

文章编号: 1000-5862(2019)05-0462-07

基于 OpenGL 的目标物 3 维位姿估计研究

张磊^{1,2} 郭志阳¹ 李红兵³ 戴丽娟^{1,4} 张兴国¹

(1. 南通大学机械工程学院, 江苏 南通 226019; 2. 约克大学拉德松学院机械工程系, 加拿大 多伦多 M3J 1P3;
3. 上海交通大学仪器科学与工程系, 上海 200240; 4. 德州大学阿灵顿分校心理系, 美国 德克萨斯州阿灵顿 76019)

摘要: 针对目标物在 3 维空间中的位姿估计存在获取困难、稳定性差和精度偏低等问题, 采用将虚拟信息与实际图像信息相对应的方式, 提出了一种基于 OpenGL 的目标物 3 维位姿估计方法。从理论上分析了虚拟与实际成像模型之间的关系, 利用 OpenGL 构建了虚拟空间的模型; 在技术上基于 VC 环境引入 OpenGL 库, 通过虚拟视觉跟踪将目标物的 OpenGL 模型在显示窗口的投影匹配于实际目标图像, 以匹配状态下的虚拟模型的位姿作为实际目标物的 3 维位姿。通过“最小化重投影误差”目标函数优化 3 维位姿值。实验结果表明, 该方法可借助虚拟空间获取目标物位姿, 可避免采用标定法实现的繁琐与不稳定问题, 而且位姿估计精度相比传统方法更高, 达到了优良水平。

关键词: OpenGL; 虚拟现实; 位姿估计; 外参数标定

中图分类号: TP 391 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2019.05.04

0 引言

计算机视觉中的一项重要任务是获得目标物在物理空间中的位置与姿态。对于一台摄像机而言, 目标物 3 维位姿是指目标物在摄像机坐标系下的位置与姿态, 一般以目标物相对摄像机的平移向量和旋转矩阵表达。目标物的位姿估计在计算机视觉涉及的较多领域中有着重要应用价值。无人驾驶中需要实时对障碍物进行位姿的定位, 从而有效避让^[1]。波士顿动力机器人 Atlas 通过对带有标记的目标纸箱进行位估计进而实现抓起动作^[2]。在军事上, 无人战斗机需要不断对目标进行识别和跟踪^[3]。在 3D 电影中, 通过虚拟现实技术将虚拟的特效模型实时跟踪真人动作进而实现炫目的特效^[4]。

关于 3 维位姿估计方法的研究历程, 最早在 19 世纪 80 年代, M. A. Fischler 等^[5]首先提出了 PnP 问题, 该问题是为了求解给定控制点与投影中心的距离, 利用摄像机采集一幅目标图像求解物体坐标系与摄像机坐标系之间的位置姿态情况。到 19 世纪 90 年代, 以美国 B. Bhanu^[6]为代表的学者基于 AutoCAD 绘图技术重构目标物的 3D 轮廓, 来估计目标

物 3 维位姿。接着, C. Harris^[7]提出了 RAPID 方法, 建立基于模型投影与图像边缘特征最小化数学模型进行位姿估计。到了 21 世纪, Song Kaitai 等^[8]提出了一种新的利用 CAD 模型来估计目标物的方法。目前这类方法在 3 维估计研究中受到更多关注和应用。如 C. H. Chen^[9]、D. Mehta 等^[10]利用卷积神经网络捕捉人体的 3D 骨骼轮廓进行人体姿态估计。

在 3 维估计中, 也有一类是采用标定的外参数估计法进行 3 维估计^[11-12], 然而存在需要解决靶标与目标物如何对应的问题, 操作有些繁琐; 其次标定法往往对目标物的图像特征精度敏感, 即图像特征的稍微定位误差会使估计的目标物 3 维位姿误差明显增大, 使得基于标定法的 3 维估计稳定性较差, 带来较大误差。因此探讨精度高、鲁棒性强、操作方便的 3 维估计方法具有很高的实用性。

由于在虚拟空间中信息可较为容易获取, 且信息稳定, 偏差较小, 能较好地避免目标图像信息获取时所带来的误差。OpenGL 作为一种可视化的 3D 虚拟图形库, 其强大的 3 维仿真功能既可以创建虚拟空间又可以实时获取虚拟模型在虚拟空间中的位姿。因此, 利用 OpenGL 创建虚拟空间来估计目标物 3 维位姿具有较好的可行性。

