期、远期 2 个参考帧后, 重建图象的信噪比有所提高。

## 2 相对静止宏块的判定与宏块的分割

背景区和相对静止区在可视电话图象中占有相当大的比例。在很多情况下, 序列图象中有一半以上的宏块处在背景区或相对静止区 (例如, Claire 序列和 Missa 序列分别占到了 67.7% 和 51.5%), 这些宏块即使经过运动补偿、DCT、量化后仍不编码, 因此宜首先用简单算法将其选出, 从而免去运动估值等复杂的计算, 提高整个系统的运行速度。我们对典型的可视电话序列进行统计分析, 得出如下结论: 设  $f_t(i, j)$  为当前帧的  $16 \times 16$  宏块中的  $(i, j)$  象素点的亮度,  $f_{t-1}(i, j)$  为前 1 帧的  $(i, j)$  象素点的亮度, 亮度取值范围为  $[0, 255]$ , 如果同时满足下列 2 个条件:

$$|f_t(i, j) - f_{t-1}(i, j)| < 5 \quad 1 \leq i, j \leq 16$$

$$SAD(0,0) = \sum_{i=1}^N \sum_{j=1}^N |f_t(i,j) - f_{t-1}(i,j)| < 200$$

则可判定该宏块为背景区或相对静止区,不必编码。

将背景宏块和相对静止宏块剔除后,再进行运动估值过程。一个宏块中一般不会出现多于 2 个方向的运动,因此只要正确地将处于 2 种不同运动方向的对象沿边缘处分开,就能找到合理的匹配块,得到较好的运动估值效果。我们以 Claire 和 Missa 序列为例,分析了宏块中各运动对象的形状,并以 SAD 作为匹配准则对各种分割方法进行了统计,选出了其中 8 种分割方式,如图 1 所示。在运动估值过程中,自适应地选择其中的一种将宏块一分为二,分别进行运动估值。实验表明,采用 8 种分割方式比采用 4 种分割方式(图 1 中前 4 种)匹配效果要好。虽然传送分割方式要用 3 比特,比采用 4 种分割方式时多 1 比特,但好的匹配效果所节省下来的码字可以补偿这一开销,并且信噪比得到了提高。

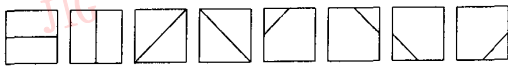


图 1 8 种分割方式

运动估值时,选用 2 帧作为参考帧,如图 2 所示。

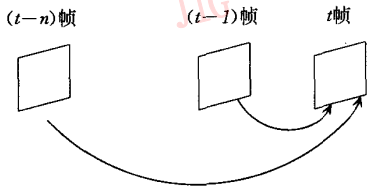


图 2 2 个参考帧

对  $t$  帧进行运动估值时,分别在  $(t-1)$  帧和  $(t-n)$  帧中寻找匹配块,运动估值过程见第 3 部分。 $(t-1)$  帧称为近期参考帧,是当前帧的前一解码重建帧; $(t-n)$  帧称为远期参考帧(亦为解码重建帧),一般情况下比当前帧早 10 帧,但对某帧进行运动估值时,如果有大于 10% 的宏块采用帧内编码,则自下 1 帧开始以此帧替换远期参考帧,以便尽可能地减少累积的量化误差。

### 3 运动估值过程

对每一宏块的运动估值过程如下:

(1) 对每一宏块进行整体的运动估值。以半象

素精度分别在近期和远期 2 个参考帧中搜索最小的 SAD 值与相应的运动矢量,定义为  $SF(SAD)$ 、 $LF(SAD)$  和  $SF(x,y)$ 、 $LF(x,y)$ 。如果  $\min\{SF(SAD), LF(SAD)\} > 5000$ , 则采用帧内编码,否则采用帧间编码。如果决定对某宏块采用帧间编码,则进入第 2 步。

(2) 依次按 8 种不同的分割方式将宏块分成 2 部分,用符号  $P_{i1}$  和  $P_{i2}$  表示,  $i=1,2,\dots,8$ 。再依次在 2 个参考帧中分别对宏块的部分以半像素精度寻找匹配块。得到 4 个匹配块的 SAD 值,分别定义为  $P_{i1}(SF, SAD)$ 、 $P_{i2}(SF, SAD)$ 、 $P_{i1}(LF, SAD)$ 、 $P_{i2}(LF, SAD)$ 。对于每一宏块的每种分割方式,有 4 种参考帧组合方式(以下简称组合方式)产生 4 个 SAD 值,它们分别为:

