

一种针对 H. 264 的快速编码模式选择算法

吴菊英¹, 李春林², 王选明³

(1. 空军工程大学, 西安 710077; 2. 西安电子科技大学 计算机网络与信息安全教育部重点实验室, 西安 710071;

3. 中国科学院软件研究所 计算机科学重点实验室, 北京 100080)

摘 要: 在 H. 264 编码算法中, 为了给图像的各个区域选择合适的编码模式, 需要对所有可能的模式分别进行计算, 这使编码模式选择过程异常复杂. 针对此问题, 本文提出了一种快速的编码模式选择算法, 该算法利用时间和空间相关性预先对图像各区域可能采用的编码模式进行判定, 从而使模式选择的过程大大简化. 实验结果表明: 在编码器的信息率—失真性能没有明显下降的前提下, 该算法最多可以使编码速度提高 20% 以上.

关键词: H. 264; 视频编码; 模式选择

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2005) 03-0496-04

A Fast Mode Decision Algorithm for H. 264 Video Coding Standard

WU Ju-ying¹, LI Chun-lin², WANG Xuan-ming³

(1. The Air Force Engineering University, Xi'an, Shaanxi 710077, China;

2. Key Laboratory of Computer Network and Information Security, Xidian University, Ministry of Education Xi'an, Shaanxi 710071, China;

3. Key Lab of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing 100080, China)

Abstract: In the encoding algorithm of H. 264, each possible coding mode should be checked to choose the best one for each part of the picture, thus make the process of mode decision rather complex. In this paper, a fast mode decision algorithm is proposed to reduce the complexity of mode decision procedure in H. 264 encoder, in which possible modes are judged beforehand based on the temporal and spacial correlation and the whole procedure of mode decision is simplified then. The results show that the algorithm can increase the encoding speed by more than 20% at most with little degradation in rate-distortion performance.

Key words: H. 264; Video Coding; Mode Decision

1 引言

H. 264/AVC 是 ITU-T 和 MPEG 新近联合制定的一个国际性视频编码标准, 它依然以传统的混合编码框架为基础, 但是提供了更多可供选择的编码模式^[1]. 在 H. 264/AVC 中, 编码器依据内容为图像各部分选择合适的编码模式, 使其整体编码性能相对于先前的标准得到了显著的提高^[2]. 目前, H. 264 编码模式的选择方法主要有两种, 一种是以待编码数据与其预测值的绝对差值之和为主要依据; 另一种则采用了率失真优化的方法^[3]. 然而, 无论是哪一种方法, 都需要对每一种可能的模式分别进行检查, 其中后一种方法还需要按照各种可能的模式做一次实际编码, 以计算该模式下的失真和比特率. 这使得编码过程异常复杂, 大大增加了编码所需要的时间.

然而, 在实际的编码过程中并没有必要对所有的编码模式进行检查. 由于待编码图像中某一区域所采用的编码模式依赖于该区域的内容^[3], 而在摄像机拍摄到的视频序列中, 图像内容的分布又必然在时间上和空间上具有很强的相关性, 完全可以利用这种相关性预先对可能采用的编码模式进行预

测, 从而在其后的编码过程中仅仅考虑这些模式即可. 这样, 势必能够大大减少模式选择过程所耗费的时间. 基于以上思路, 本文提出了一种快速的编码模式选择算法, 该算法利用时间和空间的相关性对可能的编码模式进行预测, 简化了模式选择的过程.

本文首先对 H. 264 提供的编码模式及其模式选择过程进行了分析, 并根据编码模式与图像内容的关系将图像分为背景区域、亚背景区域和运动前景三部分, 然后, 给出了一种根据时间的相关性和空间的相关性对背景区域及亚背景区域进行预测的方法, 利用这种方法, 大大简化了 H. 264 的编码模式选择过程. 最后, 给出相应的计算机实验结果及结论.

