

多视点视频编码快速帧间模式选择算法

王凤随, 沈庆宏*, 都思丹

(南京大学 电子科学与工程学院, 南京 210023)

(* 通信作者电子邮箱 mynhsiu@gmail.com, qhshen@nju.edu.cn)

摘要:为解决多视点视频编码(MVC)可变尺寸块的模式选择计算复杂度过大的问题,提出了基于模式复杂度的多视点视频编码快速模式选择算法。该算法首先分析了多视点视频编码参考模型(JMVC)中各尺寸块的分布特点。然后,提出模式复杂度的概念以确定当前宏块的模式特征。最后,将宏块分成3个不同的模式类型:如果当前宏块属于简单模式,那么仅 16×16 分块被检查,其他分块均跳过;若当前宏块属于中等模式,则 8×8 块被跳过;若当前宏块属于复杂模式,所有模式分块都要检查。这样算法对那些不必要的模式选择过程就可以提前终止,从而使得计算量大大减少。实验结果表明:所提算法能保持同JMVC中全搜索算法几乎相同的编码效率,同时使计算复杂度减少62.75%。

关键词:多视点视频编码;模式选择;模式复杂度;变尺寸块;帧间预测

中图分类号: TN919.8 **文献标志码:** A

Fast inter-mode decision algorithm for multi-view video coding

WANG Fengsui, SHEN Qinghong*, DU Sidan

(School of Electronic Science and Engineering, Nanjing University, Nanjing Jiangsu 210023, China)

Abstract: In order to solve greatly computational complexity for the variable block mode decision in Multi-view Video Coding (MVC), a fast inter-mode decision algorithm based on mode complexity for multi-view video coding was proposed. First, the characteristics for each variable block size of the Joint MVC (JMVC) were analyzed in the proposed algorithm. Then, the mode complexity was presented to determine the mode characteristics of the current macroblock. Finally, macroblocks were divided into three different mode classes: for macroblocks with simple mode, only mode size of 16×16 was checked, and other mode sizes were skipped; for macroblocks with medium mode, mode size of 8×8 was skipped; for macroblocks with complex mode, all mode sizes were tested. As a result, the unnecessary mode decision process could be early terminated in the method and computational load can be greatly reduced. The experimental results have demonstrated that the proposed method is able to significantly reduce the computational load by 62.75%, while keeping almost the same coding efficiency, compared with the full mode decision in the reference software of MVC.

Key words: Multi-view Video Coding (MVC); mode decision; mode complexity; variable block size; inter-frame prediction

0 引言

多视点视频(Multi-View Video, MVV)因其能给人以强烈的立体感和交互性,能更加生动地再现场景^[1-2],在工业和学术界越来越多地受到关注。多视点视频编码在自由视点电视(Free-view point Television, FTV)、三维电视(Three Dimensional Television, 3DTV)、家庭娱乐、教育、虚拟现实及远程医学诊疗等领域有着广泛的应用前景。同传统的单视点视频相比,多视点视频的数据量巨大,且会随着摄像机数目的增加而成倍增加。巨大的数据量对存储和传输提出了更高的要求,已经成为制约多视点视频编码(Multi-view Video Coding, MVC)实际应用的瓶颈。因此,为了将多视点视频应用到实际系统中,必须对MVV进行有效的视频压缩编码。

由视频编码专家组(Video Coding Experts Group, VCEG)和运动图像专家组(Motion Picture Experts Group, MPEG)组成的联合视频组(Joint Video Team, JVT)提出了H.264的扩展标准——多视点视频编码标准^[3],在JVT第27次会议上,公布了标准的多视点视频编码软件(Joint Multi-view Video Coding, JMVC),作为多视点视频编码的研究平台。在JMVC中,除了具有同一视点的运动估计外,还引入了视点间的视差估计以消除视点间的冗余。此外,为获得更高的图像质量, JMVC采用分层B帧(Hierarchical B Pictures, HBP)结构^[4]、全搜索算法和率失真优化函数等,这些方法可获得较高的编码图像质量,但同时也引入了极高的计算复杂度,影响了其编码速度,阻碍了多视点视频的应用。为降低计算复杂度,国内外已有学者提出各种模式选择算法^[5-9]。Ding等^[5]提出

