

文章编号:1002-2082(2014)05-0811-06

基于 Kinect 传感器多深度图像融合的 物体三维重建

郭连朋,陈向宁,刘彬,刘田间

(中国人民解放军装备学院,北京 101400)

摘要:物体的三维重建技术一直是计算机视觉领域研究的热点问题,提出一种利用 Kinect 传感器获取的深度图像实现多幅深度图像融合完成物体三维重建的方法。在图像空间中对深度图像进行三角化,然后在尺度空间中融合所有三角化的深度图像构建分层有向距离场(hierarchical signed distance field),对距离场中所有的体素应用整体 Delaunay 三角剖分算法产生一个涵盖所有体素的凸包,并利用 Marching Tetrahedra 算法构造等值面,完成物体表面重建。实验结果表明,该方法利用 Kinect 传感器采集的不同方向 37 帧分辨率 为 640×480 的深度图像完成目标物体的三维重建,仅需要 48 s,并且得到非常精细的重建效果。

关键词:机器视觉;三维重建;深度图像融合;有向距离场;Delaunay 三角剖分;Marching Tetrahedra 算法

中图分类号:TN911, TP391.4

文献标志码:A

doi:10.5768/JAO201435.0502003

3D-object reconstruction based on fusion of depth images by Kinect sensor

Guo Lianpeng, Chen Xiangning, Liu Bin, Liu Tianjian

(Equipment Academy of PLA, Beijing 101400, China)

Abstract: 3D reconstruction of object is an interest subject in computer vision. A method for 3D reconstruction of object was proposed by integrating a set of depth maps obtained by Kinect sensor. To aggregate the contributions of the depth images at their corresponding scale, the depth images were triangulated in image space firstly, and the next step was to insert the triangulated depth images into the hierarchical signed distance field, then the global Delaunay tetrahedralization was applied to all the voxel positions yielding a convex hull that covers all the voxels, and the marching tetrahedra algorithm was applied to the resulting tetrahedral mesh for extracting the surface. Experimental results show that this method can make use of 37 depth images by Kinect sensor at different directions with the resolution of 640×480 to extract high-quality surfaces, which only costs 48 s.

Key words: computer vision; 3D reconstruction; depth images fusion; signed distance field; Delaunay tetrahedralization; Marching Tetrahedra

引言

物体三维重建技术是近几十年来研究的热点

问题,在虚拟现实、文化遗产保护、反向工程和多媒体产业模型设计等诸多领域都有广泛的应用,

收稿日期:2014-05-15; 修回日期:2014-06-26

基金项目:军队探索项目(7131145)

作者简介:郭连朋(1990—),男,山东聊城人,硕士研究生,主要从事机器视觉和三维重建技术研究。

E-mail:guolianpeng1990@126.com

其目的是从拍摄的真实物体的影像中还原出其三维模型。目前,主要的重建算法有基于体素的方法^[1]、基于多边形网格变形的方法^[2]和基于多视图匹配的方法^[3]。基于体素的重建方法的重建精度取决于立方栅格的划分细度,但划分细度的增加,运算量会以三次方的速度增加;基于多边形网格变形的重建方法的精度和运算时间依赖于初始表面的质量;基于多视图匹配的重建方法无论从精度还是从完整性上,都有突出的表现。随着 Kinect 传感器等深度信息获取设备的出现,基于多深度图像融合^[4-7]的三维重建算法引起了国内外学者的高度关注,有极大的理论意义和实际应用价值。自 2010 年 Kinect 发布以后,许多学者便对利用 Kinect 进行三维重建的方法进行了深入的研究。在国外,2010 年,Peter Henry^[8] 等人尝试使用 Kinect 对室内场景进行三维重建,得到的三维模型较为粗糙,速度也只有 2 f/s 左右。2011 年,Shahram Izadi、Richard A. Newcombe 等人构建了 KinectFusion^[9] 系统,对小范围静态场景和单独物体进行三维重建,相对于 Peter Henry 等人的方法,其结果更加精细,但重建效率有待进一步提高。Engelhard 等人^[10] 利用 Kinect 提供的 RGB-D 相机实现了一个实时的视觉 SLAM 系统,取得了很好的效果。Izadi 等人^[11-12] 提出了一种基于 GPU 并行计算的实时定位与重建系统,并实现了动态场景的三维重建,但该系统重建结果依赖于实时的 ICP 配准,系统稳定性容易受到错误匹配的影响。在国内,中国科学院自动化研究所模式识别国家重点实验室刘鑫等人^[13] 实现了一种基于 Kinect 传感器的快速物体重建方法,对于有遮挡物体的重建,也取得了较好的重建效果。清华大学智能技术与系统国家重点实验室^[14] 根据 Kinect 系统中采集的 RGB 图像和红外图像,实现了对机器人 6 自由度的运动估计和场景建模。南京大学^[15] 使用 HCR 家用机器人开源平台、Kinect 摄像机和 PC 机组装成机器人,并利用该机器人实现在未知室内环境下基于视觉系统的三维地图创建。此外,大连理工大学^[16]、上海交通大学^[17] 和北京交通大学^[18] 等高校都对三维物体和场景重建进行了研究。