收稿日期: 2019-03-16

基金项目: 江苏省“六大人才高峰”(2015-ZBZZ-023), 上海市自然科学基金(18ZR1421300), 上海市高新技术课题(181511108202), 2017 年度江苏省政府留学奖学金和 2018 年度江苏省政府留学奖学金资助项目。

作者简介: 张磊(1973-)男, 江苏南通人, 副教授, 博士, 主要从事机器人视觉和触觉的研究。E-mail: zhangleint@ntu.edu.cn

1 基于 OpenGL 的 3 维估计研究方案

本文设计的 3 维估计方案主要由 OpenGL 虚拟空间的搭建、目标物图像特征检测、虚拟现实视觉匹配以及位姿的矩阵获取 4 个部分组成。首先, 标定出摄像机内参数, 并以此参数结合虚拟与现实 2 者模型的相关参数关系配置 OpenGL 投影参数, 构建与实际投影完全一致的虚拟投影空间。通过 3DMax 软件 3 维重构目标物 3DS 模型, 将虚拟模型导入虚拟空间。接着, 由摄像机采集图像, 采用线搜索式的角点检测方法^[13]提取角点, 并在 OpenGL 虚拟成像面上将其实际目标特征 3 维重构。然后, 虚拟视觉跟踪控制虚拟模型以基于角点坐标的匹配策略匹配实际目标特征点坐标, 获取最优匹配下虚拟模型 3 维位姿 M 。图 1 中的摄像机外参数 M_w 是通过目标物与标定板相帖合得到, 作为参照精确值与本文提出方法的 3 维估计值 M 相比较。具体流程如图 1 所示。

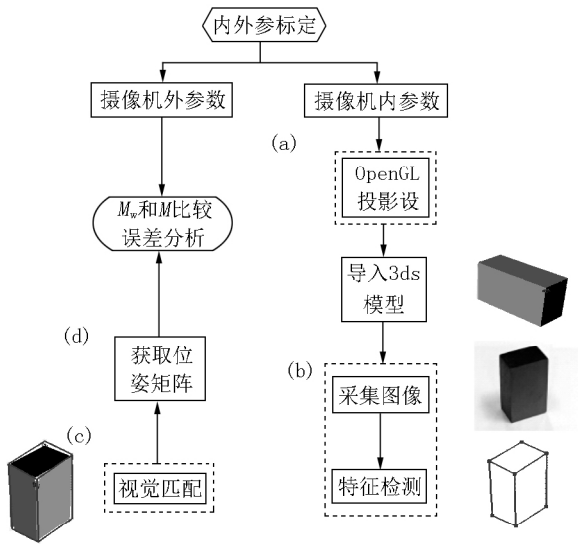


图 1 3 维估计整体方案流程

2 基于 OpenGL 的 3 维估计模型对比分析

本文从 OpenGL 透视投影和针孔摄像机成像的 2 者模型着手, 探讨 OpenGL 虚拟成像与实际成像之间的关系。由于摄像机成像模型常基于针孔模型^[14] 空间某点的世界坐标与图像像素坐标一般表示为

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} k_x & 0 & u_0 & 0 \\ 0 & k_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_{in}^c M_w [x_w \ y_w \ z_w \ 1]^T. \quad (1)$$

在 OpenGL 中, 一个虚拟的物体也存在类似的模型变换与投影关系, 其原理主要实现步骤如图 2 所示。3 维物体上的坐标系上的一点 $(x_w \ y_w \ z_w \ 1)^T$ 经过模型变换、投影变换、透视剔除以及视图变换得到显示在计算机区域的屏幕像素坐标为 $(u \ v \ 1)$ 。

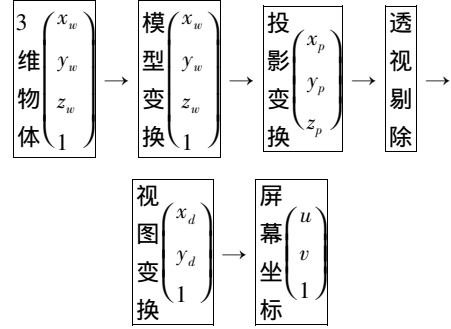


图 2 OpenGL 透视投影成像过程
OpenGL 透视投影模型最终可表示为

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = V \cdot P_0 \cdot M_0 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}, \quad (2)$$

