

文章编号: 0253-2239(2008)01-0062-05

# 无线视频传感阵列低复杂度多视点视频编码方案

蒋刚毅<sup>1,2</sup> 姜浩<sup>1</sup> 郁梅<sup>1</sup> 蒋志迪<sup>1</sup> 刘尉悦<sup>1</sup>

(<sup>1</sup> 宁波大学电路与系统研究所, 浙江 宁波 315211  
<sup>2</sup> 北京大学视觉与听觉国家重点实验室, 北京 100871)

**摘要** 基于网络中心节点的运动矢量外推技术,提出了一种无线视频传感阵列的低复杂度多视点视频编码方法。该方法考虑到密集型视频传感阵列各视点间通信复杂、布线繁重、且位于各相机节点内的编码器由于计算能力、功耗等限制难以完成复杂的编码过程等特点,利用运动矢量外推逼近技术将大量的运动估计运算从视频编码端移到了网络中心节点,使得新编解码框架下编码器的运动估计的计算复杂度只有传统全搜索运动估计运算的 0.3%,降低了系统传感阵列编码端功耗。实验结果表明,该方法的率失真性能比 H.264-I 帧高出 4 dB 以上,接近 H.264-P 帧编码,优于基于 Wyner-Ziv 理论的分布式多视点视频编码方法。

**关键词** 图像处理; 多视点视频编码; 无线视频传感阵列; 运动矢量外推; 低复杂度视频编码

**中图分类号** TN918.81 **文献标识码** A

## A Low Complexity Multi-View Video Coding Scheme for Wireless Video Sensor Array

Jiang Gangyi<sup>1,2</sup> Jiang Hao<sup>1</sup> Yu Mei<sup>1</sup> Jiang Zhidi<sup>1</sup> Liu Weiyue<sup>1</sup>

(<sup>1</sup> Institute of Circuits and Systems, Ningbo University, Ningbo, Zhejiang 315211, China  
<sup>2</sup> National Key Laboratory of Vision and Auditory, Peking University, Beijing 100871, China)

**Abstract** A novel multi-view video coding scheme with low complexity is proposed for wireless video sensor array based on motion-vector extrapolation at central node of network. It is noted that communication among cameras is difficult for implementation in dense video sensor array, and the encoder embedded in the camera is unable to accomplish complex encoding process due to the limitation of computational capability and power consumption. Therefore, motion-vector extrapolation is utilized to move the motion estimation out of the encoder to the central node of network, so that the complexity of motion estimation at the encoder is reduced to 0.3% compared with the traditional exhaustive block matching algorithm, and the power consumption at the encoder of the sensor array is also reduced. Experimental results show that the proposed method outperforms H.264 intra-frame coding more than 4 dB and approaches to H.264 inter-frame coding, and it also performs better than Wyner-Ziv coding schemes.

**Key words** image processing; multi-view video coding; wireless video sensor array; motion-vector extrapolation; low-complexity video coding

## 1 引言

随着计算机与网络技术的发展,视频信息处理与应用得到了广泛重视<sup>[1,2]</sup>。多视点视频技术是目前视频领域的一个重要研究课题<sup>[3~5]</sup>,密集型视频传感阵列是其中的一个典型应用。由于资源受限,如何实现从相机节点到网络中心节点之间的低复杂度、低功耗以及高压缩率的稳健视频编码与通信,是密集型视频

传感阵列视频编码系统的关键技术之一。

文献[6]由 100 部相机组成的一维密集型视频传感阵列系统不仅规模庞大,而且功耗极高,加之实际应用中还存在多维视频传感阵列的需求,使得该技术很难应用于多场景实时编码传输。可见密集型视频传感阵列繁琐的布线直接影响多视点流媒体的应用。而基于网络中心节点的无线通信方式可以解

收稿日期: 2007-02-02; 收到修改稿日期: 2007-09-27

基金项目: 国家自然科学基金(60472100、60672073)、教育部新世纪人才支持计划项目(NCET-06-0537)、教育部科学技术研究重点项目(2060059)和浙江省自然科学基金(Y105577)资助课题。

