

应用于多视点视频压缩的多参考 B 帧快速编码算法

郁梅^{①②} 徐秋敏^① 蒋刚毅^{①②} 陈恳^①

^①(宁波大学信息科学与工程学院 宁波 315211)

^②(北京大学视听信息处理国家重点实验室 北京 100871)

摘要: 基于 H.264 平台就多视点视频压缩中多参考帧技术存在计算复杂度高的现状, 根据多视点视频序列特性, 提出了一种针对多参考 B 帧的快速多视点视频编码压缩算法。新算法改进了多参考帧列表策略, 并结合多视点视频序列的时空域相关性特性, 提出了适用于多视点视频编码的参考帧列表调整策略。此外, 还针对多种块模式编码引入了块模式预判策略。新算法在保证率失真性能的基础上, 有效地降低了计算复杂度, 提高了多视点视频编码速度。虽然本文是基于 M-Picture 编码结构实现了新算法, 但它同样适用于其它应用多参考 B 帧预测技术的多视点视频编码结构中。

关键词: 多视点视频编码; 多参考 B 帧; 参考帧列表调整; 块模式预判

中图分类号: TN919.8

文献标识码: A

文章编号: 1009-5896(2008)06-1400-05

A Fast Coding Algorithm for Multiple Reference B-frames in Multi-view Video Compression

Yu Mei^{①②} Xu Qiu-min^① Jiang Gang-yi^{①②} Chen Ken^①

^①(Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China)

^②(National Key Laboratory of Machine Perception, Peking University, Beijing 100871, China)

Abstract: Considering high complexity of multiple reference frames in Multi-view Video Compression (MVC), a fast coding algorithm for MVC is presented based on the traits of multi-view video. The propose algorithm improves the design of reference-lists in H.264 standard that adapts to traditional video coding, and adjusts the old reference-lists using the temporal and spatial correlations of multi-view video. Additionally, the strategy of mode pre-decision is also used to reduce the computational complexity. The experimental results show that the proposed method can reduce encoding complexity and speed up encoding process effectively. Though the new algorithm is implemented on M-Picture prediction structure, it can be also used in other multi-view video coding structures with multiple reference B-frames.

Key words: Multi-view Video Coding (MVC); Multiple reference B-frames; Reorder reference-lists; Mode pre-decision

1 引言

多视点视频是一种新型的具有立体感和交互操作功能的视频^[1, 2]。多视点视频信号可以通过多视点视频成像、编解码与传输系统进行采集、压缩、传输、解码与显示。视点间相关性是多视点视频信号的重要特性, 它与相机阵列形式、相机间距、相机与拍摄对象间距离等存在着极大的联系, 直接地反映在同一时刻相邻视点两幅图像的视差上。在进行多视点视频编码时, 采用结合时空域的多参考帧进行联合运动估计和视差估计的预测方法可以更有效地进行编码压缩, 如采用顺序视点预测编码结构^[3]、基于 M-Picture 的编码

案^[4, 5]、基于 HBP(Hierarchical B Picture)的多视点编码方案^[6]等等。

M-picture 编码方案通过高复杂度的率失真优化与时空域联合预测来减少预测残差和帧内块数, 以提高编码效率, 但在计算复杂度上付出较大代价。它采用 21 种帧模式: 帧内(1 种), 单向帧间预测(4 种), 单向帧间预测兼残差编码(4 种), 双向预测(6 种)和双向预测兼残差编码(6 种)。图 1 给出了 M-Picture 编码预测结构, 图中的 25 个图像帧构成一个图像组(GOP, Group of Picture), 图中 GOP 边缘帧采用传统视频编码的 I, B 和 P 帧, 而中间的图像为 M 帧(如图中的 M6, M7, ..., M18 等), 它采用多参考 B 帧的编码方式, 以充分利用相关联 4 个 GOP 边缘帧的相关性。多视点视频编码时一般将多个视点的视频信息按一定顺序形成单一视频流, 图 1 中的数字就代表了图像帧在视频流中的编码顺序。现有的多参考 B 帧预测编码需要对多种编码模式进行遍