收稿日期: 2013-07-05; **修回日期:** 2013-09-16。 **基金项目:** 国家自然科学基金资助项目(61100111);江苏省科技支撑项目(BE2011169);江苏省交通科研计划项目(08Y29-7);南京市科技计划项目(201103051)。

作者简介: 王凤随(1981-),男,安徽宿州人,博士研究生,主要研究方向:图像通信、视频编码、计算机视觉; 沈庆宏(1969-),男,江苏无锡人,副教授,博士,主要研究方向:视频图像处理、嵌入式系统; 都思丹(1967-),女,江苏南京人,教授,博士,主要研究方向:信号处理、嵌入式系统。

一种基于视点相关性的内容感知模式选择方法,根据对应宏块的编码信息来决定视差估计是否启用,但该算法仅考虑相邻视点的相关性,并未考虑其他相关性。Huo 等^[6]提出一种可伸缩的预测结构,此方法可以自适应地跳过运动估计过程,然此算法只对单一模式进行优化,其计算量的减少有限。Zhu 等^[7]基于纹理分割和视点相关性提出快速帧间模式选择算法,实验结果表明算法使得计算量得到有效降低。但该算法需要借助 Intra 模式率失真代价来区分不同的纹理区域,增加了额外的比特率,使用范围也受到一定的限制。Peng 等^[8]基于对不同运动测试序列的统计测量值的分析,提出用三个阈值来提前结束模式选择的三步策略。然而该算法主要利用视点内模式间的相关性,并没有进一步考虑利用视点之间编码信息的相关性来降低计算复杂度。Shen 等^[9]根据相邻视点的模式分布相关性,提出一种快速帧间模式选择算法,但此算法仅利用相邻视点模式之间的相关性信息,并未考虑空间和时间内模式之间的相关性,且相邻宏块对当前宏块的不同影响也未考虑。

基于对现有算法的分析,本文在分析了 JMVC 中各模式分布情况的基础上,提出一种基于模式复杂度的快速模式选择算法。本文算法充分利用了空间、时间和视点间的相关性信息,同时考虑了不同位置宏块对当前宏块影响程度的不同。算法根据所定义的模式复杂度将宏块分成 3 个不同的模式类型,每种类型仅检查相对应的模式分块,其余不必要的模式选择过程可以提前终止,从而使计算量得到有效的降低。实验结果表明,该算法在保证图像质量基本不变的情况下,能显著地减少模式选择的数目,从而提高多视点视频编码的编码速度。

1 MVC 模式选择及分布情况

1.1 JMVC 模式选择算法

为了达到更高的编码效率, MVC 提供多个变尺寸块预测模式以消除多视点视频之间的时间、空间及视点间的冗余。对可变块运动估计 (Motion Estimation, ME) 和视差估计 (Disparity Estimation, DE), 如图 1 所示, JMVC 中将一个宏块分成 16 × 16、16 × 8、8 × 16 和 8 × 8, 每个 8 × 8 块又进一步被划分成 8 × 4、4 × 8 和 4 × 4 子块。一般而言, 大的宏块模式划分较适合于编码纹理简单均匀或者运动缓慢的区域; 而小的宏块模式划分则较适合压缩编码复杂纹理或快速运动的区域。此外, H. 264 中的增强帧内预测也被加入到 MVC 中, 这对消除空间冗余大有帮助。

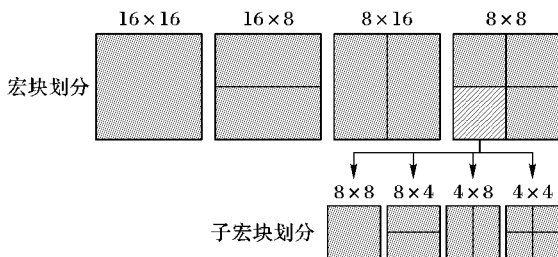


图 1 MVC 可变块运动估计和视差估计

为了获得更高的编码图像质量, JMVC 用拉格朗日率失真优化函数作为模式选择准则^[10], 对所有预测模式进行全搜

索, 检查并计算每一种模式的率失真代价 (Rate Distortion cost, RD cost) 值, 最后选择具有最小 RD cost 的模式作为最优模式。由于 JMVC 支持多种预测模式且每种模式 RD cost 的计算又十分耗时, 全搜索算法的计算复杂度是非常大。因此, 提出一种快速模式选择算法减少计算复杂度是十分必要的。