$$M_i(1, SAD) = P_{i1}(SF, SAD) + P_{i2}(SF, SAD)$$

$$M_i(2, SAD) = P_{i1}(LF, SAD) + P_{i2}(LF, SAD)$$

$$M_i(3, SAD) = P_{i1}(SF, SAD) + P_{i2}(LF, SAD)$$

$$M_i(4, SAD) = P_{i1}(LF, SAD) + P_{i2}(SF, SAD)$$

这 4 个值中的最小值,即为该宏块的第  $i$  种分割方式的最小 SAD 值  $M_i(SAD)$ 。并得到该分割方式下预测宏块的组合方式  $FM_i$  ( $FM_i=1,2,3,4$ )。

$$M_i(SAD) = \min\{M_i(1, SAD), M_i(2, SAD),$$

$$M_i(3, SAD), M_i(4, SAD)\}$$

(3) 8 种分割方式中最小的  $M_i(SAD)$ , 则为该宏块采用分割运动估值时最小的 SAD 值,记为  $M(SAD)$ 。所对应的分割方式  $i$  则为该宏块最终所采用的分割方式。

$$M(SAD) = \min\{M_i(SAD), i=1,2,\dots,8\}$$

(4) 根据  $\min\{SF(SAD), LF(SAD)\}$  和  $M(SAD)$  来决定对该宏块采用何种编码模式,是分割预测模式(PPM)还是整块采用近期参考帧预测模式(SPM)或远期参考帧预测模式(LPM)。

当  $M(SAD) < \min\{SF(SAD), LF(SAD)\} - 50$  时,采用 PPM, 否则,

当  $SF(SAD) < LF(SAD) - 25$  时,采用 SPM, 否则,采用 LPM。

在宏块的头部信息中,需要传送所采用的编码模式。如果采用 PPM 模式,还需传送分割方式  $i$  和组合方式  $FM_i$ 。码流结构与 H. 263 基本相同,不同的是:(1)编码模式采用变字长编码,SPM 用“0”表示,LPM 用“10”表示,PPM 用“11”表示,放在 H. 263 码流中“MCBPC”之前。(2)分割方式  $i$  用 3 位定字长表示,组合方式  $FM_i$  用 2 位定字长表示,放在 H. 263 码流中第一个运动矢量之前。(3)采用 SPM

和LPM时,只有一个运动矢量。用PPM模式时,有2个运动矢量。

### 4 匹配块的快速搜索

运动估值的进一步优化加重了运算的复杂度,寻找一种快速搜索算法变得更为重要。块穷举搜索算法(ESA)虽然能获得最佳的匹配效果,但运算量也是最大的。为了减少ESA的计算复杂度,很多的快速搜索算法,如三步法、对数法、分层法、交叉法已经被研究。这些方法的共同特点是,通过一定的算法减少搜索的位置以减少搜索次数,提高搜索速度,但其获得的最好的匹配结果往往只是局部最优的,与ESA搜索的结果往往不尽相同,换句话说,它们都是以牺牲运动估值的精度来换取运算量的减少。我们采用了一种新的快速搜索算法<sup>[4]</sup>,搜索结果与ESA完全相同,但运算量却减少了许多。

该算法的基本原理是从数学不等式  $| |a| - |b| | \leq |a-b|$  出发,推出:

$$R - SAD(m,n) \leq M(x,y) \leq R + SAD(m,n) \tag{1}$$

式中  $M(x,y)$  为参考帧中运动矢量为  $(x,y)$  时的匹配块的所有象素之和,  $R$  表示当前帧中待匹配块的所有象素之和,  $SAD(m,n)$  表示参考帧中已搜索到的运动矢量为  $(m,n)$  时的块与当前帧中待匹配块的对应象素的差的绝对值之和。运动矢量为  $(x,y)$  的块只有满足条件(1)才使

$$SAD(x,y) \leq SAD(m,n)$$

成立,从而成为更好的匹配块。所以,在得到某候选的运动矢量  $(m,n)$  所对应的  $SAD(m,n)$  后,只需搜索满足条件(1)的那些块,只有  $M(p,q)$  满足式(1),

才求  $SAD(p,q)$ ,并迭代(1)中的  $SAD(m,n)$ ,使得待搜索块进一步减少。这样要求  $SAD$  的块的个数比全搜索时整个搜索窗的块的数量少很多。 $R$  只需计算一次。而求  $M(x,y)$  比求  $SAD(x,y)$  要简单得多,可以用迭代的方法在搜索前将所有块的  $M(x,y)$  求出,设块的大小为  $N \times N$ ,对于大小为  $W \times H$  的搜索窗,只需  $T = 4WH - (H - N)(N + 3) - 3W(N + 1)$  次加法操作,该搜索窗中共有  $B = (W/N)(H/N)$  块,平均每块所需的计算量为:










$$G = T/B = [4 - (H - N)(N + 3)/(WH) - 3(N + 1)/H]N^2 < 4N^2$$

计算一个  $SAD(x,y)$  值则需要  $2N^2 - 1$  次加法操作,所以对满足式(1)计算所需的开销只相当于 2 次搜索,比节省下来的搜索次数要少得多。因此,运动估值的速度得到显著提高。