作者简介: 蒋刚毅(1964—),男,浙江绍兴人,教授,主要从事多媒体信息传输、图像处理与视频信号编码、基于视觉的智能控制等方面的研究。E-mail: jianggangyi@nbu.edu.cn

加布线所带来的困扰。此外,由于密集型视频传感阵列间通信复杂,难以满足系统的实时性要求,因此视点间独立编码十分必要。

多视点视频编码与传统的单通道视频编码相比,要处理成倍的数据量,而且考虑到视频传感器节点的计算能力、程序空间、内存空间及功耗等资源受限问题<sup>[7]</sup>,必须采用编码复杂度更低且具有高压缩率的视频编码压缩技术,才能对多视点视频数据进行实时有效的存储与传输。H. 264 是目前最新的视频编码国际标准<sup>[8]</sup>,有很高的编码压缩性能,然而其编码复杂度十分巨大。文献<sup>[9]</sup>针对 H. 264 的 JM8.6 校验模型进行了单通道视频编码计算复杂度分析,测试序列为 foreman, CIF@30fps, 4:2:0。结果表明,编码器中运动估计(ME)模块占用了多于 80% 的计算量。

为了实现低复杂度视频编码,人们提出了基于 Wyner-Ziv 理论的分布式编码算法<sup>[10]</sup>以及基于形状自适应的分层树集合划分(SA-SPIHT)分区域的处理方案<sup>[11]</sup>等方法;前者虽然编码器的计算复杂度很低,但编码效率不高,且在编码端无法得到重建图像,错误累积可能影响整个编码器的性能。

根据 H. 264 编码计算复杂度分析结果,本文提出通过网络中心节点获取外推运动矢量,将运动估计的大量运算移出视频编码端,以降低多视点视频编码器的计算复杂度的无线视频传感阵列多视点编解码框架。同时,为提高外推运动矢量精度,提出了基于块特征分析的两种外推矢量求精的方案。

## 2 无线视频传感阵列的多视点编解码框架

本文中无线视频传感阵列的多视点视频编码方案是通过网络中心节点对运动矢量进行外推逼近,将运算量巨大的运动估计移出多视点视频编码端。整个新的多视点视频编解码框架如图 1 所示:对于多视点视频编码端,各视点间不进行通信,而是利用网络中心节点反馈的外推运动矢量独立进行视频编码,为进行运动矢量外推,每个图像组(GOP)初始时可考虑采用两个 H. 264-I 帧;对于网络中心节点,利用已解码的视频图像,采用非线性外推法获取外推运动矢量,然后将相关信息反馈给多视点视频编码端;对于多视点视频解码,利用接收到的编码信息以及逼近后的运动矢量,进行视频序列持续解码。上述通过网络中心节点产生运动矢量的方法并不增

加解码的计算量。

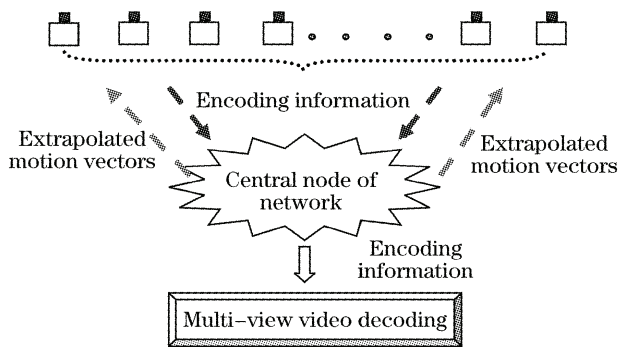


图 1 无线视频传感阵列多视点视频编解码框架

Fig. 1 Framework of proposed multi-view video coding for wireless video sensor array

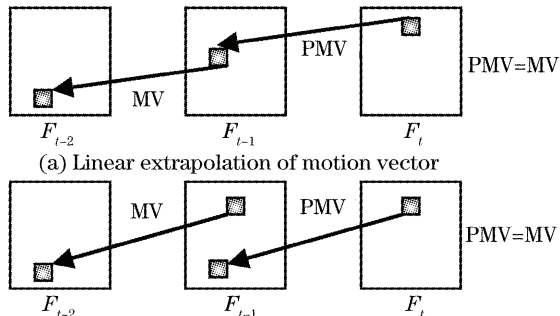
## 3 运动矢量外推及外推矢量求精

### 3.1 运动矢量的外推估计

运动矢量外推方法即通过前面已解码帧预测当前帧编码所需要的运动矢量。与传统运动估计不同,通过外推方法得到运动矢量并不需要对当前帧进行操作,这就使得外推可以脱离编码器独立进行,从而可以将高复杂度的运动估计移出编码端。