1.2 JMVC 模式分布情况

帧间预测之所以采用变尺寸块预测模式的其中一个原因就是能够更好地捕获自然序列的真实运动信息或视差信息, 这样就可以更准确地表示目标运动或深度信息。一般来说, 具有活跃运动或丰富纹理的宏块其预测模式具有多样化; 另一方面, 具有均匀运动与纹理的宏块其预测模式通常都是较大的分块^[11]。从上述分析不难看出, 对于处于均一运动和均一纹理区域的宏块, 它们的模式尺寸数目实际是很少的, 并没有必要对所有的分块都逐一进行检查。如果可以根据当前宏块所处区域的运动和纹理情况, 仅检查可能性最大的一个或几个模式分块, 那么, 其他不必要的分块就不用检查, 这样就可以减少大量 RD cost 的计算过程。

通过对 4 个不同类型的 MVV 序列进行实验, 得到全搜索模式下 MVC 各个分块模式的统计分布情况。统计结果见表 1。实验中的参数设置如下: 1) 图像组 (Group Of Pictures, GOP) 为 16; 2) 量化参数 (Quantization Parameter, QP) 为 32; 3) 启用率失真优化 (Rate Distortion Optimization, RDO) 技术; 4) ME 和 DE 搜索范围为 ± 64。

表 1 MVC 各分块模式统计分析 %

测试序列	Flamenco1	Ballroom	Akko&Kayo	Uli	平均值
16 × 16	85.76	93.27	89.23	78.30	86.64
16 × 8	5.46	2.28	4.45	6.42	4.65
8 × 16	6.34	2.86	5.28	7.26	5.44
8 × 8	2.44	1.59	1.04	8.02	3.27

从表 1 可看出, 大部分宏块最终都选择 16 × 16 分块作为最优模式, 平均为 86.64%。另一方面, 从表 1 也可看出, 选其他分块为最优模式的宏块比例非常低, 平均低于 4.5%。应当指出的是, 全搜索算法是对所有模式进行检查, 找出 RD cost 最小的模式作为最优模式, 整个过程的计算复杂度非常大。所以, 如果能够设计出一种方法, 可以提前确定当前最优模式分块是否为 16 × 16 分块, 那么对其余的模式分块就不必检查, 这些模式的 RD cost 的计算便可省去, 这样就可以有效地降低模式选择的计算量。

2 本文算法

2.1 模式复杂度

从 1.2 节的分析不难看出, 如何确定当前宏块的模式特征是算法实现的关键。为此, 本文提出模式复杂度的概念, 用以确定当前宏块所处的区域类型, 模式复杂度可以反映 MVV 序列的属性 (包括运动和纹理信息、预测模式、RD cost 等)。由于多视点视频包含大量的空间、时间和视点相关性, 模式复杂度的定义也来自于帧内、帧间及视点之间, 定义如下:

如图 2 所示, 当前宏块为 MB0, 空间相邻宏块分别是左 MB1、上 MB2 和右上 MB3 宏块; MB4 为当前宏块在先前已编

码图像里处于同一位置的宏块;MB5 是当前宏块在相邻视点图像里的对应宏块,它们由全局视差矢量(Global Disparity Vector, GDV)来识别^[12],MB $i(i=6,7,\dots,13)$ 为 MB5 的 8 个相邻宏块。由于相邻视点图像是在同一时刻不同角度拍摄的,通常呈现相同的内容,对应宏块之间呈现高度的相关性。模式复杂度 δ_{MC} 定义如式(1)所示:

$$\delta_{MC} = \frac{1}{N} \cdot \sum_{i=1}^N K_i \cdot \alpha_i \quad (1)$$

其中: N 是宏块总数,这里为 13; α_i 是宏块加权因子,其值如表 2 所示; K_i 是模式加权因子,其值的大小根据每种模式的复杂度而分配。在本文算法中,将帧间模式分块根据复杂程度分为 3 类:大尺寸(16×16),中等尺寸(16×8 和 8×16)及小尺寸(8×8)。通常情况下,大部分具有均匀运动与纹理的宏块最后都选择大尺寸模式;而具有复杂纹理与剧烈运动的宏块很可能选择小尺寸模式。每种帧间模式的模式加权因子 K_i 值如表 3 所示。

