

一种新的深度视频处理方法

毕 伟, 彭宗举*

(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

摘要: 深度视频的不连续性会影响其编码效率以及虚拟视点的绘制质量. 为增强深度视频的连续性, 提出了一种新的基于时空变换的深度视频处理方法: 首先检测出深度视频中的运动对象; 其次将彩色视频和对应的深度视频进行时空变换, 对运动区域和非运动区域在时间方向上进行不同强度的平滑处理; 最后对平滑过的深度视频进行重建. 实验结果表明, 提出的处理方法在保证虚拟视点绘制质量的同时, 能够节省 14.89%~47.48%的码率.

关键词: 深度视频; 运动对象; 时空变换; 深度视频处理

中图分类号: TN919.81

文献标志码: A

文章编号: 1001-5132 (2013) 02-0033-05

随着采集和显示技术的发展, 人们已经不满足于二维视频所提供的简单视觉信息, 希望从更加全面, 更加立体的角度观赏视频. 自由视点视频 (Free Viewpoint Video, FVV) 是当前先进视觉媒体模式. FVV 系统能让人们通过自由选择视点欣赏场景体验到身临其境的感觉^[1]. 目前国际标准化组织 (Joint Video Team, JVT) 制定的标准草案中采用多视点彩色视频和深度视频 (Multiview Video Plus Depth, MVD) 来表示 3D 场景和实现 FVV 系统^[2]. 在基于 MVD 的 FVV 系统中, 深度视频信号一般由服务器端传送给用户端, 进而降低用户端接受设备的生产成本.

由相机获取或者深度估计软件估计的深度视频不是很准确, 大大降低了深度视频时间、空间、视点间相关性, 进而影响了其压缩效率和绘制的质量^[3]. 为了提高深度视频的压缩效率, 需要在编码前对深度视频进行预处理操作. Silva 等^[4]提出了基于 JNDD 模型的方法, 去掉一些由于深度估计产生的视觉噪声以及不被观看者感知的深度细节. Zhao 等^[5]提出了一个深度无合成误差模型 (Depth No-Synthesis-Error, D-NOSE) 来求出在不引起虚拟

视点几何偏移的条件下, 深度可接受失真范围. Liu 等^[6]通过下采样/上采样的多视点深度编码方案来提高压缩效率, 然而该方法需要在解码端对深度视频进行上采样, 对解码端设备的处理能力具有较高的要求. 上述这些方法都没有从时域上对深度视频进行研究和处理. 因此, 笔者提出了一种基于时空变换的深度视频处理方法, 增强深度视频时域上的相关性. 处理后的深度视频在保证虚拟视点绘制质量的同时能够大大地降低编码的码率.

1 深度视频的特性分析

MVD 这种三维场景的表示方式能够更好地支持立体视频, 具有重建质量高, 灵活性好以及兼容性好等优点, 逐渐成为 3DTV 和 FVV 等多媒体系统中主流数据表示格式. 其中深度视频代表的是对应的彩色视频场景到摄像机成像屏幕的距离信息. 深度视频可以由深度相机获取, 也可以由深度估计软件获取. 由于深度获取技术的局限性, 得到的深度视频不是很准确. 例如, 同一彩色视频中的静止区域和背景区域应当是固定不变的, 而在对应的深度视频中, 这些区域的深度值却是变化的.

收稿日期: 2012-12-22.

宁波大学学报 (理工版) 网址: <http://nbdxlg.paperopen.com/>

基金项目: 国家自然科学基金 (61171163); 宁波市自然科学基金 (2012A610041).

第一作者: 毕 伟 (1987-), 男, 安徽休宁人, 在读硕士研究生, 主要研究方向: 三维视频编码与处理. E-mail: biwei1987@163.com

*通信作者: 彭宗举 (1973-), 男, 四川南部人, 博士/副教授, 主要研究方向: 多媒体技术、三维视频信号处理与编码等.

E-mail: pengzongju@nbu.edu.cn

图1给出了Leave Laptop的彩色视频和深度视频的帧差图。其中图1(a)为Leave Laptop彩色视频的第9视点的第30帧和第31帧的时域帧差图,图1(b)为Leave Laptop深度视频的第9视点的第30帧和第31帧的时域帧差图;白色区域表示时域上发生变化的点。比较图1(a)和(b)中2个方框可知,彩色视频的背景部分几乎没有变化,而深度视频的背景部分发生了很大的变化。这就意味着通过深度相机拍摄或者深度估计软件获取的深度视频不太准确,即在时间方向上存在不一致性。



