

基于 3D-SPIHT 的立体视频图像压缩编码¹

张勇东 * ** 李桂苓 *

*(天津大学电子信息工程学院 天津 300072)

** (中国科学院计算技术研究所 北京 100080)

摘 要 该文提出一种新的立体视频编码方案:在辅助序列中进行视差补偿预测和三维等级数集合分区(3D-SPIHT)编码。3D-SPIHT 算法建立在真三维小波分解基础上,通过定义一种新的时空方向树结构,实现了静止图像 SPIHT 算法的三维扩展。实验结果表明该方案的编码性能略高于传统方案,具有较低的计算复杂度。所产生的嵌入式辅助序列码流,可根据通道带宽自适应调整输出码率,最大限度地提高辅助序列的质量。

关键词 立体视频编码, 三维小波变换, 等级数集合分区

中图分类号 TN911.73

1 引 言

近年来,三维显示技术的进步,再次掀起了研究立体视频技术的热潮。立体视频增加了对景物深度信息的表征,增强了视觉现实感和逼真感,在电视、自动导航、虚拟现实等领域有广泛应用前景。立体视频由两个同步的视频序列组成,其数据量两倍于单目视频,必须研究更有效的压缩技术来推进其实用化。消除立体运动图像左右两视角序列各自的空间、时间及彼此间的交叉相关性,利用人的视觉系统允许两序列编码质量不一的特性^[1],可提高压缩效率。目前立体视频编码一般采用基于 MPEG-2 的方案(S-MPEG)^[2],对一个视角序列(主序列)采用 MPEG-2 进行独立高质量编码;而另一视角序列(辅助序列)相对主序列进行较低质量编码,其宏块预测选取来自主序列的视差补偿预测(Disparity Compensated Prediction, DCP)和来自本序列的运动补偿预测中误差较小者。

替代传统的运动补偿预测编码,应用 3D 小波变换进行视频序列编码近年来得到广泛关注,已发表一些有效算法^[3]。在静止图像的小波编码领域中,等级树集合分区(SPIHT)^[4]算法公认是一种有效的算法。该算法简单,产生的码流是嵌入式的。引入嵌入式 SPIHT 编码技术进行的 3D 小波视频编码称为 3D-SPIHT 算法。采用该算法进行视频编码的主要优点是:产生的嵌入式码流可在任意点截断,从而维持尽可能好的质量;容易实现码率的可扩展性和精确控制码率。

本文提出一种新的立体视频编码方案,引入 3D-SPIHT 算法对辅助序列进行编码,给出了实验结果并进行了讨论。

2 基于 3D-SPIHT 的立体视频编码方案

本文提出的基于 3D-SPIHT 的立体视频编码方案如图 1 所示。对主序列,采用 MPEG(或 H.26X)进行独立编码,实现与单通道视频编码系统的兼容。同时,通过视差估值 DE(Disparity Estimation)获得两序列间对应图像的视差向量 DV(Disparity Vector),再由 DV 和解码的主序列图像,通过重叠块视差补偿 OBDC(Overlapped Block Disparity Compensation)获得辅助序列预测图像。进而对辅助序列的源图像与预测图像的差值图像进行 3D-SPIHT 编码,加之编码 DV 的码流,形成辅助序列码流。本算法与 S-MPEG 的主要区别在于对辅助序列的编码上,其

¹ 2001-01-15 收到, 2001-07-11 定稿
国家自然科学基金资助

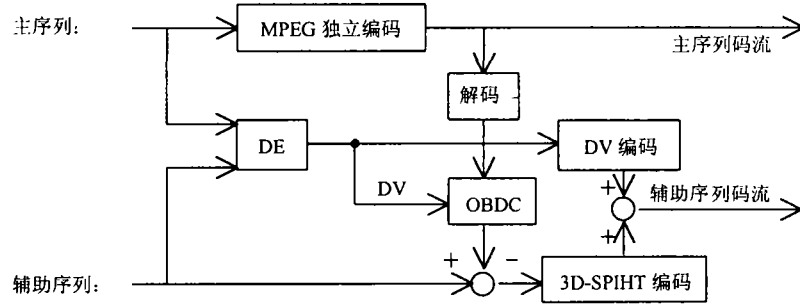


图 1 本文提出的立体视频编码方案

主要出发点是：在有限带宽的传输通道内，首先保证主序列的传输质量，然后根据带宽情况，确定辅助序列的码率。采用本算法所产生的嵌入式辅助序列码流具有码率的连续可扩展性，可对码流进行任意截断，从而根据带宽变化，自适应地调整输出码率，尽可能好地提高辅助序列的图像质量。

