

多目标背景下双目视觉深度信息优化研究及其实现^①

林 川, 郝卫东, 林承德

(桂林电子科技大学 机电工程学院, 桂林 541004)

摘 要: 针对机器人领域应用视觉进行目标物体抓取问题, 提出了一种针对多目标背景下, 新的深度优化处理方法. 通过设定一个阈值块, 以遍历成块的深度信息用类似聚类的方法, 提出目标物体的具体坐标, 传递给机器人手臂, 完成准确的抓取操作. 依次介绍了双目视觉原理、摄像机标定、双目矫正和双目匹配等内容, 以及呈现出原始的深度信息图以及优化后的深度信息图, 比较它们的差距. 最后在实验中给出了证明: 此种深度信息优化方法能够有效的提高机器人抓取目标物体的成功率. 最后, 还在文章最后给出了下一步的研究方向.

关键词: 双目视觉; 深度信息优化; 机器人抓取

Optimization of the Depth Information Based on Stereo Vision in Multi-Target Scenarios

LIN Chuan, HAO Wei-Dong, LIN Cheng-De

(Department of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: Aiming at the filed of robot grasping using binocular vision, a new method for dealing with the depth information is proposed in the multi-target scenarios. By setting a threshold in pixels and traversing the depth information image similar to clustering method can obtain the coordinates of the object and pass the coordinates to the robot arm to grasp the object accurately. In this paper, the principles of binocular vision, camera calibration, stereo recertification and stereo matching are described. Their difference between the original depth information and the depth information after optimization can be acquired by utilizing the optimization algorithm. The approach for grasping object for robot in multi-target scenarios is proved to be effective in the experiments. Research directions for future improvements are also presented.

Key words: stereo vision; depth information; optimization; robot grasping

随着计算机视觉以及机器学习的发展, 现在机器人变得越来越智能. 机器视觉技术在机器人领域变得越来越重要^[1]. 人类慢慢的从机械繁琐的工作中脱离出来. 让机器人去仓库或者其他目的地去抓取目标物体, 成为一件考验机器人智能的事情. 此时, 在复杂的环境中, 准确定位抓取成为机器人视觉中一个有待解决的问题. 基于双目视觉的机器人要准确抓取物体, 这个工作有两项内容: 第一, 是获取场景中物体的深度信息; 第二, 是使用阈值设定, 以聚类的方式来确定目标物体的坐标信息, 从而传给机械臂. 获取三维场景中的深度信息是机器视觉研究的一个基本内容之

一. 现在有很多方法可以得到场景中的深度信息, 比如: 深度摄像机, 普通的双目摄像机等. 双目视觉是机器视觉的一个非常重要的一个分支, 同时也是智能机器人的一个重要的指引器官^[2]. 但是, 由普通的摄像机实现目标定位具有速度快, 成本低, 结构简单, 精度适宜等特点, 因此能很好的定位目标物体并且获取场景的深度信息. 用双目视觉定位技术, 可用于机器人导航, 工业检测, 三维测量, 虚拟现实等其他应用^[3].

从上个世纪八十年代开始, 双目视觉已经有了广泛的应用. 例如, 日本大阪大学自适应机械系统研究

^① 收稿时间:2011-12-13;收到修改稿时间:2012-04-16

院研制了一种自适应双目视觉伺服系统;日本东京大学将实时双目立体视觉和机器人整体姿态信息集成,开发了仿真机器人动态行走导航系统.另外,麻省理工学院计算机系提出了一种用于智能交通工具的传感器融合方式,融合雷达与双目立体视觉信息. Chia-Hung Chen 运用美国加利福尼亚州 Videre Design 公司生产的双目设备来实现的机器人手臂随动系统,抓取已知物体.

在实际场景中,通常很多因素影响着机器人双目视觉定位的实时性与准确性.在多目标场景中,运用两个普通摄像头来实现机器人准确抓取目标物体.我们不仅要获得双目视觉的准确深度信息,还要解决在多目标场景中,机器人如何通过伺服电机来准确定位目标物体的位置,从而实现准确的抓取这一问题.

