

## 一种多目分布式视频编码的边信息生成算法\*

林昕<sup>1,2</sup>, 刘海涛<sup>1</sup>, 魏建明<sup>1</sup>

(1. 中科院上海微系统与信息技术研究所, 上海 200250;

2. 上海交通大学技术学院, 上海 201101)

**摘要:** 针对多目视频传感器网络, 提出了一种新的空间边信息生成算法, 该算法将体现图像空间信息的三焦点张量引入边信息的计算。三焦点张量的匹配点计算采用了改进的基于 SIFT (Scale Invariable Feature Transform) 特征描述符的匹配算法, 在保持对匹配有旋转、尺度缩放、亮度变化不变性的优点同时, 大大提高了计算速度。仿真结果表明, 无论从率失真表现还是主观质量表现角度进行比较, 该算法明显优于常用的边信息生成算法、基于运动的外推法 MC-E 和基于运动的内插法 MC-I。

**关键词:** 多目; 分布式视频编码; 边信息; 视频传感器网络

中图分类号: TP393.03

文献标识码: A

## Side information generation for multi-view distributed video code

LIN Xin<sup>1,2</sup>, LIU Hai Tao<sup>1</sup>, WEI Jian Ming<sup>1</sup>

(1. Shanghai Institute of Micro System and Information Technology, Chinese Academy of Sciences, Shanghai 200250, China;

2. School of Technical, Shanghai Jiaotong University, Shanghai 201101, China)

**Abstract:** A novel and efficient side-information frame generator of video sensor network is proposed in this paper, which is based on trifocal tensor. An improved SIFT (scale invariable feature transform) descriptor matching method is used to compute trifocal tensor, it performs much faster with keeping its original advantage: invariable to rotation, multi-scaling and brightness. Simulation results show that the proposed techniques provide potential rate-distortion performance advantages to the MC-E and MC-I method visual performance is also improved.

**Key words:** multi-view; distributed video coding; camera sensor networks; side information

最近, 身边涌现出很多具有崭新特点的多媒体应用, 例如无线视频传感器监控网络、无线 PC 摄像机、移动摄像手机、一次性摄像机和便携式摄像机等, 它们在存储容量、计算能力和功率资源等方面都受到很大的限制, 有些由电池供给能量, 有些则是即用即抛, 因此这些新兴的多媒体技术需要简单的编码器以节省资源。分布式视频编码将耗时的运动估计、运动补偿从编码端移到解码端, 从而得到简单的编码器, 使得简单的视频编码在技术上成为可能。

多个无线摄像节点从不同角度获取同一场景的交叠的图像, 这样的节点组成的网络被称为多目视频传感器网络。多目可以给很多基于视觉的技术带来方便。这种网络数据量巨大, 不同摄像机获取的图像具有很强的相关性。在多目视频传感器网络中应用分布式视频编码技术, 对多个信源

应用独立编码器进行编码。得到的多个独立码流送入同一个联合解码器, 这个解码器充分利用码流之间的相关性对所有输入码流进行解码。这样避免了传感器之间的信息交互, 简化了传感器网络分布结构的设计, 节省了信息交互带来的带宽需求。

应用多目分布式视频编码, 理论上<sup>[1]</sup>可以达到和联合编码联合解码相同的编码率。如图 1 所示, 即使摄像节点间互不通信, 视图间的相关性也可以在解码端使用, 获得简单而又高效率的编码器。多目分布式视频编码的主要优势在于它不仅能满足视频传感器网络节点低功耗、低复杂度的限制, 还可以满足节点间互不通信的要求。

## 1 多目边信息的研究进展

目前, 分布式视频编码研究的一个热点问题就是如何

\* 基金项目: 中国高技术研究发展计划(863 计划)资助(2006AA01Z216)

