

一种用于机器人抓取的单目实时三维重建系统

吉白冰, 曹其新

(上海交通大学机械与动力工程学院, 上海 200240)

摘要:为支持家庭用服务机器人抓取任务,设计了一种针对家庭环境中常见物体的实时三维重建系统。系统假设相机在围绕物体运动的同时采集图像。首先借助从运动恢复结构算法求解相机在采集图像序列时的运动轨迹。根据运动轨迹中的位置信息,基于几何约束为相机运动轨迹中的每一帧选择最合适的配对图像。提出基于虚拟双目的立体匹配方法恢复每组配对中两帧图像共视区域的三维结构。利用L-R检查和抛物线拟合等后处理方法剔除误差较大的测点。最后用不同的图像配对计算出的三维结构生成最终的物体模型。实验结果表明生成的物体模型中位数相对误差在1.0%左右,满足机器人抓取任务中对物体模型的精度要求。同时系统平均运行速度超过25帧/秒,满足家庭用物体快速建模的要求。

关键词:服务机器人;单目视觉;深度估计;从运动中恢复深度

中图分类号:TH16;TP391 **文献标识码:**A **文章编号:**1001-3997(2021)09-0287-04

DOI:10.19356/j.cnki.1001-3997.2021.09.064

Real-Time Monocular 3D Reconstruction System for Robot's Grasping

JI Bai-bing, CAO Qi-xin

(School of Mechanical Engineering Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: A real-time 3D reconstruction system for common objects in home environment is proposed to provide support for grasping task of service robots. System assumes that the camera keeps moving and captures images in the same time. Trajectory of camera motion is first estimated by applying Structure from Motion algorithm. Based on position information of camera trajectory, most suitable matched image is selected for each camera frame under geometric constraints. Virtual stereo camera based motion stereo method is proposed to recover 3D structure of covisible region for each two matched image pair. L-R check and parabola fitting and other post-processing method are applied to cull error measurements. 3D structure measurements from different image pairs are then used to generate object model. Experiments show median error of object 3D model is about 1.0% and the accuracy meets the requirement of grasp task. Fast household object construction requirement is also fulfilled as system process over 25 frames per second.

Key Words: Service Robot; Monocular Vision; Depth Estimation; Motion Stereo

1 引言

三维重建系统在机器人、生物医学工程、建筑测绘等领域均有广泛的应用。在机器人领域当中,三维重建主要有两类应用,一类是移动机器人的定位导航地图构建,另一类是机器人的抓取任务中,待抓取物体的三维模型构建。在机器人抓取任务中,无论是传统的PPF方法^[2]还是基于深度学习的物体姿态估计方法^[3]都将待抓取物体的三维模型作为已知信息。目前基于结构光原理^[4-5]的三维扫描仪虽然可以重建较高精度的物体模型,但此类三维扫描仪价格昂贵且操作繁琐,在家庭环境中难以被普通用户接受和使用。视觉传感器成本低且通用性强的传感器。目前已有应用单目传感器解决工业机器人识别问题的研究^[9],基于视觉

的三维重建系统,对家用服务机器人抓取任务也有很大意义。

针对基于视觉传感器的三维重建问题,文献^[10]提出了一种基于运动恢复的双目视觉三维重建系统,将原有抛物线拟合的方法加入到三维重建系统,提高了三维重建系统精度。但并未考虑到双目图像错误立体匹配产生的三维结构,重建结果中噪点较多,因此上述方法仅在形状规则的物体中取得较好的效果。面向家庭环境的机器人视觉抓取的三维重建系统具有以下几个特点:一是涉及到的场景较小,一般为桌面上玩偶、书本等物体,太大或太小的物体都难于使用抓手抓取;二是重建使用的硬件和三维重建软件系统需简单易用,方便家庭用户的使用;三是对精度与实时性有一定要求。

来稿日期:2020-07-23

作者简介:吉白冰,(1994-),男,辽宁人,硕士研究生,主要研究方向:视觉三维重建;