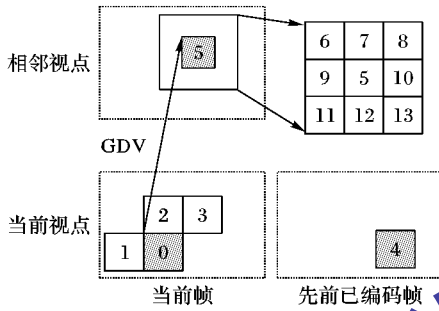


图 2 当前宏块的空间、时间及视点间相邻宏块

表 2 相应宏块的宏块加权因子

MB _{<i>i</i>}	1,2,4,5	3,7,9,10,12	6,8,11,13
α_i	1.65	1.04	0.30

表 3 相应模式的模式加权因子

模式	16×16	16×8	8×16	8×8	Intra
K_i	1	2	2	3	4

模式加权因子越大,对应宏块的复杂度就越高。实验中设置两个阈值 Th_1 和 Th_2 来确定当前宏块所处的模式类型,分类如式(2)所示:

$$\text{宏块模式特征} = \begin{cases} \text{简单模式,} & \delta_{MC} \leq Th_1 \\ \text{中等模式,} & Th_1 < \delta_{MC} \leq Th_2 \\ \text{复杂模式,} & Th_2 < \delta_{MC} \end{cases} \quad (2)$$

其中: Th_1 为假定当前宏块的对应宏块及相邻宏块均为 16×16 块(即 $K_i = 1$)时的值; Th_2 为假定当前宏块的对应宏块及相邻宏块均为 16×8 块或 8×16 块(即 $K_i = 2$)时的值。

2.2 算法步骤

本文算法的具体步骤如下:

第 1 步 检查当前宏块是否处于关键图像上,若是,则执行全搜索模式选择以确定其最优模式;否则,进入第 2 步。

第 2 步 根据式(1)计算出当前宏块的模式复杂度 δ_{MC} 。

第 3 步 根据式(2)确定当前宏块所处的模式类型。

第 4 步 如果 $\delta_{MC} \leq Th_1$,即当前宏块属于简单模式,仅 16×16 块被检查,其他分块被跳过;如果 $Th_1 < \delta_{MC} \leq Th_2$,当

前宏块属于中等模式, 8×8 块被跳过;如果 $Th_2 < \delta_{MC}$,则当前宏块属于复杂模式,所有模式分块都要被检查。

3 实验结果与分析

为了进一步评估算法的性能,实验以 JMVC8.0 为测试平台,对本文算法在 7 个不同 MVC 视频序列进行了测试。这些序列分别具有 $320 \times 240, 640 \times 480$ 和 1024×768 各种不同的分辨率,同时包含缓慢运动、剧烈运动等各种不同的运动特征和简单纹理、复杂背景等各种不同的纹理特征。实验参数设置如下:1) GOP 长度为 16;2) QP 值分别设置为 24,28,30 和 32;3) 启用 RDO 和基于上下文的自适应算术编码(Context-Adaptive Binary Arithmetic Coding, CABAC)技术;4) ME 和 DE 的搜索范围: ± 64 。

为进一步验证本文算法的有效性,将本文算法和文献[9]算法进行了比较,结果如表 4 所示。应当指出的是,每种算法都是与 JMVC 中全搜索模式选择方法做比较的结果。表 4 中各参数的定义如式(3):

$$\begin{cases} \Delta T = (T_{\text{proposed}} - T_{\text{JMVC}}) / T_{\text{JMVC}} \\ \Delta PSNR = PSNR_{\text{proposed}} - PSNR_{\text{JMVC}} \\ \Delta B = (Bitrate_{\text{proposed}} - Bitrate_{\text{JMVC}}) / Bitrate_{\text{JMVC}} \end{cases} \quad (3)$$

其中 T_{JMVC} 、 $PSNR_{\text{JMVC}}$ 和 $Bitrate_{\text{JMVC}}$ 分别表示 JMVC 中全搜索模式下的编码时间、峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)值和比特率; T_{proposed} 、 $PSNR_{\text{proposed}}$ 和 $Bitrate_{\text{proposed}}$ 分别表示快速模式选择算法(如文献[9]算法、本文算法)下的编码时间、PSNR 值和比特率。