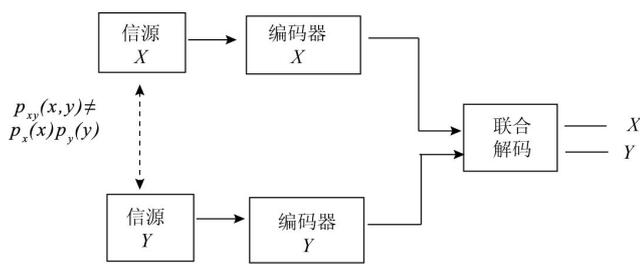


图1 多目分布式视频编码系统结构

准确地估计边信息。解码端边信息帧的存在将视频编码纳入 Wyner-Ziv 编码的理论框架,这使得利用分布式编码的思想重新思考视频编码成为可能。在分布式视频编码框架中,边信息帧的生成质量是 DVC 系统率失真性能的基础。因此,SIFG 研究具有很强的理论和实践研究价值。边信息用于接收端的信道解码和重建,是对当前 WZ 帧的预测,相当于传统编码中对 P 帧或 B 帧的预测,只不过这里的预测是在解码端进行的。预测精度越高,说明边信息与 WZ 帧越接近,相当于虚拟信道的误码越少,经过信道解码后误码率就越低。另外,信道解码后得到的只是 WZ 帧的量化版本,还需通过重建过程才能得到 WZ 帧图像。重建过程通常也要用到边信息,且边信息的预测精度越高,重建后的效果越好<sup>[2-3]</sup>。图2为 Wyner-Ziv 编码器示意图,其中 X 为信源, Y 为边信息。



图2 Wyner-Ziv 编解码器

多目分布式视频编码的边信息,可以利用同一摄像机所获得的图像序列的时间相关帧和不同摄像机所获得的图像序列的空间相关帧得到。图3是分布式编码的帧码结构,其中 C 帧表示用传统视频编解码得到的帧, I 帧表示帧内编解码得到的帧, WZ 则为待解码的分布式编码帧。ZHU Xiao Qing<sup>[4]</sup>提出根据空间的几何关系重建边信息。此外,OUARET M<sup>[5]</sup>则提出由单应性生成空间边信息再与时间边信息融合的边信息重建方法。而 ARTIGA X<sup>[6]</sup>的方法则是分别将前后 2 帧时间边信息和左右 2 帧空间边信息内插然后再将内插得到的图像融合。MAUGEY T<sup>[7]</sup>等人则对对称编码系统提出了对于长 GOP 的对角线边信息估计方法。

## 2 基于三焦点张量的边信息生成算法

### 2.1 算法概述

本文主要根据图3给出的 Wyner-Ziv 编码的帧码结构。首先采用 WZ 帧的前一帧和左右相邻摄像机的当前帧作为 3 个参考帧,通过 Ransac 算法获取三焦点张量。由于多目视频系统主要的分析对象是运动物体,为便于分析,文中应用运动检测与空域灰度分割相结合的方法进行当前摄像机的

运动对象提取。再应用其空间相关性,对临近摄像机的运动边界进行计算和运动对象提取。对运动对象应用三焦点张量,将左右 2 幅图像匹配到 WZ 帧位置,再将匹配好的图像用内插法生成前景的空间边信息。然后用基于运动补偿内插的时间边信息作为背景合成空间边信息。

### 2.2 基于三焦点张量的空间边信息

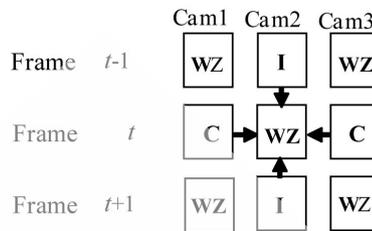


图3 分布式编码的帧码结构

#### 2.2.1 基于 SIFT 特征描述符的立体匹配

在以往的二焦点张量的计算中,一般采用 Harris 角点作为图像匹配中的特征描述,然后利用 RANSAC 方法对其进行匹配计算,其优点在于 Harris 角点的计算比较方便、快速。本文采用 SIFT 特征向量代替 Harris 角点,优化的 SIFT 生成方法(12 维),使得其在计算速度上毫不逊色。同时, SIFT 特征描述符的引入,使其对物体运动、遮挡、视角变化等有良好的匹配效果,并且提取的特征点数量较 Harris 有大幅度的提高。另外,其引入允许摄像机之间可以有较大的角度间距和图像的尺度缩放。基本 SIFT 特征描述符的匹配算法对局部图像的匹配结果如图4所示。