(a) Leave Laptop 彩色视频时域帧差图



(b) Leave Laptop 深度视频时域帧差图

图1 彩色和对应深度视频的时域帧差图

目前,主流的多视点视频编码结构都采用了以时间方向参考为主的预测方式,如分层B帧(Hierarchical B Pictures, HBP),利用这类编码结构对深度视频进行编码,深度视频时间方向上的不一致性会影响深度视频的编码效率。因此,在编码前对深度视频进行处理以提高其在时间上的一致性是很有必要的。

2 深度视频预处理总体框架

人眼视觉系统(HVS)的研究表明,人眼对不同图像区域的感觉是不同的,人眼更关注于深度不连续区域以及场景变换比较大的区域,也就是运动对象区域。为了保证虚拟视点的绘制质量,应重点保护运动对象区域。诸如以上讨论,我们提出—

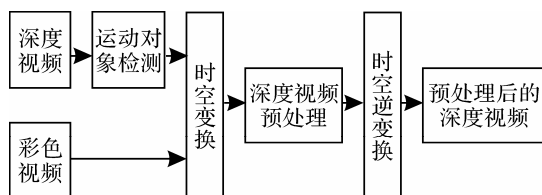


图2 深度视频处理总体框架

种基于时空变换的深度视频处理方法,该算法的流程如图2所示。

2.1 运动对象检测

深度视频中的像素值代表彩色视频场景到摄像机成像平面之间的距离。在深度图中,背景和对像内部的深度信息是逐渐变化的^[7],而在对象边缘部分深度信息是跳变的,这就有利于运动对象的分割。笔者采用文献[8]中的对象提取方法。图3为Dog深度视频第39视点第61帧的运动对象检测结果。



图3 Dog 深度视频运动对象检测结果

运动对象检测主要分以下3个步骤:

(1) 分别在水平和垂直方向上对深度视频进行时空变换。

(2) 对变换后的图像序列利用动态阈值进行二值化,然后将二值化过后的图像序列进行水平和垂直方向上的重建,取对应的2幅重建图像的共同部分,从而得到一幅二值掩膜图。

(3) 对掩膜图像进行滤波、去除小面积区域等后处理,得到最终的运动对象区域。

2.2 提出的深度视频预处理方法

2.2.1 对深度视频和彩色视频进行时空变换

对于深度视频 $D(k, m, n)$, 其中 k 、 m 和 n 分别表示帧数、水平分辨率和垂直分辨率。依次取出 $D(k, m, n)$ 中每一帧中的第 i ($i=1, \dots, n$) 行的 m 个像素,按序组合成新的一帧,分辨率为 $m \times k$ 。这样,深

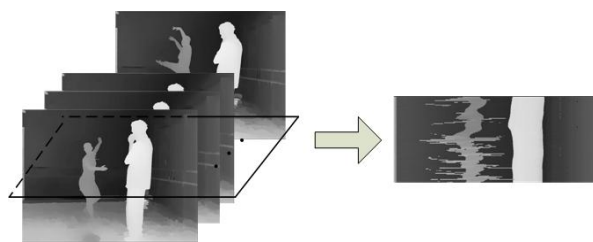


图4 深度视频时空变换效果图

度视频 $D(k, m, n)$ 就转换成为 n 帧, 分辨率为 $m \times k$ 的深度视频 $D'(n, m, k)$. 彩色视频 $C(k, m, n)$ 采用同样方法进行变换, 变换后的彩色视频为 $C'(n, m, k)$. 图4为深度视频时空变换效果图.

2.2.2 对变换后的深度视频进行时间方向上的平滑处理

对变换过后的深度视频 $D'(n, m, k)$ 中每一帧图像的每一列的 k 个像素进行分类, 分类的依据是当前待分类像素的像素值和前面相邻一类所有像素值的均值进行比较, 通过设定阈值来判定.

$$|c(i, j) - c_{\text{aver}}(i, j-1)| < T_1, \quad (1)$$

$$|d(i, j) - d_{\text{aver}}(i, j-1)| < T_2, \quad (2)$$

式中: $c(i, j)$ 和 $d(i, j)$ 分别表示变换后的彩色视频 $C'(n, m, k)$ 和对应的深度视频 $D'(n, m, k)$ 一帧图像中第 i ($i=1, 2, \dots, m$) 列的第 j ($j=1, 2, \dots, k$) 位置处的像素值. 而 $c_{\text{aver}}(i, j-1)$ 和 $d_{\text{aver}}(i, j-1)$ 分别表示当前待分类的彩色和深度像素前面相邻一类所有像素的像素值的均值. 彩色视频的分类是用来对深度视频的分类进行优化. 如果当前深度像素值和对应的彩色像素值和前一类中的像素值相比, 分别满足公式(1)和(2), 则说明当前深度像素与前面相邻一类像素属于同一类; 否则, 即当前像素值为下一类的第1个像素. 最后将每一类中的所有元素

取均值处理. 得到深度视频 $D''(n, m, k)$. 经研究分析, 彩色视频的阈值 T_1 设为5, 而深度视频中, 如果当前像素属于运动区域, 阈值 T_2 设置为10, 如果当前像素属于非运动区域, 则阈值 T_2 设置为20. 由于深度视频是用来辅助虚拟视点绘制, 而人眼对运动区域更加敏感, 为保证最终虚拟视点的绘制质量, 运动区域的阈值比非运动区域的阈值小.