本方案的 3D-SPIHT 编码模块由 3 部分构成：三维小波分解，3D-SPIHT 和自适应算术编码。下面分别加以阐述。

2.1 三维小波分解

视差预测误差图像序列仍存在空间和时间相关性^[5]，可采用三维小波分解加以消除。三维小波分解是二维分解在时间轴上的扩展，通过对图像序列的三维分解，来消除其在空间和时间方向上的相关性。文献 [2] 的 3D-SPIHT 算法所使用的三维小波分解是把图像序列以 N 帧为一组进行划分，以每一图像组为一个变换单元，先沿时间轴进行 k 级小波分解，再对所得各时间子带图像帧进行 m 级空间分解。因其帧内分解与时间轴的分解是绝对分离，互不相关的，故属准三维小波分解。本文遵循一维和二维小波分解的 Mallat 算法，对图像序列进行真正意义上的真三维小波分解，分解结构如图 2 所示，与准三维小波变换相比，降低了计算量，提高了分解效率。

本方案所使用的真三维小波分解需把图像序列以 2^m ($m = 1, 2, 3, \dots$) 帧为一组进行划分，折衷考虑帧延迟和编码效率两方面，在辅助序列中取 8 ($m = 3$) 帧连续图像为一组，把每组内的 8 帧预测误差图像作为一个变换单元进行 3 级三维小波分解。分解之后，一个变换单元内的所有系数在时空方向上构成一个金字塔型结构。本文采用 Daubechies 的 9/7 双正交小波做空间分解^[4]，Harr 小波做时域分解。

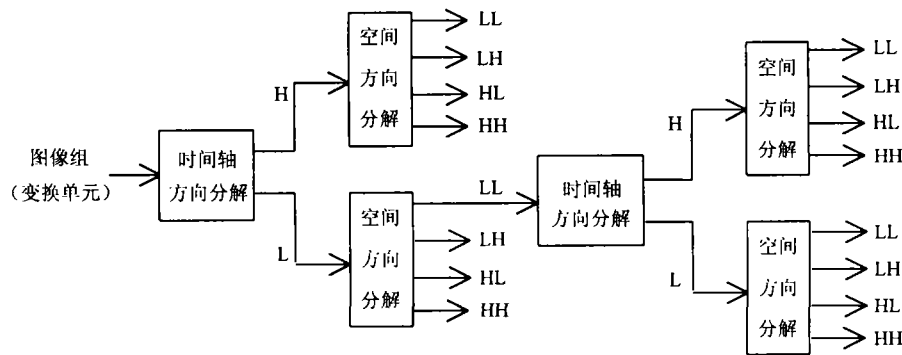


图 2 三维小波分解 (2 级分解) 结构示意图

2.2 3D-SPIHT 算法

SPIHT 算法^[4]是一种嵌入式图像编码算法。所谓嵌入式编码就是编码器将待编码的码流按重要性进行排序,根据目标码率或失真度大小的要求随时结束编码。SPIHT 算法主要包括排

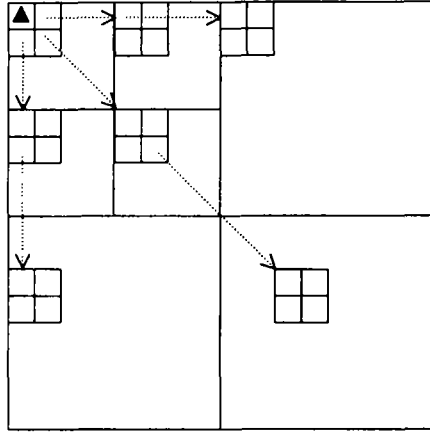


图 3 空间方向树及各结点间的父子关系

3D-SPIHT 是 SPIHT 算法的三维扩展。SPIHT 算法本身无维数限制,但为适应本文采用的三维小波分解结构,须为 3D-SPIHT 定义一种新的方向树结构以及相应结点间的父子关系。在 2D-SPIHT 算法中,空间方向树形成方法是:它的一个结点对应金字塔型结构中的一个系数,每一结点或者有 4 个直接后继结点(一组 2×2 相邻系数),或者没有后继结点。在金字塔的最高级上的系数是树的根,以 2×2 相邻系数为一组对它们进行划分,在每组中有一个系数(在图 3 中以 ▲ 表示)没有后继结点。在三维小波分解后,一个变换单元内的所有系数在时空方向上形成金字塔型结构,在此基础上本文定义了一种新的方向树——时空方向树。

