

文章编号:1001-5132 (2009) 03-0317-05

面向三维视频系统的深度图压缩方法研究

朱 波¹, 彭宗举^{1,2}, 张 云¹, 邵 枫¹

(1. 宁波大学 信息科学与工程学院, 浙江 宁波 315211; 2 浙江省综合信息网技术重点实验室, 浙江 杭州 310027)

摘要: 理论分析了基于 Mesh、JPEG、模式函数等静态深度图压缩算法, 然后比较了基于 H.264 深度视频独立压缩算法、彩色和深度视频运动矢量共享的深度视频压缩算法, 给出 2 种深度视频压缩算法的实验对比, 并对深度图压缩算法发展做了相关的预测和展望.

关键词: 三维视频系统; 深度图压缩; 视频压缩标准; 3D-mesh

中图分类号: TP391.41

文献标识码: A

三维(Three-Dimensional, 3D)视频和自由视点视频是当前先进的视觉媒体模式, 含有传统二维视频所没有的视听功能, 3D 视频能够反映出场景的深度信息, 其自由视点视频也能满足用户从任意角度选择和操作视听对象^[1-3]. 联合视频专家组(Joint Video Team, JVT)组织并提出了 Multi-view Video plus Depth(MVD)结构^[1]. 在 MVD 的结构中, 每幅彩色图像都含有相对应的深度图, 此举极大地增加了传输的数据量^[4-6]. 因此, 需要研究有效的深度图压缩算法以缓解带宽压力, 而其中的深度图压缩是三维视频系统中的核心技术.

深度图所代表的是场景到摄像机成像平面的距离信息, 它将实际深度值量化到 $[0, 255]$ 以获得代表深度信息的灰度图^[7-8], 通过深度图以及其对应的彩色图可以用来绘制相邻虚拟视点图像^[9]. 由于遮挡、暴露等原因, 虚拟视点的质量将会随着基线距离的增大而降低. 而在 MVD 的结构下, 是通过相邻的 2 个视点来绘制中间的视点, 此法可以有效地避免此问题^[1].

深度图的表示方法主要有 2 种: (1) 3D-mesh 表示深度图; (2) 将深度图认为是 2D 图像. 基于 3D-mesh 的压缩主要是静态压缩^[9], 2D 图像压缩既有 JPEG、JPEG-2000 等静态压缩标准, 又有 MPEG-2、MPEG-4、H.264 等动态压缩标准. 在现有的压缩算法中, H.264 则体现出良好的压缩性能. 因此, 深度图压缩正逐步转向基于 H.264 视频编码标准的压缩^[1, 10-12].

1 静态深度图的压缩方法

静态深度图的压缩仅利用空间相关性来进行, 没有利用深度图时间上的相关性. JPEG2000 编码器使用基于小波变换的算法来进行深度图的压缩, 但它的缺点是会在边缘像素位置上产生环形的效果^[13]. 深度图最终目的是用来合成中间图像, 深度图边缘的失真是一种几何失真, 对于绘制结果产生的影响比较大. 因此, Krishnamurthy 做了些改进, 提出了基于感兴趣区域的编码^[14]. 通过边缘检测,

求取深度边界作为感兴趣区域对其做特殊的处理,即采用无失真、压缩的方式保证边缘深度的准确性.另外,在深度图的编码中,小深度失真的影响大于大深度失真,这是因为小深度往往是前景区域,是人们比较关注的区域,所以需要特别的保护.因此,该文献中也提出了重塑深度的动态范围以反映不同深度的不同重要性.

Morvan 通过分析深度图,发现平滑区域很多,仅仅对象边界会产生突变的深度信息,基于此特征,采用模式函数来近似深度图的内容^[13].在其框架中主要使用分段常函数和分段线性函数这 2 类模式函数.对于固定深度区域用常函数表示,而对于渐进深度变化区域则用线性函数来表示.该算法比 JPEG2000 的编码效率有所提高,但由于只考虑了深度图的空间冗余,没有考虑时间冗余,所以其压缩效率还是不够高.

3D-mesh 的编码方案在 MPEG-4 GMC 就有涉及,如文献[15-17]对其进行了改进.但一般 mesh 方案并没有考虑到深度图的特殊性.为保证深度边界的准确性,将深度图转换为 3D-mesh 的思想应是基于非等距、内容自适应的采样过程^[18-19],即在平坦区域采样点数比较少,而在对象边界需要更多的采样点.为重建采样点之间的区域信息,需要采用线性插值的方式.所以需要对采样后的图像构建成三角形网格,然后通过像素与三角形 3 个顶点之间的线性关系插值三角形内部的像素.