目前基于 Kinect 传感器的物体三维重建和场景三维重建方法,都是先根据深度图像获取单个场景或物体的三维点云,然后对点云进行配准,完

成整个物体表面和大场景的三维重建,该方法的重建结果依赖于点云的配准结果,但由于 Kinect 传感器深度图像的噪声较大,重建出的单个场景和物体的三维点云很不稳定,因而重建效果并不是特别理想。本文基于 Kinect 传感器得到的一系列深度图像,充分利用冗余信息来抑制或平滑单幅深度图像的噪声,并融合各个角度拍摄的深度图像到同一坐标系下,得到高质量的深度图像,然后重建出一幅完整的物体三维表面,能有效克服 Kinect 传感器深度图像噪声较大的缺点。本文算法的主要特点有:1) 是一种全自动的物体三维重建方法,通过一系列的深度图像,可直接生成物体的三维表面;2) 鲁棒性高,能有效解决 Kinect 传感器噪声较大的问题,通过多深度图像融合可得到高质量的深度图像。

1 Kinect 传感器深度图像获取原理

Kinect 深度图像是由红外相机根据主动结构红外光,通过光学三角测量^[19] 得到真实世界的以 mm 为单位的深度图像,简而言之,利用实时获取的红外图像和存储在 Kinect 传感器中的参考图像通过三角测量来计算出场景的深度信息,得到整个场景的深度图像。

令 (x, y) 为图像平面, z 为图像中表示深度信息的坐标轴。激光发射器向场景中投射结构化的红外激光散斑,通过红外相机接收,并与存储在 Kinect 传感器中的参考图像进行对比。参考图像是由红外相机拍摄距离 z_0 处的参考平面得到的红外图像。如果目标与参考平面的距离不同,则红外图像上的激光散斑位置会发生偏移 d , 利用 Kinect 传感器自带的图像相关程序,根据偏移量 d 可以生成深度图像。

假定目标 k 在目标平面上,令 b 为红外发射器和红外相机之间的基线长度, f 为红外相机的焦距, d 为测量得到的偏移量, z_0 为用于计算深度图像的参考距离。如图 1 所示,为了计算目标以 mm 为单位的深度 z_k , 目标点 k 和参考点 o 之间的距离为 D , 则有:

$$\frac{D}{b} = \frac{z_0 - z_k}{z_0} \quad (1)$$

并且:

$$\frac{d}{f} = \frac{D}{z_k} \quad (2)$$

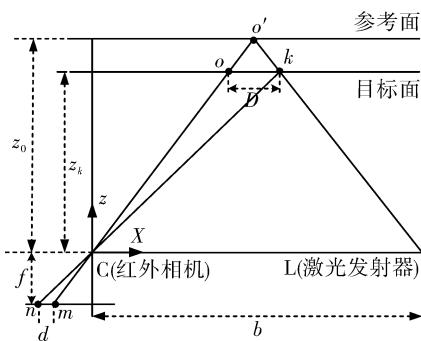


图 1 深度值与偏移量关系图

Fig. 1 Relation between relative depth and measured disparity

根据方程(1)和(2)得到偏移量到 mm 单位的深度值的变换方程:

$$z_k = \frac{z_0}{1 + \frac{z_0}{bf}d} \quad (3)$$

2 深度图像融合完成物体三维重建

为了融合不同尺度的深度图像,根据文献[7],以 SDF(signed distance field)的形式表示目标的几何信息并在尺度空间进行融合,并选择一个合适的尺度进行最终表面提取。尺度空间即 3D 欧氏空间加上一个尺度,每一个尺度不但可用于表示尺度空间中的尺度参数,还可用于表示一幅深度图像。深度图像中的重叠区域很少具有相同的尺度,无法进行信息的融合,通常的做法是把尺度空间离散为八度,每一尺度生成一个层级表示,并把所有影像整合到同一尺度下进行表面重建。把所有深度图像转换到某一确定的尺度下可导致在靠近尺度的边缘位置出现伪像,因此,要把几何信息从粗略的八度中转化成精细的八度,对于不精确的数据(比如从掠射角拍摄到的物体表面图像),进行剔除,最后,在分层有向距离场(hSDF)中应用 Marching Tetrahedra 算法^[20]构造等值面,完成物体的表面重建。

2.1 构建分层有向距离场

对于一组通过 Kinect 传感器得到的深度图像数据,在图像空间中对其进行三角化将其变为三维数据,如果生成的 2 个三角形之间的深度差大于某阈值,我们认为其深度不连续,移除此三角形。阈值的大小由像素大小 F 决定。像素大小是指像素 (u, v) 在图像中的范围。

然后,把三角化的深度图像映射到分层有向

距离场,每个层级对应于一个原始的八叉树,其中每个立方体的顶角上有 8 个体素,并且每个顶角被分为 8 个子立方体,这些子立方体生成 27 个新的体素,其中 8 个体素与父节点重合,保留这些重复的体素以表示不同层级的体素信息。

不断对深度图像进行三角化,每插入 1 个三角形,对其八叉树的位置进行一次判定。三角形 T 的八叉树的层次由(4)式确定:

$$l_T = \lfloor \log_2 \left(\frac{F_{\square}^R \lambda}{F_{\Delta}} \right) \rfloor \quad (4)$$

式中: F_{Δ} 为参考像素面积,其值由最小像素大小决定; λ 为采样率; F_{\square}^R 为根节点的像素面积;[x] 表示对 x 向上取整。

2.2 修正分层有向距离场

分层有向距离场包含了在各种尺度上所有输入的深度图像,由于 Kinect 传感器自身条件的限制,生成的深度图像不可避免存在噪声和无法获取深度信息的孔洞,因此,需要通过传输距离和可信度检测提高样本的可信度。对 hSDF 按粗体素到精细体素的顺序进行遍历,寻找可信度 w_l 小于一定阈值 T_0 的体素。对于层级 l 上的子可信体素,如果该层级上的所有体素都被占用,则在较粗的 $l-1$ 层级上插入距离值。

根据文献[21],对传输距离和可信度进行加权融合,距离为 \tilde{d}_l ,可信度为 \tilde{w}_l 的体素 x ,在层级 l 上有:

$$\tilde{d}_l = \frac{d_l w_l + d_{l-1} (T_0 - w_l) \min(1, -\frac{w_l - 1}{T_0})}{w_l + (T_0 - w_l) \min(1, \frac{w_l - 1}{T_0})} \quad (5)$$

$$\tilde{w}_l = w_l + (T_0 - w_l) \min(1, \frac{w_l - 1}{T_0}) \quad (6)$$

如果混合可信度 w_l 小于阈值 $T_1 \leq T_0$,则认为其不可用于重建,需要将其从八叉树中删除。对所有体素处理完之后,在不同层级中仍存在重复的体素,因此,以精细体素到粗体素的顺序进行遍历,删除较底层级中重复的体素。

2.3 提取等值面

通过以上步骤删除不可信和重复的体素,把分层有向距离场转变为分散有向距离场,每个体素位置都有一个与之相关联的距离值及相关属性。

对所有体素位置,应用整体 Delaunay 三角剖

分算法^[22],产生一个四面体,即涵盖了所有体素的凸包。该方法的缺点在于没有考虑到数据的形状,并且一些凸包把距离场中互不相关的部分联系到了一起,容易导致不相关的区域之间存在错误插值并产生幻影。为了删除这些四面体,根据体素的相邻关系,如果在非近邻体素之间存在至少有一个边,则删除该四面体,该方法不仅能检测出四面体之间的不相邻部分,还能通过通用的四面体剖分技术填充表面孔洞。