### 5 实验结果

对宏块进行分割的根据是宏块中各对象的形状,我们对这些形状进行了分析、归类,找出将象块一分为二的 16 种分割方式。分别对宏块的两部分进行运动估值,求出各自的  $SAD$ 。并以 2 个  $SAD$  之和的大小为判据,选出其中  $SAD$  之和最小的一种做为该宏块的分割方式。经过统计、对比,从该 16 种分割方式中选用图 1 所示的 8 种进行自适应地分割时,从信噪比和码字两方面考虑效果最佳。在对 Claire 序列和 Missa 序列剔除出不需编码的背景块和静止块后,其它宏块采用 9 种自适应方式(包括整块方式和 8 种分割方式)进行运动估值的数目所占的百分比见表 1。

表1 采用各种分割方式的宏块数所占的百分比

分割方式									
Claire	5.3%	4.2%	3.4%	3.6%	3.0%	2.6%	8.5%	9.4%	60%
Missa	2.8%	4.9%	2.6%	2.1%	4.3%	4.1%	11.9%	15.6%	51.7%

我们采用表 2 中所列的各种运动估值方法对 Claire 和 Missa 序列进行编码,序列图象的格式为 QCIF。“基本模式”和“先进模式”分别指 H. 263 的 TMN5 测试模型<sup>[5]</sup>中的运动补偿基本模式(无选项)和先进预测模式;“1 帧 4 种”是指只取近期参考帧,取图 1 中前 4 种分割方式,自适应地选取一种将宏块一分为二;“1 帧 8 种”则为本文中详细介绍的 8

种分割方式,但只取近期参考帧;“2 帧”指在 TMN5 的基本模式基础上,引入近期、远期 2 个参考帧;“2 帧 4 种”指不仅采用 4 种分割方式,而且引入 2 个参考帧;“2 帧 8 种”则为 2 个参考帧,8 种分割方式。为了说明各种方法对运动估值的影响,我们将目标帧率固定为 8.33 帧/秒,量化步长固定为 20,各种方法的编码结果如表 2、表 3 所示。表 4、表 5 为恒定码

率为 10k 比特/秒时的实验结果。

表 2 帧率、量化步长固定时 Missa 序列  
编码结果(平均每帧)

	标准 模式	先进 模式	1 帧 4 种	1 帧 8 种	2 帧 2 帧	2 帧 4 种	2 帧 8 种
信噪比	37.85	38.13	38.19	38.31	38.02	38.25	38.42
Y 所用码字	611	365	376	313	457	347	308
mv 等所用码字	300	439	422	444	361	447	463
总码字	1 362	1 256	1 277	1 247	1 276	1 251	1 261
整宏块数	46	35	29	27	45	29	27
分割宏块数	0	10	16	18	0	16	18

表 3 帧率、量化步长固定时 Claire 序列  
编码结果(平均每帧)

	标准 模式	先进 模式	1 帧 4 种	1 帧 8 种	2 帧 2 帧	2 帧 4 种	2 帧 8 种
信噪比	37.72	37.99	37.99	38.12	37.87	38.02	38.19
Y 所用码字	737	449	439	368	524	375	360
mv 等所用码字	164	291	281	307	235	329	327
总码字	1 339	1 187	1 182	1 140	1 218	1 139	1 147
整宏块数	29	21	17	15	29	17	15
分割宏块数	0	8	12	14	0	12	14

表 4 码率为 10kbps 时 Missa 序列编码结果(平均每帧)

	标准 模式	先进 模式	1 帧 4 种	1 帧 8 种	2 帧 2 帧	2 帧 4 种	2 帧 8 种
信噪比	36.97	37.33	37.36	37.47	37.28	37.48	37.64
Y 所用码字	321	213	195	144	286	156	138
mv 等所用码字	267	385	418	442	298	449	486
总码字	970	985	988	998	973	989	982
整宏块数	43	34	29	27	42	29	27
分割宏块数	0	8	13	15	0	13	15

表 5 码率为 10kbps 时 Claire 序列编码结果(平均每帧)

	标准 模式	先进 模式	1 帧 4 种	1 帧 8 种	2 帧 2 帧	2 帧 4 种	2 帧 8 种
信噪比	37.04	37.39	37.45	37.57	37.26	37.62	37.80
Y 所用码字	435	304	296	264	381	279	256
mv 等所用码字	144	253	275	301	171	302	329
总码字	937	931	924	926	928	925	921
整宏块数	26	19	15	14	26	15	14
分割宏块数	0	7	11	12	0	11	12