表 4 本文算法与文献[9]算法的性能比较

测试序列	方法	$\Delta PSNR/\text{dB}$	$\Delta B/\%$	$\Delta T/\%$
Flamenco1	文献[9]算法	-0.04	-0.20	-45.46
	本文算法	-0.06	-0.22	-59.50
Race1	文献[9]算法	-0.02	-0.14	-63.14
	本文算法	-0.02	-0.10	-67.98
Ballroom	文献[9]算法	-0.01	0.53	-55.46
	本文算法	-0.02	0.55	-63.29
Exit	文献[9]算法	-0.01	0.17	-60.22
	本文算法	-0.01	0.27	-66.89
Akko&Kayo	文献[9]算法	-0.01	0.10	-49.81
	本文算法	-0.02	0.12	-61.00
Rena	文献[9]算法	-0.01	-0.23	-62.73
	本文算法	-0.02	-0.23	-65.81
Uli	文献[9]算法	-0.03	0.38	-36.75
	本文算法	-0.06	0.50	-54.79
平均值	文献[9]算法	-0.02	0.09	-53.37
	本文算法	-0.03	0.13	-62.75

从表 4 可看出,本文算法能够显著地减少计算复杂度,平均减少 62.75%,而编码效率几乎保持不变,PSNR 仅降低 0.03 dB,比特率仅增加了 0.13%。同文献[9]算法相比,本文算法减少 9.38% 的计算复杂度。总之,本文算法对各种不同测试序列都能有效地降低计算复杂度。

为更好地比较,图 3 给出了序列 Ballroom 的 RD 曲线。从图 3 可看出,本文算法能达到和 JMVC 全搜索几乎相同的编码效率。图 4 是本文算法与文献[9]算法的时间节省率的

比较,容易看出,对各种不同类型的测试序列,本文算法总能比文献[9]算法更大程度地减少计算复杂度。

同本文算法相比,文献[9]算法仅利用视点间的相关性定义模式复杂度,而实际上,模式复杂度不仅与视点间模式有关,也同时与空间和时间上的模式有关,本文算法兼顾了这三个方面的相关性,因此取得了比文献[9]算法更好的结果。此外,文献[9]算法仅考虑了各个模式的加权因子,并没有考虑到处于不同位置的宏块对当前宏块的影响程度是不一样的,也即必须要考虑宏块加权因子,这也是本文算法优于文献[9]算法的另一个原因。

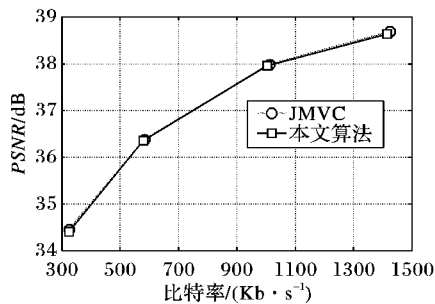


图3 Ballroom序列RD曲线

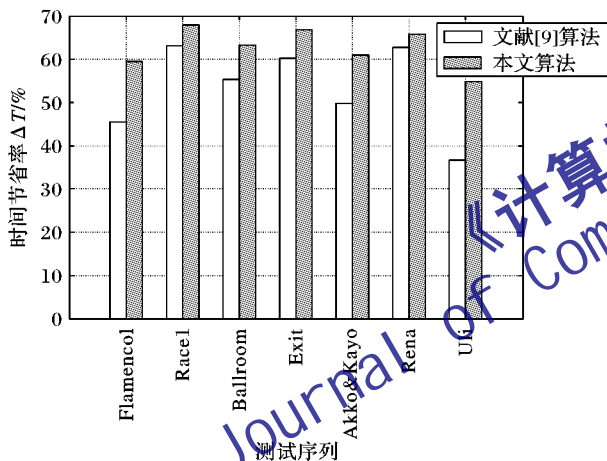


图4 本文算法与文献[9]算法计算复杂度比较

4 结语

本文提出一种有效的多视点视频编码帧间快速模式选择方法,计算复杂度减少主要是因为模式选择的过程中跳过不必要的模式分块。在本文算法中,首先计算当前宏块的模式复杂度,然后根据其值的大小确定宏块所处的模式类型,每种类型仅选择指定的模式分块,其余模式分块则不必检查。

