

基于 DIBR 和图像融合中的任意视点绘制

刘占伟 安 平 刘苏醒 张兆扬

(上海大学通信与信息工程学院, 上海 200072) (新型显示技术及应用集成教育部重点实验室, 上海 200072)

摘 要 虚拟视点生成是3D视频会议等应用领域中的关键技术,为了快速高质量地进行任意视点绘制,提出了一种基于深度图像绘制(DIBR)和图像融合的新视点生成方法,该方法首先对参考图像进行预处理,包括深度图像的边缘滤波和参考图像规正,以减少目标图像中产生的较大空洞和虚假边缘;然后利用3D图像变换生成新视点图像,并用遮挡兼容算法对遮挡进行快速处理;接着再对两幅目标图像进行融合得到新视点图像;最后用插值法填充剩余的较小空洞。实验证明,该新方法能获得令人满意的绘制效果。

关键词 基于图像的绘制 基于深度图像的绘制 3D图像变换 遮挡兼容算法

中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2007)10-1696-05

Arbitrary View Rendering Based on DIBR and Image Mergence

LIU Zhan-wei, AN Ping, LIU Su-xing, ZHANG Zhao-yang

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

(Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai 200072)

Abstract It is a key technique to generate a virtual view in 3D Video conference. In this paper, a new algorithm based on DIBR and image mergence is proposed. Firstly the depth images are smoothed by an edge dependent filter and the reference images are normalized so as to reduce the big holes and the artificial edges in desired image. Then an arbitrary virtual view is generated by making use of 3D Image Warping. In order to deal with the visibility problem quickly, the pixels in reference image are processed in an occlusion-compatible order. Then two generated images are merged to eliminate the big holes in the desired view. Finally the remaining holes are filled with an interpolation method. Experimental results show that the proposed method can obtain satisfactory rendering results.

Keywords image based rendering, depth image based rendering, 3D image warping, occlusion-compatible order

1 引 言

基于图像的绘制(image based rendering, IBR)是近十年来新兴的一门学科,它是从一系列已知的参考图像出发生成新视点图像。与传统的绘制技术相比,它具有不需要复杂的3D场景建模、绘制速度快、既能用于合成场景,又能用于真实场景、对计算机要求不高等优点。但是单纯利用已知图像生成新视点图像,其所需处理的图像数据是非常庞大的,例

如光场绘制技术^[1]。基于深度图像的绘制(depth image based rendering, DIBR)是一种利用彩色图像及其对应的深度图像生成新视点图像的方法,其核心是3D image warping 方程。DIBR 由于将场景的深度信息引入到 IBR 中,从而大大减少了参考图像的数目,是 IBR 中比较有发展前途的一门技术,其在虚拟现实、军事指挥、建筑业、影视特效、娱乐、医学图像处理等方面有着广泛的应用前景。

利用 DIBR 技术生成新视点图像,国内外已有许多学者对此进行了研究。McMillan 最早提出了

基金项目: 国家自然科学基金项目(60672052 和 60572127);上海市重点学科项目(T0102);上海市曙光计划资助项目(06SG43)

收稿日期: 2007-05-30; **改回日期**: 2007-06-15

第一作者简介: 刘占伟(1983 -),男。2004 年获聊城大学学士学位,现为上海大学通信与信息工程学院硕士研究生。主要研究方向为基于图像的绘制。E-mail: weizhanliu@163.com

3D image warping 方程^[2],其原理是根据参考图像中与各像素点对应的深度信息,将参考图像上的像素点映射到目标图像上,但对于 DIBR 中由于遮挡关系发生变化而产生的空洞问题没有提出一个很好的解决办法。Mark 等人提出的 post-rendering warping^[3]则使用多幅参考图像来解决空洞问题,但其只是对计算机合成的图像进行处理,没有涉及到真实图像,而且当参考图像数目增多时,计算量随之增加。Shade 等人提出层次深度图像法(layered depth image, LDI)^[4]中,其一个 LDI 不仅存储了参考图像中可见的像素信息,还存储了那些不可见的像素信息,而且将那些不可见的像素用来填充,会由于深度不连续而引起空洞,且该方法复杂度高。Kumar 提出了一种 depth + texture 表示方法^[5],即利用多个视点的深度图像和纹理图像来绘制复杂场景,但是其在处理遮挡问题时,仍是采用传统的 z-buffer 算法,这不仅需要比较深度值,使绘制速度慢,而且该方法对图像融合后产生的虚假边缘也没有提出解决办法。Fehn 提出了一种特殊的 DIBR 算法^[6],用来生成虚拟立体视点,但没有涉及到生成任意视点的情况。

