

视频传感网中鲁棒多视角分布式视频编码

蔡述庭, 王钦若, 谢云, 尹明

(广东工业大学自动化学院, 广州 510006)

摘要: 针对传统无线视频传感器网络中存在的信息丢失和传感器之间需要建立通信等问题, 提出一种鲁棒多视角分布式视频编码系统。通过采用改进的 PRISM 编码结构, 在利用视角间冗余信息提高编码增益的同时将冗余用于增加编码在有损信道中传输的鲁棒性。实验结果表明, 该系统在利用信息冗余获得视频压缩的基础上, 对无线信道固有的数据包丢失也具有较好的鲁棒性。

关键词: 多视角; 分布式视频编码; 鲁棒性; 无线传感器网络

Robust Multi-view Distributed Video Coding for Video Sensor Networks

CAI Shu-ting, WANG Qin-ruo, XIE Yun, YIN Ming

(Faculty of Automation, Guangdong University of Technology, Guangzhou 510006)

【Abstract】 To solve the problem of transmitting in the lossy channel and the requirement communication between the sensors for traditional wireless video sensor networks, a robust multi-view distributed video coding system is proposed. By using modified PRISM codec schemes, the correlation of all views to is explored to get more gains of compression. The redundancy also is used to enhance the system robust for intrinsic wireless lossy channel. Experimental results show that the proposed system accesses the compression gain, and system is also robust to wireless lossy channel.

【Key words】 multi-view; distributed video coding; robust; wireless sensor networks

1 概述

无线视频传感器网络典型的一个应用场合是高密度的视频监控网络系统, 这些系统在军事、安防以及医疗领域都得到了大量应用。在这种视频监控网络系统中, 往往需要从不同角度来获取同一场景的重叠图像, 把这种情况称为多视角。

多视角的视频传感器网络如 CCD 摄像机阵列, 相比非多视角的情况, 针对同一场景获得的视频数据量要大很多。这就需要更大的存储空间和无线传输速率将数据保存和发送到服务器。通常由于受限于传感器的功耗、体积, 以及计算能力, 每个传感器保存和传输这么大的数据量不容易实现或者实现代价太高。这就要求对获取的数据进行大量的压缩。

分布式视频编码的兴起为多视角视频编码提出了一种崭新的解决方案, 基于 Slepian-Wolf^[1]和 Wyner-Ziv 理论^[2]。

在多视角分布式视频编码中, 边信息能够被摄像机的视觉内的运动补偿时间插值来产生, 或者通过临近的摄像机的视角间插值。

在文献[3]中, 一种融合技术被提出, 是基于像素差异和运动矢量幅度。然而, 这些方法都没有能与 MCTI 相比获得相应的编码增益。

在文献[4]中边信息由 MCTI 或者单应性矩阵补偿视角间插值(Homography Compensated Inter-view Interpolation, HCII)来产生。提出一种融合技术来组合所有预测方法, 使得编码增益与传统带 MCTI 的分布式视频编码(Distributed Video

Code, DVC)相比提高 0.2 dB~0.5 dB。

文献[5]将庞大的摄像机阵列划分为传统摄像机和“Wyner-Ziv”摄像机。在这种方法中反馈非常重要, 直到解码成功才会停止请求校验位。当传感器网络非常庞大时, 这么多反馈回路显然是不现实的。

2 分布式视频编码

无线视频、移动摄像电话和无线视频监控摄像机系统等应用需求日益增长, 使得分布式视频编码的研究日益受到重视。Slepian-Wolf 和 Wyner-Ziv 的理论表明, 可以很容易地将一部分的计算量移到解码端实现, 达到构造低复杂度编码器的目标, 即 DVC。

现在最为典型的几个分布式视频编码模型包括: 斯坦福大学 Girod 等人提出的斯坦福模型, 伯克利大学 Ramchandran 等人提出的伯克利模型, 以及欧盟信息技术协会下的未来和新兴技术组(IST-FET)2005 年开始支持的 DISCOVER 项目。

为了适应无线信道易丢失数据的特点, 采用了伯克利的 PRISM 模型。

PRISM 的基本编码思想是使用基于简单的对小的块长度都十分有效的 trellis 码机进行 Syndrome 编码。假设 X 和 Y

基金项目: 广东省自然科学基金研究团队基金资助项目(8351009001000002); 广东工业大学青年基金资助项目(082038)