2 H. 264 中的编码模式及模式选择方法

与现有的视频编码标准 H. 263、MPEG4 等类似, H. 264 依然采用了块变换和运动补偿相结合的混合编码技术. 所不同的是, H. 264 用 4×4 整数变换代替了以前的 DCT, 同时还提供了多种不同的编码模式. 在 H. 264 中, 编码的基本单位仍然是 16×16 的称作宏块的区域, 编码过程中, 编码器首先将图像

分成若干个宏块,然后针对每个宏块选择一种合适的编码模式并按照该模式对其进行编码。

H.264 的编码模式主要分为帧间预测和帧内预测两大类。其中,帧内预测利用图像的空间相关性,以待编码区域空间上相邻像素的值为预测值。亮度信号的帧内预测模式共有 13 种,其中针对 16×16 区域的有 4 种,针对 4×4 区域的有 9 种。帧间预测则利用视频序列时间上的相关性,把经过运动补偿的上一帧图像的数据作为预测值^[4]。帧间预测模式是根据运动预测过程中匹配块的尺寸来划分的。H.264 仍然采用块匹配的方法^[4]来进行运动预测,但是其匹配块的尺寸是可变的,共有 16×16 、 16×8 、 8×16 、 8×8 、 8×4 、 4×8 、 4×4 七种。依据匹配块尺寸的不同,宏块的帧间模式分为四种,前三种模式分别按照一个 16×16 块、两个 16×8 块和两个 8×16 块来进行运动预测;最后一种模式记作 P8 \times 8,在 P8 \times 8 模式下,一个宏块被分为 4 个 8×8 的子块,而每一个子块又有 4 种可能的子模式,分别按照一个 8×8 块、两个 8×4 块、两个 4×8 块及四个 4×4 块进行运动预测(图 1)。

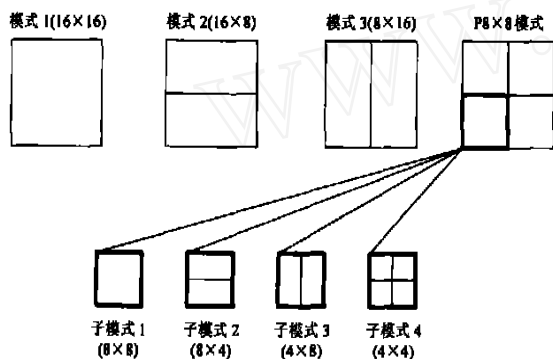


图 1 H.264 中的帧间预测模式及子模式

需要说明的是,上面提到的编码模式并非可以用于任何一帧图像。一些图像只允许使用帧内模式,称为帧内编码图像;一些图像既可以使用帧内模式也可以使用帧间模式,称为帧间编码图像;通常,为了保证图像的质量,需要定期地使用帧内编码图像。

依照目前采用的编码模式选择方法,编码器需要对每个宏块所有可能的模式分别进行检查,然后才能确定合适的编码模式。以帧间编码图像为例,每个宏块要按照以上提到的四种模式分别进行运动预测,其中对于 P8 \times 8 模式,还要对四个 8×8 的块——按照四种可能的子模式分别进行运动预测,然后,根据各种模式下运动预测的结果选出一个最佳的模式,并将其与帧内预测模式作比较,最后决定所要采用的编码模式。如果采用率失真优化的方法,还需要按照各种模式及子模式分别对宏块进行编码,从而得到相应的码率和信噪比,以此为根据来进行模式选择。显然,这种方法是相当耗时的,难以达到实时视频通信的要求,因此,需要将待编码图像各区域可能采用的编码模式预先计算出来,从而节省大量的编码时间。

3 基于背景预测及亚背景预测的快速模式选择算法

图像采用的编码模式与其内容有很大的关系,一般,静止

区域和运动简单的区域(如背景)采用匹配块较大的帧间编码模式,运动复杂的区域(如图像中的运动物体)采用匹配块较小的帧间编码模式^[5]。由此,本文按照编码模式将图像分为三个部分并给出以下定义:第一部分为采用 16×16 帧间预测模式或 16×16 帧内预测模式的区域,本文将此定义为背景区域,第二部分为采用 16×8 及 8×16 帧间预测模式的区域,本文将此定义为亚背景区域,另一部分为采用 8×8 帧间模式和各种子模式的部分,本文将此定义为运动前景。

在通常的视频序列中,背景区域或亚背景区域在空间上和atory 上的相关性是很强的,尤其是一些可视电话或视频会议应用中常见的头肩像序列,背景区域的位置随时间的变化很小而且具有很强的空间相关性。因此,完全可以利用这些相关性,在编码前对背景区域及亚背景区域预先进行预测。