应用 Marching Tetrahedra 算法^[22]提取等值面,该算法是自适应的,在信息较少的区域生成较少的三角形,在信息较多的区域生成较多的三角形,但由于算法的限制,重建结果中含有大量无效的三角形,对提高表面精度没有作用。为了减少重建结果中大量的无效三角形,根据文献[23],如果过零点非常接近四面体网格中边的顶点,则把该边上的顶点定义为零点,并令其距离值为零,可有效减少无效的三角形。

3 实验结果与分析

实验所采用的硬件包括 Kinect 传感器 1 台及配套的 USB 数据线及电源等;笔记本电脑 1 台,操作系统为 64 位 Microsoft Windows7,处理器为 Intel 酷睿 i5 3230 M,主频为 2.6 GHz,内存 4 G,硬盘 1 T,显卡型号为 NVIDIA GeForce GT 750 M,显存容量为 2 GB。

首先选用表面纹理较为复杂的玩具车作为重建对象,用 Kinect 传感器采集其不同方向的 37 幅分辨率为 640 像素×480 像素的深度图像,如图 2 所示。由于 Kinect 传感器自身条件的限制,深度图像噪声较大并且存在无法获得深度信息的孔洞,用本文算法对其进行融合处理,通过修正的分散有向距离场克服深度图像中噪声和孔洞的影响,并对所有体素位置应用整体 Delaunay 三角剖分算法生成涵盖了所有体素位置的凸包的四面体,利用 Marching Tetrahedra 算法构造等值面,生成三角网格如图 3 所示。最终重建出目标的三维表面如图 4 所示。实验中,令匝道大小因子 $\gamma=4$,采样率 $\lambda=1$, $T_0=T_1=0$ 。由图 4 可以看出,基于多深度图像融合重建的方法,能有效地利用深度图像完成物体的三维重建,得到较好的重建效果,但对于小车底部,由于 Kinect 传感器无法准确

