面向虚拟视点图像绘制的深度图编码算法*

朱 波[1], 蒋刚毅[1**], 张 云[1], 郁 梅[1,2]
(1. 宁波大学 信息科学与工程学院, 浙江 宁波 315211; 2. 南京大学 计算机软件新技术国家重点实验室, 江苏 南京 210093)

摘要: 针对不同区域对绘制虚拟视点图像质量产生不同的影响, 以及深度估计不准确导致时域抖动效应影响压缩效率的问题, 提出了一种面向虚拟视点图像绘制的深度图压缩算法。通过彩色图像帧差、深度图边缘提取等相关处理过程, 提取深度图的静态区域、边缘区域以及动态区域。对深度图边缘区域使用了较低的量化系数, 以提高深度图边缘区域编码质量; 根据深度图各个区域的编码模式特点, 仅对部分编码模式而不是所有模式进行率失真优化搜索, 以提高深度图的编码速度; 对于深度图 P 帧的静态区域, 合理地采用了 SKIP 模式, 以消除由于深度估计算法的局限性导致时域抖动效应对深度图压缩的影响。实验结果表明, 与传统的 H.264 编码方案相比, 本文方案在传输码流大小基本不变的前提下提高了最终虚拟视点图像边缘区域的绘制质量, 其余区域主观质量相近, 而深度图编码时间则节省了约 77 ~ 87%。

关键词: 自由视点视频; 深度图; 虚拟视点图像绘制
中图分类号: TN919.8 文献标识码: A 文章编号: 1005-0086(2010)05-0718-07

Virtual view rendering oriented depth map coding algorithm

ZHU Bo[1], JIANG Gang-yi[1**], ZHANG Yun[1], YU Mei[1,2]
(1. College of Information Science and Engineering, Ningbo University, Ningbo 315211, China; 2. State Key Lab. of Software New Technology, Nanjing University, Nanjing 210093, China)

Abstract: Different regions of depth map may make different impacts on synthesized virtual views, and temporal jitter at the background of depth map due to inaccuracy of depth estimation algorithm may decrease the coding efficiency of depth map. Based on the facts, a virtual view rendering-oriented depth map coding algorithm is proposed. Frame differences of color image, edge detection on depth map are firstly utilized to segment still regions, edge regions and motion regions in depth map. The smaller quantization parameter is assigned to edge regions to protect the depth accuracy of such regions. According to features of encoding modes at different regions, partial encoding modes instead of all modes will be traversed by rate-distortion optimization process to speed up the encoding time. For still regions, SKIP mode is used reasonably to eliminate the jitter effect. Experimental results show that compared with traditional H.264 codec, the proposed algorithm improves the quality of edge regions, preserves the quality of other regions of synthesized virtual views at almost the same bitrate, and saves up to 77% to 87% of encoding time.

Key words: free viewpoint-video; depth map; virtual view rendering

1 引 言

自由视点视频是一种先进的视觉媒体模式, 能够给观众提供身临其境的立体感知[1,2]。自由视点视频系统一般由多视点视频信号采集、编码、网络传输、解码、绘制和显示等组成, 其中虚拟视点图像绘制主要采用了基于深度图的绘制(DIBR, depth image based rendering)技术, 它通过场景深度信息、相机内外参数来计算用户需要观看的视点图像[3,4]。但随着多视点成像的相机数目增多, 不仅多视点彩色视频数据量成倍增加[5,6], 同时相应的深度图数据量也成倍增加[7]。因此, 需要研究有效的深度图压缩算法, 在保证最终观看虚拟视点视频质量的前提下, 有效缓解带宽压力、降低复杂度以提高编码压缩速度。为此, 人们提出了基于网格[7]、基于小波变换[8]以及基于分段线性函数[9]等方法对深度图进行压缩编码。但是这些算法在压缩效

①收稿日期: 2009-06-29
* 基金项目: 国家自然科学基金资助项目(60872094, 60832003); 国家"863"计划资助项目(2009AA01Z327); 教育部博士点基金资助项目(20081646003)
** E-mail: jianggangyi@126.com

率、可实现性上比不上基于视频编码标准(如 MPEG-X、H.26X)的深度图像压缩编码方法[10~12]。现有的深度图压缩算法大多着重于提升深度图编码的率失真性能[5,11],即在相同码率下提高编码深度图的信噪比。而深度图并非直接用于显示,它主要应用于虚拟视点图像绘制,所以深度图的压缩不仅要关注其率失真性能,还应该注重其对绘制虚拟视点图像质量的影响[8]。