在表 2 至表 5 中,“Y 所用码字”指亮度信号所用的码字,“mv 等所用码字”指运动矢量和编码模式、分割方式、组合方式总共所用的码字。“整宏块数”是指以  $16 \times 16$  的整块进行运动估值的宏块数,“分割宏块数”是指将宏块分成 2 部分或 4 部分分别进行运动估值的宏块数。

从 4 个表中都可以看出,采用“2 帧 8 种”时,与“先进模式”相比,信噪比有明显的提高,而码字基本持平;与“2 帧 4 种”相比,也有改进。采用 4 种方式和 8 种方式分割时,根据 SAD 判决整宏块/分割预测模式的偏置值为 50(见本文第 3 节中的(4)),而“先进模式”中相应的偏置值为 100<sup>[5]</sup>,所以前者的“分割宏块数”较多,这是信噪比提高、传送亮度信号所需码字减少的一个重要原因。但因为采用 4 种方式和 8 种方式分割时,每个分割块只需传送 2 个运动矢量,比“先进模式”中的分割块少 2 个,所以“总码字”并没有随“分割宏块数”的增加而增加。

## 6 结 论

在极低码率视频编码进行运动估值时,自适应地选用 8 种分割方式中的一种对宏块进行一分为二的分割,并在近期、远期 2 个参考帧中寻求匹配块,与 H. 263 的 TMN5 中的“先进预测模式”相比,在同样的码率、帧率下信噪比有明显的提高。通过在运动估值前去除了不需编码的宏块,并引用逐次递减快速搜索算法,显著减少了由于运动估值算法的改进带来的运算量的增加。

## 参 考 文 献

- 1 Draft ITU-T Recommendation H. 263 Video Coding for Narrow Telecommunication Channels at  $<64\text{ kbit/s}$ , 1995.
- 2 Sikora T. MPEG-4 Very Low Bit Rate Video, ISCA'97, Hang Kang. 3 Fukuhara T. etc., Very Low Bit-Rate Video Coding with Block Partitioning and Adaptive Selection of Two Time-Differential Frame Memories, PCS'96, Melbourne, Australia.
- 4 Li W, et. al, Successive Elimination Algorithm for Motion Estimation, IEEE Transactions on Image Processing, 1995, 4(1): 105 ~ 107.
- 5 ITU-T Standardization Section, Study Group 15, Working Party 15/1, Expert's Group on Very Low Bitrate Visual Telephony, Video Codec Test Model, TMN5, 1995.



王辉柏 博士生,天津大学电子信息工程学院,主要研究方向为图象压缩编码,图象信息的处理与传送。



张春田 教授,博士生导师,中国图象图形学会常务理事,天津图象图形学会副理事长。

## Motion Estimation with Adaptive Block Partitioning in Very Low Bit-Rate Video Coding

Wang Huibai, Zhang Chuntian

(Faculty of Electronic Information Engineering, Tianjin University, Tianjin 300072)

**Abstract** In the process of motion estimation for very low bit-rate video coding, a macroblock is partitioned adaptively with a pattern which matches the shape of edges of the objects in the macroblock, and two time-differential reference frames are used. To speed up the process of motion estimation, the macroblocks which need not to be coded are distinguished before the process, and a fast search algorithm is used during the process. Simulation results demonstrate that, compared with the advanced prediction mode in H. 263, the proposed method for motion compensation makes higher signal-noise ratio (SNR) for the reconstructed image.

**Keywords** Video coding, Motion estimation, Motion compensation, H. 263, MPEG-4.

书 讯

# 从速购买

## 《中国图象图形科学技术新进展》

### 第九届全国图象图形科技大会论文集

《中国图象图形科学技术新进展》一书是第九届全国图象图形科技大会(1998年5月11日~14日)学术论文的汇总,共收录了143篇论文。反映了2年来我国图象图形科技工作者在图象处理、图象分析与分割、图形学、计算机视觉、模式识别、多媒体技术、仿真与模拟、三维几何模型与地图、工程应用等领域的最新研究成果。

该论文集对于图象图形科技工作者、相关领域的研究、开发人员、大专院校师生了解当前该学科研究水平、前沿发展、理论方法等极具参考价值。

该书共565页,定价80元。可以邮购,只需将书款汇至北京海淀区花园路6号,中国图象图形学学会收(邮编100088)即可。银行汇款,开户行:北京工商银行北太平庄分理处;帐号:100144354-82;户名:中国图象图形学学会,款到即寄书及发票。

由于印数较少,欲购从速。咨询电话:(010)62378784