获取其深度信息,重建效果不好。

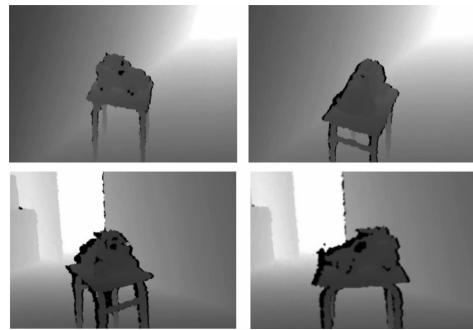


图 2 玩具车深度图像
Fig. 2 Depth images of toy cars

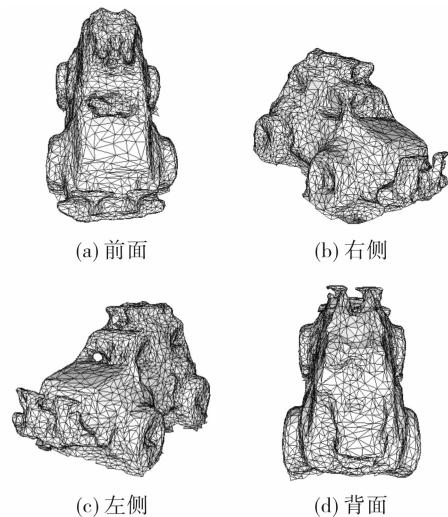


图 3 本文算法重建的三维网格
Fig. 3 Triangle meshes reconstructed by this method

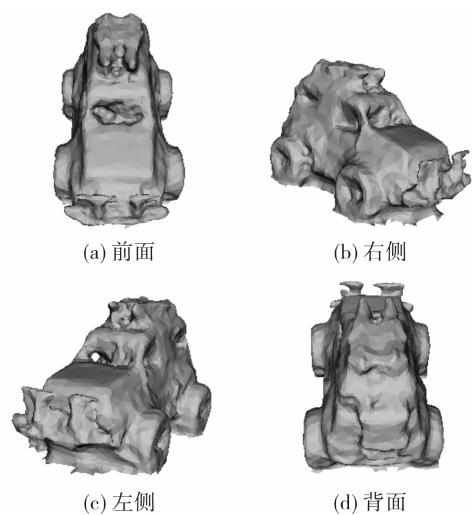


图 4 本文算法重建结果
Fig. 4 Reconstruction of car by this method

为了验证基于 Kinect 传感器的多深度图像融合算法的有效性,作为对比,本文利用普通彩色摄像机拍摄小车模型一组 72 幅分辨率为 640 像素×

480 像素的彩色图像作为原始数据,如图 5 所示。根据文献 [24],利用 PMVS(patch multi-view stereo)算法完成物体三维重建。首先检测每幅图像的 Harris 和 DoG(difference of gaussian)特征点,然后通过特征点匹配重建出稀疏三维点云,再进行扩散得到物体的稠密点云,最后过滤掉错误点云,并对其进行 Poisson 表面重建,得到重建结果如图 6 所示。由图 6 可以看出,整体重建结果与本文的算法重建结果相差不大,但对在左侧车轮处,由于无法得到足够多的匹配特征点,出现了信息的缺失,PMVS 算法的重建结果依赖于物体表面特征点的提取,容易受到表面纹理、光照等因素的影响,纹理较少的物体重建效果不好。



图 5 玩具车及花盆彩色图像

Fig. 5 Color images of toy cars and flowerpot

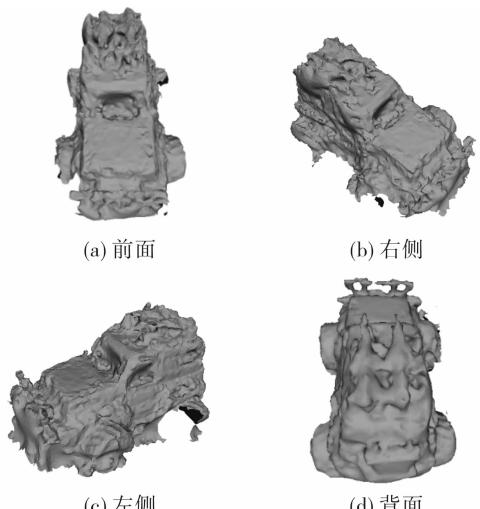


图 6 Poisson 表面重建结果

Fig. 6 Reconstruction of car by Poisson

本文基于 Kinect 传感器多深度图像融合的物体三维重建算法,由于输入的深度图像分辨率是固定的,消耗的时间主要取决于深度图像的数量。表 1 列出了 2 组实验的运行时间,PMVS 算法无论在重建精度还是在重建完整性上,表现都非常突出,代表了当前多视重建达到的最高水平,但该

算法时间和空间复杂度都比较大,通过对比我们可以看出本文算法的重建时间相对于 PMVS 算法有较大优势。

表 1 PMVS 与本文算法运行时间对比表

Table 1 Time costs of PMVS and this method

	影像数量/幅	分辨率/pixel	重建时间/s
PMVS	72	640×480	384
本文算法	37	640×480	48

4 结论

本文针对 Kinect 传感器能方便获取深度图像的特点,将多深度图像融合算法应用于利用 Kinect 传感器获取的深度图像,完成了单个物体的三维重建,能有效克服 Kinect 传感器深度图像噪声较大和尺度不一致的缺点,得到较为理想的重建模型,重建出相对比较精细的细节特征,为利用 Kinect 传感器实现物体的三维重建提供了一种新方法和新思路,尤其是对纹理较少或光线较暗甚至黑暗条件下的物体,仍能鲁棒重建出物体的三维模型。由于本文算法依赖于深度图像,缺乏物体的彩色纹理。因此,需要进一步利用 Kinect 传感器的彩色相机得到的彩色图像实现纹理映射,得到带彩色纹理的三维模型。

参考文献:

- [1] Paris S, Sillion F X, Quan L. A surface reconstruction method using global graph cut optimization[J]. International Journal of Computer Vision, 2006, 66(2): 141-161.
- [2] Zaharescu A, Boyer E, Horaud R. Transformesh: a topology-adaptive mesh-based approach to surface evolution[M]. Berlin Heidelberg : Computer Vision-ACCV 2007 Springer, 2007: 166-175.
- [3] Lhuillier M, Quan L. A quasi-dense approach to surface reconstruction from uncalibrated images [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 418-433.
- [4] Bradley D, Boubekeur T, Heidrich W. Accurate multi-view reconstruction using robust binocular stereo and surface meshing [C]//2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008: 1-8.
- [5] Merrell P, Akbarzadeh A, Wang L, et al. Real-time visibility-based fusion of depth maps[C]//11th Inter-