深度图可以通过深度相机[13]和视差估计算法[14]获取,由于深度相机价格昂贵且深度检测距离较小,现阶段主要通过视差估计算法获取深度信息。而深度图序列和彩色视频之间的差异性使得直接采用彩色视频的编码方式对深度图序列进行编码压缩不尽合理,原因在于:其一,深度图不同区域对绘制图像质量产生的影响不同,其中以对象边缘对绘制的影响较大[3,6,15],因而需要对对象边缘加以保护;其二,深度图的平坦性以及与彩色图像的相关性决定了可以采用合理的算法设计以降低深度图的编码复杂度;其三,由于深度图获取算法局限性导致的时域抖动效应会降低深度图压缩效率,应尽量设计合理算法以降低其影响。

本文针对如何保护深度图边缘区域、降低深度图编码复杂度以及减小抖动效应对深度图压缩效率的影响等问题,提出了面向虚拟视点图像绘制的深度图压缩快速算法。在提高绘制虚拟视点图像主观质量和客观质量相近的前提下,新算法减少了 77~87%的编码时间。

2　面向虚拟视点图像绘制的深度图压缩算法

图 1 给出了彩色图像视频与深度图序列的联合编码框架,其中彩色图像视频压缩延用现有的视频编码方法,本文着重讨论如何有效地对深度图序列进行快速编码压缩。

2.1　深度图序列相关性分析

通过视差估计算法所获取的深度图序列背景区域可能存在着时域抖动效应。图 2 为"Breakdancers"序列[4]第 4 个相机成像视频的第 0、1 时刻的彩色图像及其对应的深度图。对比彩色图像视频以及深度图序列各自的帧差图,显然,(c)中的彩色图像帧差图反映了彩色图像视频有较好的前后帧背景一致性;而(f)所示的深度图帧差图反映出深度图序列存在背景区域深度不一致的现象,这意味着前后帧深度图背景部分的深度