2006-11-14 收到, 2007-05-25 改回

国家自然科学基金(60672073), NSFC-KOSEF(中韩)合作研究项目, 教育部科学技术研究重点项目(2060059), 教育部新世纪优秀人才计划项目(NECT-06-0537)和浙江省自然科学基金(Y105577)资助课题

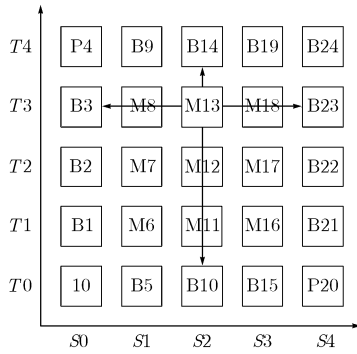


图 1 M-Picture 编码预测结构

历^[7]: 包括双向预测的直接模式, 7 种块模式的运动估计, 帧内块编码, 另外基于 7 种块模式分割遍历两个参考帧列表 (List0, List1) 内所有的参考帧, 通过块匹配搜索完成对 B 帧的率失真优化编码。多参考 B 帧的率失真优化策略不仅要获得宏块各种编码模式的编码代价, 采用帧间模式时还需要计算多参考帧的前向、后向和双向预测的编码代价, 最终通过编码代价确定编码宏块的最优编码模式以及相应的参考帧选择和运动矢量信息等。

多视点视频编码中多参考 B 帧编码, 要比普通的单个视点视频压缩中的多参考 B 帧编码更为复杂, 因为它不仅要考虑多视点视频序列时间相关性, 而且需要进一步考虑多视点视频序列视点相关性。本文分别从调整参考帧列表和块模式预判两个方面提出了基于 H.264 的多视点视频编码中多参考 B 帧的快速编码算法, 并选用 M-Picture 编码结构作为本文算法与原有算法的对比平台, 分别从率失真和计算复杂度两个角度进行性能对比, 说明了新方法的有效性。

2 多视点视频编码中的多参考 B 帧快速预测编码算法

2.1 基于多视点视频序列特性的参考帧列表改进策略

在 H.264 编码标准中, 多参考帧列表是基于单通道视频编码帧与前后向最临近的参考帧相关性最大这一原则来进行多参考帧列表的设置。多参考 B 帧是根据它们在视频流中的顺序参数 Picture Order Count (POC) 来确定其在参考帧列表中的位置, 具体设置如下: 参考帧列表 List0 起始于编码帧时间上最临近的前向参考帧, 此后依次按 POC 降序排列前向参考帧, 最后按 POC 升序排列后向参考帧; 类似地, 参考帧列表 List1 起始于编码帧时间上最临近的后向参考帧, 此后依次按 POC 升序排列后向参考帧, 最后按 POC 降序排列前向参考帧, 如图 2 所示。不同于普通视频的编码, 在多视点视频编码中进行时空域联合预测, 多参考帧之间的相关性会大大降低, 增加多参考帧数量能够更为明显地带来编码率失真性能的提高。

根据 H.264 的多参考帧设置策略^[7], 编码 B 帧时实际上对所有参考帧进行了两次遍历, 重复操作使得块匹配计算量增加了一倍。率失真代价函数决定了参考帧列表中首位参考

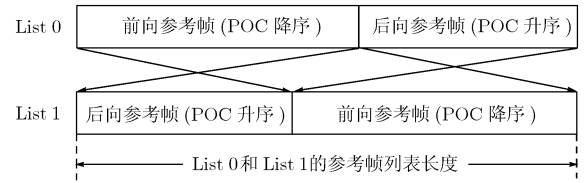


图 2 B 帧的多参考帧列表设置方式

帧具有相对更小的参考帧和运动矢量代价, 两个参考帧列表的重复参考帧对于编码效率的提高效果并不明显, 重复的参考帧被选为最优参考帧的概率也十分之小。所以, 同时在参考帧列表 (List0, List1) 内存放前后项参考帧并进行遍历的方案并不合理, 它不仅大大增加了计算复杂度, 也没有达到有效提高编码效率的目的。另一方面, 多视点视频编码中多参考帧不仅可以分为前后向参考帧, 还有时间域参考帧和空间域参考帧之分, 那么, 多视点视频序列的时空域相关性分布特性也将直接影响着这两类参考帧与编码帧之间的相关性。相机间距、视频内容变化情况、帧率都是影响视频序列相关性的主要参数, 不同多视点视频序列可能具有多种不同的相关性特性, 即使同一视频序列随着视频内容的变化、帧率的改变也会导致序列相关性的变化^[9]。

