

基于多幅深度图像的三维重构技术

章 勤 魏 伟

(华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

摘要: 提出基于多幅深度图像的三维重构方法. 采用 6 幅固定视点采样图像, 分别代表三维空间中立方体包围盒的 6 个面; 通过对这些二维的采样集合进行三维信号还原, 生成一个空间采样样本, 可以在视点任意改变的情况下, 输出近似的重构图像. 该方法由于采样集合固定, 经过采样集筛选, 背景剔除筛选和棱台视锥体裁减后, 可以极大减少待变换的样本集合, 采用 z -buffer 算法解决图像折叠问题. 当视点在采样点附近任意变换时, 可以重构出比较满意的结果, 而且该方法与场景复杂度无关. 在纯软件模式下在普通 PC 机上达到 20 帧/s 的速度, 在通过硬件加速情况下可以达到 30 帧/s.

关键词: 多幅深度图像; 三维重构; 立方体固定采样; 筛选; 场景复杂度

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1671-4512(2007)12-0088-04

Depth images-based approach for 3D reconstruction

Zhang Qing Guo Wei

(College of Computer Science and Technology, Huazhong University of
Science and Technology, Wuhan 430074, China)

Abstract: Depth images-based approach for 3D reconstruction which employs six pictures with the fixed viewer is put forward in this essay. They stand for the faces of the cube model respectively. For converting 2D sample to 3D signal, the image of any viewing can be reconstructed. Through the fixed sampling of scene, it can get 20fps in a common pc after a series of filtration by software and can get up to 30fps by hardware. The z -buffer is used to solve the image fold. The better outcome can be get when the viewing is not very difference to the sample viewing. And it is independent to the complexity of scene.

Key words: depth images; 3D reconstruction; fixed sampling; filtration; complexity of scene

目前的多幅图像渲染(IBR)算法大多数都是基于单幅深度图像, 仅考虑单幅图像的前提下提高精度, 提高采样率, 改进反走样算法^[1~4]. 实践表明: 如果要想取得逼真的输出视觉效果, 那么新视角的改变不能超过 5° , 否则会产生明显视觉错误. 上述问题的存在, 极大限制了基于单幅深度图像的 IBR 算法的应用范围.

另外, 目前的基于多幅图像的渲染, 仍然主要采用的是基于 3D warp 方法, 它需要在参考图像之间互相变换, 以解决像素间的遮挡关系, 增加了

很多额外的计算量. 虽然输出图像视觉效果较单幅有很大提高, 但是它导致绘制速度较慢^[5~9], 难以满足用户实时交互和浏览的要求.

本文基于多幅图像的三维重构技术, 在视点任意改变的情况下, 以较快的速度, 绘制出较令人满意的三维重构结果. 基于多幅图像的三维重构方法, 只需对渲染的场景进行一次采样, 便可多次重构任意视点的输出图像. 并可不依赖于硬件, 以纯软件方式进行渲染, 在普通 PC 机上能达到 20 帧/s. 若结合硬件对它进行加速, 则速度将会有大

收稿日期: 2006-09-11.

作者简介: 章 勤(1955-), 女, 教授; 武汉, 华中科技大学计算机科学与技术学院 (430074).

E-mail: qzhang@hust.edu.cn

基金项目: 国家高技术研究发展计划资助项目 (2006AA02Z347).

的提高.

1 图像空间到投影空间的变换

三维重构方法使用的参考图像序列, 是基于三维长方体模型的 6 个表面的采样图片以及这些采样图片对应的深度图片. 在采样图像序列预处理中, 要将这些参考图像变换成可用的空间三维采样点集合, 采用了比较常规的反投影方法来实现. 由于这个预处理过程只需要一次计算, 计算结果可以反复利用, 因此不用考虑预处理过程的计算量.

定义 3 个空间: 图像空间, 投影空间和世界坐标系空间. 其中投影空间是图像空间到世界坐标系空间的中介.

要完成图像空间到投影空间的变换, 首先要将图像空间变换到视窗空间. 假设图像空间为一个 $n_{\text{cols}} \times n_{\text{rows}}$ 的像素阵列, 视窗空间为 $[-w, w] \times [-h, h]$ 的窗口, 得出线性映射 (如图 1 所示).

$$\begin{cases} u = -w + 2w \cdot c / n_{\text{cols}}; \\ v = -h + 2h \cdot r / n_{\text{rows}}; \end{cases}$$

式中: $r = 0, 1, \dots, n_{\text{rows}} - 1$; $c = 0, 1, \dots, n_{\text{cols}} - 1$. 图 1 中 (c, r) 为图像空间中的任意一点, (u, v) 为其在视窗空间中的对应点.