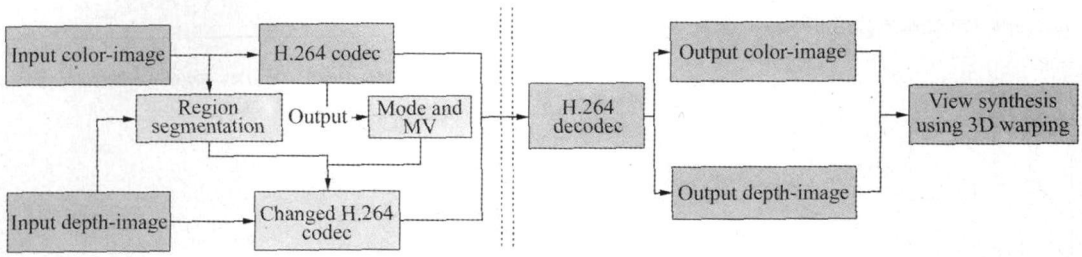


图 1　彩色图像视频与深度图序列的联合编码框架

Fig 1　Framework for united coding of color image and depth map sequences

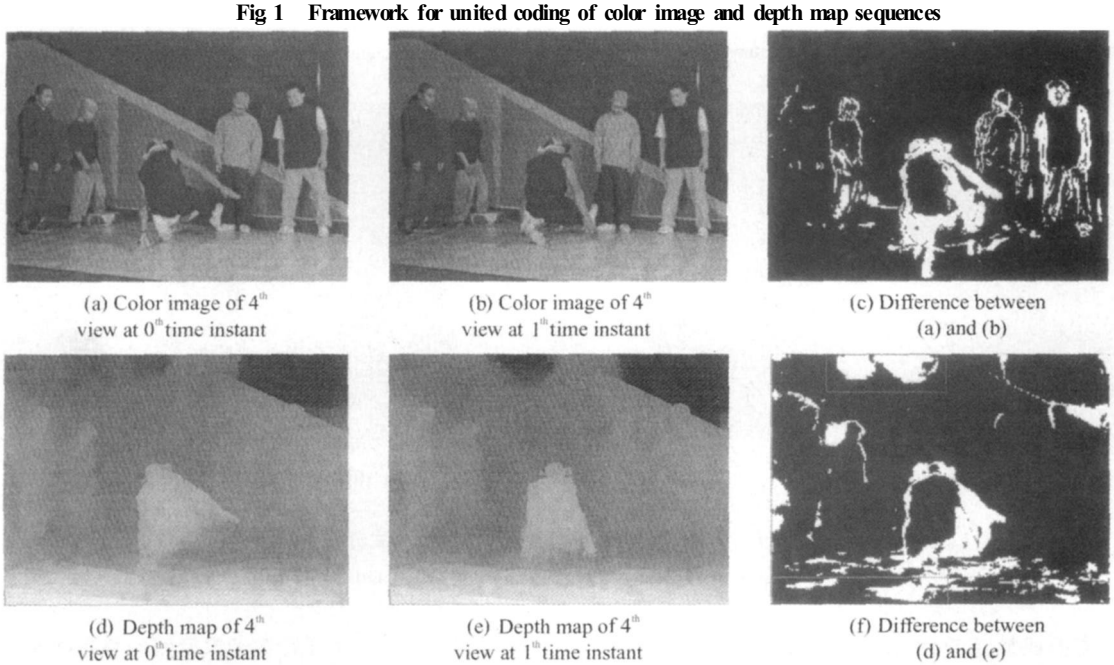


图 2　"Breakdancers"序列彩色图像和深度图像帧差结果比较

Fig. 2　Frame differences between successive frames of Breakdancers' color images and depth maps

估计并不准确，这是由前后帧深度图求取过程中的误差所导致的。进一步地，图 3 比较分析了彩色图像视频和深度图序列中部分背景区域的帧间相关系数 R，这里相关系数 R 定义为

$$R=\sum_{i=1}^{n}[1/(n-1)][(v_i-u_v)/S_v][(k_i-u_k)/S_k] \quad (1)$$

其中，S_v 和 S_k 分别为前后帧的标准差；u_v 和 u_k 为均值；v_i 和 k_i 则分别代表前后帧中第 i 个像素的像素值；n 为像素个数。图 3 中，“Ballet”序列[4]的彩色图像视频和深度图序列的相关系数均值分别为 0.995、0.979，而“Breakdancers”序列的彩色图像视频和深度图序列的相关系数均值分别 0.910、0.829。由此可见，由于深度估计算法的局限性，导致深度图背景存在较明显的不精确现象，进而使得深度图序列的时间相关性减弱。

2.2 深度图和彩色图像的编码模式分析

考虑到深度图序列和彩色图像视频有相似的相关性特征，图 4 进一步分析了采用 H.264 编码标准对深度图序列和彩色图像视频各自编码时的预测模式。图中红、黄、蓝与绿色宏块，粗边宏块分别表示了编码压缩中的 SKIP 模式、帧间子宏块模式、帧内模式与帧间宏块模式。观察图 4 可以发现，在静止区域（除地板等具有时间抖动效应的区域以外）彩色图像的编码模式和深度图的编码模式具有较强的相关性；在运动较为剧烈的对象边缘区域，如 Ballet 中的舞者右手右腿区域，彩色图像和深度图大多采用了帧内模式编码。但彩色图像的编码模式和运动矢量与深度图的编码模式和运动矢量又并非完全相同，例如在对象（人物）边缘的帧间模式上也并非完全一致，所以不能直接把彩色图像块编码预测模式用作为对应位置深度图块编码预测模式。

深度图与彩色图像的区别在于，深度图只在对象边缘处存在深度跳变，在其它区域较平坦。所以深度图各个区域的编码模式具有一定的规律性。图 4 结果表明，深度图的帧间模式主要分布在对象边缘区域以及一些出现时域抖动效应的深度区域，而对于背景以及对象内部主要采用了 SKIP 模式和帧内模式。统计深度图对象内部区域以及边缘区域模式，深度图对象内部区域采用帧内或者 SKIP 模式的概率较高。分析统计表明，在“Ballet”序列中，帧内或者 SKIP 模式占 97.6%，而在“Breakdancers”序列该比例为 94.8%。利用深度图编码的规律性，可以加快深度图编码速度。