其中 V 为视图矩阵, P_0 为 OpenGL 中模型向视图投影的透视投影矩阵, M_0 为模型本身在其空间的姿态, 类似于摄像机外参数矩阵, 具体含义参见文献 [15], 此处直接列出矩阵为

$$V = \begin{bmatrix} s_{crX}/2 & 0 & u_0 & 0 \\ 0 & s_{crY}/2 & v_0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$P_0 = \begin{bmatrix} \cot(f_{xy}/2) & 0 & 0 & 0 \\ a_{xy} & \cot(f_{xy}/2) & 0 & 0 \\ 0 & 0 & \frac{f_{cr} + n_{cr}}{f_{cr} - n_{cr}} & -2\frac{f_{cr} \cdot n_{cr}}{f_{cr} - n_{cr}} \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad (3)$$

$$M_0 = R \cdot T = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}.$$

设摄像机成像的焦距为 f , 成像平面的物理宽高为 $l_x \times l_y$, 单一像素点的物理表达单一像素点的物理表达为 $dX \times dY$, 水平像素点数可表达为 l_x/dX , 以及垂直方向的的像素点数表达为 l_y/dY . 其中在摄像机模型中 $k_x = f/dX$, $k_y = f/dY$, 则 (1) 式可转换为

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/dX & 0 & u_0 & 0 \\ 0 & f/dY & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} M_w \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_{in}^c \cdot$$

$${}^cM_w [x_w \ y_w \ z_w \ 1]^T. \quad (4)$$

根据图3的OpenGL透视投影成像图,为保证实际摄像机成像与 OpenGL 成像参数一致,首先设置 OpenGL 的透视投影矩阵 P_0 中 $n_{ear} = f$ 假设图像中心 O_1 图像物理坐标为 $(0 \ 0)$ 像素坐标表示为 $(u_0 \ v_0)$.

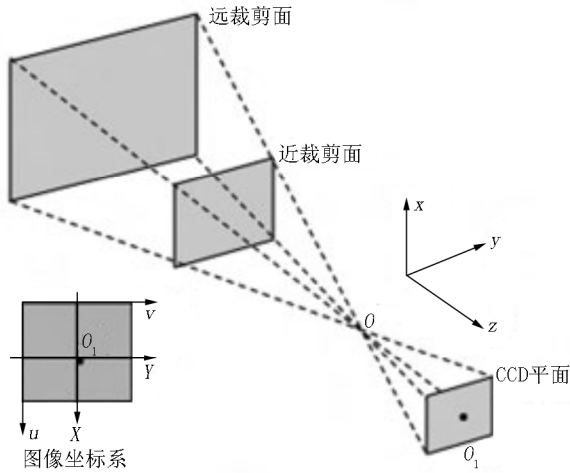


图3 OpenGL 透视投影成像模型

参数 $u_0 \ v_0 \ \mu_{spect}$ 可分别表示为

$$u_0 = l_x/2dX \ v_0 = l_y/2dY \ \mu_{spect} = l_x/l_y, \quad (5)$$
$$\cot(f_{ovy}/2) / \mu_{spect} = 2f/l_x \ \cot(f_{ovy}/2) = 2f/l_y.$$

将(5)式代入(3)式,则透视投影矩阵 P_0 可表示为

$$P_0 = \begin{bmatrix} 2f/l_x & 0 & 0 & 0 \\ 0 & 2f/l_y & 0 & 0 \\ 0 & 0 & -\frac{f_{ar} + n_{ear}}{f_{ar} - n_{ear}} & -2\frac{f_{ar} \cdot n_{ear}}{f_{ar} - n_{ear}} \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad (6)$$

从而视图矩阵 V 可表示为

表1 目标物角点图像坐标与虚拟重构坐标

	0	1	2	3	4	5	6
图像坐标	64	96	106	75	174	206	214
	322	294	228	254	386	364	296
OpenGL	-192.0	-168.0	-160.5	-183.7	-109.5	-85.5	-79.5
虚拟坐标	61.5	-19.5	-9.0	10.5	109.5	93.0	42.0

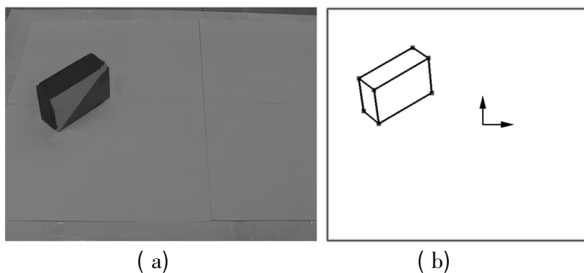


图4 目标物的图像角点检测效果和 OpenGL 的重构特征线

$$V = \begin{bmatrix} l_x/2dX & 0 & u_0 & 0 \\ 0 & l_y/2dY & v_0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

将(6)~(7)式代入(2)式,最终 OpenGL 成像模型可表达为

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f/dX & 0 & u_0 & 0 \\ 0 & f/dY & v_0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} M_0 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (8)$$

比较实际摄像机模型(4)与 OpenGL 透视投影成像模型(8)可发现二者高度一致,只需适当配置虚拟投影参数即可使得与实际投影参数一致.故 OpenGL 透视投影完全可用于模拟实际摄像机成像.