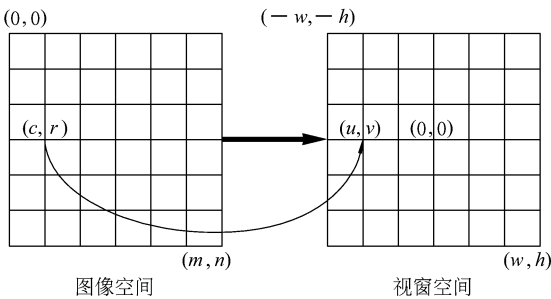


图 1 图像空间和视窗空间的对应关系

其次, 要将视窗空间变换到投影空间. 要完成此变换, 需讨论一般的照相机透视模型. 假设照相机的焦距为 f , 近平面和远平面距离分别为 d_{near} 和 d_{far} . 认为投影平面在焦平面上.

在得到这个投影变换之前, 对于归一化深度还要做一定变换. 归一化深度规定近平面的深度为 0, 远平面的深度为 1. 对于任意归一化的深度 d , 作如下变换将其还原为其真实深度,

$$D = (az + b) / -z, \tag{1}$$

式中: $a = -(d_{\text{far}} + d_{\text{near}}) / (d_{\text{far}} - d_{\text{near}})$; $b = -2 \cdot d_{\text{far}} d_{\text{near}} / (d_{\text{far}} - d_{\text{near}})$; $D = 2d - 1$.

对式 (1) 变形可得对应归一化深度 d 的真实深度值

$$z = \frac{-2d_{\text{far}} d_{\text{near}}}{[(2d - 1)(d_{\text{far}} - d_{\text{near}}) - (d_{\text{far}} + d_{\text{near}})]}. \tag{2}$$

在照相机透视模型下, 假设视窗平面为距离原点为 f 的焦距处, 通过调节照相机的水平视角 θ_{fov} 以及可见窗口的长宽比 α_a , 可以得到此视窗平面的长宽. 将 θ_{fov} 带入公式可得图像空间到焦平面的映射方程

$$\begin{cases} u = -f \tan(\theta_{\text{fov}}/2) + 2cf \tan(\theta_{\text{fov}}/2) / n_{\text{cols}}; \\ v = -(f / \alpha_a) \tan(\theta_{\text{fov}}/2) + (f / \alpha_a) \cdot \tan(\theta_{\text{fov}}/2) 2r / n_{\text{rows}}. \end{cases} \tag{3}$$

对于图像空间的任意一点 (c, r) , 假设这点的深度为 d_1 , 它对应焦平面上一点 $p(u, v)$, 综合式 (2) 和 (3) 可得其在投影空间点 P 的坐标

$$\begin{cases} x = |z| u / |f|; \\ y = |z| v / |f|; \\ z = \frac{-2d_{\text{far}} d_{\text{near}}}{[(2d_1 - 1)(d_{\text{far}} - d_{\text{near}}) - (d_{\text{far}} + d_{\text{near}})]}. \end{cases} \tag{4}$$

2 从投影空间到世界空间的变换

投影空间是以视点为原点, 以及视见方向为坐标轴的坐标体系, 可以通过照相机相关输入参数获得其在水世界坐标系中的位置. 假设视见坐标系原点位置为 P_{eye} , 其坐标系基向量组为 $[U, V, N]$, 如图 2 所示.