设一个变换单元的尺寸为 $H \times V \times T$ (H, V, T 分别为变换单元的水平、垂直和时域尺寸)。对变换单元进行 l 级三维小波分解后,时空方向树的每一结点对应变换单元中的一个系数,在此共有 3 种类型的结点:根结点 (RN)、中间结点 (MN) 和叶结点 (LN),对它们定义如下:

对变换单元中的每一系数 (i, j, k) ($0 \leq i < H, 0 \leq j < V, 0 \leq k < T$):

- (1) 若 $i < H/2^l$ 且 $j < V/2^l$ 及 $k = 0$, 则 $(i, j, k) \in RN$;
- (2) 若 $i \geq H/2$ 或 $i \geq V/2$ 或 $k \geq T/2$, 则 $(i, j, k) \in LN$;
- (3) 若 (i, j, k) 不满足以上两个条件, 则 $(i, j, k) \in MN$ 。

时空方向树根结点、中间结点和叶结点分别对应于时空方向上的金字塔结构的最高级、中间级和最低级。在金字塔的最高级,把相邻的 2×2 根结点划分为一组,每组有一个结点(图 4 中以 ▲ 表示)有 4 个后继结点,其它结点有 8 个后继结点。根节点与其后继结点的关系如图 4 所示。每一中间结点 (i, j, k) 有 8 个后继结点,设 $O(i, j, k)$ 为其后继结点的集合,则

$$O(i, j, k) = \{(2i, 2j, 2k), (2i + 1, 2j, 2k), (2i, 2j + 1, 2k), (2i + 1, 2j + 1, 2k), \\ (2i, 2j, 2k + 1), (2i + 1, 2j, 2k + 1), (2i, 2j + 1, 2k + 1), (2i + 1, 2j + 1, 2k + 1)\}$$

叶结点没有后继结点。

序和细化两个过程。首先对小波系数建立空间方向树的结构(图 3),排序过程就是沿空间方向树对系数进行显著性测试的过程。SPIHT 算法基于如下假设:如果在低分辨率子带上系数不显著,则其后继结点很可能也不显著。细化过程是对显著的系数,从高位到低位依次输出各位值的过程。SPIHT 算法的实现步骤如下:

- (1) 输出幅值最大的系数 $c(i, j)$ 的最高位次 n , 即 $n = \lfloor \log_2(\max_{(i, j)} \{|c(i, j)|\}) \rfloor$;
- (2) 对于所有 $2^n \leq |c(i, j)| \leq 2^{n+1}$ 的系数,输出坐标和符号位(排序过程);
- (3) 对于所有 $|c(i, j)| \geq 2^{n+1}$ 的系数 $c(i, j)$,输出第 n 次位(0 或 1)(细化过程);
- (4) n 减 1, 跳回步骤(2)。

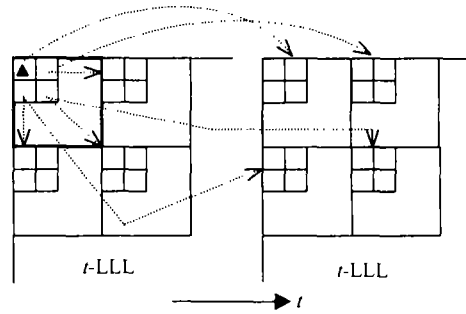


图 4 时空方向树的根结点与其后继结点的关系

建立起时空方向树后, 即可按 2D-SPIHT 算法完成 3D-SPIHT 编码的其余过程。

2.3 自适应算术编码

为了提高压缩效率, 需对 3D-SPIHT 编码输出码流用自适应算术编码进一步压缩。与文献 [4] 不同, 本文采用了基于上下文模型的自适应算术编码器对各结点的显著性状态 (0 或 1) 进行压缩。因相邻结点的显著性状态有很强的相关性, 因此对于某一结点, 可选择与其在同一子带内的水平和垂直方向的各两个相邻结点作为它的上下文模型, 按这 4 个相邻结点的显著与否可分为 5 种上下文状态, 状态 0 表示 4 个相邻结点都不显著, 状态 1 表示有一个显著, 以此类推, 状态 4 表示 4 个相邻结点都显著。对于 3D-SPIHT 的其它输出值仍采用不带上下文模型的自适应算术编码器 [6] 进行压缩。