依据多视点视频信号特性^[8]将其分为两类, 即以时间相关性为主和以空间相关性为主。若仍采用 H.264/AVC 视频编码标准的参考帧设置策略, 多视点视频编码过程中就不一定满足相关性最强的参考帧放置于参考帧列表的首选位置, 无法获得合适的多参考帧列表, 这就需要针对不同序列的相关性特性来调整时空域参考帧在参考帧列表中的位置。基于多视点视频序列相关性特性分析, 针对编码 B 帧的参考帧列表, 提出如下基于多视点视频序列特性的改进策略 (以 4 个参考帧为例):

(1) 舍去原有参考帧列表设置方案中两个参考帧列表的重复参考帧部分, 即 List0 只存放前向多参考帧, 而 List1 中只存放后向多参考帧。

(2) 修改原有参考帧列表之后, 依据多视点视频序列的相关性特性进一步调整参考帧列表。首先需要确定序列特性, 针对 M-Picture 编码结构可以通过两个 P 帧参考 I 帧进行视差估计或运动估计编码时所获得的 I 块数量作为判断当前序列时空域相关性的参数, I 块 (即 H.264 中进行帧内预测编码的块) 数量越多则表明编码帧与参考帧之间的相关性愈差; 反之, 则相关性愈佳^[8], 通过两个 P 帧的 I 块数量就可以直接有效地反映出序列的相关性特性。这里设置阈值 $T = P_{t_i} / P_{s_i}$ 作为序列相关性特性的分界值。 P_{t_i} 表示对应于图 1 中的时间域 (S0, T4) 帧的 I 块数量, P_{s_i} 表示对应于图 1 中的空间域 (S4, T0) 帧的 I 块数量。经大量统计实验分析, 阈值 T 小于 1.5 时, 则认为该序列时间相关性更为突出; 当阈值 T 大于等于 1.5 时, 则认为该序列以空间相关性为主。

(3) 对于时间相关性为主的视频序列, 其参考帧列表调整

为前向时间域参考帧置于 List0 的首位，其后为前向空间域参考帧；后向时间域参考帧置于 List1 的首位，其后为后向空间域参考帧。

(4)对于空间相关性为主的视频序列，将前后向空间域参考帧分别置于参考帧列表 List0、List1 首位，其它则根据参考帧的前后项分类分别置于两个参考帧列表中。这样两个参考帧列表长度均减半。

2.2 块模式预判决策略

B 帧的双向直接预测模式分为两类：时间直接预测模式和空间直接预测模式。时间直接预测模式依赖于前后向参考帧之间的预测关系以及连续性^[9]，而多视点视频编码的多参考帧已经不能满足于时间上参考帧间的密切连续性，根据其视点间相关性，有时需要空间直接预测模式加以补充，以提高预测性能。

7 种块模式的块匹配是造成编码计算复杂度增加的主要原因之一，然而，经分析发现率失真优化最终确定为直接模式的宏块在一帧图像中占有较大比例，若能够预先判断出当前编码块是否采用直接模式，就能够大幅度降低块匹配的搜索次数。对于平坦区域多采用直接模式或宏块级的块模式分割，而纹理丰富区域则多采用亚宏块级的块模式分割，而相邻宏块的块模式选择存在着一定的相似性，因此通过相邻宏块的块模式预先判定当前块模式的类型具有其合理性，并能够在一定程度上降低计算复杂度。

多视点视频编码的宏块编码模式预测方案为：首先，对于没有相邻宏块信息可以利用的图像边界宏块(上边界以及左边界)采用遍历 7 种块模式及所有参考帧的编码方式进行边界处理；而非图像边界宏块均有相邻宏块的模式信息作为编码宏块模式的预判信息，这里选用如图 3 所示的相邻有效宏块(左宏块 A，上宏块 B 以及右上宏块 C)的块模式信息，对编码宏块 D 的预测块模式(Pred_mode)进行分类：