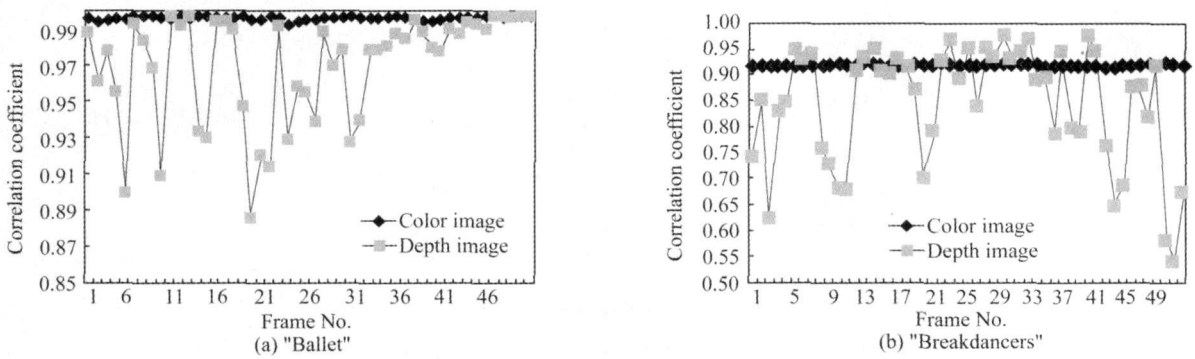


图 3 深度图和彩色图像背景区域帧间相关系数比较

Fig 3 Comparison of inter-frame correlations at backgrounds of color images and depth maps

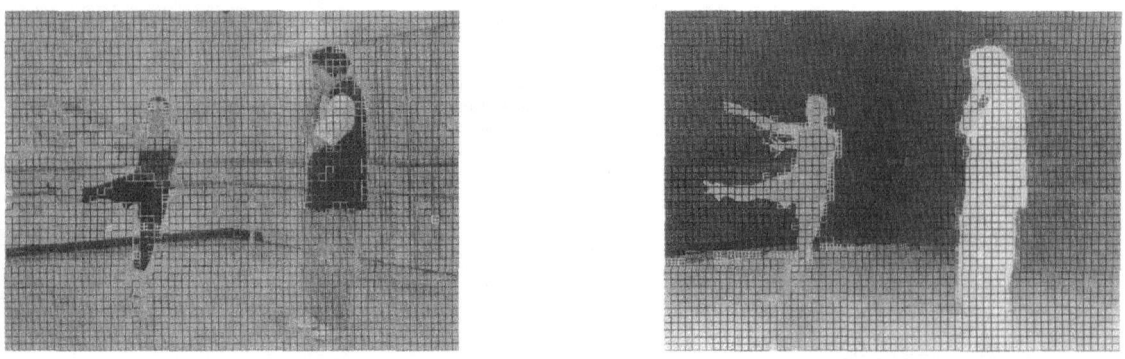


图 4 “Ballet”深度图和彩色图像的编码模式分析

Fig 4 Coding mode analysis of ballet's color image and depth map

2.3 深度图编码算法流程

通过对深度图序列和彩色图像视频的编码模式分析表明，在深度图中，帧间模式主要分布在对象边缘区域中，而在其余区域则以帧内预测和 SKIP 模式为主；通过相关性分析发现，时域抖动效应主要存在于深度图的静止区域。通过以上两点并考虑到深度图边缘对绘制的影响，可以将深度图分为静止区域、动态区域以及对象边缘区域，以达到保护深度图对象边缘，消除静止区域的抖动效应以及加快确定动态区域的

编码模式的目的。

首先利用深度图序列边缘检测、彩色图像视频求帧差的方法分别提取深度图的对象边缘区域 A 和运动区域 B, 进而区分运动对象内部区域 $C=B-(A\cap B)$ 以及区域 $D=I-(A\cup B)$。对对象边缘区域 A 通过调节 ΔQP(即对象边缘编码量化参数 QP 与其它区域编码 QP 之间的差值)进行编码以保证该区域深度值的准确性; 同时, 考虑到深度图边缘和彩色图边缘的模式相关性, 当彩色图像宏块采用帧内模式编码时, 深度图对应宏块也直接采用帧内模式。对于运动对象内部区域 C, 基于深度图编码模式的统计, 对该区域采取率失真遍历帧内模式和 SKIP 模式的方式。区域 D 包含了部分对象内部区域以及静止背景区域, 基于彩色图像中静止区域的深度值时域上保持不变的假设(固定相机情况下), 对于区域 D 可利用彩色图像的宏块编码模式和运动矢量进一步区分动态区域和静止区域, 对于静止区域直接采用 SKIP 模式以消除抖动效应对压缩效率的影响, 而对于动态区域则仅率失真遍历 SKIP 或帧内模式。