当把深度图转换成三维网格之后,网格信息需要被传输编码.编码主要有几何编码和拓扑编码 2 个部分,几何信息指顶点的位置坐标,拓扑信息指多边形网格顶点之间的连接信息. Morvan 在对象边界处的每个点都分配 2 个深度值,并且保证网格边界沿着对象的轮廓,防止同个三角形网格中既含有前景,又有背景部分,而影响绘制效果^[13].这种传统的建立三角形网格很难寻找其规律,所以需要进行拓扑编码.而 Kim 提出采用栅格方式来建立三角形网格,这样就无需传输拓扑信息,并

且采用分层方式以保证对象边缘深度的准确性^[5].为减少时间上的冗余度, Grewatsch 在相邻两帧匹配搜索得到结点的运动矢量,仅仅编码两帧中变化的部分以减小码率^[19]. Grewatsch 还分别运用 H.264 和 mesh 压缩深度图,结果显示 mesh 的压缩方式和 H.264 帧内编码的效率相近,但低于 H.264 帧间的结果.

2 深度图序列的压缩方法

采用 H.264/avc 对深度图序列进行压缩,体现出良好的压缩性能.为进一步减小深度图压缩后的码率,对深度图进行下采样^[11],通过 H.264 压缩编码后的上采样,这样会大大减少码流.但是由于采样影响了深度图边界的准确性,对绘制的结果产生影响,增加了后处理的难度.文献也提出先对深度图进行 7×7 均值滤波,再经 H.264 压缩,与原始的相比,码率减少了很多,并且绘制效果没有明显的不同.虽然 H.264 编码器有良好的压缩性能,但是其编码复杂度很高, Oh 提出用彩色图像的编码信息来辅助深度图的编码^[12],并采用一种彩色视频和深度视频运动矢量共享的编码结构,既减少了编码时间,同时又提高了压缩效率.

深度图序列和彩色图像序列之间具有很强的相关性.深度图序列经 H.264 编码后,序列中主要包含有超过 40% 运动参差信息.彩色图序列和深度图序列含有相似的特征——对象边界的一致性,对象运动方向的一致性公式如下:

$$dist_{\text{frame}} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|t_{i,j} - d_{i,j}\|, \quad (1)$$

其中, t 为彩色图像 4×4 块的运动矢量; d 为深度图像 4×4 块的运动矢量. $M \times N$ 表示图像中总共的 4×4 块数目.统计 INTERVIEW、Breakdancers、Ballet 序列运动矢量的平均差额分别为 0.59, 0.78, 0.96 个像素.可见,深度序列和彩色序列的运动矢量存在很强的相关性,所以我们可以用彩色图像

的运动信息来辅助深度图序列的编码. 具体的编码框图如图 1 所示, 它是在 H.264 编码器的基础上改变了运动矢量的获取方式和获得最佳模式的率失真代价函数, 以获取深度图最佳模式, 主要包含 3 部分: 彩色图像模式和运动矢量的解码、深度图候选模式以及运动矢量的获取.

彩色图像模式和运动矢量的解码: 在深度图编码的过程中并不进行运动估计的过程, 而是利用彩色图的运动矢量, 所以首先需要利用 H.264 解码器得到彩色图像的运动矢量和模式. 在获得彩色图的运动矢量之后, 再对运动矢量做进一步的处理. 彩色图像中当前宏块的编码模式为帧内, 则通过相邻块计算运动矢量. SKIP 模式的运动矢量应该与其预测运动矢量相同, 所以直接输出预测运动矢量, 相应公式如下:

$$(p_x, p_y) = \text{MEDIAN}(A, B, C), \quad (2)$$

其中, A, B, C 分别表示为当前宏块左边块、上边块和右边块的运动矢量; MEDIAN 表示取均值. 得到彩色图像编码的最优模式后, 需要经过合并、拷贝和分裂等操作来产生深度图压缩的候选模式和相对应的运动矢量, 如图 2 所示.

图 2 中, 当候选模式与彩色图像的最佳模式相同时, 直接拷贝运动矢量. 合并后的块或宏块的运动矢量为平均运动矢量, 分裂块的运动矢量为彩色图像中当前分裂块以及其左边和上边块的平均运动矢量. 此处, 最优模式通过(3)式的率失真代价函数在帧间候选模式、SKIP 模式和帧内模式间

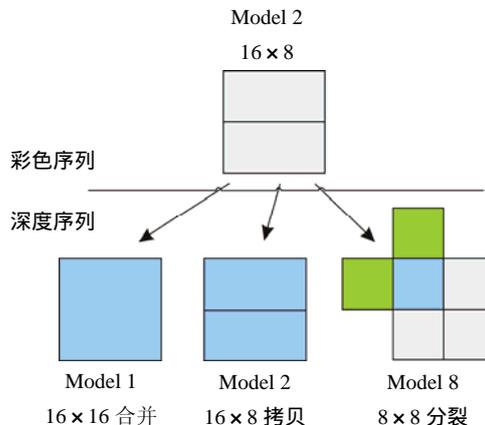


图 2 候选模式

选取.

$$I^* = \arg \min D(S, I | \lambda) + \lambda R(S, I | \lambda), \quad (3)$$

其中, I^* 为最优模式; S 为需要编码的块; R 为码率; λ 为拉格朗日乘子.