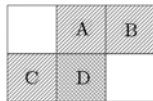


图3 编码宏块与其相邻参考宏块

(1)若相邻宏块中存在直接模式，则 Pred_mode=0：计算编码宏块 D 在直接模式下预测残差的绝对差和(sum of absolute difference, SAD)，若满足阈值条件则不再进行其它块模式的块匹配搜索；否则还需要再实施宏块级块模式的块匹配搜索以及帧内编码模式；

(2)若相邻宏块中只存在宏块级块模式(16×16, 16×8, 8×16)，则 Pred_mode=1：只进行宏块级块模式的块匹配搜索以及帧内编码模式；

(3)若相邻宏块中只存在亚宏块级块模式(8×8, 8×4, 4×8, 4×4)，则 Pred_mode=2：只进行亚宏块级块模式的块

匹配搜索以及帧内编码模式；

(4)若相邻宏块中同时存在宏块级与亚宏块级块模式，则 Pred_mode=3：同时进行宏块级和亚宏块级块模式的块匹配搜索以及帧内编码模式；

其中，如何选择直接模式的 SAD 阈值是影响预判宏块直接模式正确与否的重要因素。由于 SAD 值反映的是预测重构宏块与原图像相应宏块亮度值差异的累加和，图像本身的纹理信息将会影响 SAD 值，一般纹理丰富宏块的 SAD 值会明显大于平滑区域宏块的 SAD 值，因此，选择硬阈值将无法很好适应序列内容的变化。本文中通过统计当前帧内已编码宏块中采用直接模式的 SAD 均值来作为当前编码宏块的 SAD 阈值，此阈值也将随着采用直接模式编码宏块的纹理信息自适应调整。图 4 为本文多视点视频编码算法的总流程图。

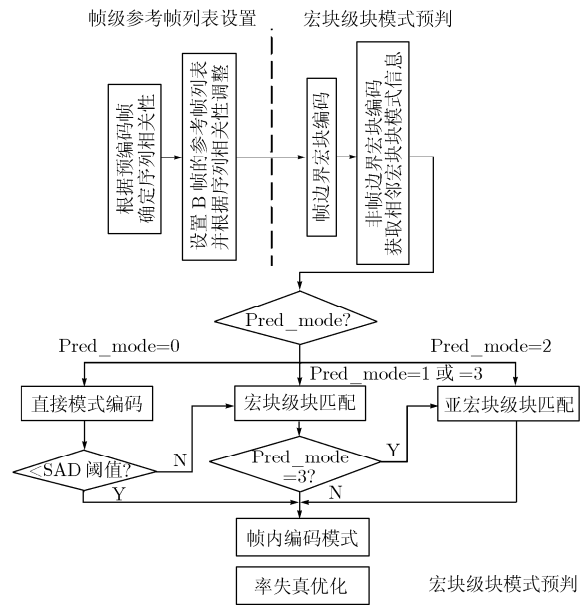


图4 本文算法实现总流程

3 实验分析

采用了 H.264 的 JM8.6 校验模型实现本文算法，选取图 5 所示的 Flamenco1、Race2 以及 Xmas 等 3 个多视点标准测试序列^[10]进行实验对比分析。其中，Flamenco1 与 Race2 的图像分辨率为 320×240，相机间距 200mm；Xmas 序列属于稠密序列，其图像分辨率为 640×480，我们选取间距 9mm 的视点作为测试序列。3 个序列测试具有代表性，Flamenco1 序列为时间相关性突出的视频内容，Race2 序列选用时空相关性基本一致的一段视频内容；而 Xmas 序列的特点在于相机间距小，它的视点间相关性远远强于时间相关性。实验编码参数如表 1 所示。这里通过率失真曲线图来对比本文算法与原方案的编码效率，并通过块匹配时间和编码时间来反映本文算法在编码计算复杂度方面的情况，利用参数 η 体现