3　实验结果与分析

深度图序列与传统彩色图像视频不同的是其并非直接用于终端显示, 而是作为虚拟视点图像绘制的参数使用。因此, 评价一个深度图编码压缩算法的性能, 应该从最终绘制的虚拟视点的质量来评价[3,7]。将同一码率下最终绘制的虚拟视点图像的主、客观质量以及编码时间 3 方面进行比较。采用 H.264 标准参考平台 JM10.1 进行编码, 编码结构采用 IPPP 形式、YUV 4∶0∶0 格式和 CABAC 熵编码模式。

3.1　深度图编码时间和码流的对比

图 5 给出了原始 H.264 编码方案以及本文方案在量化参数 QP 为 20、24 和 28 时对深度图进行编码的结果。图中, PRO 为本文方案的结果, ORG 为原始 H.264 编码方案的结果。本文方案背景部分编码的 QP 与原始 H.264 方案相同, 而在对象边缘部分则采用了相对较低的 QP 以保证对象边缘深度值的准确性。这里, 背景和对象边缘所采用的不同的 QP 之间的差值记为 ΔQP。为合理比较本文方案与 H.264 编码深度图序列的效果, ΔQP 的取值在 6、4 和 2 中加以调整以使 2 个方案的编码码率相近。如图 6 所示, 在编码码率相近的情况下, 对于"Ballet"序列本文方案的编码速度提高了约 77%, "Breakdancers"序列则提高了约 87%。需要指出的是, 本文方案由于对背景区域直接采用了 SKIP 模式, 屏蔽了背景区域的时域抖动效应, 因而节省了背景部分编码码率。图 5 所示采用本文方案对"Ballet"深度图序列进行编码, 在 $QP=20$、$\Delta QP=6$ 时其码率为 1082 kbits/s, 比 $QP=20$ 的 H.264 方案码率减少 140 kbits/s。

3.2　虚拟视点图像绘制的主观质量比较

考虑到深度图主要用于自由视点视频系统用户端的虚拟视点图像绘制, 因而可以通过比较虚拟视点图像的主观质量来评价深度图压缩算法。

图 7~10 给出了采用原始未经压缩的深度图和分别经 H.264 和本文方案压缩后重建的深度图绘制的虚拟视点图像及其融合后的图像, 图 10 为融合后的图像的局部放大图。从图 10 中"Ballet"的局部区域放大图可以看出, 本文方案能够较好地保证深度图的边缘, 从而由其编解码得到深度图所绘制