由于深度图编码省略了运动估计时间, 且传输的码流中没有运动矢量参差, 因此可以减小码流. 同时, 深度图编码模式中含有更多的 INTRA 模式, 可以阻断错误的运动信息所带来的误差传播问题.

3 实验和分析

实验采用 H.264 标准参考平台 JM9.7 进行编码实验, 测试序列采用微软研究院的交互视觉媒体组提供的序列. 图 3 和图 4 分别表示了由相机捕获的“Breakdancers”序列和“Ballet”序列的第一时刻参考图像及其对应的深度图像. 深度图编码时间数

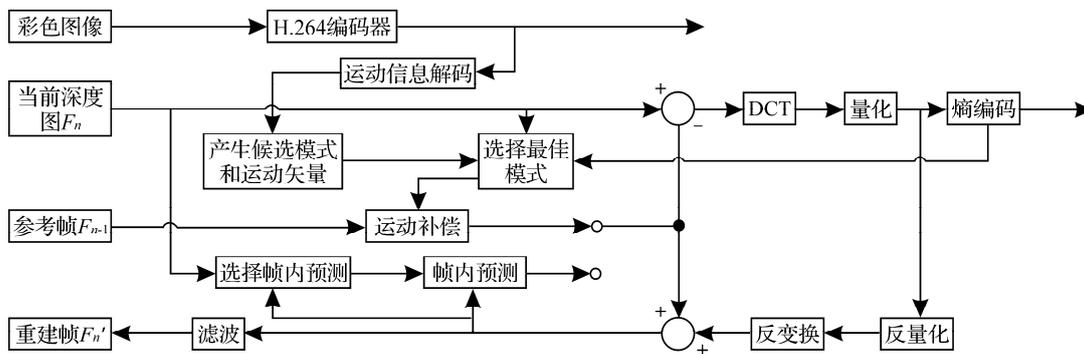


图 1 编码框图

据见表 1. 实验结果对比如图 5 所示.



图 3 测试序列“Breakdancers”及其深度图



图 4 测试序列“Ballet”及其深度图

表 1 深度图编码时间

测试序列	原始/s	文献方案/s	比率/%
Breakdancers	3 876.54	668.03	5.80
Ballet	3 821.26	665.43	5.83

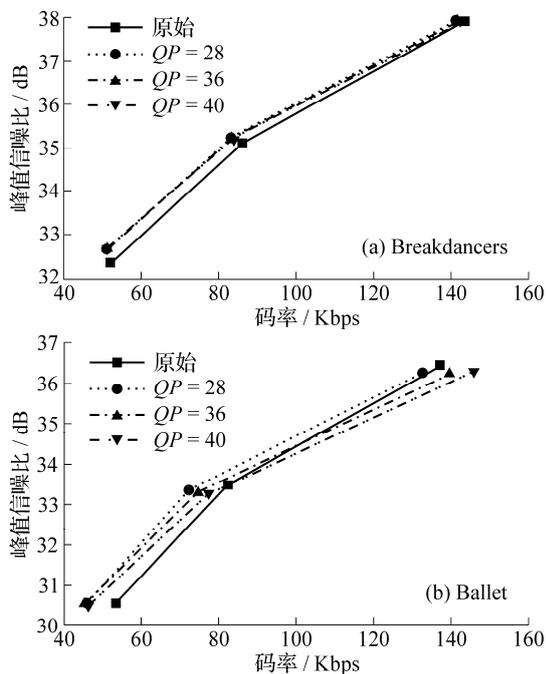


图 5 实验结果对比

图 5 中实线表示深度图在 QP 分别为 41、45、49 时 H.264 的编码结果, 虚线、单点虚线及双点虚线分别表示利用彩色图像 QP 在 28、40、36 编码所得的 MV 来编码深度图的结果. 可以发现当 QP 增大时, 深度图的编码效果降低, 尤其是在高比特

率时, 在 Ballet 中出现相同码率下 PSNR 小于原始的方案. 主要原因当 QP 增大时, 彩色图像的运动矢量不精确, 使得编码参差的增加大于减小的运动矢量参差. 该方案只在低比特率情况下可行, 但这样就不能保证其绘制的质量, 一般利用 H.264 来编码深度图时, $QP < 40$ 已保证深度图的准确性.

4 结论和展望

深度图压缩是当前视频压缩的研究热点, 基于上述分析讨论, 进一步可从以下方面开展研究:

(1) 深度图压缩正逐步转向基于视频编码标准的压缩, 现有的基于 H.264 的压缩算法以及一些改进算法都有其缺陷, 需要进一步研究.