图 2(a)所示的线性矢量外推法利用前一帧  $F_{t-1}$  各块所得到的运动矢量 MV 来推算当前帧  $F_t$  的运动矢量 PMV。该方法中运动矢量的获得是在假定前一帧各块的运动走向在当前帧仍然保持的前提下。利用线性外推得到的运动矢量来预测当前帧图像时会出现重叠区域和空白区域。

非线性矢量外推法利用前一帧各块所得到的运动矢量来推算当前帧的运动矢量,如图 2(b)所示。该运动矢量获得的前提是假定当前帧中各块与前一帧位置相同各块的运动矢量相同。显然,这一假设对于背景块来说是成立的;而对于运动块,尽管当前帧中运动块与前一帧中位置相同的块并不相同,但在运动幅度不大或视频信号采样频率较高的情况



(b) Linear and nonlinear extrapolation of motion vector

图 2 线性与非线性运动矢量外推示意图

Fig. 2 Linear and nonlinear extrapolation of motion vector

下,当前帧中多数运动块与前一帧中位置相同的块仍然属于同一运动物体,即它们拥有相近的运动矢量。由此可见,非线性矢量外推所得到的运动矢量偏差主要集中在图像的遮挡和暴露区域。而由于物体的非线性运动所带来的外推偏差无论是线性外推还是非线性外推都是难以避免的。

由于非线性外推在获得预测图像时不会出现空白区域,所以其相对于线性外推来说更加适用于视频编码。

### 3.2 外推运动矢量求精

由上述非线性外推得到的运动矢量与真实运动矢量之间存在一定的偏差,需要进一步求精。

根据前述分析可知,对于当前帧中的背景块,其外推运动矢量与真实运动矢量的偏差较小。而对于具有不同特征的运动块,其情况则不尽相同:若运动块较为平坦,即使运动矢量略有偏差,该平坦运动块的预测残差也相对较小;若运动块为纹理块,或者该运动块位于遮挡或暴露区域(通常在物体边缘附近),即使外推运动矢量偏差较大,其预测残差也相对较大。因此,从降低编码复杂度的角度出发无需对背景块和平坦运动块的外推运动矢量进行修正求精,而从提高编码压缩率的角度则应对位于纹理区和遮挡暴露区的运动块的外推运动矢量进行修正求精。

由于位于纹理区和遮挡暴露区的运动块的预测残差相对较大,因此可以利用绝对差(SAD)值(用 $A_{SAD}$ 表示)的大小来判断块的特征:将相应的 $16 \times 16$ 宏块划分成四个 $8 \times 8$ 块,分别记为 $S_i (i=1, \dots, 4)$ ,若 $A_{SAD12} < T_{SAD}$ 且 $A_{SAD23} < T_{SAD}$ 且 $A_{SAD34} < T_{SAD}$ ,则该块为背景块或平坦块,反之为纹理块或遮挡暴露块,其中 $A_{SADij} = A_{SAD}(S_i) - A_{SAD}(S_j)$ 。这里,上述块特征的判定可以在网络中心节点进行,此时,由于网络中心节点没有当前帧,因此块特征可如图2(b)所示由 $F_{t-1}$ 帧中与当前帧 $F_t$ 中当前块同位置的块相对于 $F_{t-2}$ 帧中匹配块的 $A_{SADij}$ 来判定。块特征的判定也可以在编码端完成,即利用外推运动矢量直接计算当前帧 $F_t$ 中当前块相对于前一帧 $F_{t-1}$ 的对应块的 $A_{SADij}$ 来判定。块特征判定在网络中心节点进行的优点是进一步降低了编码器的计算复杂度,但块特征需要反馈给编码端,增加了通信的开销,而且块特征判定的误差相对后一种方法也更大,因而其率失真性能也有一定的损失。