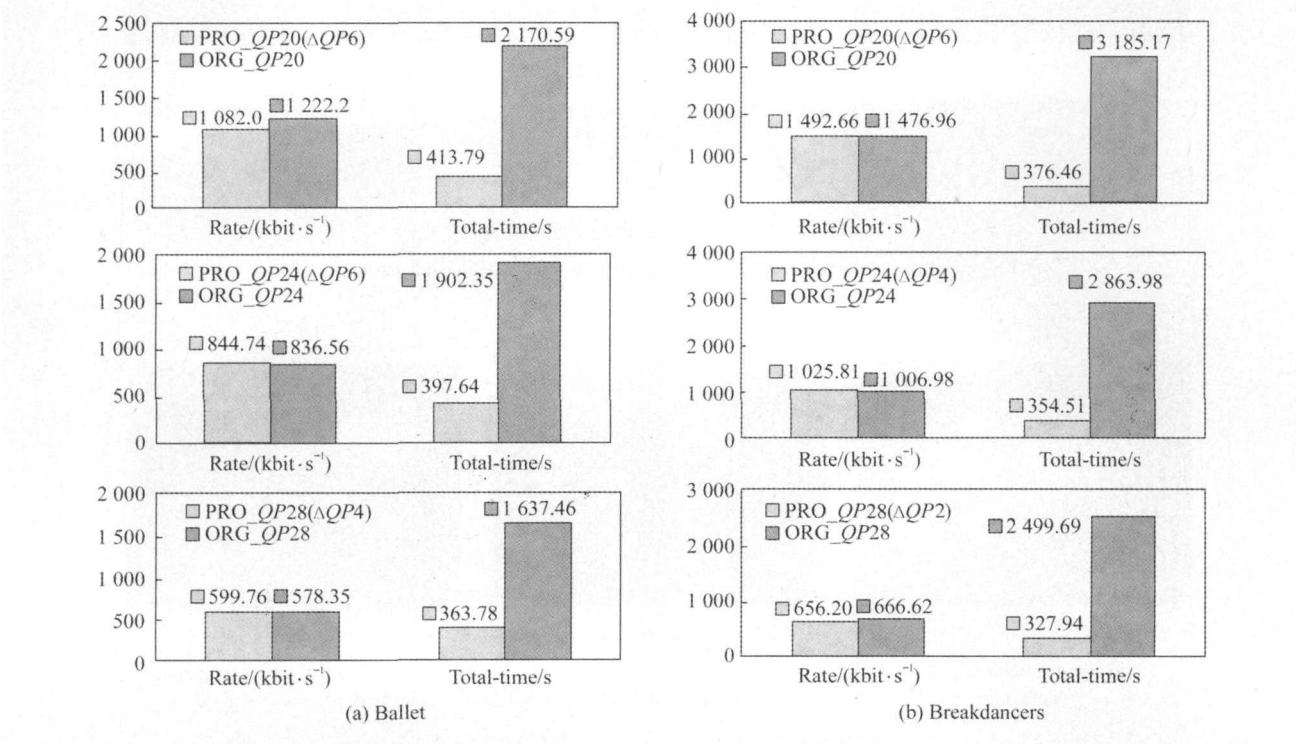
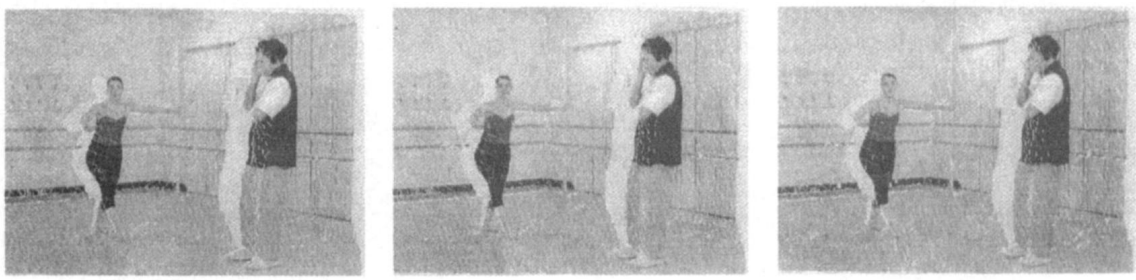


图 5　不同 QP 下的编码时间的比较

Fig. 5　Comparison of bit rate and coding time under different QP

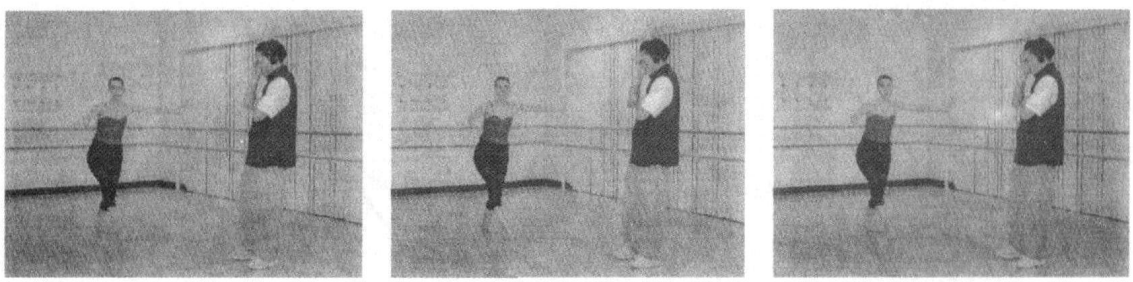
的虚拟视点图像对象边缘更接近于采用未压缩深度图绘制的图像的效果。“Breakdancer”序列其绘制的虚拟视点图像在主观质量上的差别不大。本文方案虽然略微增加了背景空洞，但融合算法能有效地填补空洞，所以背景部分空洞的增加并不影响最后虚拟视点图像的质量。



(a) Warping with original depth map　(b) Warping with map coded with H.264　(c) Warping with depth map coded with proposed method

图 6 “Ballet”的虚拟视点图像绘制比较

Fig. 6 Comparison of warping results of “Ballet” sequence



(a) Warping with original depth map　(b) Warping with map coded with H.264　(c) Warping with depth map coded with proposed method