- national Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007: 1-8.
- [6] Tylecek R, Sara R. Depth map fusion with camera position refinement [C]//Computer Vision Winter Workshop. Paris, France: Computer Vision, 2009: 59-66.
- [7] Fuhrmann S, Goesele M. Fusion of depth maps with multiple scales[J]. ACM Transactions on Graphics, 2011, 30(6): 148.
- [8] Peter H, Michael K, Evan H, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments [J]. International Journal of Robotics Research, 2012, 31(5): 647-663.
- [9] Newcombe, Richard A, Izadi S, et al. Kinectfusion: Real-time dense surface mapping and tracking[C]// 10th IEEE/ACM International Symposium on Mixed and Augmented Reality. Basel, Switzerland: IEEE, 2011: 127-136.
- [10] Engelhard N, Endres F, Hess J, et al. Real-time 3D visual SLAM with a hand-held RGB-D camera[C]// 24th Annual ACM Symposium on User Interface Software and Technology. Robotdalen, Sweden: Association for Computing Machinery, 2011: 559-568.
- [11] Izadi S, Newcomber A, Kim D, et al. Kinectfusion: real-time dynamic 3D surface reconstruction and interaction [C]// ACM SIGGRAPH 2011 Talk. Vancouver, Canada: ACM SIGGRAPH, 2011: 449-458.
- [12] Izadi S, Kim D, Hilliges O, et al. Kinectfusion: real-time 3D reconstruction and interaction using a moving depth camera[C]//24th Annual ACM Symposium on User Interface Software and Technology. Santa Barbara, CA: ACM, 2011: 559-568.
- [13] Liu Xin, Xu Huarong, Hu Zhanyi, et al. GPU based fast 3D-object modeling with kinect[J]. Acta Automatica Sinica, 2012, 38(8): 1288-1297.
刘鑫,许华荣,胡占义,等.基于 GPU 和 Kinect 的快速物体重建[J].自动化学报, 2012, 38(8): 1288-1297.
- [14] Yang Dongfang, Wang Shicheng, Liu Huaping, et al. Scene modeling and autonomous navigation for robots based on kinect system[J]. Robot, 2012, 34 (5): 581-586.
- 杨东方,王仕成,刘华平,等.基于 Kinect 系统的场景建模与机器人自主导航[J].机器人,2012,34(5): 581-586.
- [15] Huang Yuqiang. Vision based mapping of a mobile robot in unknown environment[D]. Nanjing: Nanjing University, 2012: 56-57.
黄玉强.移动机器人在未知环境下的基于视觉系统的地图创建[D].南京:南京大学,2012:56-57.
- [16] Zhang Cuihong. 3D indoor scene reconstruction with kinect depth camera[D]. Daling: Dalian University of Technology, 2013: 48.
张翠红.基于 Kinect 深度相机的室内三维场景重构[D].大连:大连理工大学,2013:48.
- [17] Chen Xiaoming. Research of 3D reconstruction and filtering algorithm based on depth information of kinect[D]. Shanghai: Shanghai Jiao Tong University, 2013: 53-54.
陈晓明.基于 Kinect 深度信息的实时三维重建和滤波算法研究[D].上海:上海交通大学,2013:53-54.
- [18] Li Guozhen. Research and implementation of kinect based 3D reconstruction[D]. Beijing: Beijing Jiaotong University, 2012: 47.
李国镇.基于 Kinect 的三维重建方法的研究与实现[D].北京:北京交通大学,2012:47.
- [19] Arieli Y, Freedman B, Machline M, et al. Depth mapping using projected patterns: US. 8150142 B2 [P]. 2012-04-03[2014-05-06]. <http://www.freepatentonline.com/8150142.html>.
- [20] Akio D, Koide A. An efficient method of triangulating equi-valued surfaces by using tetrahedral cells [J]. Ieice Transactions on Information and Systems, 1991, 74(1): 214-224.
- [21] Gortler S J, Grzeszczuk R, Szeliski R, et al. The lumigraph[C]//Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. New Orleans : ACM, 1996: 43-54.
- [22] Akio D, Koide A. An efficient method of triangulating equi-valued surfaces by using tetrahedral cells [J]. IEICE, 1991, 74(1): 214-224.
- [23] Schaefer S, Warren J. Dual marching cubes: primal contouring of dual grids[C]//Computer Graphics and Applications. Houston, TX, USA: IEEE, 2004: 70-76.
- [24] Frahm J M, Fite G P, Gallup D, et al. Building rome on a cloudless day[M]. Berlin Heidelberg: Computer Vision-ECCV Springer, 2010: 368-381.