文章主要由三部分构成.第一部分,介绍双目视觉的原理以及双目定位的关键步骤与技术.第二部分,运用阈值设定,遍历深度信息图用类似聚类的方法提取目标物体的有效坐标,总结对比实验数据,得出结论.第三部分,利用 SIAT 机器人实现上述功能,在多目标背景下实现物体的抓取^[1].机器人在多目标背景下抓取物体大致有 8 个步骤组成,在图 1 中可以看到大致框架.

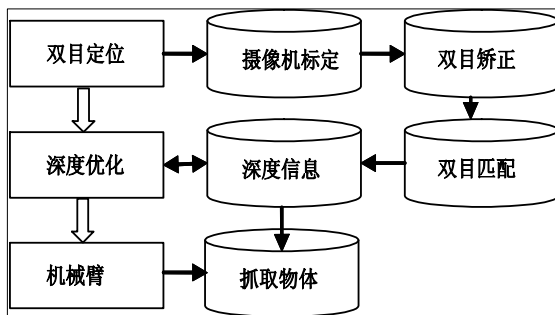


图 1 机器人抓取目标物体流程

1 系统概述

本系统总体分成四部分:双目定位、深度优化、抓取目标物体.系统框架如图 1 所示.

双目定位主要是给机器人提供场景中的三维坐标信息,主要包含摄像机标定、双目矫正、双目匹配、深度信息获取这四个方面.

深度信息优化部分是主要将双目系统获取的深度信息进行优化,然后将优化后的准确的深度信息传输给机器人来准确定位目标物体.

机器人将物体的坐标信息提供给机械臂,驱动伺服电机来完成准确的定位,从而完成抓取动作.

2 双目定位

双目定位的基本原理是三角测量原理.

图 2 讲的是双目视觉模型,摄像机焦距 f 是摄像机的内部参数.通过摄像机标定以及优化校准后,就可以得到最终的摄像机焦距 f .通过双目标定后,从旋转矩阵中可以读取两个摄像机的光心距离 T .深度信息可以通过三角测量原理,解三角方程得到. $X_l(X_r)$ 是物体 P 在焦平面上所成的像点到左(右)光轴的距离.所以基于三角测量原理,那么深度坐标信息 Z 可以有如下公式计算得到:

$$\frac{T - (X_l - X_r)}{Z - f} = \frac{T}{Z} \Rightarrow Z = \frac{f * T}{X_l - X_r} \quad (1)$$

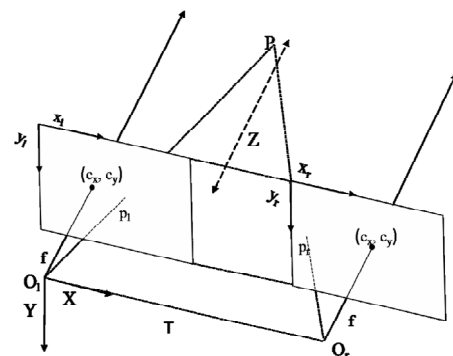


图 2 双目视觉模型

2.1 摄像机标定

摄像机标定准确是后续双目精确定位的基础,所以摄像机标定这一步非常关键.双目摄像机标定首先需要将两个摄像机进行单独的标定;然后左右摄像机同时进行标定^[5].双目摄像机进行标定后,不仅能够得到经过优化的各个摄像机内外参数,而且还能够得到两个摄像机之间的位置关系,即右摄像机相对于左摄像机的旋转矩阵 R .本文采用了 8×8 的棋盘来作为矫正参照物.