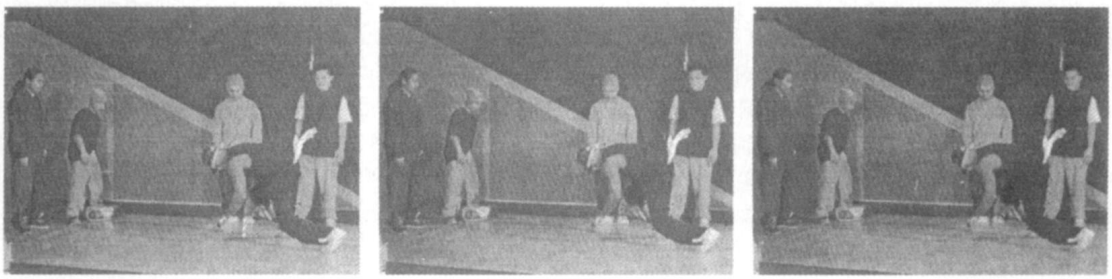
图 7 “Ballet”序列融合后的图像比较

Fig. 7 Comparison of fused virtual view of “Ballet” sequence

(a) Warping with original depth map　(b) Warping with map coded with H.264　(c) Warping with depth map coded with proposed method

图 8 “Breakdancers”的虚拟视点图像绘制比较

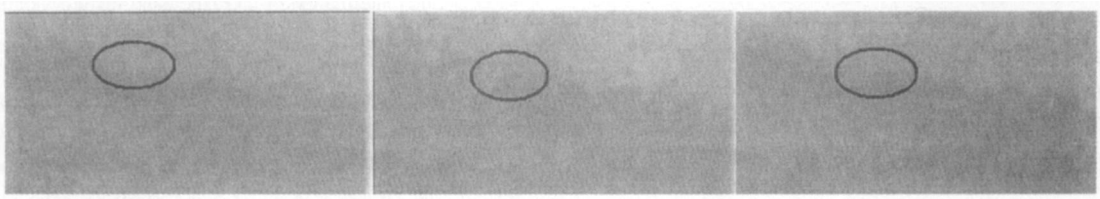
Fig 8 Comparison of warping results of “Breakdancers” sequence



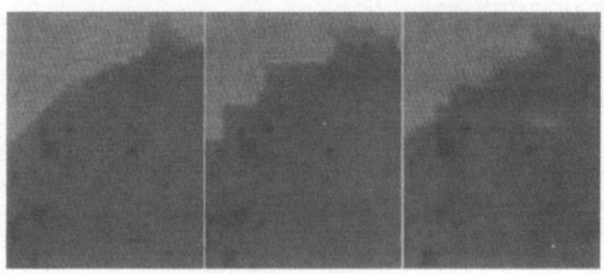
(a) Warping with original depth map　(b) Warping with map coded with H.264　(c) Warping with depth map coded with proposed method

图 9　“Breakdancers”融合后的图像

Fig 9　Comparison of fused virtual view of “Breakdancers” sequence



(a) Enlarged local of virtual view of Ballet



(b) Enlarged local areas of virtual view of Breakdancers

图 10　测试结果的局部区域放大

Fig 10　Enlarged local areas of virtual view

3.3　虚拟视点图像绘制的客观质量比较

影响虚拟视点图像质量的主要因素有两个: 深度图质量和绘制算法。为减小绘制算法对视点质量的影响, 本文仅计算经 3D-warping 后所得的虚拟视点图像质量, 其中空洞部分的 *PSNR* 不进行计算。表 1 为利用 H.264、本文方案压缩重建深度图绘制的虚拟视点图像客观质量的比较。表中 ΔHole 表示采用本文方案编码重建深度图进行绘制所得虚拟视点图像空洞数目增加的比例, *PSNR*_org、*PSNR*_pro 分别为采用 H.264 和本文方案压缩重建的深度图绘制所得虚拟视点图像的 *PSNR* 值; Δ*PSNR* 为二者的差值。本文方案略微增加了虚拟视点图像绘制过程中的空洞数目, 但对于虚拟视点图像的 *PSNR* 影响不大。

表 1　利用 H.264 与本文方案压缩重建深度图绘制的虚拟视点图像客观质量的比较

Tab.1　Objective quality comparison of virtual view with respect to H.264 and the proposed method

Sequence	*QP*	ΔHole/%	*PSNR*-org	*PSNR*-pro	Δ*PSNR*
Ballet	20	1.4	32.10	32.13	0.03
	24	1.2	32.00	32.06	0.06
	28	1.1	31.74	31.88	0.15
Breakdancers	24	7.4	31.65	31.64	−0.01
	28	6.6	31.55	31.57	0.02
	20	6.3	31.42	31.42	0.00