3.1 利用时间的相关性进行背景区域预测

本文用 $M_n(x, y)$ 表示第 n 帧图像中的一个宏块,该宏块左上角的像素位于图像的 (x, y) 处。由于视频序列在时间上具有很强的相关性,宏块 $M_n(x, y)$ 的内容必然和上一帧同位置的宏块 $M_{n-1}(x, y)$ 的内容有很大的关系,考虑到运动的影响, $M_n(x, y)$ 的内容与 $M_{n-1}(x, y)$ 相邻宏块的内容也有一定的关系。

为此定义一个宏块集:

$$\mathbf{Mref}_n = \{ M_{n-1}(x, y) \mid |x| < MV_{xm}, |y| < MV_{ym} \}$$

并把它称为宏块 $M_n(x, y)$ 的参考区域。其中, MV_{xm} , MV_{ym} 分别是 x 方向和 y 方向的运动矢量的最大值。

如果参考区域 \mathbf{Mref}_n 中的所有宏块都采用了 16×16 的帧间预测模式或者针对 16×16 区域的帧内预测模式(该模式适合于图像中的平坦背景区域),则 \mathbf{Mref}_n 属于背景区域,可以认为宏块 $M_n(x, y)$ 也属于背景区域。在这种情况下,编码器忽略 16×16 以外的所有帧间预测模式,而只对其按照 16×16 的块尺寸进行运动预测。

在实际的编码过

程中,对于帧内编码帧和第一个帧间编码帧,必须对所有可能的编码模式进行检查,以得到一个精确的模式选择结果,并依此将图像的内容划分为背景、亚背景和运动前景三部分;而对于其他的帧间

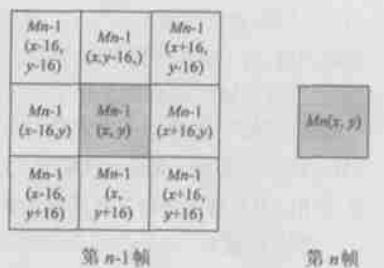


图 2 利用时间的相关性进行背景预测 (其中 $MV_{xm} = MV_{ym} = 16$)

编码帧,就可以采用以上方法根据上一帧图像背景区域的分布来对当前图像中的背景区域进行预测了。

图 3 给出背景区域预测的仿真实验结果。其中, (a) 为第 1 帧图像的编码模式选择情况; (b) 为根据第 1 帧的编码模式选择情况对第 2 帧图像进行背景区域预测的结果(白色部分为预测背景区域); (c) 为不采用背景区域预测时第 2 帧图像的模式选择结果; (d) 为采用背景区域预测时第 2 帧图像的模式选择结果。

3.2 利用空间的相关性进行背景及亚背景预测

利用本文 3.1 中给出的算法,可以比较精确地将大面积

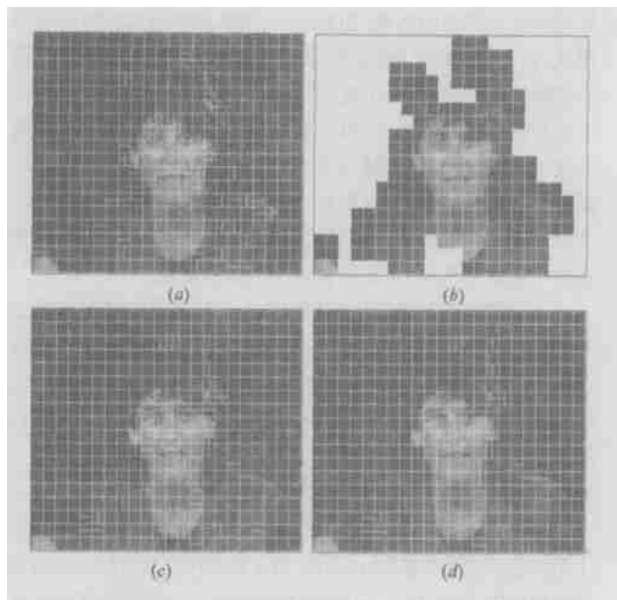


图3 背景区域预测示意 (MissA_cif)