C_x 、 C_y 、 f_x 以及 f_y 是摄像机的内部参数.其中,单个像素点在低价成像仪器上是矩形而不是正方形,所以引入了两个不同的焦距 f_x 和 f_y ; C_x 和 C_y 为摄像机中心在光轴的偏移量.

$$s \begin{pmatrix} X \\ Y_f \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & C_x & 0 \\ 0 & f_y & C_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = M_1 M_2 \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (2)$$

M_1 是 3×4 的内部参数矩阵, s 是一个尺度因子. R 是一个 3×3 的旋转矩阵, t 是一个三维平移向量. M_2 是一个 4×4 的摄像机外部参数矩阵. 其中, 有三种坐标系: (X_w, Y_w, Z_w) 是世界坐标系; (X, Y, Z) 是摄像机坐标系; 而 (X, O, Y) 是图像坐标系, 是用毫米来刻画的; (X_f, O_f, Y_f) 是基于图像坐标系, 是用像素来刻画的. 等式(2)将对应的点从图像坐标系转换到世界坐标系中^[6]. 图 3 为摄像机标定示意图.

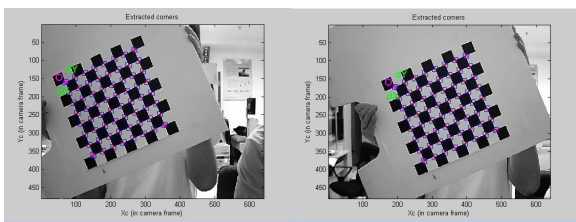


图 3 摄像机标定示意图

2.2 双目校正

两台摄像机完全准确的共面和行对准的成像平面, 现实中是不存在. 所以就显示了双目矫正的目的: 两台摄像机的图像平面重投影, 使得它们精确落在同一个平面上, 而且图像的行完全地对准到前向平行的结构上. 通过这样的校正, 两台摄像机的光轴(或者主光线)就是平行的, 即所谓在无穷远处相交^[16].

双目校正主要有两种方法: 一是 Hartley 算法, 它只使用基础矩阵来生成非标定立体视觉; 二是 Bouguet 算法^[6], 它使用两台标定摄像机的旋转和平移参数. Hartley 算法可以通过单个摄像机记录的运动推导出立体结构. 在可以使用标定模式的情形下, 如机器人臂或者安全摄像机装备上, Bouguet 算法更简单自然. 我们应用在机器人抓取上, 首选 Bouguet 算法^[12].

为了使图像重投影畸变最小化, 将右摄像机图像平面旋转到左摄像机图像平面的旋转矩阵 R 被分离

成图像之间的两部分, 我们称之为左右摄像机的两个合成旋转矩阵 r_l 和 r_r . 为了计算将左摄像机极点变换到无穷远并使极线水平对准的矩阵 R_{rect} , 创建一个由极点 e_1 方向开始的旋转矩阵. 让主点 (c_x, c_y) 作为左图像的原点, 极点的方向就是两台摄像机投影中心之间的平移向量方向:

$$e_1 = \frac{T}{\|T\|}, \quad e_2 = \frac{[-T_y T_x 0]^T}{\sqrt{T_x^2 + T_y^2}}, \quad e_3 = e_1 \times e_2 \quad (3)$$

这个矩阵将做图像绕着投影中心旋转, 使得极线变成水平, 并且极点在无穷远处. 计算校正后的左右摄像机矩阵 M_{rect_l} 和 M_{rect_r} , 与投影矩阵 P_l 和 P_r 一起返回:

$$P_l = M_{rect_l} P_l \begin{bmatrix} f_{x_l} & a_l & c_{x_l} \\ 0 & f_{y_l} & c_{y_l} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (4)$$

$$P_r = M_{rect_r} P_r \begin{bmatrix} f_{x_r} & a_r & c_{x_r} \\ 0 & f_{y_r} & c_{y_r} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

其中, a_1 和 a_2 是像素畸变比例, 现在的摄像机中几乎为 0.

应用 Bouguet 校正方法即可生成图 1 中的理想立体构型. 为旋转图像选择新图像中心和边界从而使叠加视图的面积最大化. 大体上来说, 这正好生成一个相同的摄像机中心和两个图像区域 共有的最大高度和宽度作为新的立体视图平面^[12]. 图 4 与图 5 为左右摄像机分别捕捉到的图片. 图 6 就是双目摄像机获取图像的校正图.