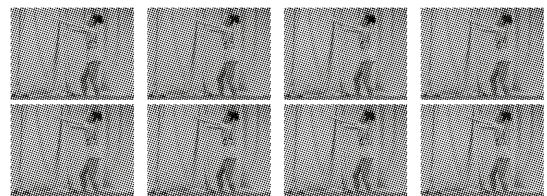
由于运动矢量具有较强的空间相关性,因此可以先利用当前块的八个相邻块的外推运动矢量对当前块外推矢量进行一维的 $x$ 、 $y$ 方向的中值滤波修

正,以排除孤立块运动矢量所引起的严重偏差。然后根据当前块特征决定是否对其外推运动矢量进行求精:对于背景块及平坦块直接采用中值滤波后的运动矢量;而对于纹理运动块或遮挡暴露块,则采用十字形二次匹配法对其运动矢量进行求精。十字形二次匹配法是指对由外推运动矢量所指示的点及与其同行或同列的四个相邻点进行块匹配搜索,从中选择具有最小残差的点计算得到运动矢量的方法。对编码后残差仍然较大的块,通过与相应I块的率失真优化(RDO)计算比较,适时地用I块进行替换。

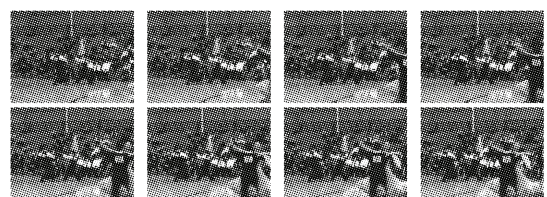
虽然十字形二次匹配法需要在编码端才能完成,但由于该算法只有部分宏块需要对步长为 $\pm 1$ 的十字形的五个点进行搜索,因此与传统运动估计相比,计算量与读取存储器的次数都有了极大的降低;此外,在匹配算法中还加入了提前终止算法,在保持编码性能不变的前提下亦可进一步降低计算复杂度。

## 4 实验结果与分析

为评价本文方案的率失真性能,使用联合视频专家组(JVT)的JM.86作为实验平台,采用日本名古屋大学提供的Rena和三菱电气研究实验室(MERL)提供的Ballroom多视点测试序列进行分析。这2个序列的图像分辨率均为 $640 \text{ pixel} \times 480 \text{ pixel}$ ,平行相机各视点间隔分别为 $5.0 \text{ cm}$ 和 $20.0 \text{ cm}$ ,图3为其部分视点图像。图4给出了Rena第45视点和Ballroom第6视点各90帧图像的测试结果,并与H.264-I帧和H.264-P帧进行了率失真性能比较。图中横坐标为某视点视频信号编码压缩后的码率,纵坐标为其解码重建信号的峰值信噪比(PSNR)。实验中,图像组为30,编码帧率为 $30 \text{ frame/s}$ ,



(a) The 41<sup>th</sup>~48<sup>th</sup> viewpoints of Rena sequence



(b) The 1<sup>th</sup>~8<sup>th</sup> viewpoints of Ballroom sequence

图3 多视点测试序列

Fig. 3 Multi-view video test sequences



Proposed (1)和 Proposed (2)分别代表块特征判定在网络中心节点进行(阈值  $T_{SAD}$  为 30)和在编码端进行(阈值  $T_{SAD}$  为 130)2 种方案。上述 2 方案阈值的差异主要是由于 Proposed (1)进行比较的 2 帧均为解码帧,而 Proposed (2)中进行比较的 2 帧中一帧为当前帧的原始帧,另一帧为解码帧。为公平比较各方案率失真性能,各方案运动估计均只采用  $16 \times 16$  宏块模式。实际应用中本文方案也可根据图像纹理进行多种块模式下的运动估计。从图 4 中的

率失真曲线可以看出,本文方案的编码性能明显优于 H.264-I 帧,且高出 4~5 dB 以上,在低码率下与 H.264-P 帧相接近。图 5 为采用 Proposed (1)方案时的解码视频图像,其中 QP 为量化参量(Quantization parameter)。对于分辨率为  $320 \text{ pixel} \times 240 \text{ pixel}$  的 Rena 测试序列,本文编码算法比文献 [12]中所提到的多视点 Wyner-Ziv 编码算法还要高出 1.5 dB 以上,如图 6 所示(图像组为 4<sup>[12]</sup>),图中坐标含义与图 4 相同)。