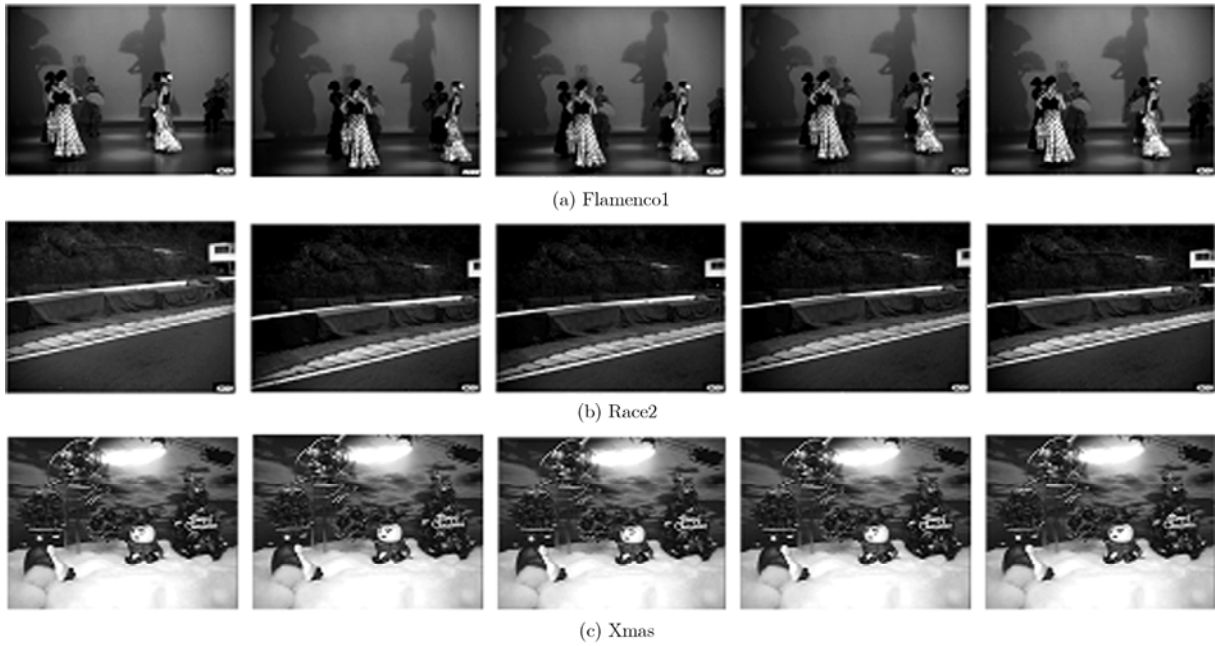


图 5 3 组多视点标准测试序列中某一时刻的 5 个视点图像

本文算法对比原方案编码时间的降低幅度, 参数 η 定义见式 (1), 其中 T_{Proposed} 表示本文算法的编码时间, T_{Original} 则表示原方案的编码时间。

$$\eta = \left(1 - \frac{T_{\text{Proposed}}}{T_{\text{Original}}} \right) \times 100\% \quad (1)$$

表 1 编码参数

| | | | |
|-----------|-----------|----------------|-------------|
| RDO | YES | 搜索范围 | ± 32 |
| FrameRate | 30 | UseFME | YES |
| 熵编码 | CABAC | DirectModeType | Spatial |
| 精度 | 1/4 Pixel | QP | 18,24,30,36 |

首先, 对帧级参考帧列表设置的改进策略进行实验分析。图 6 为两种编码方案的率失真性能对比, 其中 M-Picture 是基于 H.264 多参考帧设置方式进行多视点视频编码的率失真曲线, Lists 表示为本文在 JM8.6 平台上实现的多参考帧列表改进方案的编码率失真曲线。从实验结果可以看到, 经过调整的参考帧列表虽然减少了部分参考帧, 但并没有造成编码效率的损失, 无论是在低码率端还在高码率端, 该方案依然保持了很好的编码压缩效率。与此同时, 如表 2 所示, 多参考帧列表调整方案的计算复杂度得到了显著的降低, 块匹配时间和总编码时间分别降低了 40%与 25%左右。表 2 中分别用参数 η_{BS} 和 η_{AE} 表示块匹配时间和总编码时间的降低程度。