曹其新,(1960-),男,浙江人,博士研究生,博士生导师,教授,主要研究方向:机器视觉,工业机器人

针对上述特点,提出了一种应用于家庭环境中机器人抓取任务的单目三维重建系统,可以基于目前消费电子产品中主流的单目视觉传感器实时运行,降低了传统三维重建系统对传感器的要求。同时,在传统单目三维重建系统上增加了基于连通区域滤波器后处理模块,用于消除立体匹配中错误恢复的三维结构,保证重建效果和精度。此外,算法基于GPU并行处理技术^[11]设计和实现,加速立体匹配过程,确保三维重建系统实时性。

2 单目三维重建系统算法框架

双目三维重建系统建立在的理想双目相机成像模型下,即左右两个相机位于同一平面(光轴平行),且相机固有参数(如内参矩阵 K)一致。可推出如下公式,如图1所示。

$$Z = \frac{f \cdot b}{x_1 - x_2} \quad (1)$$

式中: $x_2 - x_1$ —视差, b —相机左右目的光心 O_1 和 O_2 间的距离, p_1 —空间点 P 在左相机成像的像素, p_2 —空间点 P 在右相机成像的像素。空间点 P 的深度 Z 与视差成反比,即深度图的计算与视差图的计算具有反比对应关系。

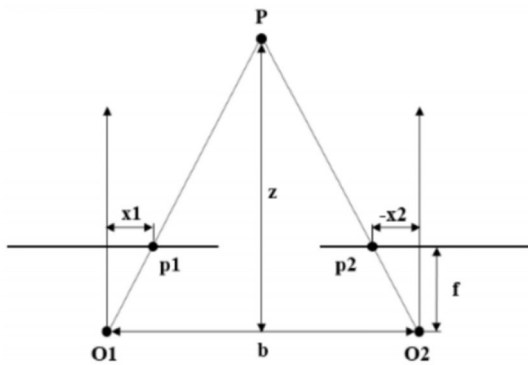


图1 理想双目模型
Fig.1 Ideal Stereo Model

对于围绕物体运动并采集图像的单目相机而言,虽然缺少物理意义上的双目结构,但可以组合当前相机所拍摄图像与历史某时刻相机所拍摄的图像组成从运动中构造双目(motion stereo)。如图2所示,公式(1)在从运动中构成双目的相机配置下不再成立,且从运动中构造双目的相机相对位置关系未知且并不固定。因此不能简单应用双目三维重建方法解决单目三维重建问题。

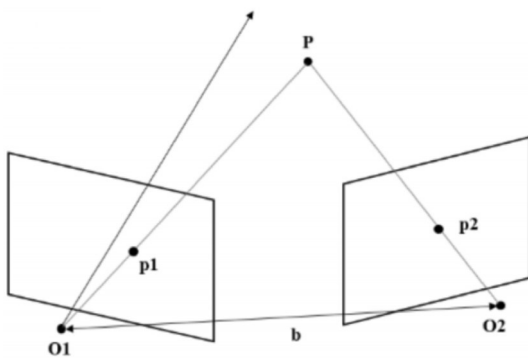


图2 从运动中构造双目的双目相机位置关系
Fig.2 Motion from Stereo

针对单目三维重建问题的上述特点,提出了的三维重建系

统算法框架。在单目三维重建系统中,首先根据SFM算法^[7-8]恢复不同时刻相机间的相对运动关系,根据几何关系选择相机当前帧和过去某一帧组成虚拟双目。根据虚拟双目间的几何关系和图像像素相似性,计算当前帧视场内的场景点云。经过坐标系变换将不同时刻相机帧重建点云拼接为完整的物体模型。系统基于SFM算法恢复相机运动,后应用TSDF方法^[6]完成坐标系对齐功能。上述方法已有成熟的研究结果,因此不再赘述相应原理。下文中将详细说明基于虚拟双目构造、立体匹配及后处理过程,如图3所示。