图5左为通过 Matlab 仿真后计算出的 SIFT 特征向量,

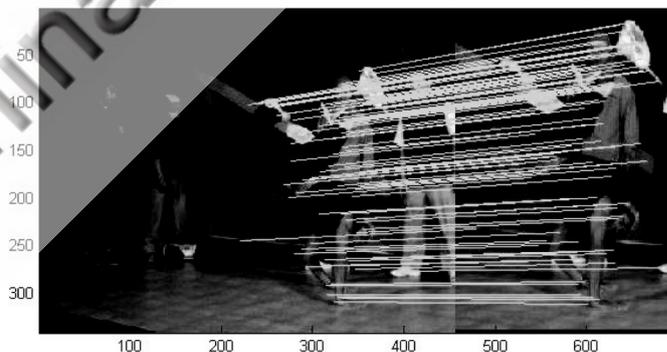


图4 基于 SIFT 特征描述符的匹配算法对局部图像的匹配结果

右图为同一幅图像检测出的 Harris 角点。通过对同一幅图像的检测可以发现,在相同精度下,基于 SIFT 特征描述符的匹配算法可以找到更多的特征点,可以最大限度地为 2 幅不同角度图像匹配提供可能的对应点,这样在 Ransac 鲁棒估计计算三焦点张量的算法中提供更多的采样点,从而提高计算精度。

#### 2.2.2 三焦点张量的估计

三焦点张量是描述 3 幅像间几何关系的数学量,它对 3 幅图像上的同名点间、同名线间以及同名点与同名线间都存在约束关系,因此能够应用三焦点张量把 2 幅视图的对应点转移到第 3 幅视图的对应点。因此,借助于三焦点张量



图5 单幅图像 SIFT 特征向量和 Harris 角点的比较

可以计算非标定的无线视频网络的边信息。

具体而言,若空间有一点  $M$ , 3 个不同摄像机成像分别为  $m_1, m_2, m_3$ , 其次坐标分别为  $(x, y, 1)^T, (x', y', 1)^T, (x'', y'', 1)^T$ , 则  $M$  在 3 幅图像中的像点之间存在如下的三线性约束关系:

$$\begin{cases} x''(\alpha_1 x + \alpha_2 y + \alpha_3) + x'x'(\alpha_4 x + \alpha_5 y + \alpha_6) \\ + x'(\alpha_7 x + \alpha_8 y + \alpha_9) + \alpha_{10} x + \alpha_{11} x + \alpha_{12} x = 0 \\ y''(\beta_1 x + \beta_2 y + \beta_3) + y'y'(\beta_4 x + \beta_5 y + \beta_6) \\ + x'(\beta_7 x + \beta_8 y + \beta_9) + \beta_{10} x + \beta_{11} x + \beta_{11} = 0 \end{cases} \quad (1)$$

式中  $\alpha_i, \beta_i$  均为待定系数。

上述三线性约束关系可由一张量  $T$  来表示, 称为三焦点张量, 它是一个  $3 \times 3 \times 3$  的张量。利用此三焦点张量, 当已知 3 幅图像中其中 2 幅之间的对应点(匹配点)的位置时, 则在第 3 幅图像中的对应点的位置可由(2)式确定<sup>[3]</sup>:

$$x'_i = x'_l \sum_{k=1}^3 x_k T_{kij} - x'_j \sum_{k=1}^3 x_k T_{kil} \quad (2)$$

其中  $i, j, l=1, 2, 3$ , 上式被称为像素转移公式。

最后采用 Ransac 算法作为搜索引擎, Ransac 算法可通过自适应的  $N$  次采样, 将数据划分为内点(最大一致性集)和野值(数据集的余下部分), 并由所有的内点重新估计模型, 进而可计算得到三焦点张量  $T$ 。例如对应图 6 的 3 帧图像可得到其对应的三焦点张量  $T$  为:

-1.9626797395679853e+002 -2.131447576917091e+000 -6.3940054861666884e-001  
-8.7273603801625050e+000 -9.1385710498959771e-002 1.1162255078547268e-003  
-5.2297824667762050e-001 -5.4840173104039281e-003 -1.2327727500501868e-006  
-2.2896423479003225e-002 -1.2797702398520130e+002 -1.4054459754935461e-002  
-6.8061301619729377e+001 -9.5825455256971424e+000 -6.4511532228745661e-001  
-2.0174646722899703e-003 -5.1989272500940586e-001 -7.5306213711935302e-005  
3.3356889728621536e+002 2.1892788778986687e+002 -1.2195859688534651e+002  
3.1742464406226770e+002 1.8453418183734129e+001 -5.5940279505344392e+000  
-6.6971556070410216e+001 9.8625108480717927e-002 -1.1420974218547471e+000

### 2.2.3 运动对象提取

针对多目视频系统的特性, 本文首先采用运动检测与空域灰度分割相结合的方法进行当前摄像机的运动对象的提取。运动检测是对多帧的运动信息, 运用高阶统计检测方法, 提取完整连续的运动区域。空域灰度分割采用改进后的分水岭算法, 该方法既能够减少传统分水岭算法的过分割现象, 同时也能降低计算复杂度, 而且空域分割仅对运动区域及其附近部分进行, 这样又能有效降低算法的复杂度。

46



图6 用于计算三焦点张量的三幅参考图片

利用相邻摄像机的空间相关性, 借助三焦点张量的立体匹配约束关系, 由当前摄像机的边界经过视差匹配及形态学处理, 计算出左右摄像机的边界, 对左右摄像机的相关帧进行背景分离。这种方法比较简单, 不需要重新进行复杂的分割, 计算量较小。

### 2.3 边信息生成

前景的运动物体多为人物、动物等复杂的立体, 如果进行复杂的立体匹配或是进行可视外壳的构建, 会引入大量的计算。所以本文算法在生成中间摄像机边信息时, 将左右摄像机的前景图像作为一个平面匹配到其在中间图像的位置, 再通过匹配后的图像内插来消除误差。当然, 如果视频系统的基线过长, 误差会相应增加。边信息生成过程如图 7 所示。

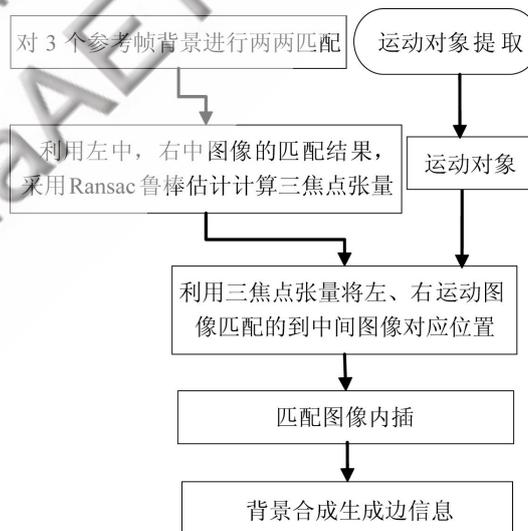


图7 边信息生成过程

前景边信息基本能反映出 WZ 帧中运动对象的动作, 而背景则采用同一摄像机的前后 2 帧做基于块的运动估计, 也就是说这里可以将基于运动补偿内插的时间边信息作为背景与前景空间边信息合成, 得到完整的空间边信息。

### 3 仿真结果

实验采用了多目视频序列 Breakdancer 和 Ballet (1024x768, 15 fps, 100 帧) 进行仿真。首先针对 Ballet 序列的摄像机 1 和摄像机 3 采用传统的 H.264 编解码, 摄像机 2 采用分布式编解码, 以获取摄像机 2 的 WZ 帧边信息。实验采用 H.264 作为参考传统编码进行编解码, 参考软件为 JM15.1。