作者简介: 蔡述庭(1979 -), 男, 讲师、博士研究生, 主研方向: 视频编码; 王钦若, 教授、博士生导师; 谢云, 教授、博士; 尹明, 讲师、博士

收稿日期: 2009-06-30 **E-mail:** shutingcai@126.com

是长度为 3 位的二进制数, 都可以看成是 8 个可能的二进制三元组, 且 X 和 Y 是相关的随机变量。 X 和 Y 之间的相关性使得 X 和 Y 之间的 Hamming 距离总是 1。 X 的码本空间分为 4 个集合, 每个集合包含 2 个码字, 命名为陪集 1~陪集 4。 编码器将集合的索引(index)传送而不是单个码字。 由于有 4 个集合, 因此用 2 位就够了。 而在解码端, 在收到陪集索引后, 使用 Y 来获得正确的 X (最接近 Y 的码字解码为 X)。 源码本空间的划分以及划分后的陪集索引标签(Syndromes)能够通过陪集编码高效计算来完成。 因此, 针对边信息的编码系统可以有很低的编码复杂度和鲁棒性。

3 多视角视频编码

多视角视频系统通常产生大量数据。 因此, 高压缩率成为成功的首要条件。 不同视角间具有很强的相关性, 但摄像机传感器网络的功耗受到严格控制, 必须有一个低复杂度编码器来避免复杂的内部节点间通信。

3.1 多视角分布式视频编码

文献[6]提出了一种多视角 DVC 系统结构, 如图 1 所示, 包括 3 个摄像机并且假设是静止的。 两边视角的编码使用传统的 AVC 帧内编码。 中间视角采用 MDVC。 从中间视角来看, 可以假设奇数帧是使用 AVC 编码的关键帧, 而偶数帧采用 WZ。

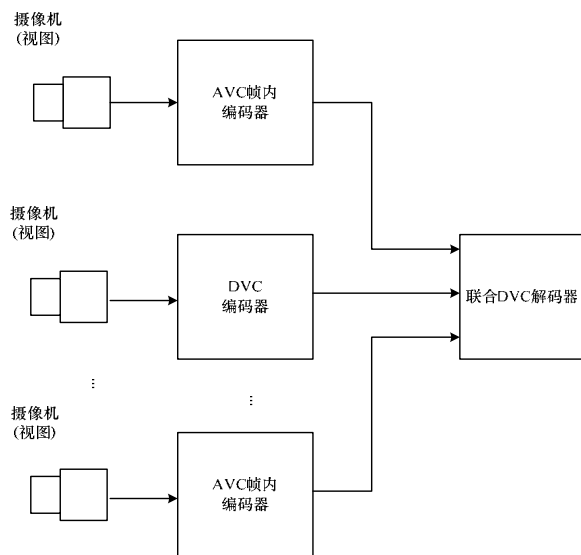


图 1 多视角分布式视频编码

这种系统结构的边信息分为 2 个部分: 一种是视角内的, 采用 MCTI 方式来估计产生; 另外一种为视角间的, 采用单应性矩阵补偿视角间插值 HCII 边信息估计, 获得较好的效果^[6]。 但是从图 1 中可以看出, AVC 帧内编码的数据量占了一半, 也就是说这种结构中有一半的传感器采用的还是传统的 H.264/AVC 编码, 虽然是帧内编码方式, 编码器无需运动补偿, 但相对 WZ 编码, 数据量还是要大很多, 要想办法尽量减少 AVC 编码。

提出的编解码系统如图 2 所示。 这种方法摄像机端具有低功耗、低复杂度编码特点, 且不需要摄像机间通信, 每个摄像机的视频流均单独采用 PRISM 的 WZ 编码, 通过无线信道发送到解码端。 解码端接收到陪集索引后通过运动搜索和单应性矩阵补偿时间插值进行融合解码。