与以上这些算法相比,本文提出的新视点生成算法的特色为:(1)对目标摄像机的位置没有限制,适用于生成真实场景的任意新视点的图像;(2)利用 McMillan 的遮挡兼容算法处理遮挡问题,由于不需要比较深度值,故速度快;(3)采用预处理深度图像和图像融合相结合的方法填充空洞,取得了较好的效果;(4)通过对参考图像进行归正处理,减少了图像融合后产生的虚假边缘。

2 DIBR 技术简介

DIBR 是利用参考图像及其对应的每个像素的深度信息来合成场景中虚拟视点的过程^[7],该过程可以分为以下两步:(1)利用深度信息,将参考图像上所有的像素点重投影到它们对应的 3 维空间中的位置;(2)这些 3 维空间点再被投影到目标图像平面。这个从 2 维到 3 维的重投影以及从 3 维再到 2 维的投影称为 3 维图像变换(3D image warping)。

3 基于 DIBR 的任意视点绘制算法

图 1 为本文基于 DIBR 任意视点生成的算法框

图,包括基于深度的边缘滤波器、图像归正、3 维图像变换、图像融合、空洞填充 5 个部分。由于实验主要生成的目标视点和参考视点之间的距离比较远(约 20cm),其产生的空洞将会比较大,因此可首先利用边缘滤波器预处理深度图像,以减少目标图像中可能出现的较大的空洞,同时对参考图像进行归正,以减少图像融合后产生的虚假边缘,接着通过 3D image warping 生成新视点图像;然后利用两幅图像融合的方法进一步减少空洞;最后通过邻域像素插值填补剩余的较小的空洞。

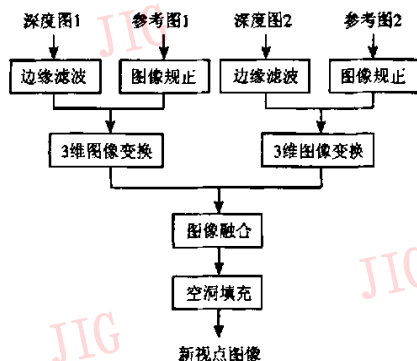


图 1 基于 DIBR 新视点生成算法框图

Fig. 1 Framework of the new view creation based on DIBR

3.1 预处理深度图像

一幅深度图像是一个以图像坐标为自变量,以对应的 3 维空间点的深度值为函数值的 2 维函数。深度图像中深度值有较大变化的区域(主要发生在对象的边缘部分)在 3D image warping 之后可能会产生较大空洞,通过平滑深度图像能避免在目标图像中出现较大的空洞或者使之变小。为了消除生成的立体图像中存在的较大空洞,文献[6]简单地用高斯滤波器平滑深度图像,文献[8]则提出用不平衡滤波器平滑深度图像。但是由于它们都是对整个深度图像进行平滑,因此在去除空洞的同时,降低了图像的质量。文献[9]提出利用基于深度的边缘滤波器平滑深度图像,由于这个滤波器只平滑深度图像中可能会产生空洞的区域(对象的水平边缘部分),从而取得了较好的平滑效果。

本文对文献[9]方法进行了改进,使之适用于生成任意视点的情况,即平滑深度图像中深度值有较大变化的所有区域(所有对象的边缘部分,而不仅仅是水平边缘部分)。为了减少虚像,在平滑深度图的时候,应当保持前景物体的深度值不变,因为

人眼对前景对象的变化比对背景的变化要敏感。

3.2 图像归正

由于光照、图像噪声等因素的影响,致使两幅图像的对应点可能有不同的颜色值。设两幅图像对应点的灰度值 I_1 和 I_2 之间的关系为

$$I_2(x, y) = AI_1(x, y) + B \quad (1)$$

其中, A 和 B 为系数,文献[10]提出了一种通过计算分块累积直方图来归正灰度图像的方法,在此基础上,作者对两幅参考图像进行了归正,以减少图像融合后产生的虚假边缘。

3.3 3 维图像变换(3D image warping)

3D image warping 是利用深度图来将参考图像中的像素点映射到目标图像中。在世界坐标系和第 1 个摄像机坐标系重合的假设下,文献[6]给出了一个适用于真实场景的 3D image warping 方程,即

$$\hat{Z}\hat{m} = \hat{Z}\hat{A}RA^{-1}m + \hat{A}T \quad (2)$$

m 和 $\hat{m}$ 分别是参考图像和目标图像中的像素坐标, Z 和 $\hat{Z}$ 分别是 3 维空间点在第 1 个和第 2 个摄像机坐标系下的深度值, 3×3 矩阵 A 和 $\hat{A}$ 是指第 1 个和第 2 个摄像机的内参矩阵, 3×3 矩阵 R 和 3×1 矩阵 T 是指第 2 个摄像机的旋转矩阵和平移矩阵。