图 4 左摄像头图片

图 5 右摄像头图片

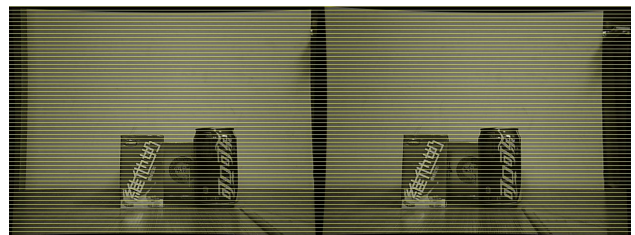


图 6 校正图片

2.3 双目匹配

双目匹配是利用两个摄像机的视图,将两个摄像机视图的公共部分(即视图重叠区域)才可以匹配计算出视差^[4].在双目匹配之前获取图像特征点.没有统一的理论可以应用在特征提取上.双目匹配采用了很强约束匹配方法^[7],用一个 SAD(“绝对误差累计”)窗口,快速水平遍历两个摄像机视图的公共部分,从而寻找匹配点.这个算法只是查找两幅图像中的强匹配点,也就是说在弱纹理场景中(例如室内的走廊),只有少数匹配点需要计算^[9].对于处理非畸变的校正立体图像,块匹配^[8]立体匹配算法如下图 7.

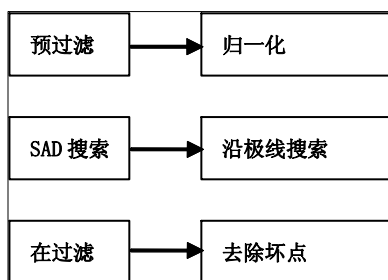


图 7 块匹配步骤

2.4 深度信息获取

双目匹配会获得目标物体像素点在左右图像中的视差.为了得到 X 和 Y 坐标,我们需要知道 C_x 和 C_y .我们能够通过双目标定以及双目校正获得一些参数的初始值 f , T , C_x 和 C_y ^[10].根据公式(1)即可推出场景中的深度信息图.在我们能到了三维坐标后,机器人能够根据坐标信息抓取相应的目标物体.



图 8 经过双目匹配得到的深度信息



图 9 经过优化后的深度信息

3 深度信息的优化以及分析

经过第二部分的研究,我们可以得到场景中的三维场景深度信息(如图 8 所示),实现过程中执行的速度较快,但是深度信息并不是非常理想.如果上面得到深度信息反馈给机器人,那么势必造成机器人的很多误操作^[13].

为了减少机器人的误操作,现在我们提出了一种方法对场景中的深度信息进行优化清理,算法思路大致如下:

1) 双目匹配算法中运用了 BM(Block Matching)算法得到视差信息^[9];

2) 将视差值做归一化,利用重投影矩阵将每个像素 (x, y) 坐标和视差 (x, y, d) 转换到匹配的三维点坐标;

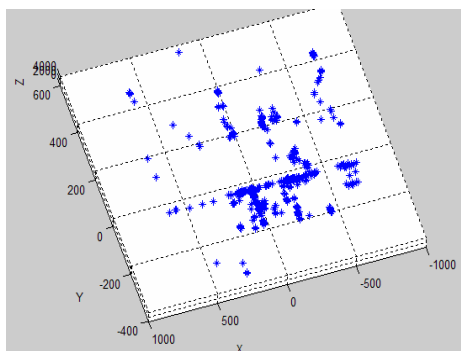
3) 遍历整个原始深度图(图 7),图像大小为 640×480 ,即将会产生 307200 个点云,将像素对应的三维物体点坐标 (X, Y, Z) 的 Z 坐标值存在矩阵中;设定一个阈值 $d=100$,并去掉 100cm 以后的所有坐标点的取值,并且赋予无穷远;

4) 设定一个误差范围 $\beta \in [0, 2]$,按照实验值进行设定块大小的值 $s=800$,设定一个初值变量 b ,将第一个像素值赋给变量 b ,从第一个像素值开始查找它周围的 8 个像素值,与初值进行减法,如果它们的差值的绝对值在 β 范围之内,则计数加 1.在整幅深度图中,用这个方法进行遍历,若计数值大于 s ,则这个区域的深度值保留,且为连通的;相反,如果计数值小于 s ,则将块中的数据全部深度值赋成无穷远.