3 目标物的 OpenGL 3 维虚拟重构

在本文的实验中,采集目标物在任意一位姿下的图像并通过线搜索式角点检测算法进行特征角点检测,得到角点图像坐标.将相关角点图像坐标转换为通过 OpenGL 中所表达的虚拟特征点并以点线连接形式将目标物轮廓和角点虚拟重构,形成 OpenGL 重构的图像特征线.目标物图像角点及其相对应的 OpenGL 虚拟角点坐标值见表1.图像处理的效果如图4(a)所示;OpenGL 重构的图像特征线如图4(b)所示.图5为所开发的 VCOpenGL 创建虚拟环境的界面,本文研究的目标是找到虚拟模型的位姿矩阵使虚拟模型(如图5为 OpenGL 创建的虚拟目标物模型)的投影与图5所示的重构图像特征线相匹配.

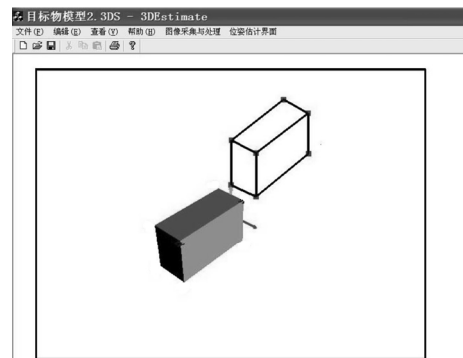


图5 VC 环境下 OpenGL 虚拟重构的模型与图像特征线

4 位姿矩阵的求解

目标物初始的位姿矩阵可由目标物的图像特征点根据传统 Faugeras 方法初步确认, 由该初始位姿矩阵, 以及目标物各角点世界坐标, 可以获得初始投影. 开始时虚拟投影与图像特征线存在偏差, 如图 6 所示. 故需要建立一个匹配原则来实现位姿匹配. 在经典的基于边缘的 RAPiD 方法^[16]中, 该方法是基于模型与实际中目标物边缘的误差最小化建立目标函数. 由于该方法存在边缘点之间难以准确对应, 因此精度受到影响. 本文在该方法的基础上, 提出了一种基于模型角点坐标与实际目标物角点图像坐标建立最小二乘目标函数的方法, 通过虚拟视觉跟踪控制模型角点逼近实际图像角点, 求解目标函数最小值.

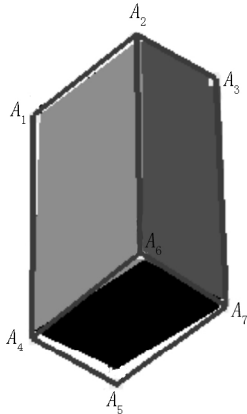


图 6 OpenGL 创建的虚拟模型与重构的特征线相匹配
首先建立目标函数

$$F(x) = \sum_{i=1}^N r_i^2 = \|g(C_w) - b\|^2,$$

其中 r_i 为每个控制点的模型投影与实际对应图像点之间的距离; i 为控制点 (本文将角点作为控制点) 序号, 在此目标物中控制点数共 $N = 7$ 个, C_w 表示 (8) 式中的世界坐标, $g(C_w)$ 表示模型上的所有控制点在 (8) 式表达的射影变换下在图像上的投影; b 表示图像上的所有控制点图像坐标, 其值由本文角点提取方法^[13]得到. 本文目标即利用“最小化重投影误差 (Bundle adjustment)^[17]”最小得到优化的位姿矩阵 M , 可表达为

$$M = \arg \min_{M_0} \sum_{i=1}^N r_i^2 = \arg \min_{M_0} \|g(C_w) - b\|^2, \quad (9)$$

$\arg \min$ 表示符合最小值的位姿 M , 与反函数性质类似. 将其转化为待求的 M 中未知参数向量 p 的函数表达形式 $f(p) = g(C_w)$, 则 (9) 式转化为

$$p = \arg \min_p \|f(p) - b\|^2. \quad (10)$$

直接解析求解式 (10) 的未知数 p 是有困难的,

为方便求解位姿矩阵 M 中的旋转矩阵 R 和位移向量 t , 首先, 将 (8) 式表达为 $m = KMC_w/s$, 其中 K 为摄像机内参数, 投影矩阵 $P = KM/s$ 以动态的形式表达为 $P = KM_t M_t$, 其中 M_t 为初始估计位姿矩阵,

M_t 为位姿变化矩阵, 也采取 $\begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix}$ 的表达形式. 于是当前更新位姿矩阵与前一步的位姿关系可表达为

$M_t = M_{t-1} M'_{t-1}$, 此处 t 代表当前更新的位姿矩阵编号. M'_{t-1} 为优化的每步之间的位姿变化矩阵. 问题转化为求解 M'_{t-1} 中的旋转矩阵与位移向量的参数.