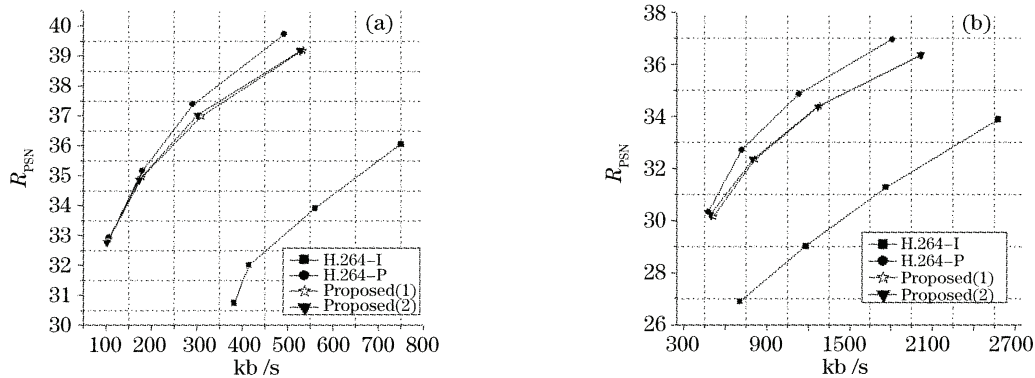


图 4 率失真性能比较

Fig. 4 Comparison of rate-distortion performance

(a) The 27<sup>th</sup> frame in 45<sup>th</sup> viewpoint of Rena(b) The 27<sup>th</sup> frame in 6<sup>th</sup> viewpoint of Ballroom

图 5 解码图像

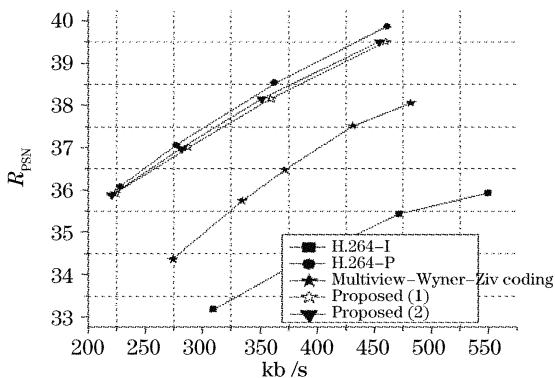
Fig. 5 Decoded frames.  $R_{PSN}$  is peak signal-to-noise ratio图 6 与多视点 Wyner-Ziv 编码算法<sup>[12]</sup>的率失真性能比较

Fig. 6 Rate-distortion performance compared with one by Wyner-Ziv coding scheme in Ref[12]

我们以运动估计搜索点数来衡量运动估计复杂度。以整像素全搜索运动估计搜索点数为基准,定义本文方案运动估计复杂度相对于全搜索方案的百分比  $\eta$  为

$$\eta = \frac{5\mu N_b}{(2w+1)^2 N_b} = \frac{5\mu}{(2w+1)^2}, \quad (1)$$

其中  $w$  为全搜索的搜索窗大小,  $N_b$  为一帧中的总块数,  $\mu$  为本文方案一帧中需要进行十字形二次匹配的块数相对于总块数的平均百分比。表 1 给出了全搜索的搜索窗大小  $w$  为 16 时,本文两种方案与整像素全搜索的运动估计复杂度比较,从中可见,新编解码框架下编码器的运动估计计算复杂度不到全搜索的 0.3%。当然在网络中心节点完成运动矢量外