连续背景区域预测出来(图3),但是对于某些情况,如 $M_n(x, y)$ 的参考区域 M_{ref_n} 中至少有一个宏块不属于背景区域,3.1中的算法便无法判断宏块 $M_n(x, y)$ 是否属于背景区域了.基于这一问题,本文给出了一种利用空间相关性进行预测的方法,用于以上提到的情况.在这种方法中,需要利用 16×16 帧内预测模式下宏块中各点数据与其预测值的绝对差值之和 (SAD_{mb}) 来推测宏块属于背景区域、亚背景区域还是运动前景.对于宏块 $M_n(x, y)$,其 SAD_{mb} 值由下式得到:

$$SAD_{mb}(M_n(x, y)) = \sum_{i=0}^{15} \sum_{j=0}^{15} |Y(x+i, y+j) - \hat{Y}(x+i+u, y+j+v)|$$

其中, $Y(x^*, y^*)$ 为当前图像在 (x^*, y^*) 处的像素值, $\hat{Y}(x^*, y^*)$ 为参考图像在 (x^*, y^*) 处的像素值, (u, v) 为宏块的运动矢量.对于整幅图像来说, SAD_{mb} 的大小与宏块属于背景区域、亚背景区域还是运动前景并没有直接的关系,但在局部的小区域(例如相邻的若干个宏块)中,属于背景区域的宏块的 SAD_{mb} 值明显要低于属于运动前景的宏块的 SAD_{mb} 值.因此,本文采用以下方法来进行背景区域和亚背景区域的预测:

如图4,宏块 $M_n(x, y)$ 为当前待编码宏块, $M_n(x-16, y)$ 、 $M_n(x-16, y-16)$ 、 $M_n(x, y-16)$ 、 $M_n(x+16, y-16)$ 是与 $M_n(x, y)$ 相邻的四个已编码宏块, $SAD_{mb}(M_n(x, y))$ 、 $SAD_{mb}(M_n(x-16, y))$ 、 $SAD_{mb}(M_n(x-16, y-16))$ 、 $SAD_{mb}(M_n(x, y-16))$ 和



图4 利用空间相关性进行背景及亚背景预测

$SAD_{mb}(M_n(x+16, y-16))$ 分别为五个宏块对应的 SAD_{mb} 值.若 $M_n(x-16, y)$ 、 $M_n(x-16, y-16)$ 、 $M_n(x, y-16)$ 、 $M_n(x+16, y-16)$ 中有 $n(n \geq 1)$ 个宏块采用了 16×16 的帧内预测模式或者 16×16 的帧内预测模式,依照下式得

$$SAD_{AVG} = \frac{1}{n} \sum_{i=0}^{n-1} SAD_{mb}(X_i)$$

其中宏块 $X_i \in \{M_n(x-16, y), M_n(x-16, y-16), M_n(x, y-16), M_n(x+16, y-16)\}$ 且 X_i 采用 16×16 帧内模式或 16×16 帧内预测模式.

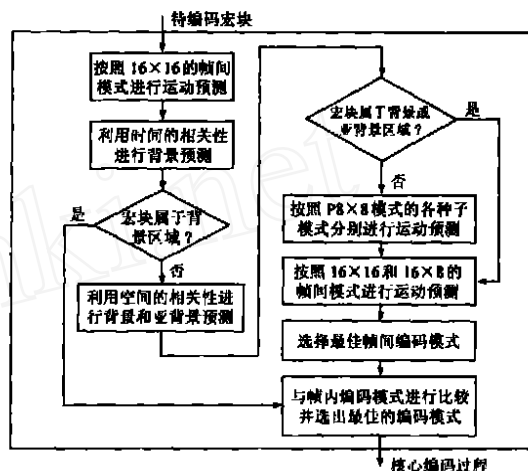


图5 简化后的模式选择过程框图

然后,将 $SAD_{mb}(M_n(x, y))$ 与 SAD_{AVG} 做比较,如果 $SAD_{mb}(M_n(x, y)) < SAD_{AVG}$,则认为宏块 $M_n(x, y)$ 属于背景或亚背景部分,编码过程中不予考虑 8×8 的帧内预测模式及其各种子模式.否则,所有可能的编码模式都需要考虑.

如果 $M_n(x-16, y)$ 、 $M_n(x-16, y-16)$ 、 $M_n(x, y-16)$ 、 $M_n(x+16, y-16)$ 中没有任何一个宏块采用 16×16 的帧内预测模式或者 16×16 的帧内预测模式,便无法对 $M_n(x, y)$ 进行背景及亚背景预测,在编码过程中依然需要考虑所有的编码模式.