其次, 引入李群代数表达形式^[21], 将 M'_{t-1} 基于群 $SE(3)$ 的 4 阶方阵形式, 以 6 维的李群来表达. $SE(3)$ 确切地表达了系统输出的姿态空间, 构建代表 x, y, z 轴移动与旋转的一组基如下:

$$\begin{aligned} G_1 &= \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, & G_2 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \\ G_3 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, & G_4 &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \\ G_5 &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, & G_6 &= \begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

将 M'_{t-1} 视为群中的元素, 由这些基通过指数映射 $M'_{t-1} = \exp(\sum_r \alpha_r G_r)$ 来获得, 此处 i 表示每个基及其因子的编号. 考虑到 M'_{t-1} 表示的是相邻姿态下的变化, 这种变化很微小, 可以用 1 阶线性表达的方式, 即 $M'_{t-1} \approx I + \sum_r \alpha_r G_r$, 此刻问题转化为求解未知参数向量 $\alpha = (\alpha_1 \ \alpha_2 \ \alpha_3 \ \alpha_4 \ \alpha_5 \ \alpha_6)^T$.

设第 i 控制点的当前图像像素坐标为 m_i , 有 $m_i = KM_{t-1} M'_{t-1} C_{w_i}$. 前后 2 次位姿之间控制点的距离 $l_i = n_i^T KM_{t-1} (\sum_{r=1}^6 \alpha_r G_r) C_{w_i}$. 对于 N 个控制点, 该线性变换简记为 $A\alpha = l$, 其中距离向量 l 是由 N 个 l_i 构成的列向量. 于是问题转化为超定方程 $A\alpha - l = 0$ 的线性最小二乘求解问题, 有关目标函数为

$$\min F(\alpha) = \min \sum_{i=1}^N d_i^2 = \min (A\alpha - l)^T (A\alpha - l), \text{ 可}$$

通过广义逆求解最小二乘解得

$$\alpha = (A^T A)^{-1} A^T l. \quad (11)$$

α 获得后, 即获得 M'_{t-1} , 而后更新 $M_t = M_{t-1} M'_{t-1}$. 此后可以重新量取距离向量 l , 更新 A 矩阵, 再依据 (11) 式更新 α , 直至 l 不再发生变化 (前后 2 步之间小于给定误差值), 得到最终的 M . 因此前述方法实际实施时一般采取迭代优化过程. 本文在位姿匹配算法中, 通过模型与图像点的匹配来获取目标物位姿, 当前仅在线性最小二乘方法基础上优化, 非线性优化将在今后的研究中深入展开.

5 实验数据的采集与对比验证

5.1 目标图像特征检测与位姿匹配效果

图 7 是目标物在 4 个不同姿态下的图像; 图 8 是 OpenGL 的特征线重构效果, 图 9 是 4 组目标物在前述基于视觉跟踪的位姿矩阵求解方法下的 OpenGL 虚拟模型与目标物 (特征线) 匹配后的效果图.