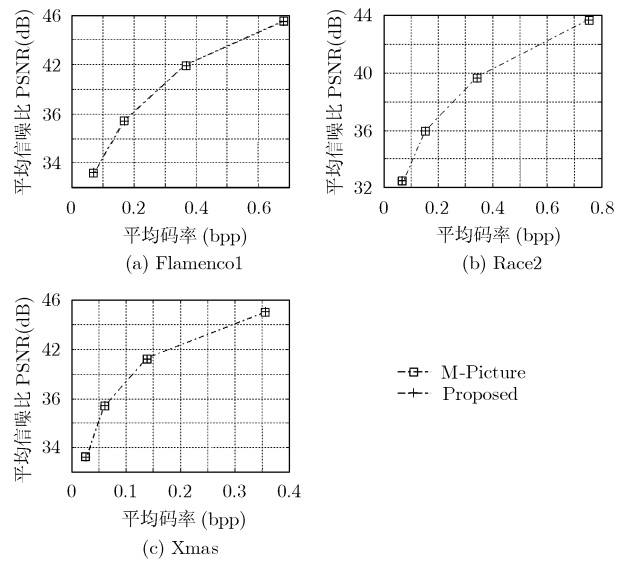


图 6 帧级参考列表设置改进策略与原方案的率失真性能对比

表 2 帧级参考列表设置改进策略的编码时间降低幅度(%)

| 序列 | Flamenco1 | | Race2 | | Xmas | |
|-------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| | η_{BS} | η_{AE} | η_{BS} | η_{AE} | η_{BS} | η_{AE} |
| QP=18 | 42.72 | 22.41 | 42.69 | 25.84 | 35.99 | 21.23 |
| QP=24 | 42.07 | 25.75 | 43.60 | 28.69 | 39.32 | 23.72 |
| QP=30 | 39.90 | 26.79 | 42.29 | 29.03 | 32.30 | 21.91 |
| QP=36 | 42.09 | 26.80 | 37.67 | 27.34 | 38.68 | 26.46 |
| 平均 | 41.70 | 25.44 | 41.56 | 27.73 | 36.57 | 23.33 |

图7给出了本文所提出的参考帧列表调整算法和块模式预判算法的实验分析, M-Picture 仍表示基于 H.264 多参考帧设置方式进行多视点视频编码的率失真曲线, 本文方案(即图中的 Proposed)则结合了本文提出的两种改进算法所获得的率失真曲线。图7中反映出对于时间相关性强的测试序列(如 Flamenco1 和 Race2)在高码率端的率失真性能下降约 0.2dB, 而对于空间相关性强的测试序列(如 Xmas)在中低端码率率失真性能也下降约 0.2dB, 由前一个实验结果了解到率失真性能的降低主要源于块模式预判。虽然 3 个测试序列在率失真性能方面都存在不同程度的损失, 但同时在计算复杂度方面有了进一步的提高, 在编码总时间方面平均降低了 50%左右, 运动估计和视差估计运算的累加时间减少更达到 70%以上。表 3 给出了 3 个序列在不同量化步长下采用本文方案(Proposed)所获得的完全编码时间降低幅度 η_{AE} 和块搜索时间降低幅度 η_{BS} , 本文算法在不同量化步长编码时的结果相近, 有较好的适应性。

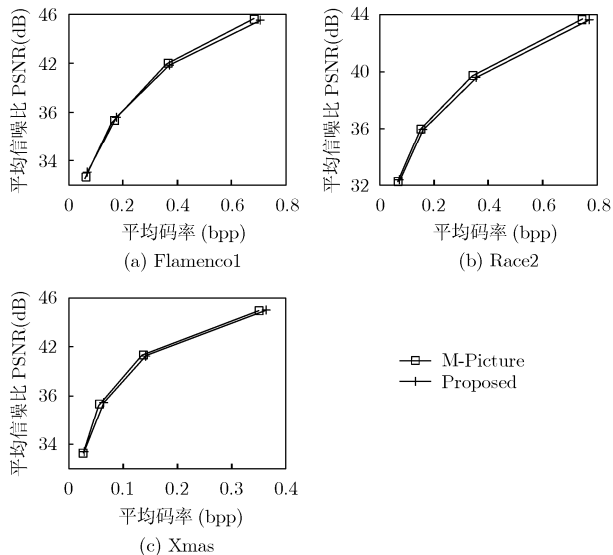


图7 本文两种策略结合方案与原方案的率失真性能对比

表3 图7中本文方案(Proposed)编码时间降低幅度(%)

