

# 一种密集相机阵列的低复杂度视频压缩方法

金智鹏<sup>1)</sup> 郁梅<sup>1,2)</sup> 蒋刚毅<sup>1,2)</sup> 刘尉悦<sup>1,2)</sup> 蒋志迪<sup>1)</sup>

<sup>1)</sup>(宁波大学信息科学与工程学院, 宁波 315211) <sup>2)</sup>(北京大学视听信息处理国家重点实验室, 北京 100871)

**摘要** 为了降低多视点视频压缩的复杂度,根据密集相机阵列系统与 Wyner-Ziv 编码的特点,提出了基于 Wyner-Ziv 编码的密集相机阵列低复杂度视频压缩算法。该算法首先在各相机之间相互独立地采用基于感兴趣区提取的低复杂度来进行 Wyner-Ziv 编码,然后在中心解码端利用各视点间的相关性进行联合解码。该算法是通过 DCT 量化系数进行相似性判断来提取感兴趣区,以有效地避免对背景和平坦等区域进行编码,从而降低了编码复杂度。实验证明,该算法具有极低的编码复杂度特性和良好的率失真性能,编码复杂度仅为 H.264 帧间预测编码的 1/22。

**关键词** 视频压缩 密集相机阵列 Wyner-Ziv 编码 低复杂度编码

**中图分类号**: TN919.81 **文献标识码**: A **文章编号**: 1006-8961(2006)11-1592-05

## Low Complexity Video Compression in Dense Camera Array

JIN Zhi-peng<sup>1)</sup>, YU Mei<sup>1,2)</sup>, JIANG Gang-yi<sup>1,2)</sup>, LIU Wei-yue<sup>1,2)</sup>, JIANG Zhi-di<sup>1)</sup>

<sup>1)</sup>(Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211)

<sup>2)</sup>(National Key Laboratory of Machine Perception, Peking University, Beijing 100871)

**Abstract** In this paper, a new low complexity video coding method in camera array is proposed, based on the ROI-based Wyner-Ziv coding. The proposed method encodes each view independently at the senders, and performs joint decoding with side information interpolated from adjacent views at the receiver. Through comparing the quantized DCT coefficients, the proposed ROI-based Wyner-Ziv coding method can avoid to encoding the regions with little changes. Experimental results show that, the average encoding time of the proposed scheme is only 1/22 of the H.264 P coding.

**Keyword** video compression, dense camera array, Wyner-Ziv coding, low encoding complexity

## 1 引言

多视点视频是一个能够提供立体感和交互操作功能的新型视频技术,它能够提供给场景对象在不同观察角度下的视频图像信息<sup>[1]</sup>。密集相机阵列系统是典型的多视点视频系统,而如何实现从相机节点到网络中心节点之间的低复杂度、低功耗、高压缩率的稳健视频编码与通信则是密集相机阵列视频系统的关键技术之一。

现有的视频压缩标准(如 MPEG、H.26x)都是利用帧间预测技术来消除时域冗余<sup>[2-4]</sup>;而 MPEG

组织在 3 维音视频(Three-dimensional acoustics video, 3DAV)中提出的多视点视频压缩技术则进一步利用视差估计与补偿来消除视点间冗余<sup>[5-7]</sup>。然而由于运动估计、视差估计的使用,造成此类标准中编码器的计算复杂度是解码器的 5~10 倍<sup>[2]</sup>,因此,现有的视频编码标准不能很好适应于密集相机阵列系统的低复杂度视频编码与通信。

Wyner-Ziv 编码(Wyner-Ziv video coding, WZC)突破了传统的视频编码框架的束缚,它采用帧内编码+帧间联合解码的技术方案来最大可能地降低编码复杂度<sup>[8]</sup>。为解决密集相机阵列视频系统编码与传输的问题,Grid 等人提出了基于 WZC 的低复

**基金项目**:国家自然科学基金项目(60472100、60672073);教育部科学技术研究重点项目(206059);浙江省自然科学基金(RC01057, Y105577);浙江省科技攻关项目(2004C31105)资助

**收稿日期**:2006-06-15; **改回日期**:2006-08-05

**第一作者简介**:金智鹏(1982~),男,2004 年获中国科学技术大学电子信息工程专业学士学位,现为宁波大学信息科学与工程学院硕士研究生。主要从事视频编码压缩方面的研究。E-mail:jianggangyi@nbu.edu.cn

杂度多视点视频压缩方法<sup>[9]</sup>,但该方法的压缩效率还有待提高。

从降低系统编码复杂度的角度出发,本文提出了一种基于 WZC 的密集相机阵列低复杂度视频压缩算法和改进的基于感兴趣区提取的 WZC 算法。实验证明,本文算法对该型系统视频信号的处理是有效的,且编码复杂度极低。