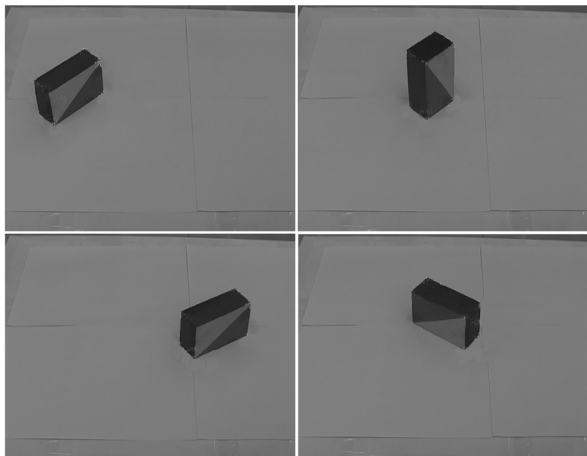


图 7 目标图像及角点检测效果图

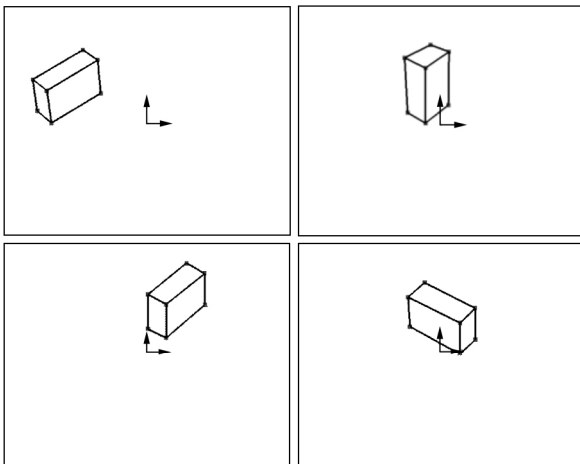


图 8 目标图像 OpenGL 虚拟重构

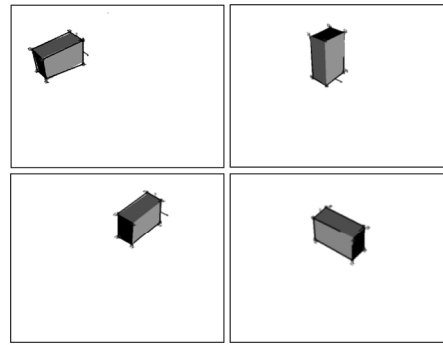


图 9 位姿匹配效果图

5.2 位姿结果验证方法

为验证经过虚拟视觉跟踪匹配后所获得的目标物位姿精度, 通过与张正友外参数标定法^[18]获得的参数相对比验证本文方法 3 维位姿估计的精度. 由于张正友标定算法只能对标准靶标进行标定, 并不是对目标物的位姿测量, 但通过将目标物与标定板相贴合并设置与目标物坐标系一致的外参数坐标系, 可以以张正友的标定结果作为标准参考值来验证所估计的目标物位姿. 本文将目标物与标定板位置贴合并设置与目标物坐标系一致的外参数坐标系如图 10 所示; 去掉目标物以张正友方法求解外参数的某位姿如图 11 所示. 目标物世界坐标系按图示 xyz 轴指定.

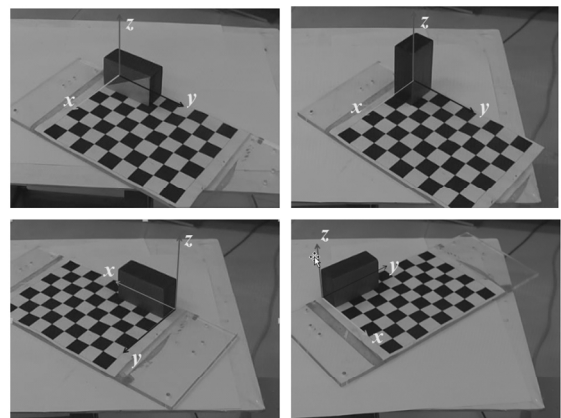


图 10 标定板与目标物位置贴合并设置

image points(+) and projected grid points(0)

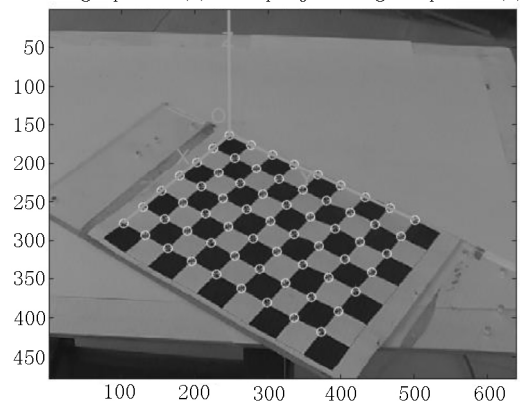


图 11 标定靶标处于位姿一的外参数标定

5.3 位姿矩阵的参数值比较

表 2 和表 3 分别为本文方法下求解的各目标物相对摄像机坐标轴的旋转角度 α 、 β 、 γ 值与位移量 x 、 y 、 z 以及张正友方法下求解的各角度值与位移量. 在表 4 中, 本文的绝对误差的计算是以张正友标定方法作为参照标准, 即将本文的估计结果与张正友方法得出的结果相减得到的偏差作为绝对误差. 由于不同文献中因目标物不同, 绝对误差不宜直接

比较, 但相对误差可以比较. 故在表 4 中, 将本文方法的相对误差与文献 [19] 中经典 RAPID 方法以及文献 [20] 中基于颜色的方法各自获得的实验结果作为对比. 从比较结果可见, 本文方法的相对误差均优于 2 种传统方法. 研究过程中发现, 由于本文方法是以已接近目标的虚拟特点与图像特点逼近, 所以实时产生的数据不发生明显抖动, 稳定性较好, 且算法很快收敛于终值.