推并将其反馈给编码端会增加一定的通信开销,但由于运动矢量是以宏块或块为单位且数值较小,因

此运动矢量的数据量很小,即通信开销很小。

表 1 运动矢量获取的复杂度分析比较

Table 1 Complexity comparison of motion vector computation

chemes	equences	Ballroom			Rena		
		QP	$\mu / \%$	$\eta / \%$	QP	$\mu / \%$	$\eta / \%$
Proposed(1)	28	36.1	0.17	32	38.8	0.18	
	32	36.3	0.17	36	37.3	0.17	
	36	37.8	0.17	40	36.9	0.17	
	40	39.8	0.18	44	36.6	0.17	
Proposed(2)	28	33.3	0.15	32	37.9	0.17	
	32	43.3	0.20	36	42.6	0.19	
	36	54.2	0.25	40	52.5	0.24	
	40	66.7	0.31	44	59.2	0.27	

## 5 结 论

基于网络中心节点的运动矢量外推技术,提出了一种无线视频传感阵列的低复杂度多视点视频编码方法。实验结果表明,即使是从多视点视频编码端移出了运动估计之后,还是能够通过网络中心节点获得较高的编码性能。与现有其它的低复杂度多视点视频编码方案相比,该方案具有整体复杂度低、功耗低的特点。

## 参 考 文 献

- 1 Yong Fang, Chengke Wu, Hong Zhu. Error detection using mode information of macroblocks for video transmission [J]. *Chin. Opt. Lett.*, 2005, **3**(10): 582~584
- 2 Liu Guixi, Shao Mingli, Liu Xianhong *et al.*. Video moving object auto-extraction in real scene [J]. *Acta Optica Sinica*, 2006, **26**(8): 1150~1155  
刘贵喜,邵明礼,刘先红等. 真实场景下视频运动目标自动提取方法[J]. *光学学报*, 2006, **26**(8): 1150~1155
- 3 Sun Lifeng, Li Fang, Zhong Yuzhuo *et al.*. Multiview video based virtual teleconferencing synthesizing [J]. *Acta Electronica Sinica*, 2004, **33**(2): 193~196  
孙立峰,李放,钟玉琢等. 基于多视点视频的虚拟会议显示与合成[J]. *电子学报*, 2004, **33**(2): 193~196
- 4 Shao Feng, Jiang Gangyi, Yu Mei *et al.*. A multi-view video automatic color correction system [J]. *Acta Optica Sinica*, 2007, **27**(5): 830~834  
邵枫,蒋刚毅,郁梅等. 一种多视点视频自动颜色校正系统

[J]. *光学学报*, 2007, **27**(5): 830~834

- 5 Mao Chongde, Wang Yuanqing. Multi-view auto-stereoscopic projection system [J]. *Opto-Electronic Engineering*, 2006, **33**(4): 59~62  
毛崇德,王元庆. 多视点自由立体投影系统[J]. *光电工程*, 2006, **33**(4): 59~62
- 6 ISO/IEC JTC1/SC29/WG11, MPEG2006/W8019, Description of Core Experiments in MVC[S]. Montreux, Switzerland, April 2006
- 7 Ma Huadong, Tao Dan. Multimedia sensor network and its research progresses [J]. *J. Software*, 2006, **17**(9): 2013~2028  
马华东,陶丹. 多媒体传感器网络及其研究进展[J]. *软件学报*, 2006, **17**(9): 2013~2028
- 8 Thomas Wiegand, Gary J. Sullivan. The H. 264/AVC video coding standard [J]. *IEEE Signal Processing Magazine*, 2007, **24**(2): 148~153
- 9 Yun Cheng, Kui Dai, Zhiying Wang *et al.*. A fast motion estimation algorithm based on diamond and simplified square search patterns [C]. *LNCS*, 2005, **3773**: 440~449
- 10 Limin Liu, Yuxin Liu, Edward J. Delp. Network-driven Wyner-Ziv video coding using forward prediction [C]. *Proc. SPIE*, 2005, **5685**: 641~651
- 11 S. Yaman, G. AlRegib. A low-complexity video encoder with decoder motion estimator [C]. *International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, 2004, **III**: 157~160
- 12 Jin Zhipeng, Yu Mei, Jiang Gang yi *et al.*. Low complexity video compression in dense camera array [J]. *J. Image and Graphics*, 2006, **11**(11): 1592~1595  
金智鹏,郁梅,蒋刚毅等. 一种密集相机阵列的低复杂度视频压缩方法[J]. *中国图像图形学报*, 2006, **11**(11): 1592~1595