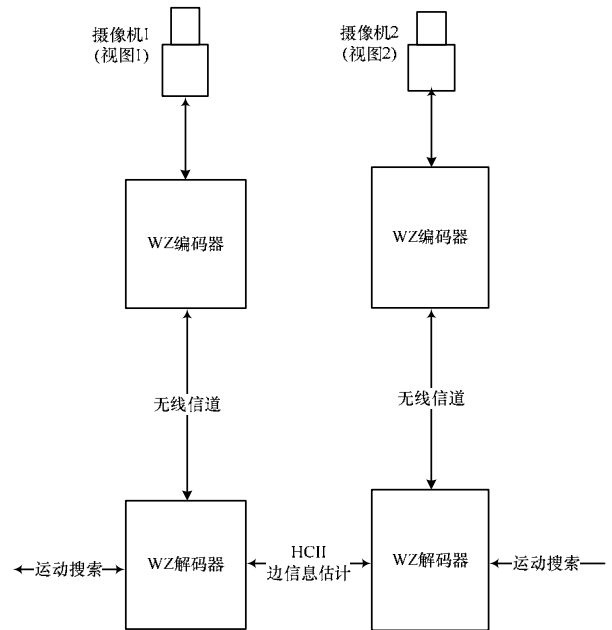


图 2 多视角 PRISM 编解码框图

3.2 HCII 边信息估计

临近视角的差异可以通过单应性矩阵建模, 也就是所谓的投影变换^[7]。 这种边信息产生的方法复杂度要比运动估计低一些, 因为单应性矩阵只需计算一次。 单应性矩阵是一个 3×3 的矩阵, 该矩阵是单应性矩阵坐标系中从一个视角变换到另一个视角的转换系数。 该矩阵有 8 个参数 a, b, c, d, e, f, g, h 。 这样, 从一个视角中每一个点 (x_i, y_i) 映射到另一个视角中点 (x'_i, y'_i) , 变换比例系数为 λ , 公式如下:

$$\lambda \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, x'_i = \frac{ax_i + by_i + c}{gx_i + hy_i + 1}, y'_i = \frac{dx_i + ey_i + f}{gx_i + hy_i + 1}$$

当 $a = e = 1, b = d = g = h = 0$, 这个模型是一个单纯变换。 当 $g = h = 0$, 该模型为 Affine 变换。 其他情况为投影变换。 式中, (x'_i, y'_i) 表示中间视角的像素坐标, (x_i, y_i) 是临近视角的对应坐标, a, b, c, d, e, f, g, h 是变换参数。 这个模型有效的前提条件是场景能被光滑表面近似。 2 个变换 $H1$ 和 $H2$ 分别计算中间和左边视角、中间和右边视角的像素转换关系。

8 个参数能够用于整个序列的计算, 单应性矩阵参数估算利用了全局运动估计技术^[7], 这些参数可以通过求以下达式的最小值而获得:

$$E = \sum_{i=1}^N e_i^2, e_i = I'(x'_i, y'_i) - I(x_i, y_i)$$

其中, $I'(x'_i, y'_i)$ 和 $I(x_i, y_i)$ 分别代表中间和临近视角的图像像素值, 求和是针对 N 对有图像边界的像素执行的。 这个非线性问题被 Levenberg-Marquart 梯度下降法解决, 通过迭代估计这些参数:

$$a^{(t+1)} = a^{(t)} + H^{-1}b, H_{kl} = \frac{1}{2} \sum_{i=1}^N \frac{\partial^2 e_i^2}{\partial a_k \partial a_l}, b_k = -\sum_{i=1}^N e_i \frac{\partial e_i}{\partial a_k}$$

为增加奇异值的鲁棒性, 使用一种截尾二次鲁棒估计器:

$$\sum_{i=1}^N \rho(e_i)$$

$$\text{其中, } \rho(e_i) = \begin{cases} 0 & |e_i| > T \\ e_i^2 & |e_i| \leq T \end{cases}$$

HCI 被用于视角间边信息的获取,如图 3 所示。摄像机 2 的 T 时刻 WZ 帧与左边视角进行 $H1$ 单应性矩阵转换像素,与其右边视角使用 $H2$ 单应性矩阵。摄像机 2 的 T 时刻 WZ 帧同时对同视角的 $T-1$ 时刻 WZ 帧进行运动搜索解码,故解码方式是一种融合方式。运动搜索范围的选择取决于从一帧到下一帧可能的运动量,而 HCI 计算量不会太大,因为单应性矩阵参数只需计算一次,所以对原始 PRISM 解码增加的额外计算负担较少。