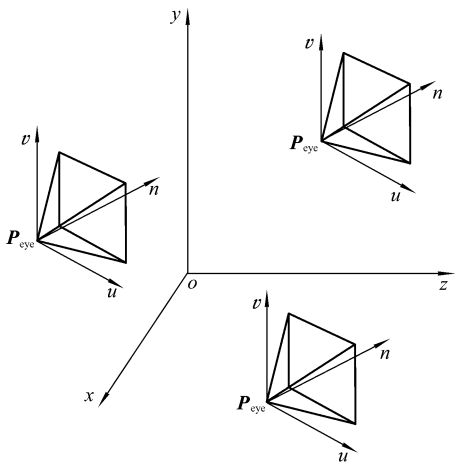


图 2 投影空间和世界坐标系

通过向量几何学, 投影空间到世界坐标系空间的变换方程为

$$P_0 = P_{\text{eye}} + nN + uU + vV. \tag{5}$$

投影空间到图像空间的变换, 其实就是从投影空间到世界空间坐标系的变换的一个逆过程,

也就是正投影. 要实现正投影, 可以将式(5)写成

$$[U, V, N][u, v, n]^T = P_0 - P_{eye}, \quad (6)$$

由式(6)得

$$[u, v, n]^T = [U, V, N]^{-1}(P - P_{eye}) = [U, V, N]^T(P - P_{eye}).$$

由式(4)可得到对应图像空间中的 (c, r) 坐标

$$\begin{cases} c = (u + f \tan(\theta_{fov}/2))n_{cdis}/ \\ [2f \tan(\theta_{fov}/2)]; \\ r = [v + (f/\alpha_a) \tan(\theta_{fov}/2)]n_{cols}/ \\ [2(f/\alpha_a) \tan(\theta_{fov}/2)]. \end{cases} \quad (7)$$

最后将式(6)变形获得的 (u, v) 带入式(7), 即可获得最终的 (c, r) 坐标值, 完成投影变换.

3 采样点集合的选取及像素遮挡

采用多幅源图像输出新的图像, 只要源图像的采样足够多, 理论上可以重构出新的任意视图的较逼真的视图. 但是由于采样点集合较大, 若不进行筛选, 则会带来实现速度慢的问题. 因此对采样点集合进行合理的筛选与选取是必须的. 眼睛位置相对于场景位置的改变, 导致了一些在源采样图像中不在同一个位置的像素, 而在新生成的图像中被投影到了同一个位置, 因此找出那个离新视点最近的点是解决像素遮挡的关键.

采用长方体与视见截头体相交的方法来选取像素集. 长方体的 6 个面, 分别是 6 个方向的采样集合(上下左右前后), 这也是整个采样场景的几何直观表示. 首先判断眼睛与这 6 个采样平面的位置关系, 再选取合适的采样集合. 先给立方体的各个面编号: 前侧面 0, 右侧面 1, 后侧面 2, 左侧面 3, 下侧面 4, 上侧面 5, 则整体的采样集合 $S = \{I_0, I_1, I_2, I_3, I_4, I_5\}$. 设待选取的集合为 C , 计算 S 中的每个侧面 I_k 与眼睛的位置关系, 若眼睛在它的正侧, 则将 I_k 加入 C ; 否则丢掉.

获得采样点集合以后, 对于这所有的采样点, 再进行一次截头体筛选, 只有位于截头体内部的采样点才渲染. 具体判断过程: a. 计算截头体 6 个平面的方程(如果没有近裁剪平面, 就只有 5 个面需计算); b. 计算采样点与这个 6 个平面的关系, 若采样点在所有这 6 个平面的负侧, 则采样点位于截头体内部, 渲染此采样点; 否则放弃.

文献[6]提出了一种独特解决像素遮挡问题的算法, 但是只适用于单幅图像的 IBR 计算, 不适用于多幅图像 IBR 计算. z -buffer 算法不用考虑采样点绘制的顺序问题, 可以以任意顺序渲染, 仅通过深度比较来决定最终的结果, 而且目前几

乎所有的显卡都支持 z -buffer 算法, 所以也方便使用硬件来加速渲染. 本文采用 z -buffer 算法来解决像素的遮挡问题. 由于只讨论纯软件的实现方式, 因此没有将采样点的 z 值变换到常规的 $[0, 1]$ 范围内, 仅保存了 z 值的绝对值. 这样不仅可以避免误差, 还可以减少计算量.

4 反走样和孔洞填补

由于物体表面部分存在对不同视点张角的差异, 导致在该表面的采样分辨率的差异. 同一表面在不同图像中的投影面积差异越大, 它们对该表面的采样分辨率相差越大; 反之相近. 区域太大, 容易产生马赛克(走样); 区域太小, 又会出现非常多的黑洞(孔洞). 因此必须按照一种合理的方法来计算投影区域的大小. 由 3D warp 公式得知, 输出图像是输入图像的有理连续的可导函数, 所以这里通过计算输出图像相对于输入图像的变化率, 来计算最终投影区域的大小.