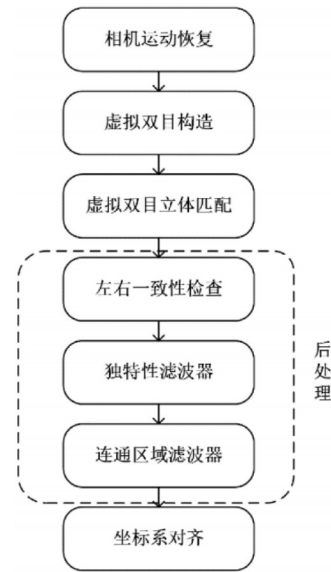


图3 算法流程图

Fig.3 Flow Chart of Proposed Algorithm

3 单目三维重建系统的三维结构恢复

在单目三维重建系统中,三维结构恢复指根据从运动中构造的双目计算相机视场内的场景三维结构。下面将从虚拟双目构造,虚拟双目立体匹配,后处理三个部分介绍基于单目相机的三维结构恢复方法。

3.1 虚拟双目构造

基于SFM方法恢复的相机相对位置关系存在一定误差,因此要保证最终场景三维结构的准确性,需要合理选择历史相机帧与当前帧配对,使得误差对深度图质量的影响较小。在配对选择的过程中,历史相机帧与当前相机帧间要保持较大的重叠和相似的朝向,这样深度图可以覆盖图像的大部分范围,同时两个相机状态间要有合理的基线,较长的基线会降低图像的重叠度,减少深度图的有效区域,较短的基线会导致深度图易受到噪声的干扰,降低深度估计的准确性。基于上述分析,提出了如下评价函数:

$$Eval(i) = n_{cov}^i \cdot \cos \theta_{pose}^i \cdot \cos \theta_{up}^i \quad (2)$$

式中: n_{cov}^i —历史关键帧与当前帧共视的特征点数量,特征点基于ORB特征点^[1]提取并匹配, $\cos \theta_{pose}^i$ —两帧对应相机光轴夹角的余弦值, $\cos \theta_{up}^i$ —两帧的 up 向量夹角的余弦值。计算所有的历史帧与当前帧的评价函数值,选取评价函数值最低的历史帧与当前帧构造虚拟双目。

3.2 虚拟双目的立体匹配

由章2所述,从运动中恢复双目得到双目图像对基本不满足理想双目模型,因此其立体匹配过程分为如下几步:

像素匹配和匹配代价计算:首先基于先验信息确定待计算深度图的图像所对应场景的深度分布区间,在得到的分布区间内确定 N 个深度假设,每个深度假设对应于一个代表深度值 d 的平面。为增强深度估计结果对于场景光照变化和图像噪声的鲁棒性,在以像素 p 为中心的方形领域内采样计算代价函数 $C(p, d)$ 。应用ZNCC算子构造像素对应代价函数 $C(p, d)$ 。

$$q = K \left(R + \frac{1}{d} t^T n \right) K^{-1} p \quad (3)$$

式中: q —右图像中与像素 p 对应的像素, K —内参矩阵, R —相机旋转矩阵, t —相机平移向量, n —平面法向量。 q 代表像素 p 在由图像上的投影,如图4所示。

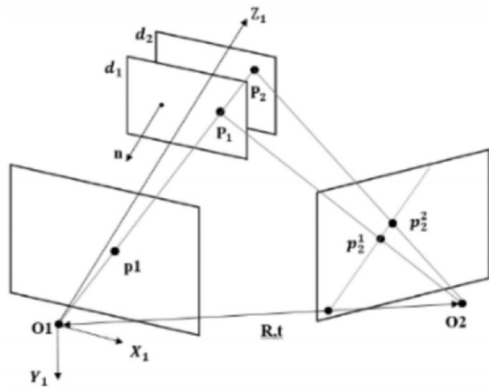


图4 像素点投影示意图

Fig.4 Illustration of Pixel Reprojection

代价聚合与深度值求精:计算得到每个深度假设 d 所对应的代价函数 $C(p, d)$ 后,按WTA规则估计每个像素 p 所对应的深度:

$$d_i = \operatorname{argmin}_d C(p, d) \quad d \in \{d_1, \dots, d_N\} \quad (4)$$