| 序列 | Flamenco1 | | Race2 | | Xmas | |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|
| | η_{BS} | η_{AE} | η_{BS} | η_{AE} | η_{BS} | η_{AE} |
| QP=18 | 66.32 | 41.30 | 68.44 | 48.73 | 74.76 | 43.96 |
| QP=24 | 67.85 | 46.13 | 71.86 | 52.30 | 79.48 | 49.99 |
| QP=30 | 72.86 | 51.71 | 75.14 | 55.54 | 85.30 | 65.56 |
| QP=36 | 78.23 | 60.30 | 77.82 | 61.36 | 86.69 | 70.78 |
| 平均 | 71.32 | 49.86 | 73.32 | 54.48 | 81.56 | 57.57 |

4 结束语

在对 H.264/AVC 视频编码标准的多参考帧列表以及多宏块运动估计进行分析研究的基础上, 根据多视点视频编码

的特点, 针对计算复杂度极大的多参考 B 帧, 提出了一种基于 H.264/AVC 的快速编码算法, 实现了基于多视点视频特性的自适应参考帧列表调整, 以及利用相邻宏块信息进行的宏块编码模式的预判操作。实验结果表明, 该快速算法以率失真性能的略微下降为代价, 大幅度降低了多视点视频编码的计算复杂度, 其总编码时间在原有基础上降低了约 50%。本文基于 M-Picture 编码结构实现了新算法, 但它同样适用于其它采用多参考 B 帧预测技术的多视点视频编码结构, 如基于 HBP 的多视点视频编码结构等。

参考文献

- [1] ISO/IEC JTC1/SC29/WG11. Report on status of 3DAV work exploration[C]. N5878, Trondheim, Norway, 2003.
- [2] ISO/IEC JTC1/SC29/WG11. Requirements on multi-view video coding v.4[C]. N7282, Poznan, Poland, 2005.
- [3] ISO/IEC JTC1/SC29/WG11. Survey of algorithms used for multi-view video coding (MVC), N6909, HongKong, China, 2005.
- [4] Oka S, Bangchang P N, and Fujii T. Dynamic ray-space coding for FTV [A]. Proc. 3D Image Conference [C], Tokyo, 2004: 139-142.
- [5] Oka S, Endo T, and Fujii T. Dynamic ray-space coding using multi-directional picture [A]. IEICE Technical Report[C], 2004, 104(493): 15-20.
- [6] Mueller K, Merkle P, and Schwarz H. Multi-view video coding based on H.264/MPEG4-AVC using hierarchical B pictures, PCS2006, CD-ROM, Beijing, China, 2006.
- [7] Flierl M and Girod B. Generalized B pictures and the draft H.264/AVC video-compression standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(7): 587-597.
- [8] Zhang Y, Jiang G, and Yi W, et al. An Approach to Multi-Modal Multi-View Video Coding. Proc. of ICSP06, Guilin, China, 2006: 1405-1408.
- [9] Tourapis A M, Wu F, and Li S. Direct mode coding for bi-predictive slices in the H.264 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, 15(1):119-126.
- [10] ISO/IEC JTC1/SC29/WG11. KDDI multi-view video sequences for MPEG 3DAV use [C], 2004.

郁梅: 女, 1968年生, 教授, 主要研究领域为多媒体信号处理、编码与通信、计算机视觉及其应用。

徐秋敏: 女, 1982年生, 硕士生, 研究领域为多视点视频信号处理、编码与传输。

蒋刚毅: 男, 1964年生, 教授, 主要研究领域为多媒体信息传输与信息安全、图像处理与视频信号编码、基于视觉的智能控制。

陈 息: 男, 1962年生, 副教授, 主要研究领域为图像处理、基于视觉的智能控制。