这里将式(2)进行扩展,使之不受上面假设的约束。为了方便运算,本文定义 2 个 4×4 的矩阵 P 和 $\hat{P}$,它们分别代表参考相机和目标相机的投影矩阵,参考相机投影矩阵 P 可以由下式计算得到:

$$P = AC = \begin{bmatrix} A_{11} & A_{12} & A_{13} & 0 \\ A_{21} & A_{22} & A_{23} & 0 \\ A_{31} & A_{32} & A_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

用 4×4 的矩阵 C 将 3 维点从世界坐标系变换到摄像机坐标系, C 包括 R 和 T ,式(4)和式(5)给出了 3 维空间点和图像像素坐标之间的变换关系:

$$\begin{pmatrix} Z_1^c \times x_1 \\ Z_1^c \times y_1 \\ Z_1^c \\ 1 \end{pmatrix} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (4)$$

$$\begin{pmatrix} Z_2^c \times x_2 \\ Z_2^c \times y_2 \\ Z_2^c \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{P}_{00} & \hat{P}_{01} & \hat{P}_{02} & \hat{P}_{03} \\ \hat{P}_{10} & \hat{P}_{11} & \hat{P}_{12} & \hat{P}_{13} \\ \hat{P}_{20} & \hat{P}_{21} & \hat{P}_{22} & \hat{P}_{23} \\ \hat{P}_{30} & \hat{P}_{31} & \hat{P}_{32} & \hat{P}_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (5)$$

(X, Y, Z) 是 3 维空间点的世界坐标, Z_1^c 和 Z_2^c 分别表示 (X, Y, Z) 在第 1 个和第 2 个摄像机坐标系

中的 Z 坐标(上角 C 代表 camera)。因此,对于参考图像中每个像素坐标 (x_1, y_1) ,如果已知其深度值 Z ,则利用式(4)就能计算出其相应的 3 维空间点坐标 (X, Y, Z) ,之后利用式(5)就能得到新视点图像中相应的像素坐标 (x_2, y_2) ;最后将新视点图像中坐标为 (x_2, y_2) 的像素的颜色值与参考图像中坐标为 (x_1, y_1) 的像素的颜色值一致,这样即可以生成带空洞的新视点图像。

3.4 可见性问题

当用 3D image warping 方程生成新视点图像时,有可能出现参考图像中的几个像素点映射到目标图像中的同一个像素点的情况,被称为可见性问题或者遮挡问题。在计算机图形学中通常是用传统的 z -buffer 算法处理遮挡问题,由于该算法需要比较深度值,因此绘制速度比较慢。1995 年,McMillan 提出了一种遮挡兼容算法^[11],该算法首先计算目标摄像机的投影中心在参考图像平面上的投影点(也叫极点);然后根据极点极性的不同,按照不同的顺序处理参考图像中的像素,即若为负极点,则按照远离极点的方向处理参考图像中的像素,若为正极点,则反之(如图 2 所示,箭头表示处理的方向)。由于这种顺序能保证对应的 3 维空间点可按照由后向前的顺序投影到目标图像中,因此较近的点自动覆盖较远的点,从而形成正确的遮挡关系。与 z -buffer 算法相比,该算法具有以下两个优点:(1)由于不需要比较深度值,因此处理速度快;(2)由于只依赖于目标相机与参考相机的相对位置,与场景信息无关,即可以按照水平方向和垂直方向分别处理参考图像中的像素,因此处理简单。