以上本文阐述了所提出编码方案的 3D-SPIHT 编码模块。DE 模块与运动估值相似, 可采用固定尺寸块匹配 (FSBM) 算法, 也可采用更复杂的 DE 算法 (如文献 [7] 所提出的算法), 目前本文采用的是最简单的 FSBM(16×16 块) 算法, 所得 DV 采用类似 MPEG-2 TM5^[8] 的运动向量编码方式来编码。为了提高视差补偿的效率, 减小块效应, 本文的视差补偿模块采用重叠块视差补偿 (OBDC) 算法, OBDC 与重叠块运动补偿^[9] 相似, 不再加以阐述。

3 实验结果及讨论

为了评价本文提出的 3D-SPIHT 编码方案, 并与 S-MPEG 编码方案比较, 选取 3 个标准立体视频测试序列: Train, Aqua 和 Tunnel 进行实验测试。各序列长 40 帧, 图像尺寸 720×576。S-MPEG 编码系统^[2] 是由 MPEG-2 TM5 编码器^[8] 经相应的修正而得。在实验中, 两种编码方案对于主序列 (右视角) 都采用 MPEG-2 TM5 编码器以 6Mbps 的码率进行编码, 对于辅助序列 (左视角) 取 8 帧图像为一组, 分别以不同的码率进行编码。表 1 列出了 3 个测试序列的辅助序列重建图像 PSNR 平均值。图 5 对比显示出序列 Train 的辅助序列各重建图像帧的 PSNR, 图 6 给出了序列 Train 的辅助序列中第 25 帧重建图像。上述结果表明: 相同码率下, 3D-SPIHT 方案的 PSNR 值接近甚至高于 S-MPEG。在较低码率下 (如 1Mbps), 3D-SPIHT 与 S-MPEG 相比, 具有更好的主观视觉质量, 基本消除了 S-MPEG 产生的一定程度的块效应。

表 1 重建图像的 PSNR 平均值

测试序列	PSNR 平均值 (dB)			
	1Mbps		1.8Mbps	
	3D-SPIHT	S-MPEG	3D-SPIHT	S-MPEG
Train	29.90	29.86	31.54	31.39
Aqua	28.53	27.90	29.23	29.05
Tunnel	30.52	30.70	32.28	32.33

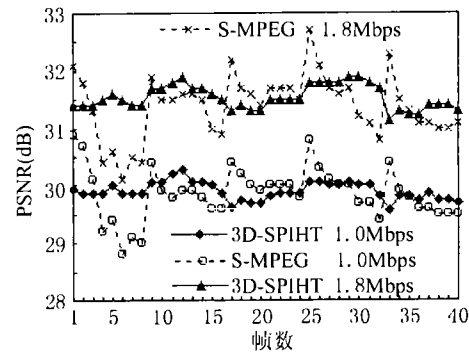
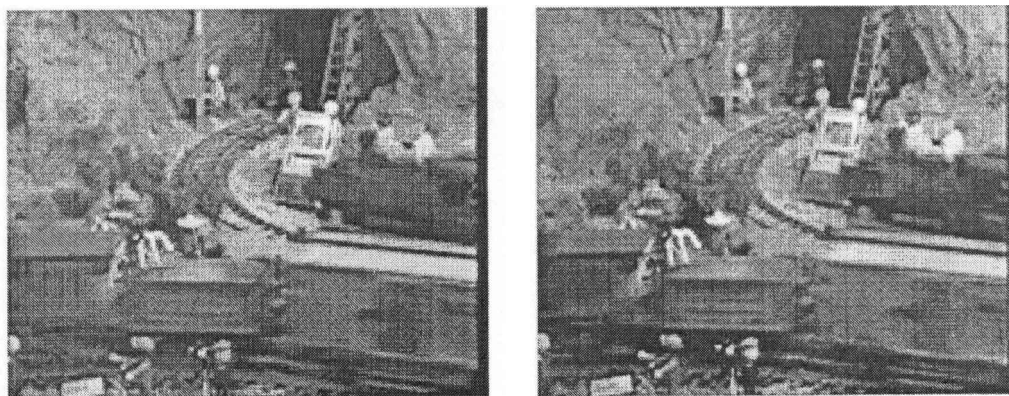


图5 每帧重建图像 (Train) 的 PSNR 值



(a) 3D-SPIHT

(b) S-MPEG

图6 第25帧重建图像 (Train)(码率: 1Mbps)