《微型机与应用》2009 年第 22 期

# 图形、图像与多媒体

Image Processing and Multimedia Technology

仿真中，主要考察不同质量的参考帧生成的边信息帧亮度分量的 PSNR 比较,即边信息算法的率失真特性。为了测试提出的边信息生成算法,选取 MC-E(运动补偿外推)和 MC-I(运动补偿内插)2 个参照的边信息生成算法。由于 MC-E 算法采用前面 2 个重建帧作为参考,为了同 MC-E 算法比较,利用相同的参考帧,在实验中对 3 种 SIFG 算法都采用 I-I-WZ 的 GOP 结构。图 8 为 Ballet 序列的 3 种边信息生成算法的 RD 曲线。

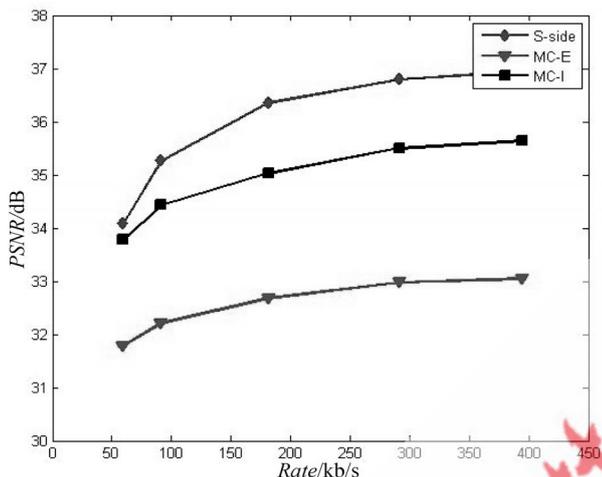


图 8 Ballet 序列的 3 种边信息生成算法的 RD 曲线

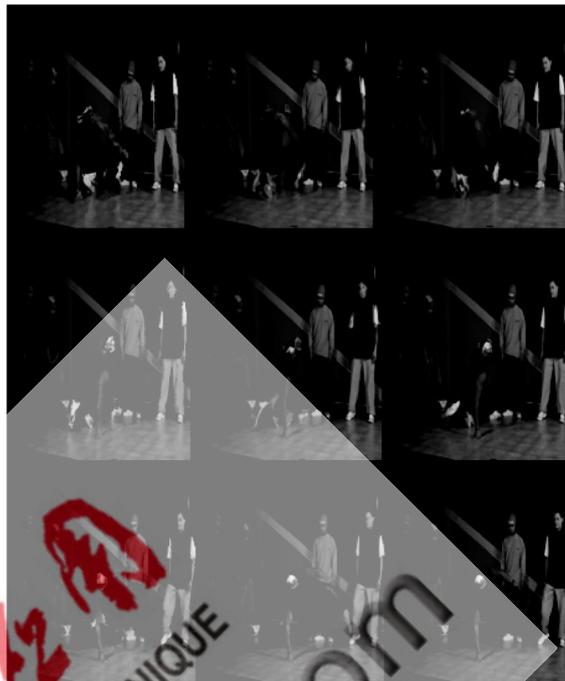
接着本文又针对 Breakdancer 序列的摄像机 2、3、4 做了仿真分析。摄像机 2 和摄像机 4 采用传统的 H.264 编解码,摄像机 3 采用分布式编解码,以获取摄像机 3 的 WZ 帧边信息。3 种 SIFG 算法生成边信息帧的主观视觉质量,如图 9 所示。图 10 和图 11 分别为摄像机 3 的率失真曲线及 QP 取 26 时的 3 种边信息的逐帧 PSNR 性能比较。

分布式视频编码的特点是编码简单、解码复杂,压缩效率高,抗误码性能好,易于分级和多描述处理。因而这种编码方式非常适合于耗电低、存储容量小、运算速度有限的无线视频终端,具有广泛的应用前景,因此逐渐成为视频编码研究领域较为关注的前沿课题。