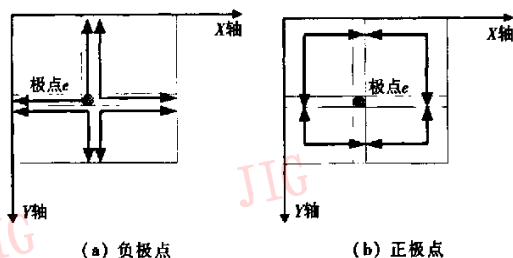


图 2 参考图像上像素的不同处理方向

Fig. 2 The different processing order of pixels in reference image

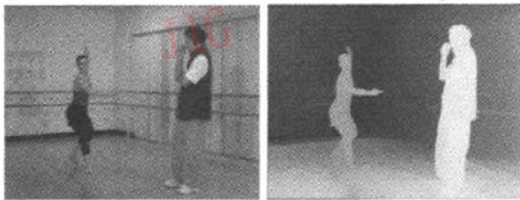
3.5 图像融合

如何填充目标图像中的空洞是 DIBR 中的一个重要问题,由于实验中要生成的目标视点和参考视

点之间的距离比较远,会使目标图像中的空洞比较大,因此仅由一幅参考图像所提供的信息是不充分的。本文方法先使用两幅参考图像分别生成目标图像,然后再将这两幅目标图像融合成最终的新视点图像。由于之前已对参考图像进行了归正处理,以减少由图像融合而产生的虚假边缘,因此生成的图像质量比文献[5]要好(文献[5]中生成的图像中的地板有明显的裂痕)。

4 实验结果

为验证本文方法的效果,采用微软研究院(Microsoft research, MSR)的交互视觉媒体组(interactive visual media group)提供的多视点视频测试序列进行了实验,实验用视频包括由 8 个摄像机捕获的 100 帧彩色图像序列和由立体对应计算得到的深度图^[12],两个摄像机之间的水平间距是 20cm,并且成 1 维弧形排列^[13],MSR 同时提供了摄像机标定参数。图 3 表示了由摄像机 5 捕获的“Ballet”序列和“Breakdancers”序列的第 1 帧参考图像及其对应的深度图像。



(a) 测试序列“Ballet”及其深度图



(b) 测试序列“Breakdancers”及其深度图

图 3 测试序列及其深度图

Fig. 3 Test sequences and their depth maps

为了验证本文提出的方法的正确性,在 PC 机上用 C 语言编写程序模拟第 5 个相机生成的图像。图 4 为只用一幅参考图像(如摄像机 4 或摄像机 6)生成的目标图像(摄像机 5),其中图 4(a)、图 4(b)、图 4(e)和图 4(f)是深度图没有平滑时所生成

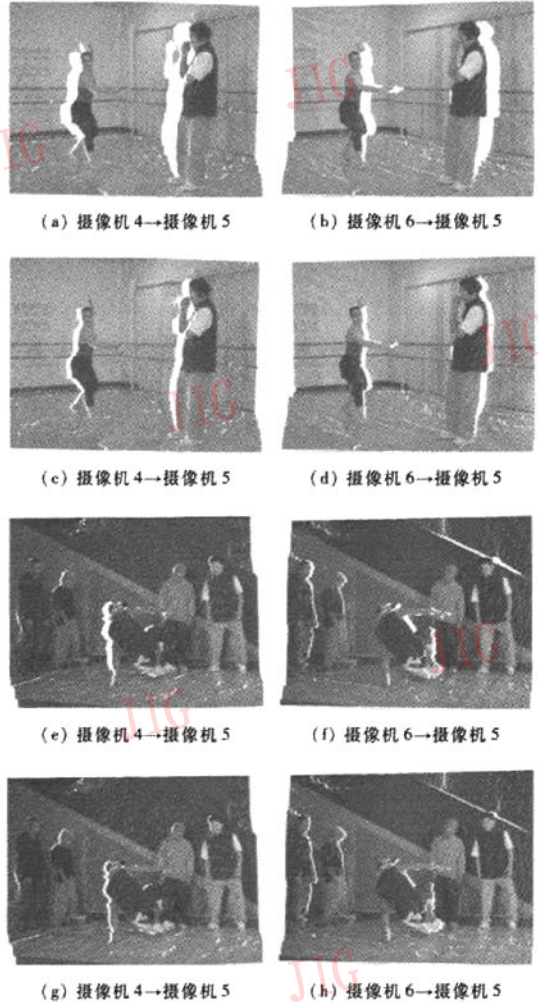


图 4 融合之前测试序列的绘制结果

Fig. 4 Rendering results of the sequences before image mergence

的目标图像,图 4(c)、图 4(d)、图 4(g)和图 4(h)表示用边缘滤波器平滑深度图后所生成的目标图像。由图 4 可见,对深度图像平滑后,目标图像中大空洞的范围大为减小。图 5 为两幅目标图像融合以后的



图 5 两幅目标图像融合以后的结果

Fig. 5 Results after merging the two target images

结果,与摄像机 5 捕获的真实图像(图 3)相比,其所生成的目标图像的质量好、真实感强。

为了利用遮挡兼容算法处理遮挡问题,还编程计算了目标摄像机在参考图像平面上的投影点 $e = (x, y, w)$,其像素坐标如表 1 所示, $w < 0$ 表示极点都是负极点,因此处理顺序都是背对着极点 e 的方向。