2.2.3 对平滑处理后的深度视频进行逆变换

依次取出处理后的深度视频 $D''(n, m, k)$ 中每一帧的同一行的 k 个像素, 按序组合成新的分辨率为 $m \times n$ 的一帧, 这样就重建出最终的深度视频 $D'''(k, m, n)$.

3 实验结果及分析

为了测试上述深度视频处理算法, 笔者在 JMVC 8.3 校验平台上进行测试, 选取由德国 HHI 公司提供的“Book Arrival”, “Door Flowers”, “Alt Moabit”和“Leave Laptop”序列, 选择8、9、11视点, GOP设为15, 编码61帧. 利用深度估计软件 (Depth Estimation Reference Software, DERS) 估计出这些序列的深度视频作为原始深度视频. 通过上述深度视频预处理算法得到预处理后的深度视频. 笔者从深度视频的压缩效率和虚拟视点的绘

表1 原始深度视频和处理后深度视频的编码码率比较

深度序列	QP	BR _{Or} /kbps	BR _{Worrall} /kbps	Saving of Worrall's/%	BR _{Pro} /kbps	Saving of Worrall's/%
Door Flowers	22	872.98	622.24	28.72	537.51	38.43
	27	385.47	338.89	12.09	252.44	34.52
	32	161.51	160.70	0.50	114.84	28.90
	37	67.39	69.60	-3.28	53.87	20.06
Book Arrival	22	1 029.93	731.85	28.94	726.79	29.43
	27	476.37	409.24	14.09	343.07	27.98
	32	210.29	202.37	3.77	161.10	23.39
	37	95.91	96.06	-0.16	80.82	15.73
Leave Laptop	22	777.44	538.84	30.69	434.63	44.09
	27	337.38	295.70	12.35	213.87	36.61
	32	143.66	143.77	-0.08	103.07	28.25
	37	62.28	64.38	-3.37	50.90	18.27
Alt Moabit	22	834.67	455.67	45.41	438.33	47.48
	27	426.59	296.01	30.61	257.52	39.63
	32	217.49	180.63	16.95	154.96	28.75
	37	110.59	101.21	8.48	94.12	14.89

制来衡量所提出的深度处理算法的性能。

在深度视频压缩效率方面,笔者所提出的预处理算法能节省 14.89%~47.48%的码率。表 1 列出了各个序列分别采用 Worrall 方法^[4]和本文提出的方法处理后得到的深度序列相对于原始深度序列的 4 个不同的量化参数(Quantization Parameters, QP)(22, 27, 32, 37)条件下编码码率的节省情况。各个序列的时间连续性和运动的剧烈程度不一样,本文提出的处理算法对码率的减少程度也不同,处理后,增强了深度视频时间上的一致性。

Worrall 提出的方法平滑了深度视频的背景区域和对象区域,同时也增强了对象的边界区域。图 5 给出了“Book Arrival”序列处理结果的边界局部效果。从图 5 可看出, Worrall 方法增强了深度视频的对象边界;所以,当量化参数 QP 设为 22 和 27 时,编码的码率可以大大降低;但是,当 QP 设为 32 和 37 时,大的量化步长会抵消深度视频滤波的平滑效果,而边界区域的增强则会消耗更多的码率。因此,在 QP 大时,边界区域的增强有可能会导致编码码率的上升。

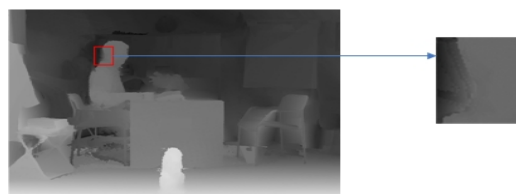
在虚拟视点绘制方面,笔者提出的深度视频处理算法对虚拟视点绘制质量没有太大影响。实验中使用的绘制软件(VSRS 3.5)绘制虚拟视点,利用第 8 视点和第 11 视点绘制第 9 视点。表 2 和表 3 分别为虚拟视点的绘制质量峰值信噪比和结构相似度的比较, $PSNR_{Ori}$, $PSNR_{Worrall}$ 和 $PSNR_{Pro}$ 分别为原始深度视频、Worrall 方法和本文方法的峰值信噪比, $SSIM_{Ori}$, $SSIM_{Worrall}$ 和 $SSIM_{Pro}$ 分别为原始深