本文提出编码方案与 S-MPEG 的主要区别在于: 对辅助序列编码用 3D-SPIHT 算法替代了运动补偿 DCT 编码, 避免了在辅助序列内进行计算负荷较大的运动估值。文献 [3] 实验结果证明 3D-SPIHT 编码的计算复杂度只有运动补偿 DCT 编码的 50%, 加之本文的 3D-SPIHT 算法与文献 [3] 的算法相比, 采用的是计算量更小的真三维小波分解, 因此可以推断在对辅助序列编码上, 3D-SPIHT 的计算复杂度明显低于 S-MPEG。此外, 还需指出文献 [3] 的算法取 16 帧为一个图像组, 时域帧延迟是本文 3D-SPIHT 算法的 1 倍。

3D-SPIHT 编码方案较 S-MPEG 方案具有下列主要优点:

(1) 辅助序列产生嵌入式码流, 码率具有连续可扩展性, 可在任意点截断码流。从而可自适应地调整辅助序列的输出码率, 保证主序列图像质量并最大限度地提高辅助序列的图像质量。

(2) 简化了辅助序列的码率控制算法, 实现了在 GOP 中比特率的全局优化分配, 重建的图像质量较平稳。

(3) 避免了在辅助序列内进行计算负荷较大的运动估值, 具有较低的计算复杂度。

4 结束语

本文提出了基于 3D-SPIHT 的立体视频压缩编码方案。它对主序列进行独立编码, 实现与单通道视频编码系统的兼容; 通过 DE 和 OBDC 模块对辅助序列图像进行预测, 用 3D-SPIHT

编码模块完成对其预测差值图像的嵌入式编码。实验表明新的编码方案的编码性能与 S-MPEG 相当甚至略有超出, 并具有 S-MPEG 所不具备的多种优点, 适用于立体视频编码。此外所提出的 3D-SPIHT 算法也适用于常规视频编码。

致谢 感谢法国电信 CCETT 为作者提供的标准立体图像测试序列。

参 考 文 献

- [1] P. R. Chang, M. J. Wu, A wavelet multiresolution compression technique for 3D stereoscopic image sequence based on mixed-resolution psychophysical experiments, *Signal Processing: Image Communication*, 2000, 15(9), 705-727.
- [2] M. Ziegler, Digital stereoscopic television-state of the European project DISTIMA, in Proc. 4th European Workshop on 3DTV, Rome, Italy, 1993, 247-256.
- [3] B. J. Kim, Z. Xiong, W. A. Pearlman, Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT), *IEEE Trans. on CAS VT*, 2000, 10(8), 1374-1387.
- [4] A. Said, W. A. Pearlman, A new, fast and efficient image codec based on set partitioning in hierarchical trees, *IEEE Trans. on CAS VT*, 1996, 6(3), 243-250.
- [5] J. S. McVcigh, Efficient compression of arbitrary multi-view video signals, [PhD dissertation], Carnegie Mellon University, 1996.
- [6] I. Witten, R. M. Neal, J. G. Cleary, Arithmetic coding for data compression, *Comm. of the ACM.*, 1987, 30(6), 520-540.
- [7] 张勇东, 李桂苓, 基于视差补偿预测的立体视频图像压缩编码, *信号处理*, 2001, 17(4), 335-339.
- [8] ISO/IEC/JTC1/SC29/WG11/93-400, MPEG-2 Test Model 5(TM5), Test Model Editing Committee, April, 1993.
- [9] C. Auyeun, J. Kosmach, M. Orchard, T. Kalafatis, Overlapped block motion compensation, *Proc. SPIE*, 1818, Nov. 1992, 561-572.

AN EFFICIENT CODING ALGORITHM FOR STEREO VIDEO WITH 3D-SPIHT

Zhang Yongdong* ** Li Guiling*

**(School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China)*

*** (Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)*

Abstract This paper presents a new strategy for efficiently coding stereo video sequences. In the scheme, the auxiliary sequence is coded by disparity compensated prediction and 3D-SPIHT algorithm. 3D-SPIHT algorithm is based on genuine 3D wavelet transform, and implements a three dimensional extension of SPIHT in still image coding by defining a new spatio-temporal orientation tree structure. The scheme can provide comparable performance to the conventional method with minimal computational complexity and produce an embedded bit stream in the auxiliary sequence. According to channel bandwidth, the bit rate can be adaptively adjusted so that the best quality of the auxiliary sequence is obtained.

Key words Stereo video coding, Three dimensional wavelet transform, SPIHT

张勇东: 男, 1973 年生, 博士, 从事视频压缩编码、计算机视觉等方面的研究工作。

李桂苓: 男, 1938 年生, 教授, 博士生导师, 从事图像处理、数字电视等方面的教学和研究工作。