## 2 基于 WZC 的密集相机阵列系统低复杂视频压缩算法

密集相机阵列系统中各相机节点到网络中心节点的视频通信有以下明显不同于普通网络视频压缩与通信的特点:

(1) 基于系统复杂度和设备成本等因素考虑,其相机节点之间一般不能进行视频通信,因此无法在相机节点处进行基于视差补偿的预测编码;

(2) 由于各相机节点的计算能力和存储容量等资源都十分有限,因此难以执行复杂的帧间预测编码过程。

鉴于此,本文提出了基于 WZC 的密集相机阵列低复杂度视频压缩算法(如图 1 所示)与改进的基于感兴趣区提取的 WZC 算法,其框图如图 2 所示。本文算法在编码端不进行运动估计和视差估计,而是

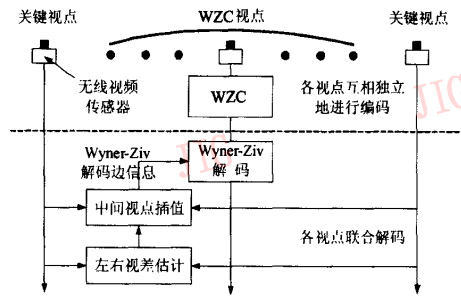


图 1 基于 WZC 的多视点视频压缩算法框图

Fig. 1 The proposed coding structure based on WZC for dense camera array

对各相机节点互相独立地进行 WZC,并将码流独立传送到中心解码端;在中心解码端,则利用各视点间视图的相关性,对各视点联合解码。本文算法具体流程如下:

- (1) 抽取部分相机作为关键视点,如:最左、最右相机,对其采用无率失真优化的帧内预测编码,以保证较低的编码复杂度和较好的编码率失真性能;
- (2) 将其余的相机作为 WZC 视点,采用本文基于感兴趣区提取的低复杂度 WZC 算法进行编码;
- (3) 在中心解码端,WZC 帧的解码边信息  $\hat{W}$  由左右关键视点依据视差场加权插值而得。

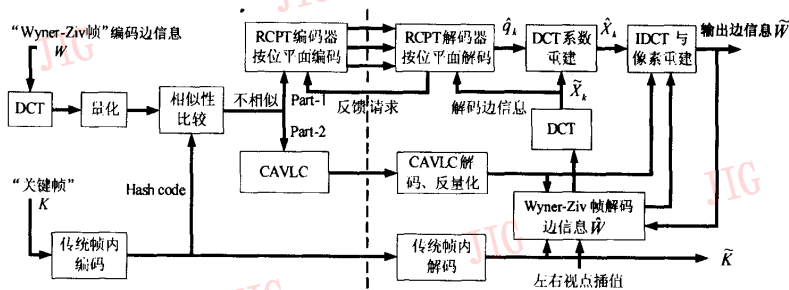


图 2 基于感兴趣区提取的 Wyner-Ziv 编解码算法框图

Fig. 2 The proposed ROI-based Wyner-Ziv coding

### 2.1 感兴趣区提取算法

$DCT$ (discrete cosine transform)量化系数不但可以很好地表示原始图像的信息,而且与存储原始图像像素值相比,存储  $DCT$  量化系数只需要很小的内存空间和计算量。由于视频序列前后帧中背景和平滑区的  $DCT$  量化系数非常的相似甚至相同,因此在编码端可通过对  $DCT$  量化系数的相似性进行判断来对提取的感兴趣区域进行编码,这样可以节省

码率和编码时间。同时,为了获得极低的编码复杂度特性,本文算法只在前后帧的相同位置块之间进行相似性判断,而不采用运动估计与补偿等复杂算法。本文将当前  $4 \times 4$  块的  $DCT$  量化系数与 Hash code 中存储的  $DCT$  量化系数进行比较,定义两者间的加权累积绝对差为

$$D = \sum_{i=0}^n R_i |q_i - h_i| \quad (1)$$

这里,  $n$  指 Hash code 所含的 DCT 系数个数,  $q_i$  为当前块的 DCT 量化系数,  $h_i$  为存储的 Hash code 中的关键帧的 DCT 量化系数,  $R_i$  为权重系数。权重系数  $R_i$  随 DCT 系数 Zig-Zag 扫描先后顺序的不同而不同, 即低频系数权重值大, 高频系数权重值小。如果两个块之间的加权累积绝对差  $D$  大于阈值  $T$ , 则将当前块划归为感兴趣区, 否则判定当前块为背景或平滑块。本文的感兴趣区提取算法简单有效, 能很好地提取真实的感兴趣区。