对于投影区域的计算将要建立一个输入图像空间到输出图像空间的一个映射, 通过计算输出图像在 u, v 方向的变化率来决定输出投影区域的大小. 由式(5)和(6)可得到输入图像到世界坐标系的方程

$$P_0 = P_{eye_1} + M_1[x_1, y_1, z_1]^T = P_{eye_1} + |z_1|/|f| M_1[u_1, v_1, f]^T.$$

而从 P 到输出图像空间的变换方程为

$$[u_2, v_2, n_2]^T = M_2^{-1}(P_0 - P_{eye_2}) = M_2^{-1}(P_{eye_1} - P_{eye_2}) + |z_1|/(|f| |M_2^{-1} M_1[u_1, v_1, f]^T). \quad (8)$$

设 $M_2^{-1} M_1 = [w_j] \ (i, j = 1, 2, 3)$. 由式(8)可得

$$\begin{aligned} \partial u_2 / \partial u_1 &= (|z_1|/|f|)(w_{11}); \\ \partial u_2 / \partial v_1 &= (|z_1|/|f|)(w_{12}); \\ \partial v_2 / \partial u_1 &= (|z_1|/|f|)(w_{21}); \\ \partial v_2 / \partial v_1 &= (|z_1|/|f|)(w_{22}), \end{aligned}$$

式中: M_2 为输出视点坐标系的基向量构成的矩阵; M_1 为输入视点坐标系的基向量构成的矩阵. 设 $|z_1|/|f| = \delta$ 计算雅可比行列式的值得得投影区域的面积

$$S = \begin{vmatrix} \partial u_2 / \partial u_1 & \partial u_2 / \partial v_1 \\ \partial v_2 / \partial u_1 & \partial v_2 / \partial v_1 \end{vmatrix} = \delta(w_{11}w_{22} - w_{12}w_{21}).$$

得到投影区域的面积以后, 将投影区域看作一个正方形, 其中正方形的边长 $l = \delta \sqrt{|w_{11}w_{22} - w_{12}w_{21}|}$. 用 Splat 方法^[3]将投影区

域看作一个圆形.通过计算投影区域,解决了图像反走样以及孔洞问题.

参 考 文 献

[1] David K, Marc L. Protecting 3D graphics content [J]. 2005 Communications of the ACM, 2005, 48(6): 234-239.

[2] Mark D, Garrett M, Jiangtao K, et al. Image appearance modeling and high dynamic range image rendering[C] //2004 Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization. Los Angeles: ACM SIGGRAPH Press, 2004: 90-100.

[3] Brian C M. Reflection space image based rendering [C] //Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1999: 121-129.

[4] Shum H, Kang S. A review of image-based rendering techniques[C] //Proceedings of IEEE/ SPIE Visual

Communications and Image Processing. San Jose: VCIP, 2000: 2~13.

[5] Leonard M. An image-based approach to three-dimensional computer graphics[D]. Carolina: Department of Computer Science, University of North Carolina, 1997.

[6] Mcmillan L, Bishop G. Plenoptic modeling[C] //An Image-Based Rendering System. Proceedings of SIGGRAPH 95. Los Angeles: SIGGRAPH Press, 1995: 899-903.

[7] Zhang Z. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1 330-1 334.

[8] Maxime L, Long Q. Image-based rendering by joint view triangulation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(11): 278-282.

[9] 李新肖. 基于采样的场景建模及其绘制技术研究 [D]. 西安: 西安交通大学计算机学院, 1999.

(上接第 87 页)

参 考 文 献

[1] Wilkin D R, Fleurat-Lessard F. The detection of insects in grain using conventional samplings spears[C] //Proc 5th Int Wkg Conf on Stored-Product Protection. Bordeaux: [s. n.], 1990: 1 445-1 450.

[2] 徐 舫, 邱道尹. 图像识别在粮仓害虫检测方面的应用研究[J]. 郑州工程学院学报, 2001, 22(1): 78-81.

[3] Mallat S, Zhong S. Characterization of signals from multiscale edges[J]. IEEE Trans on PAMI, 1992, 14(7): 710-732.

[4] Vidakovic B, Lozoya C B. On time-dependent wavelet

denoising[J]. IEEE Trans Signal Processing, 1998, 46(9): 2 549-2 551.

[5] 周 龙, 牟 恽. 二维小波变换在粮虫图像处理中的应用[J], 华中科技大学学报: 自然科学版, 2007, 35(8): 24-26.

[6] 周 龙. 小波分析在粮虫图像去噪及边缘检测中的应用[J], 华中科技大学学报: 自然科学版, 2005, 33(5): 52-54.

[7] 戴 君. 数学形态学在图像处理中的应用[J]. 佛山科学技术学院: 自然科学版, 1998. 16: 50-55

[8] 沈庭芝, 方子文. 数字图像处理与模式识别[M]. 北京: 北京理工大学出版社, 1998.

[9] 胡小峰. Visual C++/MATLAB 图像处理与识别实用案例精选[M]. 北京: 人民邮电出版社, 2004.