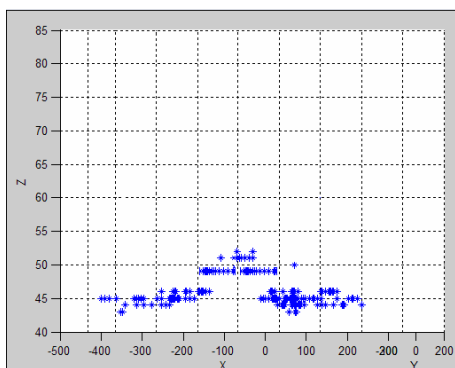
5) 再进行遍历,很容易就优化后的深度图(图 8),得到三个物体块的在场景中实际代表距离 Z_{real} ,每次将最小的 Z_{real} 值传递给机器人,上述方法从分类聚集中获得启发.

从图 8 与图 9 的深度信息图可以清楚的看到,图 9 中小噪点已经基本消除.这样就可以使机器人根据优化后的深度信息,得到这个物体前表面的深度信息,从而进行更加准确抓取操作.将上述算法用点云图(如图 10)的形式进行显示,优化与优化后点云有了明显的优化,并且可以在投影中看到物体到双目摄像头光轴的距离.清楚的可以看到可乐瓶稍微比利乐瓶近一点,它到光轴的距离大致为 45cm,而茶叶盒到光轴的距离大约为 50cm.本算法可以容易的判断出离双目摄像机水平光轴最近的 Z 距离值,从而顺利地抓取离光

轴最近的目标物体. 其成功率可以提升至 92% 以上. 具体数据可以参看下文的表 1.



(a)未进行处理的点云图



(b)经过深度优化后的点云图

图 10 进行深度优化前与优化后的点云图

4 算法实现结果以及分析

本文中的实验部分共做了 100 次机器人抓取可乐瓶的实验, 其中 50 次实验, 机器人根据没有优化的深度图中的信息, 提取坐标值进行抓取操作; 剩余的 50 次实验, 机器人采取了先优化深度信息, 再根据坐标信息抓取目标物体的方式. 实验结果列表如表 1 所示.

表 1 机器人抓取最近的目标物体的实验数据表

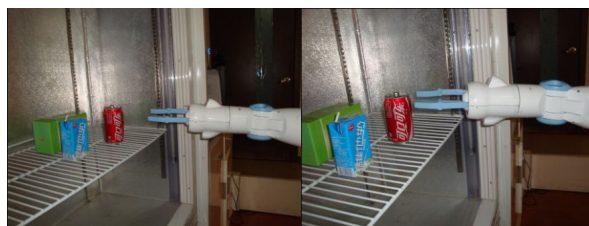
场景	成功次数	失败次数	成功率
深度信息优化的实验方案	46	4	92.5%
未经深度信息优化的实验方案	22	28	44%

第一种情况, 机器人抓取被遮挡的目标物体 21 次, 失败 28 次; 失败的主要原因是不清晰的深度图, 传递给机器人的错误目标物的坐标值; 第二种情况, 机器人抓取目标物体成功 46 次, 失败 4 次. 其中 3 次是由于实验室的光线太暗的情况下, 出现了定位物体错误以及匹配失误, 另 1 次是由于机器人的机械手臂的触点太长干扰了算法的执行(经过后续的程序调整, 之后的实验机器人都可以完成既定抓取动作).

图 10 中的机器人是我们实验室搭建的家庭服务机器人, 在上面安装了双目摄像头, 来验证我们方案的有效性与否. 图 11 是机器人在多目标场景下按照本文提出的深度优化方案来抓取目标物体的结果示意图, 抓取步骤从(a)到(e). 结合表 1 的统计结果, 我们可以看出, 在深度信息进行优化后, 机器人抓取目标物体(可乐瓶)的成功率比深度信息未处理时的成功率提高了一倍左右. 这可以充分证明, 本文提出的这种深度信息优化方法是行之有效的.