为补偿深度假设离散化造成的精度损失,采用抛物线拟合的方式获取亚像素级深度估计精度。则抛物线函数极小值点对应的深度值 d_{\min} 满足公式(5):

$$d_{\min} = d_i + \frac{C(d_{i+1}) - C(d_{i-1})}{2(C(d_{i+1}) - 2C(d_i) + C(d_{i-1}))} \cdot \Delta d \quad (5)$$

式中: d_i —最小 $C(p, d)$ 所对应的深度假设, d_{i-1} 、 d_{i+1} —与 d_i 相邻的深度假设, Δd —深度假设间距。由于间中每个像素的深度估计过程具有独立性,因此每个像素的深度估计过程可以并行进行,得到初始深度图。

3.3 后处理

立体匹配计算的深度图中存在大量错误深度估计,使最终的三维模型产生大量噪点,降低三维模型质量。为解决这一问题,提出了如下的后处理过程。

左右一致性检查:由于立体匹配过程具有计算对称性,因此估计正确的深度应满足如下所示的重投影误差公式:

$$q = K \left(D_c(p) R K^{-1} + t \right) \\ \left\| K \left(D_k(q) R K^{-1} q + t \right) - p \right\| < \delta \quad (6)$$

式中: D_c —当前帧深度图, D_k —历史帧深度图, δ —阈值,其余符号同公式(3)。

独特性比率滤波器:在被图像噪声干扰或相对运动关系的误差而产生错误深度估计的区域,其最小代价函数与次小代价函数会比较接近。根据上述观察,定义如下独特性比率过滤噪声。当像素 p 的独特性比率小于阈值时过滤对应深度:

$$r_{\text{uniq}}(p) = \frac{C(p, d_{\text{sec}})}{C(p, d_{\text{min}})} < r_{\text{thres}} \quad (7)$$

式中: $r_{\text{uniq}}(p)$ — p 像素的独特性比率, d_{sec} —次小代价函数对应的深度假设, d_{min} —最小代价函数对应的深度假设。 r_{thres} —独特性比率的阈值。

连通区域滤波器:真实场景中大部分区域在空间是连续的,因此体现在深度图上相邻区域的深度值也具有一定的连续性。错误匹配的深度值一般与其周边区域的深度值相比具有较大的不连续性,且在深度图中零散的分布与各个区域。根据如下公式判断深度图上某像素与相邻区域在深度图上的连续性关系。

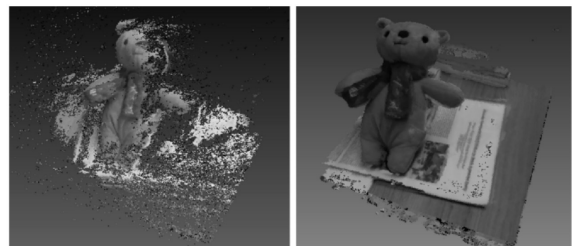
$$|d - d_r| < \eta \cdot \Delta d \quad (8)$$

式中: d —某像素对应深度, d_r —某像素领域的对应深度, Δd —深度假设间距。应用上述公式将深度图的深度连续区域的划分问题转化为一个二值图像连通区域标记问题,使用两步法将所有的深度图连通区域分类并标记。根据标记的连通区域中像素的数量,将像素数量小于一定值的连通区域从原始深度图中去除。经过上述后处理后得到的深度图经过坐标系对齐后得到最终的物体模型。

4 实验

在家庭环境中常用且易于机器人抓取的三种物体上进行了实验。实验图像分辨率为 640×480 ,程序运行于配备i7-4710MQ处理器和GTX880M显卡的笔记本电脑上。以Realsense SR300传感器测量的结果作为真值。

对比了应用后处理和不应后处理方法的三维重建结果。在实际环境中由于采集的图像中图像噪声及场景光照变化,一些在理想条件下正确匹配的像素在噪声干扰下匹配错误,进而物体模型表面不完整且中存在大量噪点。通过节3.3介绍的后处理流程可以过滤单帧深度图中的错误匹配结果。在基于运动的单目三维重建系统中,不同帧图像恢复的三维结构存在一定冗余,某一帧中被过滤的错误三维结构会被其他帧中正确估计的三维结构代替,因此后处理模块不会对完整性造成影响,如图5所示。