## 2.2 解码边信息生成算法

解码边信息  $\hat{W}$  的质量很大程度上决定着 WZC 的率失真性能。通常在遮挡暴露问题不严重的情况下, 采用基于像素点的视差补偿算法可以有效地插值生成解码边信息。本文算法首先对左右关键视点进行基于像素点的双向视差估计(采用绝对误差和 SAD 匹配准则,  $5 \times 5$  匹配窗); 然后利用视差场自身的平滑性和唯一性约束做视差矢量的有效性检测, 并采用深度一致性假设来校正和补偿误差较大的视差; 最后依据左右关键视点的解码图像和它们之间的视差场来线性插值 Wyner-Ziv 帧的解码边信息  $\hat{W}$ 。当  $QP = 30$ , 左右关键视点采用帧内编码时, 用本文边信息插值算法据左右关键视点重建图像插值得到的解码边信息如图 3 所示。

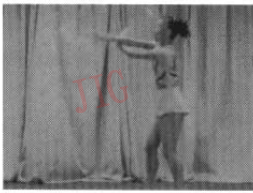


图 3 由关键视点插值得到的解码边信息  
(第 46 视点第 1 帧的解码边信息, 34.84dB)

Fig.3 The side information interpolated from key views

## 2.3 基于感兴趣提取的密集相机阵列低复杂度视频编码算法

本文基于感兴趣区提取的 Wyner-Ziv 视频编码算法首先将  $4 \times 4$  块 DCT 变换系数分为 Part-1 和 Part-2 两部分, 然后将低频 3 个 DCT 系数作为 Part-1 部分(即直流系数和两个低频交流系数), 其余量化系数作为 Part-2 部分。本文算法具体编码过程如下:

(1) 若 Part-1 部分 DCT 量化系数的加权累积绝对差  $D_1$  大于阈值  $T_1$ , 则对这部分系数采用 RCPT (rate-compatible punctured turbo code) 编码, 同时将

校验信息以  $1/2$  删余后存在编码端缓存器, 并据反馈请求逐步发送给解码端进行 Turbo 解码;

(2) 若 Part-2 部分 DCT 量化系数的加权累积绝对差  $D_2$  大于阈值  $T_2$ , 则对其采用基于上下文内容的变长编码(CAVLC 编码), 并将码流传送到解码端。

(3) 对当前  $4 \times 4$  块, 若其 Part-1 和 Part-2 部分 DCT 量化系数都未编码, 即当前块属于背景等非感兴趣区域, 则认为当前块与前面已解码的关键帧相应位置块有很强的相似性, 即可将当前块的重建像素值拷贝到关键帧相应位置块的重建像素。

## 3 实验结果与分析

为验证本文算法效果, 本文采用日本名古屋大学(Nagoya University)提供的‘Rena’多视点测试序列进行了实验, 其图像大小为  $320 \times 240$ , 各视点间隔为 5.0cm, 摄像机以水平方式排列。本文抽取第 45、46 和 47 等 3 个视点各 100 帧图像做了测试, 并从率失真性能和编码时间等方面与传统编码算法进行了比较。编码帧率为 30fps,  $GOP = 4$  (即 Wyner-Ziv 视点编码帧类型为 K-W-W-W-K, 帧间预测编码帧类型为 I-P-P-P-I)。实验所用 PC 机操作系统为 Windows2000, Pentium IV3G, DDR512M 内存, 编译软件为 VC++6.0。

本文算法由 H.264 校验模型 JM86 改造而来, 实验中阈值  $T_1 = 3$ ,  $T_2 = 2$ , Turbo 码的生成矩阵为  $[1 + D^3 + D^4, 1 + D + D^3 + D^4]$ 。Wyner-Ziv 视点的编解码实验结果如图 4 所示, 图中本文算法的峰值信噪比(peak signal noise ratio, PSNR)为所有解码图像的平均 PSNR; 曲线“dct-domain WZC”代表普通 DCT 域 WZC 的实验结果; 曲线“H.263 + IPPP”指 H.263 + P 帧编码的实验结果, 曲线“H.264IPPP、H.264 intra coding”分别指 H.264 P 帧、帧内编码的实验结果。

从图 4 中可以看出, 本文算法的率失真性能优于 H.263 的 P 帧编码(提高了 0.1 ~ 0.8dB)。此外, 与 DCT 变换域的 Wyner-Ziv 编码算法相比, 本文算法的率失真性能提高了 1.0 ~ 1.9dB。需要说明的是, H.264P 帧编码算法的高压缩效率是与其庞大的计算复杂度为代价的<sup>[2]</sup>。H.264 编码的计算复杂度约为 H.263 的 3 倍, 其解码复杂度相当于 H.263 的 2 倍。Wyner-Ziv 视频编码算法本身采用帧内编码 + 帧间联合解码的方案, 其编码计算复杂度相当