(2) 分析深度图不同区域对绘制的影响, 彩色图和深度图之间的相关性, 深度图各个视点之间的相关性, 是进一步的研究方向.

(3) 由于深度图的特殊性, 需要有一种客观的质量评价算法来评价压缩算法的优劣.

参考文献:

- [1] Smolic A, Mueller K, Merkle P, et al. Multi-view video plus depth (MVD) format for advanced 3D video systems[C]//23rd Meeting: ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-W100, USA: San Jose, 2007.
- [2] Tanimoto M, Fujii T. FTV-free viewpoint television[C]//MPEG-M8595, Kaernten: Klagenfurt, 2002.
- [3] Tanimoto M. Overview of free viewpoint television[J]. Signal Processing Image Communication, 2006, 21(6): 454-461.
- [4] Kauff P, Atzpadin N. Depth map creation and image based rendering for advanced 3DTV services providing interoperability and scalability[J]. Signal Processing Image Communication, Special Issue on 3D Video and TV, 2007, 22(2):217-234.
- [5] Kim S Y, Ho Y S. Three-dimensional natural video system based on layered representation of depth maps [J]. IEEE Trans on Consumer Electronics, 2006, 52:1 035-1 024.
- [6] Fehn C, Atzpadin N. An advanced 3DTV concept provid-

- ing interoperability and scalability for a wide range of multi-baseline geometries[C]//Proceedings of International Conference on Image Processing. GA: Atlanta, 2006: 2 961-2 964.
- [7] Tanimoto M, Fujii T, Suzuki K, et al. Multi-view depth map of Rena and Akko & Kayo[C]//ISO/IEC JTC1/SC29/WG11, MPEG-M14888, China: Shenzhen, 2007.
- [8] Tanimoto M, Fujii T, Suzuki K, et al. Experiment of view synthesis using multi-view depth[C]//ISO/IEC JTC1/SC29/WG11, MPEG-M14889, China: Shenzhen, 2007
- [9] Fehn C. Depth-image-based rendering, compression and coding strategies for compression of depth map sequences [C]//49th SPIE's Annual Meeting, 2004.
- [10] Grewatsch S, Muller E. Evaluation of motion compensation and coding strategies for compression of depth map sequences[J]. Proceedings of SPIE, 2004, 5 561:117-124.
- [11] Ho Y S, Lee C. Depth map generation and virtual view synthesis[C]//JVT-Y065 25th Meeting: China: Shenzhen, 2007.
- [12] Oh H, Ho Y S H. 264-based depth map sequence coding using motion information of corresponding texture Video [J]. Lecture Notes in Computer Science, 2006, 4 319:898-907.
- [13] Morvan Y, Farin D. Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multiview images[J]. IEEE Int, 2007, 105-108.
- [14] Krishnamurthy R, Tao H. Compression and transmission of depth maps for image-based rendering[C]//Image Processing Proceedings, 2001.
- [15] Lee S W, Kim D. Quantization-based compact representation of 3D mesh[C]//MPEG-M15392, FR: Archamps, 2008.
- [16] Mamou K, Zaharia T. A low complexity approach for static 3D mesh compression[C]//MPEG-M15438 FR: Archamps, 2008.
- [17] Kim B J, Kim D Y. Fast 3D mesh compression[C]//MPEG-M15376, FR: Archamps, 2008.
- [18] Chai B B, Sehuraman S, Hatrack P. Mesh-based depth map compression and transmission for real-time view-based rendering[C]//Proc of International Conference on Image Processing, 2001.
- [19] Grewatsch S, Muller E. Fast mesh-based coding of depth map sequences for efficient 3D-video reproduction using OpenGL[C]//Proc of International Conference on Visualization, Imaging and Image Processing, 2005:66-71.
- [20] Zitnick L, Kang S B. High-quality video view interpolation using a layered representation[J]. ACM Trans Graph, 2004, 23(3):598-606.

Depth Map Compression Algorithm in 3D Video System

ZHU Bo¹, PENG Zong-ju^{1,2}, ZHANG Yun¹, SHAO Feng¹

(1.Faculty of Information Science and Technology, Ningbo University, Ningbo 315211, China;

2.Zhejiang Provincial Key Laboratory of Information Network Technology, Hangzhou 310027, China)

Abstract: This paper starts with analysis on the static depth map algorithms based on the Mesh, JPEG, mode function, followed by investigating the independent depth sequence coding and depth sequence coding based on motion vector in combination with color sequence. The experimental comparison result of the two depth video sequence coding algorithms is presented, and the paper concludes by discussing on the development trend of the depth map compression techniques.

Key words: three dimensional video system; depth map compression; video compression standards; 3D-mesh

CLC number: TP391.41

Document code: A

(责任编辑 章践立)