(a) 无后处理的三维模型 (b) 有后处理的三维模型

图5 实验结果对比

Fig.5 Comparison of Experiment Result

基于结构光传感器的测量结果作为真值计算三维重建结果

物体模型的精度。根据计算结果绘制了误差色温图,误差表示范围从0%到100%,且色温越冷代表误差越小。上述三个实验物体的模型与结构光传感器测量真值的中位数相对误差(MAE)均在1.0%左右,从相对误差色温图中也可以看出重建物体模型的大部分区域误差非常小。对于长宽高在20cm范围内的家庭中常见物体而言,生成的模型可以满足服务机器人构建抓取物体数据库的模型精度要求,如图6所示。

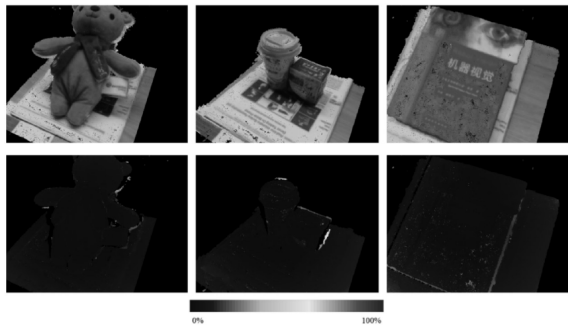


图6 三种不同物体重建模型(上)与误差色温图(下)

Fig.6 Models of Three different objects 3D Reconstruction (Up) and Corresponding Error Color Map (Down)

为验证三维重建系统实时性,记录了图6中玩偶熊模型建模过程中每一帧的处理时间并绘制了的处理时间随拍摄时间变化的曲线。帧平均处理时间为33.95ms,即29.45帧/秒的处理速度,满足实时运行的要求,如图7所示。

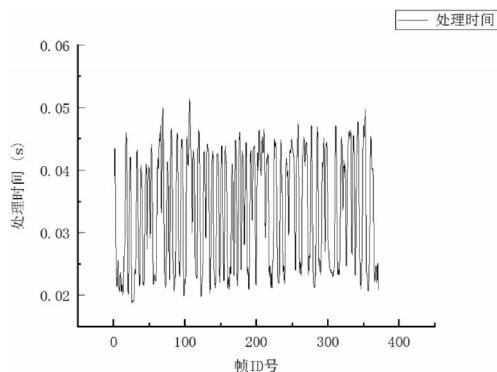


图7 三维重建系统单帧处理时间曲线图

Fig.7 Graph of Running Time of 3D Reconstruction

5 结论

介绍了一种基于运动恢复深度的单目实时三维重建系统,

针对运动中单目相机采集的图像序列,提出了一种虚拟双目相机的构造准则。并推导了虚拟双目相机的立体匹配公式。针对实际场景中图像噪声对三维重建结果产生较大影响的问题,在传统算法框架基础上增加了后处理模块。实验结果表明应用后处理模块,取得了质量更好的物体模型。针对三种不同物体的重建实验说明了应用三维重建系统的精度。系统并行处理图像数据并生成三维模型。平均处理速度为29.45帧每秒,满足服务机器人抓取任务对建模系统速度的要求。未来将在本文实验得到的模型基础上,设计基于物体模型服务机器人抓取系统,进一步验证本文模型对于服务机器人抓取任务的适用性。