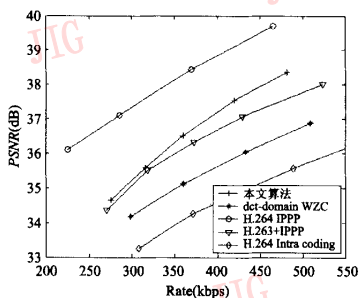


图 4 Wyner-Ziv 视点的编解码率失真曲线对比图  
ig. 4 The rate-distortion performance of Wyner-Ziv view

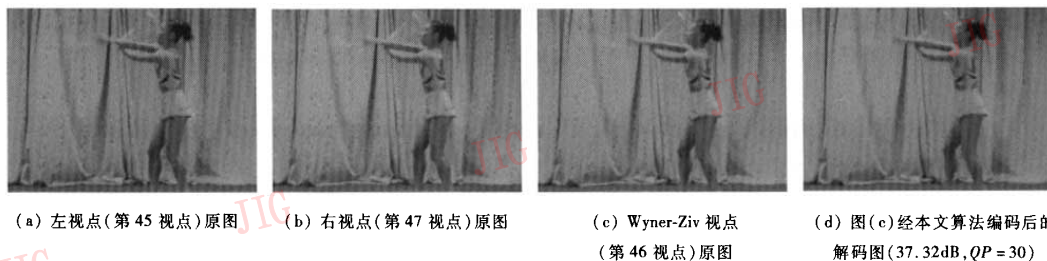


图 5 'Rena' 多视点序列各视点原图及解码图  
Fig. 5 The original and decode images of 'Rena'

## 4 结 论

多视点视频处理技术是 3DAV 中的核心技术, 本文提出了一种基于 WZC 的密集相机阵列低复杂度视频压缩算法。该算法在各相机之间互相独立地进行编码, 且通过对感兴趣区的提取, 有效地降低了编码复杂度, 并提高了压缩效率。实验证明, 本文算法的编码复杂度仅为 H. 264 帧间编码的 1/22。本文算法不但适用于密集相机阵列系统, 也适用于无线多视点视频等要求低功耗编码的视频系统。

### 参考文献 (References)

- 1 Na Banchang P, Fujii T, Tanimoto M. Experimental system of free viewpoint television [A]. In: Proceedings of SPIE Symposium on Electronic Imaging[C], Santa Clara, CA, USA, 2003, **5006**: 554 ~ 563.
- 2 Ostermann J, Bormans J, List P, et al. Video Coding with H. 264/AVC: tools, performance, and complexity[J]. IEEE Transactions on Circuits and Systems Magazine, 2004, **4**(1): 7 ~ 28.
- 3 ITU-T Rec. H. 264 | JM8. 6 [S]. ISO/IEC 11496-10 AVC, Sept.

于 H. 263 的帧内编码, 远远低于 H. 263 和 H. 264 的帧间预测编码。此外, 本文算法通过对感兴趣区域的提取, 又大幅节省了编码时间。在各量化步长下 (量化参数  $QP = 28, 30, 32, 34, 36$ ), 本文算法 Wyner-Ziv 视点的平均编码时间为 15.433s, 而该视点 H. 264 P 帧的平均编码时间为 328.125s, 即在 H. 264 校验模型 JM86 的基础比较平台上, 本文算法的编码时间仅占 H. 264 帧间编码的 4.7%。若采用 X264 视频压缩平台, 则本文算法的执行时间可进一步降低。图 5 为 'Rena' 多视点测试序列各视点第 1 帧 (Wyner-Ziv 帧) 的原图和解码图。

2004.

- 4 Wiegand T, Sullivan G J, Bjontegaard G, et al. Overview of the H. 264/AVC video coding standard [J]. IEEE Transactions on Circuits Systems Video Technology, 2003, **13**(7): 560 ~ 576.
- 5 Kimata H, Kitahara M, Kamikura K, et al. Multi-view video coding using reference picture selection for free viewpoint video communication [A]. In: Proceedings of Picture Coding Symposium [C], San Francisco, CA, USA, 2004: 499 ~ 502.
- 6 Kimata H, Kitahara M, Kamikura K, et al. System design of free viewpoint video communication [A]. In: Proceedings of International Conference on Computer and Information Technology [C], Wuhan, China, 2004: 52 ~ 59.
- 7 López J, Kim J. Block-based illumination compensation and search techniques for multiview video coding [A]. In: Proceedings of Picture Coding Symposium [C], San Francisco, CA, USA, 2004: 509 ~ 514.
- 8 Girod B, Aaron A, Rane S, et al. Distributed video coding [J]. Proceedings of the IEEE, (Special Issue on Video Coding and Delivery), 2005, **93**(1): 71 ~ 83.
- 9 Zhu X, Aaron A, Girod B. Distributed compression for large camera arrays [A]. In: Proceedings of IEEE Workshop on Statistical Signal Processing [C], St Louis, Missouri, USA, 2003: 30 ~ 33.