4　结　论

针对深度图中边缘部分对虚拟视点图像绘制影响大这一特点以及现有深度图在背景部分深度值时域上不一致的问题, 基于深度图不同区域的编码模式统计分析, 提出了面向自由视点图像绘制的深度图压缩快速算法。在深度图压缩时, 对深度图划分边缘区域、背景区域以及运动对象内部区域, 其中边缘区域采用小量化系数以达到保护边缘的目的, 其余部分则采用相对较大的量化系数以节省码率。对于包含了部分

对象内部区域以及静止背景区域的区域 D，进一步利用彩色图像的运动矢量和编码预测模式来区分区域 D 中的静态和动态区域，对其中的静态背景区域直接采用 SKIP 模式编码，以降低背景时域抖动效应对编码的影响，而对于对象内部区域则仅 RDO 遍历 SKIP 和帧内模式，以减少深度图编码复杂度，提高编码速度。实验结果表明，在略微提高绘制虚拟视点图像主观质量和客观质量相近的前提下，本文方案减少了约 77～87%的编码时间。后续的工作将进一步围绕如何提高区域分割的准确度，减少绘制虚拟视点图像的背景空洞展开；同时，引入视点间的相关性以及彩色图像和深度图像间的相关性，可进一步提高压缩效率和编码速度。

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[R]. ITU-T and ISO/IEC JTC1, JVT-W100, San Jose, USA. Apr, 2007.

[2] Tanimoto M. Overview of free viewpoint television[J]. Signal Processing: Image Communication, 2006, **21**(6): 454-461.

[3] Krishnamurthy R, Tao H, Chai B B, et al. Compression and transmission of depth maps for image-based rendering[A]. In: Proceedings of International Conference on Image Processing[C]. 2001, **3**: 828-831.

[4] Tanimoto M, Fujii T, Suzuki K, et al. Multi-view depth map of rena and akko & kayo[R]. ISO/IEC JTC1/SC29/WG11, MPEG-M14888, Shenzhen, China, Oct, 2007.

[5] SUN Zheng, YU Dao-yin. 3-D sequential reconstruction and motion estimation of coronary artery from angiograms[J]. Journal of Optoelectronics · Laser (光电子 · 激光), 2008, **19**(9): 1274-1279. (in Chinese)

[6] LIU Su-xing, AN Ping, MI Tao, et al. Virtual viewsynthesis and plane sweep based depth correction for free-view video[J]. Journal of Optoelectronics · Laser (光电子 · 激光), 2009, **20**(9): 1234-1237. (in Chinese)

[7] Kauff P, Atzpadin N, Fehn C, et al. Depth map creation and image based rendering for advanced 3-D TV Services Providing Interoperability and Scalability[J]. Signal Processing: Image Communication, 2007, **22**(2): 217-234.

[8] Daribo I, Tillier C, Pesquet-Popescu B. Adaptive wavelet coding of the depth map for stereoscopic view synthesis[A]. IEEE 10th Workshop on Multimedia Signal Processing[C]. 2008. 413-417.

[9] Morvan Y, Farin D, et al. Depth-image compression based on an R-D optimized quadtree decomposition for the transmission of multi-view images[J]. IEEE Transactions on Consumer Electronics, 2007: 105-108.

[10] Merkle P, Smolic A, Mueller K, et al. MVC: Experiments on coding of multi-view video plus depth. JVT-X064, Geneva, CH, 2007.

[11] 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, **4319**: 898-907.

[12] Merkle P, Smolic A, Mueller K, et al. Multi-View Video Plus Depth Representation and Coding[A]. IEEE International Conference on Image Processing[C]. 2007, I: 201-204.

[13] Iddan G J, Yahav G. 3D Imaging in the Studio and Elsewhere . . . , '[A]. SPIE[C]. 3D Shape Measurements'01, pp. 48-55, (San Jose, CA, USA), Jan. 2001.

[14] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International Journal of Computer Vision, 2002, 47(1/2/3): 7-42.

[15] Zitnick L, Kang S B, Uyttendaele M, et al. High-quality video view interpolation using a layered representation[J]. ACM Trans. Graphics, 2004, **23**(3): 598-606.

作者简介：

朱 波 (1984—)，宁波大学通信与信息系统硕士研究生，主要研究方向为多媒体通信与压缩.