参考文献

- [1] Rublee E, Rabaud V, Konolige K. ORB: An efficient alternative to SIFT or SURF [C]. International Conference on Computer Vision. IEEE, 2012: 2564-2571.
- [2] Abbeloos W, Goedemé T. Point Pair Feature Based Object Detection for Random Bin Picking [C]. Computer and Robot Vision. IEEE, 2016: 432-439.
- [3] Kehl W, Manhardt F, Tombari F. SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again [C]. IEEE International Conference on Computer Vision. IEEE, 2017: 1530-1538.
- [4] Whelan T, Salas-Moreno R F, Glocker B. ElasticFusion [J]. International Journal of Robotics Research, 2016, 35(14): 1697-1716.
- [5] Newcombe R A, Izadi S, Hilliges O. KinectFusion: Real-time dense surface mapping and tracking [C]. IEEE International Symposium on Mixed and Augmented Reality. IEEE, 2012: 127-136.
- [6] Curless B, Levoy M. A Volumetric Method for Building Complex Models from Range Images [C]. Conference on Computer Graphics and Interactive Techniques. ACM, 1996: 303-312.
- [7] Wu C. Towards Linear-Time Incremental Structure from Motion [C]. International Conference on 3d Vision. IEEE Computer Society, 2013: 127-134.
- [8] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision [M]. Cambridge: Cambridge University Press, 2003: 1865-1872.
- [9] 王修岩, 程婷婷. 基于单目视觉的工业机器人目标识别技术研究 [J]. 机械设计与制造, 2011(04): 155-157.
(Wang Xiu-yan, Cheng Tingting. Research on target recognition of industrial robot based on monocular vision [J]. Machine Design and Manufacture, 2017(4): 41-44.)
- [10] 王欣, 袁坤, 于晓. 基于运动恢复的双目视觉三维重建系统设计 [J]. 光学精密工程, 2014, 22(5): 1379-1387.
(Wang Xin, Yuan Kun, Yu Xiao. Design of binocular vision 3D reconstruction system based on motion recovery [J]. Optics and Precision Engineering, 2014, 22(5): 1379-1387.)

(上接第286页)

- [4] 邓娟, 杨荣松, 郑香美. 基于ANSYS的谐波减速器杯型柔轮应力分析与参数优化 [J]. 机械设计与制造, 2012, (4): 59-61.
(Deng Juan, Yang Rong-song, Zheng Xiang-mei. Stress analysis and parameter optimization of column-shaped flexspline in harmonic drive based on ANSYS [J]. Machinery Design & Manufacture, 2012, (4): 59-61.)
- [5] 张世民, 李春雷. 基于ANSYS的谐波减速器柔轮受力分析 [J]. 传动技术, 2009, 23(01): 13-17+44.
(Zhang Shi-min, Li Chun-lei. ANSYS based analysis of flexspline in harmonic drive [J]. Drive System Technique, 2009, 23(01): 13-17+44.)
- [6] 张庆茹, 顾京君, 黄迪山. 基于ANSYS/LS-DYNA的谐波齿轮减速器模拟装配与动态仿真 [J]. 计量与测试技术, 2015, 42(11): 1-3.
(Zhang Qing-ru, Gu Jing-jun, Huang Di-shan. Harmonic gear reducer assembly simulation and dynamic analysis based on Ansys/ls-dyna [J]. METROLOGY & MEASUREMENT TECHNIQUE, 2015, 42(11): 1-3.)

- [7] 庄小红. 等效椭圆谐波齿轮减速器的研究 [D]. 重庆: 重庆大学, 2008.
(Zhuang Xiao-hong. Study on the equivalent elliptical harmonic gear drive [D]. Chongqing: Chongqing University, 2008.)
- [8] 王晓伟, 郝长中. 基于ADAMS谐波减速器的模态分析 [J]. 成组技术与生产现代化, 2014, 31(02): 50-53+62.
(Wang Xiao-wei, Hao Chang-zhong. Modal analysis of harmonic reducer based on ADAMS [J]. GROUP TECHNOLOGY & PRODUCTION MODERNIZATION, 2014, 31(02): 50-53+62.)
- [9] 任毅斌. 杯形柔轮啮合传动性能分析 [D]. 内蒙古(内蒙古工业大学), 2015.
(Ren Yi-bin. Analysis of the cup-shaped flexible wheel meshing transmission performance [D]. Inner Mongolia (Inner Mongolia University of Technology), 2015.)
- [10] Kwang Seop Jeong, Dai Gil Lee. Development of the composite flexspline for a cycloid-type harmonic drive using net shape manufacturing method [J]. Composite Structures. 1995, Vol 36(557 ~ 565.)