表 2 本文方法求解的目标物在摄像机坐标系的各角度值与位移量

	回转角 $\alpha/(^{\circ})$	俯仰角 $\beta/(^{\circ})$	偏转角 $\gamma/(^{\circ})$	x 方向位移/mm	y 方向位移/mm	z 方向位移/mm
位姿 1	144.316	42.896	40.241	-97.8	-26.8	1 051.2
位姿 2	141.358	43.141	35.242	-57.1	-17.4	1 038.3
位姿 3	128.848	-25.740	-22.212	16.2	6.4	1 011.2
位姿 4	136.816	33.963	32.879	-230.4	-9.5	1 021.3

表 3 张正友方法求解的靶标外参数值

	回转角 $\alpha/(^{\circ})$	俯仰角 $\beta/(^{\circ})$	偏转角 $\gamma/(^{\circ})$	x 方向位移/mm	y 方向位移/mm	z 方向位移/mm
位姿 1	144.906	44.442	40.784	-98.364 320	-27.290 450	1 042.555 43
位姿 2	141.841	41.993	35.854	-59.163 630	-16.295 370	1 036.770 19
位姿 3	129.516	-26.280	-22.466	16.862 028	5.785 154	1 009.659 92
位姿 4	135.518	35.576	33.407	-232.648 700	-9.486 831	1 019.947 52

表 4 目标物位姿估计误差对比

位姿参数		x/mm	y/mm	z/mm	$\alpha/(^{\circ})$	$\beta/(^{\circ})$	$\gamma/(^{\circ})$	总体平均误差
RAPiD 法 ^[17]	平均相对误差/%	16.96	13.13	2.80	0.603	0.801	1.252	5.92
	平均绝对误差	2.62	3.12	9.231	0.874	0.612	0.723	
基于颜色法 ^[18]	平均相对误差/%	7.352	4.58	3.63	0.319	2.408	0.272	3.095
	平均绝对误差	2.62	3.12	9.231	0.874	0.612	0.723	
本文方法	平均相对误差/%	3.21	4.13	1.68	0.523	0.704	0.925	2.557
	平均绝对误差	2.62	3.12	9.231	0.874	0.612	0.723	

6 结论

针对传统目标物在 3 维空间的估计存在获取困难、稳定性差或精度偏低等问题, 本文着重从理论分析了摄像机模型和 OPENGL 投影模型的对应性, 从技术上实现了 VC 环境中基于 OpenGL 通过虚拟与现实相对应的方式来估计目标物在摄像机下的 3 维空间位姿. 实验结果表明: 本文的研究方案相比于 RAPID 方法和基于颜色的估计方法具有更高的精度, 能为后续视觉伺服机械手抓取目标物提供视觉基础.

本文的研究结果对于利用计算机视觉进行目标物检测、识别或操作有着良好的应用价值. 今后将进一步研究更为不规则目标物的 3 维位姿估计, 以及融入非线性的优化策略, 以提高精度.

7 参考文献

[1] Zhong Zhun, Lei Mingyi, Li Shaozi, et al. Re-ranking object proposals for object detection in automatic driving

[EB/OL]. [2018-12-19]. <https://arxiv.org/pdf/1605.05904.pdf>.

[2] Kuindersma S, Deits R, Fallon M, et al. Optimization-based locomotion planning, estimation and control design for the atlas humanoid robot [J]. Autonomous Robots, 2016, 40(3): 429-455.

[3] Kim Y, Bang H. Decentralized control of multiple unmanned aircraft for target tracking and obstacle avoidance [EB/OL]. [2018-12-23]. https://www.researchgate.net/publication/303924107_Decentralized_Control_of_Multiple_Unmanned_Aircraft_for_Target_Tracking_and_Obstacle_Avoidance.

[4] 赵沁平, 周彬, 李甲, 等. 虚拟现实技术研究进展 [J]. 科技导报, 2016, 34(14): 71-75.

[5] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [EB/OL]. [2018-12-28]. https://www.researchgate.net/publication/215458581_Random_Sample_Consensus_A_Paradigm_for_Model_Fitting_with_Applications_To_Image_Analysis_and_Automated_Cartography.