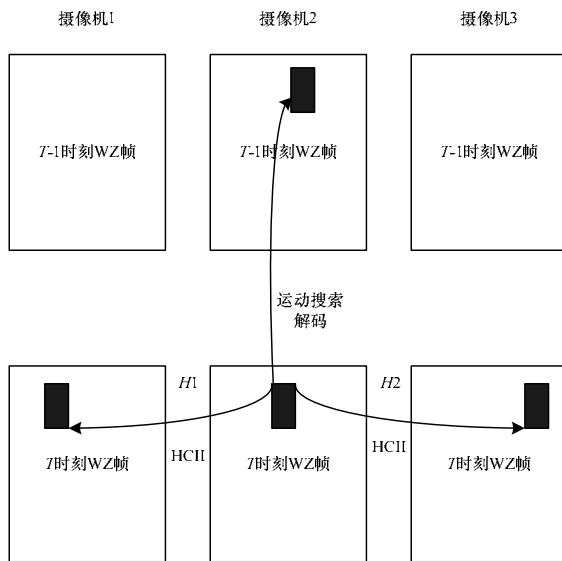


图 3 PRISM 融合解码方式

4 实验结果

仿真代码参考 DISCOVER 编码,使用 Matlab 和 C++ 混合编程。使用 Breakdancing 多视角视频序列,空间分辨率为 256×192 像素,15 帧/s。仿真编码算法包括:提出的带运动搜索和单应性矩阵补偿的 PRISM(PRISM-HCI),欧洲的 DISCOVER 编码(DISCOVER),带 FEC 的 H.263+ 编码(H.263+FEC)。算法仿真在有损信道下进行传输,使得信道具有不同比例的数据包错误,验证系统的鲁棒性和视频压缩增益的提高。

3 种编码算法在不同数据包错误率下以 15 帧/s 传输的图像 PSNR 的结果如图 4 所示。从图中可以看到,在有损信道下,提出的 PRISM-HCI 相比较 H.263+FEC 有大约 4 dB~5 dB 的增益,相比 DISCOVER 编码也有大约 1 dB~2 dB 的增益。且随着误码率的增加,PSNR 下降不明显。

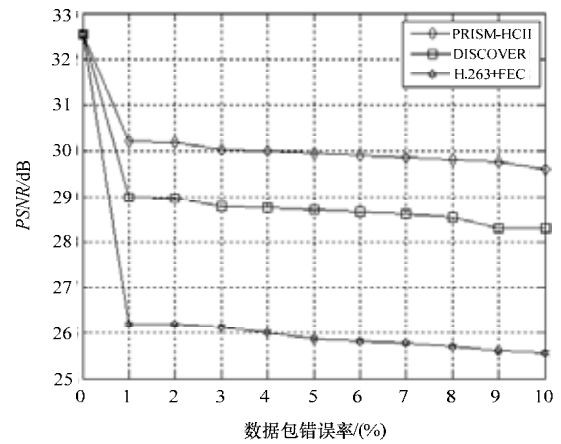


图 4 不同错误率情况下序列的 PSNR

5 结束语

本文提出了一种针对无线视频传感器网络的鲁棒多视角视频编码系统。该系统具有较低的编码复杂度,且不要求传感器之间进行通信,并利用多视角间的信息冗余来提高编码系统的鲁棒性。

参考文献

- [1] Slepian J, Wolf J. Noiseless Coding of Correlated Information Sources[J]. IEEE Trans. on Information Theory, 1973, 19(4): 471-480.
- [2] Wyner A, Ziv J. The Rate-distortion Function for Source Coding with Side Information at the Decoder[J]. IEEE Trans. on Information Theory, 1976, 22(1): 1-10.
- [3] Artigas X, Angeli E, Torres L. Side Information Generation for Multi-view Distributed Video Coding Using a Fusion Approach[C]// Proc. of the 7th Nordic Signal Processing Symposium. [S. l.]: IEEE Press, 2006.
- [4] Ouaret M, Dufaux F, Ebrahimi T. Fusion-based Multi-view Distributed Video Coding[C]//Proc. of the 4th ACM International Workshop on Video Surveillance and Sensor Networks. Santa Barbara, CA, USA: [s. n.], 2006.
- [5] Zhu Xiaoqing, Aaron A, Girod B. Distributed Compression for Large Camera Arrays[C]//Proc. of IEEE Workshop on Statistical Signal Processing. [S. l.]: IEEE Press, 2003: 30-33.
- [6] Dufaux F, Ouaret M, Ebrahimi T. Recent Advances in Multi-view Distributed Video Coding[C]//Proc. of SPIE Mobile Multimedia/Image Processing for Military and Security Applications. Orlando, FL, USA: [s. n.], 2007.
- [7] Dufaux F, Konrad J. Efficient, Robust, and Fast Global Motion Estimation for Video Coding[J]. IEEE Trans. on Image Processing, 2000, 9(3): 497-501.

编辑 顾逸斐