(a) 原始深度视频 Book Arrival S_9T_0



(b) Worrall 方法处理后的深度视频 Book Arrival S_9T_0



(c) 本文提出方法处理后的深度视频 Book Arrival S_9T_0

图 5 Book Arrival 序列处理结果的局部边界对比度视频、Worrall 方法和本文方法的结构相似度。

从表 2 和表 3 可以看出,本文提出的方法得到的深度视频与原始深度视频以及 Worrall 方法得到的深度视频绘制的虚拟视点的质量相比,基本上能够保持。结合上面的实验结果,可以认为本文提出的处理方法在保证虚拟视点绘制质量的前提下,能够大大地降低深度视频编码的码率。

表 2 虚拟视点绘制质量 PSNR 的比较

深度序列	$PSNR_{Ori}$	$PSNR_{Worrall}$	$PSNR_{Worrall} - PSNR_{Ori}$	$PSNR_{Pro}$	$PSNR_{Pro} - PSNR_{Ori}$
Door Flowers	35.85	35.59	-0.26	35.92	+0.07
Book Arrival	37.38	37.25	-0.13	37.41	+0.03
Leave Laptop	37.62	37.19	-0.43	37.61	-0.01
Alt Moabit	36.49	36.50	+0.01	36.49	+0.00

表 3 绘制质量 SSIM 的比较

深度序列	$SSIM_{Ori}$	$SSIM_{Worrall}$	$SSIM_{Worrall} - SSIM_{Ori}$	$SSIM_{Pro}$	$SSIM_{Pro} - SSIM_{Ori}$
Door Flowers	0.9618	0.9610	-0.0008	0.9619	+0.0001
Book Arrival	0.9611	0.9612	+0.0001	0.9611	+0.0000
Leave Laptop	0.9620	0.9611	-0.0009	0.9622	+0.0002
Alt Moabit	0.9727	0.9729	+0.0002	0.9728	+0.0001

4 结论

在基于MVD的FVV系统中,深度视频需要传输至三维电视终端用以辅助生成任意视点的虚拟视点图像。由于深度获取技术的限制,得到的深度视频不是很准确,这将会影响深度视频的编码效率以及虚拟视点的绘制质量。针对深度视频的特性,笔者提出一种深度视频处理方法,以增强时间上的相关性。实验结果表明,该处理方法在保证虚拟视点绘制质量的同时,能节省14.89%~47.48%的码率。

参考文献:

- [1] Tanimoto M. FTV (Free-viewpoint TV)/[C]. Bonnie L. 17th IEEE International Conference on Image Processing (ICIP). Hongkong: IEEE Press, 2010:2393-2396.
- [2] ISO/IEC, JTC1/SC29/WG11 and ITU-T SG16 Q 6 JVT-W100, Multi-view video plus depth (MVD) format for advanced 3D video systems[S].
- [3] Rana P K, Flierl M. Depth consistency testing for improved view interpolation/[C]. Stéphane P. IEEE International Workshop on Multimedia Signal Processing (MMSP). Saint-Malo: IEEE Press, 2010:384-389.
- [4] Silva D, Ekmekcioglu E, Worrall S T, et al. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling[J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(2):335-351.
- [5] Zhao Y, Zhu C, Chen Z Z, et al. Depth no-synthesis-error model for view synthesis in 3-d video[J]. IEEE Transactions on Image Processing, 2011, 20(8):2221-2228.
- [6] Liu Q, Yang Y, Ji R, et al. Cross-view down/up-sampling method for multiview depth video coding[J]. IEEE Signal Processing Letters, 2012, 19(5):295-298.
- [7] ISO/IEC JTC1/SC29/WG11 M15090, Improvement of depth map estimation and view synthesis[S].
- [8] Lu X M, Yu M, Zhang Y, et al. Motion detection based on temporal-to-spatial conversion of depth maps for multi-view video/[C]. Zhang Y F. The International Conference on Wireless Communications and Signal Processing. Nanjing: IEEE Press, 2009:13-15.

A Novel Depth Video Processing Algorithm

BI Wei, PENG Zong-ju*

(Faculty of Information Science and Technology, Ningbo University, Ningbo 315211, China)

Abstract: Inconsistency of the depth video deteriorates the quality of rendered virtual view and coding performance. To improve the consistency problem, a novel depth processing algorithm based on temporal-horizontal transposing is proposed in this paper. Firstly, the motioning objects are detected. Then, the depth video and the corresponding color video are transposed towards temporal and horizontal directions. Finally, in the transposed domain, the depth values in motion regions and non-motion regions are smoothened by different intensity. Experimental results show that the proposed algorithm can reduce the bit rate ranging 14.89%-47.48% without compromising the virtual view rendering quality.

Key words: depth video; motion object; temporal-horizontal transposing; depth processing

(责任编辑 史小丽)