- [6] Bhanu B. CAD-based Robot Vision [J]. Computer, 1987, 20(8): 13-16.
- [7] Harris C. Tracking with Rigid Objects [M]. Massachusetts: MIT Press, 1992.
- [8] Song Kaitai, Wu Chenghe, Jiang Sinyi. CAD-based pose estimation design for random bin picking using a RGB-D camera [J]. Journal of Intelligent and Robotic Systems, 2017, 87(3/4): 455-470.
- [9] Chen Chinghang, Ramanan D. 3D human pose estimation = 2D pose estimation + matching [EB/OL]. [2018-12-28]. <https://arxiv.org/pdf/1612.06524.pdf>.
- [10] Mehta D, Sridhar S, Sotnychenko O, et al. VNect: real-time 3D human pose estimation with a single RGB camera [J]. Acm Transactions on Graphics, 2017, 36(4): 44-46.
- [11] Tsai C Y. CAD model-based 3D object pose estimation using an edge-based nonlinear model fitting algorithm [EB/OL]. [2018-12-28]. <http://teacher.tku.edu.tw/StfFdDtl.aspx?tid=6887464>.
- [12] Gavrilidis A, Stahlschmidt C, Velten J, et al. Automatic extrinsic camera parameters estimation for a mobile ToF camera application [C]//IEEE International Symposium on Intelligent Control. New York: IEEE, 2015: 41-46.
- [13] 张磊, 张志胜, 史金飞, 等. 一种快速检测图像角点特征的线搜索式方法 [J]. 自动化学报, 2010, 36(4): 509-521.
- [14] 徐德, 谭民, 李原. 机器人视觉测量与控制 [M]. 北京: 国防工业出版社, 2016.
- [15] 张祖勋, 苏国中, 郑顺义, 等. OpenGL 成像机理及其与摄影测量方位元素的相关分析 [J]. 武汉大学学报: 信息科学版, 2004, 29(7): 570-574.
- [16] Lepetit V, Fua P. Monocular model-based 3d tracking of rigid objects: a survey [EB/OL]. [2018-12-28]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.84.8498>.
- [17] Miyagawa I, Arai H, Koike H. Simple camera calibration from a single image using five points on two orthogonal 1-D objects [EB/OL]. [2018-12-28]. http://med.wanfangdata.com.cn/Paper/Detail/PeriodicalPaper_PM20129859.
- [18] Zhang Zhengyou. A flexible new technique for camera calibration [EB/OL]. [2018-12-30]. <https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/?from=http%3A%2F%2Fresearch.microsoft.com%2Fen-us%2Fum%2Fpeople%2Fzhang%2Fpapers%2Ftr98-71.pdf>.
- [19] 张磊, 许映秋, 徐海黎. 基于角点特征的 3D 视觉精确跟踪研究 [J]. 光电子技术, 2012, 32(4): 251-257.
- [20] Kyriakoulis N, Gasteratos A. Color-based monocular visual-inertial 3-D pose estimation of a volant robot [J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(10): 2706-2715.
- [21] 张磊, 张兴国. 基于李群代数表达帧间位姿变化矩阵的 3D 视觉跟踪研究 [J]. 江西师范大学学报: 自然科学版, 2012, 36(5): 466-471.

The 3D Pose Estimation of the Object Based on OpenGL

ZHANG Lei^{1,2}, GUO Zhiyang¹, LI Hongbing³, DAI Lijuan^{1,4}, ZHANG Xingguo¹

(1. School of Mechanical Engineering, Nantong University, Nantong Jiangsu 226019, China;

2. Department of Mechanical Engineering, Lassonde School of Engineering, York University, Toronto M3J 1P3, Canada;

3. Department of Instrument Science and Engineering, Shanghai Jiaotong University, Shanghai 200240, China;

4. Department of Psychology, University of Texas at Arlington 76019, USA)

Abstract: In the research of pose estimation for an object in the space, most existing methods have the problem of low precision and difficulty in direct acquisition. A 3D pose estimation method based on OpenGL is proposed by matching the image information with the virtual information in this paper. The relationship between virtual and real imaging models is theoretically analyzed and the actual projection is simulated by constructing a virtual space based on OpenGL. In technical level, the actual target image is matched by virtual visual tracking model and optimized with the "Bundle adjustment". The experimental results show that the proposed method can obtain the object pose parameters from the virtual space. It avoids the instability of calibration-based methods and can achieve good accuracy in pose estimation.

Key words: OpenGL; virtual reality; pose estimation; external parameter calibration

(责任编辑: 冉小晓)