(a)



(b)

(c)



(d)

(e)

图 11 机器人抓取物体的结果图

5 结论

本文提出了对于多目标场景下双目视觉的深度信息优化处理方案, 主要涉及两个方面: 其一, 大致描述了双目定位系统的内容; 其二, 是一个新的针对多目标场景的深度信息优化方法; 其三, 结合机械臂伺服电机, 将通过双目视觉算法得到的坐标位置指令与机械臂的运动以及操作结合在一起, 完成准确的抓取

操作. 双目定位系统是机器人的智能化的一个基本构件. 优化深度信息, 能够解决机器人定位不准确的问题, 对于机器人的准确抓取定位有很大帮助, 显著的提高机器人抓取目标物体的成功率. 尤其在复杂的多目标场景下, 机器人仍然能够较好的完成目标物体抓取操作, 具有较强的现实意义.

从第 4 部分的实验中, 可以看出, 通过优化深度信息后, 管家机器人(SIAT)^[11]抓取目标物体成功率提高了一倍之多. 所以利用这种方法, 使得机器人智能抓取目标物体具有比较高的精确度. 但是文中还没有很好的解决物体的上下高度与左右宽度问题. 如果要用到更为复杂的环境中, 可以认为机器人学习还是会准确的多. 但是训练的时间会长很多, 如何权衡利弊, 这也是我后期继续要研究的方向.

参考文献

- 1 Wen G, Xilin C. Computer Vision: Algorithm and System Principle. Beijing:Tsinghua University Press, 1999.
 - 2 Li H, Chen YL, Chang T, Wu X, Ou Y, Xu Y. Binocular vision Positioning for Robot Grasping. to appear in Proc. IEEE International Conference on Robotics and Biomimetics 2011.
 - 3 Forsyth D, Ponce J. Computer Vision: A Modern Approach. Prentice Hall,2002.
 - 4 MÜhlmann K, Maier D, Hesser J, Männer R. Calculating dense disparity maps from color stereo images. an efficient implementation, international Journal of Computer Vision, 2002,47(1):79–88.
 - 5 Zhang Z. Flexible camera calibration by viewing a plane from unknown orientations. Proc. IEEE International Conference on Computer Vision, 1999, 666–679.
 - 6 Bouguet JY. Perona P, Camera calibration from points and lines in dual-space geometry. Proc. European Conference on Computer Vision, 1998. 2–6.
 - 7 Geiger A, Roser M, et al. Efficient Large-Scale Stereo Matching. Computer Vision--Asian Conference on Computer Vision, 2010,25–38.
 - 8 Wikipedia. OpenCV, <http://opencv.willowgarage.com/wiki/>.
 - 9 Tao TF, Koo JC. A Fast Block Matching Algorithm for Stereo Correspondence. Proc. International Conference on Computational Intelligence and Security. 2008. 38–44.
 - 10 Rosin PL. Measuring corner properties. Computer Vision and Image Understanding, 1999,73(2):291–307.
 - 11 Han L, Wu X, Liu G, Chen C, Ou Y, Xu Y. An efficient and low-cost robot grasping system household environment. Proc. World Congress on Intelligent Control and Automation (WCICA). 2011, 593–598.
 - 12 Gary B, Adrian K. Learning OpenCV. O'Reilly Mdia. Inc, 2008.
 - 13 Chuan L, Chen Y, et al. Occluded Object Grasping Based on Robot Stereo Vision, unpublished, WCICA,2011.
-
- (上接第 112 页)
- 11 Rao J, Su X. A survey of automated web service composition methods. Lecture Notes in Computer Science, 2005,3387: 43–54.
 - 12 Narayanan S, McIlraith A. Simulation, verification and automated composition of web service. Proc. of the Eleventh International World Wide Web Conference. Honolulu, USA,2002,77–88.
 - 13 Bucchiarone A, Gnesi S. A survey on services composition languages and models Proc. of International Workshop on Web Services Modeling and Testing, Palermo. Berlin: Springer-Verlag Press, 2006. 51– 63.
 - 14 吴哲辉. Petri 网导论.北京:机械工业出版社,2006.1–312.