模式复杂度的设计是算法的关键,提出算法综合考虑空间、时间及视点间的相关性定义模式复杂度参数。实验结果表明同JMVC的全搜索算法相比,提出算法不仅大大减小了计算复杂度,且保持了几乎不变的编码效率。

参考文献:

- [1] TANIMOTO M, TEHRANI M P, FUJII T, *et al.* Free-viewpoint TV [J]. IEEE Signal Processing Magazine, 2011, 28(1): 67-76.
- [2] MULLER K, MERKLE P, WIEGEND T. 3-D video representation using depth maps [J]. Proceedings of the IEEE, 2011, 99(4): 643-656.
- [3] PANDIT P, VETRO A, CHEN Y. JVT-AA208 Joint Multiview Video Model (JMVM) 8.0 software[R]. Geneva: JVT, 2008.
- [4] VETRO A, WIEGAND T, SULLIVAN G J. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard [J]. Proceedings of the IEEE, 2011, 99(4): 626-642.
- [5] DING L F, TSUNG P K, CHIEN S Y, *et al.* Content-aware prediction algorithm with inter-view mode decision for multiview video coding [J]. IEEE Transactions on Multimedia, 2008, 10(8): 1553-1563.
- [6] HUO J Y, CHANG Y L, LI M, *et al.* Scalable prediction structure for multiview video coding [C]// Proceedings of the 2009 IEEE International Symposium on Circuits and Systems. Piscataway, NJ: IEEE Press, 2009: 2593-2596.
- [7] ZHANG W, TIAN Y, ZHOU F, *et al.* Fast inter mode decision based on texture segmentation and correlations for multiview video coding [J]. IEEE Transactions on Consumer Electronics, 2010, 56(3): 1696-1704.
- [8] PENG Z J, JIANG G Y, YU M. A fast multiview video coding algorithm based on dynamic multi-threshold [C]// Proceedings of the 2009 IEEE International Conference on Multimedia and Expo. Piscataway, NJ: IEEE Press, 2009: 113-116.
- [9] SHEN L Q, YAN T, LIU Z, *et al.* Fast mode decision for multiview video coding [C]// Proceedings of the 2009 IEEE International Conference on Image Processing. Piscataway, NJ: IEEE Press, 2009: 2953-2956.
- [10] WIEGAND T, SCHWARZ H, JOCH A, *et al.* Rate-constrained coder control and comparison of video coding standards [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 688-703.
- [11] WANG F S, ZENG H Q, SHEN Q H, *et al.* Efficient early direct mode decision for multi-view video coding [J]. Signal Processing: Image Communication, 2013, 28(7): 736-744.
- [12] KOO H S, JEON Y J, JEON B M. JVT-W081 MVC Motion skip mode[R]. San Jose: JVT, 2007.

(上接第98页)

- [16] SONG Y, JIANG G, XU J. An epidemic spreading model in adaptive networks based on cellular automata[J]. Acta Physica Sinica, 2011, 60(12): 110-119. (宋玉蓉, 蒋国平, 徐加刚. 一种基于元胞自动机的自适应网络病毒传播模型[J]. 物理学报, 2011, 60(12): 110-119.)
- [17] CAO C, WANG L, ZHAO D. The research based on the discrete cellular ant algorithm in the geometric constraint solving[J]. Acta Electronica Sinica, 2011, 38(5): 1127-1130. (曹春红, 王利民, 赵大哲. 基于离散元胞蚂蚁算法的几何约束求解技术研究

[J]. 电子学报, 2011, 38(5): 1127-1130.)

- [18] BRINKHOFF T. A framework for generating network-based moving objects[J]. Geoinformatica, 2002, 6(2): 153-180.
- [19] ZOU Y, ZHANG Y. Location-cloaking algorithm based on grid-divided space[J]. Application Research of Computers, 2012, 29(8): 3059-3061. (邹永贵, 张玉涵. 基于网格划分空间的位置匿名算法[J]. 计算机应用研究, 2012, 29(8): 3059-3061.)