本文主要针对多目视频传感器网络,提出了一种结合了多目系统时间和空间相关性的空间边信息生成算法。该算法运用三视图对应点之间的几何关系,得到对应点之间的三线性约束关系,即三焦点张量,再由此张量和参考图像之间的对应关系完成边信息的生成。算法仅需要 4 幅参考图像,且实现起来简单、快速。仿真结果证明,该算法生成的空间边信息 RD 性能比基于运动补偿的内插法可提高 2 dB 左右,而比外推法的 RD 性能提高的可达 3 dB~4 dB。从主观效果看,该边信息生成算法也明显优于基于运动补偿的外推和内插法。

### 参考文献

- [1] GIROD B,AARON A,RANE S,et al.Distributed video coding. Proc.IEEE 93,2005(71):71-83.
- [2] AARON A,RANE S,SETTON E,et al.Transform-domain Wyner-



自上而下为第 3、6、9 帧 QP=26(左起:MC-E,MC-I,空间边信息)  
图 9 Breakdancer 序列 3 组边信息

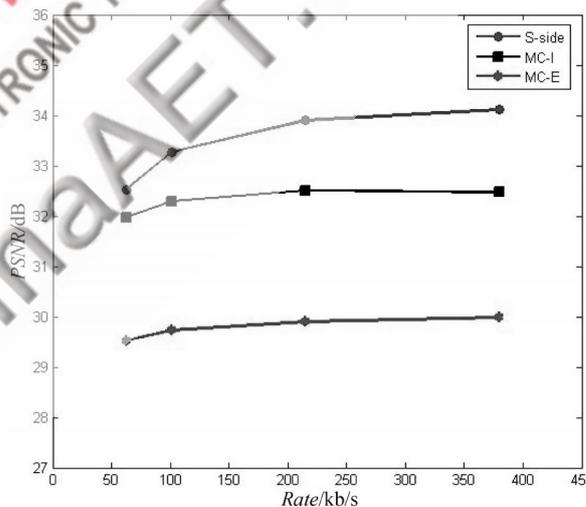


图 10 Breakdancer 序列 3 种边信息的 RD 性能曲线

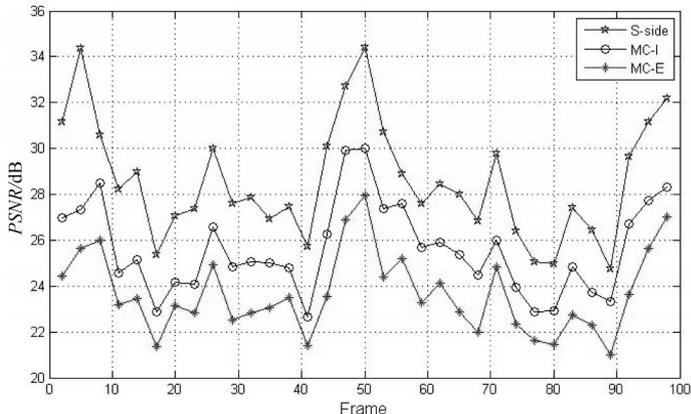


图 11 Breakdancer 序列 3 种边信息的逐帧 PSNR 性能比较

(下转第 50 页)

- Ziv codec for video. SPIE Visual Communications and Image Processing Conference, 2004(5308):520-528.
- [3] XU Q, XIONG Z. Layered Wyner-Ziv video coding. IEEE Trans. Image Process, 2006, 1(12):3791-3803.
- [4] ZHU X, AARON A, GIROD B. Distributed compression for large camera arrays. Proc. IEEE Workshop on Statistical Signal Processing, SSP-2003, St Louis, Missouri, Sept. 2003. Invited Paper.
- [5] OUARET M, DUFAUX F, EBRAHIMI T. Fusion-based multi-view distributed video coding. ACMI nternational Workshop on Video Surveillance and Sensor Networks, Santa Barbara, CA, USA, 2006.
- [6] ARTIGAS X, ANGELI E, TORRES L. Side information generation for multiview distributed video coding using a fusion approach. 7th Nordic Signal Processing Symposium, Iceland, 2006.
- [7] GAUGEY T, PESUET-POPESC B. Side information estimation and new symmetric schemes for multi-view distributed video coding. J. Vis. Commun. Image R, 2008(19):589-59.

(收稿日期: 2009-09-20)