3.3 模式选择过程的简化

本文通过采用以上两种背景及亚背景预测的方法,对 H.264 的编码模式选择过程进行了简化,简化后的模式选择过程如图5所示:

从图中可以看出,本文介绍的算法为编码模式选择的过程中引入了两个判决步骤,利用这两个判决步骤可以减少诸多不必要的运动预测运算.

表1 编码时间比较(QP=28,不使用率失真优化方法)

序列	采用本文中的算法			H.264 测试程序 JM61e					
	PSNR Y (dB)	比特率 (kbps)	时间 (ms)	PSNR Y (dB)	比特率 (kbps)	时间 (ms)	PSNR Y (降低)	比特率 (增加)	时间 节约
Container . qcif (10fps/100 帧)	35.73	25.60	96.44	35.71	25.79	82.66	0.02	0.74 %	14.29 %
MissA . cif (30fps/100 帧)	38.35	158.46	399.99	38.32	159.78	312.02	0.03	0.20 %	21.90 %
Football . cif (30fps/80 帧)	36.85	969.15	318.48	36.84	969.81	279.87	0.01	0.07 %	12.12 %

4 相应实验结果及分析

本文采用以上介绍的快速编码模式选择方法对 H. 264 的测试程序 JM61e 进行了改进,表 1 和表 2 给出相应的实验结果:根据实验结果可以看出,采用本文中的算法可以在信噪比降低小于 0.05db,比特率增加不超过 1 %的前提下,节约 10 % ~ 20 % 的整体编码时间.对于运动较小的低比特率序列,该方法的效果尤其显著.

5 结论

针对最新的国际视频编码标准 H. 264/ AVC,本文提出了一种新的快速编码模式选择算法.该算法依照编码模式和图像内容的关系,将图像分为背景区域、亚背景区域和运动前景三部分,通过对背景区域和亚背景区域进行编码前预测使模式选择的过程得到了简化,从而大大提高了编码速度,实验结果表明,采用本算法,编码器的信息率-失真性能不会有明显的下降.

参考文献:

[1] Thoms Wiegand. Draft ITU-T recommendation and final draft international standard of joint video specification[A]. (ITU-T Rec. H. 264 | ISO/ IEC 14496-10 AVC) . 7th JVT Meeting and 19th VCEG Meeting [C]. Pattaya :2003.

[2] Nejat Kamaci , Yücel Altunbasak. Performance Comparison of the Emerging H. 264 Video Coding Standard with the Existing Standards [A]. Proc IEEE Inter Conf on Multimedia and Expo (ICME2003) [C]. Baltimore :2003.

[3] Gary J. Sullivan and thomas wiegand. rate-distortion optimization for video compression [J]. IEEE Signal Processing Magazine ,Nov ,1998 , 15 (11) :74 - 90.

表 2 编码时间比较(QP = 28 ,使用率失真优化方法)

序列	采用本文中的算法			H. 264 测试程序 JM61e					
	PSNR Y (dB)	比特率 (kbps)	时间 (ms)	PSNR Y (dB)	比特率 (kbps)	时间 (ms)	PSNR Y (降低)	比特率 (增加)	时间 节约
Container . qcif (10fps/ 100 帧)	35.88	23.82	164.61	35.84	23.97	147.42	0.04	0.63 %	10.63 %
MissA . cif (30fps/ 100 帧)	38.55	130.49	644.17	38.53	130.75	534.85	0.03	0.20 %	16.97 %
Football . cif (30fps/ 80 帧)	37.27	942.52	533.45	37.24	943.78	472.79	0.03	0.13 %	11.43 %

[4] J R Jain ,A K Jain. Displacement measurement and its application in interframe image coding [J]. IEEE Trans Commun , 1981 , COM-29 (12) :1799 - 1808.

[5] M H Chan ,Y B Yu. Variable size block matching motion compensation with applications to video coding [A]. IEEE Proceedings , 1990 , 137 (4) .

作者简介:



吴菊英 女,空军工程大学,讲师,主要研究方向:密码学.



李春林 男,西安电子科技大学,计算机网络与信息安全教育部重点实验室,主要研究方向:密码学.