表 1 目标图像的投影中心在参考图像上的投影

Tab.1 Projection on the reference images of the projective centre

测试序列	Ballet		Breakdancers	
目标视点	摄像机 5	摄像机 5	摄像机 5	摄像机 5
参考视点	摄像机 4	摄像机 6	摄像机 4	摄像机 6
目标摄像机在参考图像平面上的投影点 $e = (x, y, w)$	$e_1 =$ 33 394.6 -926.9 -0.2	$e_2 =$ -8 155.4 237.8 -0.8	$e_3 =$ 17 176.7 335.0 -0.4	$e_4 =$ -5 396.4 247.4 -1.2

5 结 论

本文利用 DIBR 技术,由两幅参考图像生成了一幅任意新视点图像,并利用 McMillan 的遮挡兼容算法快速地处理了遮挡问题。为了填充目标图像中的空洞,首先利用一个边缘滤波器来平滑深度图像中深度值变化比较大的区域,然后进行 3D image warping,最后对两幅目标图像进行融合,结果表明,融合后的视点图像质量良好、真实感强。

然而在基于 DIBR 的自由视点绘制中还有很多工作要做,比如,虽然在 3D image warping 之前已对两幅参考图像进行了归正,已减少了由于图像融合而产生的虚假边缘,但还不能完全消除虚假边缘,如 Ballet 序列中女演员的左边区域。下一步打算构建一个精确的像素混合函数,并对图像规正方法做进一步优化,以消除目标图像中的边缘假象。

参考文献 (References)

- Levoy M, Hanrahan P. Light field rendering[A]. In: Proceedings of ACM SIGGRAPH'96[C], New Orleans, Louisiana, USA, 1996: 31~42.
- McMillan L. An Image-Based Approach to Three-Dimensional Computer Graphics[D]. University of North Carolina, Chapel Hill, NC, USA, 1997.
- Mark W R. Post-rendering 3D Image Warping: Visibility, Reconstruction, and Performance for Depth-Image Warping[D]. University of North Carolina, Chapel Hill, NC, USA, 1999.
- Shade J, Gortler S, He L W, et al. Layered depth images[A]. In: Proceedings of ACM SIGGRAPH'98[C], Orlando, FL, USA, 1998: 231~242.
- Narayanan P J, Sashi Kumar P, Sireesh Reddy K. Depth + texture representation for image based rendering[A]. In: Proceedings of Fourth Indian Conference on Computer Vision, Graphics and Image Processing(ICVGIP)[C], Kolkata, Indian, 2004: 113~118.
- Fehn C. Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV[A]. In: Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI[C], San Jose, CA, USA, 2004: 5291: 93~104.
- Fehn C. A 3D-TV approach using depth-image-based rendering (DIBR)[A]. In: Proceedings of Visualization, Imaging, and Image Processing'03[C], Benalmádena, Spain, 2003: 482~487.
- Zhang L, Tam J, Wang D. Stereoscopic image generation based on depth images[A]. In: Proceedings of IEEE Conference on Image Processing[C], Singapore, 2004: 2993~2996.
- Chen Wan-yu, Chang Yu-lin, Lin Shyh-feng, et al. Efficient depth image based rendering with edge dependent depth filter and interpolation[A]. In: Proceedings of IEEE International Conference, Multimedia and Expo(ICME)[C], Amsterdam, Netherland, 2005: 1314~1317.
- Cox Ingmar J, Hingorani Sunita L, Rao Satish B, et al. A maximum likelihood stereo algorithm[J]. Computer Vision and Image Understanding, 1996, 63(3): 542~567.
- McMillan L. A List-Priority Rendering Algorithm for Redisplaying Projected Surfaces[R]. UNC Technical Report TR95-005, University of North Carolina, Department of Computer Science, Chapel Hill, NC, USA, 1995.
- Zitnick C L, Kang S B, Uyttendaele M, et al. High-quality video view interpolation using a layered representation[A]. In: Proceedings of ACM SIGGRAPH[C], New York, USA, 2004: 600~608.
- Yoon Seung-Uk, Lee Eun-Kyung, Kim Sung-Yeol, et al. A framework for multi-view video coding using layered depth images[A]. In: Proceedings of 6th Pacific-Rim Conference on Multimedia[C], Jeju Island, Korea, 2005: 431~442.

1 Levoy M, Hanrahan P. Light field rendering[A]. In: Proceedings of