

基于 Kinect 的运动目标跟踪与三维测量^①

谭 艳, 王宇俊, 李飞龙, 葛耿育

西南大学 计算机与信息科学学院, 重庆 400715

摘要: 首先利用微软 Kinect 的图像传感器获得 RGB 图像, 并且利用 CamShift 算法实时跟踪一个选中的目标物体, 然后获得该目标物体的质心, 最后将该质心映射到 Kinect 深度传感器获得的深度图像中, 从而实时得到该运动目标质心的三维坐标. 实验结果表明, 在目标与背景的色彩空间有一定偏差的情况下, 该方法能够实时跟踪目标物体并获得目标物质心的三维坐标.

关 键 词: Kinect; 运动目标跟踪; 三维坐标

中图分类号: TP391

文献标志码: A

目前, 运动目标的跟踪是计算机视觉领域的一个研究热点, 在安全监控、手势识别、智能交通以及未知环境下移动机器人的视觉导航等领域都有非常广阔的应用前景. 针对这些不同的应用领域, 研究者提出了众多的运动目标跟踪算法, 总体上可以分为以下几种: 基于区域的跟踪、基于特征的跟踪、基于变形模板的跟踪和基于模型的跟踪 4 类^[1-3]. 考虑到光照、运算速度的问题, 本文选用了基于颜色特征信息的连续自适应均值漂移算法(CamShift 算法)来跟踪目标物体. 运动目标物体三维信息的获取主要有激光测距、单目视觉、双目视觉等方法, 然而这些方法都有它们各自的缺陷: 激光传感器设备价格昂贵, 单目视觉测量速度较慢, 双目视觉计算过程复杂^[4]. 由于这些缺陷, 本文利用了微软的 Kinect 来获取三维信息, Kinect 的输出是一个分辨率为 640×480 的 RGB 图像和红外深度图像^[5], 其深度测量技术是由 PrimeSense 公司提供的光编码(light coding)技术, 成本低且使用方便.

本文首先介绍了 CamShift 跟踪算法, 然后利用该算法获得 RGB 图像中跟踪目标的质心, 再将质心映射到红外深度图像中, 得到该质心的深度信息, 最后实验表明, 在目标与背景的色彩空间有一定偏差的情况下, 该方法能够实时跟踪目标物体并且得到物质心的三维信息. 该方法可以应用到移动机器人的导航、避障, 以及机械臂的移动物体抓取等实际问题中, 因此本文方法具有一定的研究及应用价值.

1 CamShift 跟踪算法

CamShift 跟踪算法主要通过视频序列图像中运动物体的颜色信息来达到跟踪的目的, 它包括背投(Back Projection)计算、均值漂移(Mean Shift)算法和搜索窗口的更新^[1,6]. 但是 RGB 颜色空间对光照亮度变化比较敏感^[7], 因此 CamShift 算法首先将 RGB 图像转换到 HSV 色彩空间中, 用 H 分量来表示颜色信

① 收稿日期: 2012-11-22

基金项目: 家蚕基因组生物学国家重点实验室开放课题研究基金.

作者简介: 谭 艳(1987-), 女, 重庆开县人, 硕士研究生, 主要从事计算机视觉的研究.

通信作者: 王宇俊, 教授.

息, 然后进行后续处理.

CamShift 跟踪算法的具体计算步骤如下^[8]:

- 1) 读入视频图像, 将其他色彩空间的值转化到 HSV 空间, 计算出 H 分量的 1D 直方图.
- 2) 根据获得的 1D 直方图将原始图像转化成色彩概率分布图像, 即背投(Back Projection)计算.
- 3) 初始化搜索窗口的大小和位置 (x_c, y_c) .
- 4) 计算此时搜索窗口内的质心位置. 假设 (x, y) 为搜索窗口中的像素位置, $I(x, y)$ 为投影图中 (x, y) 处的像素值. 定义搜索窗口的零阶矩 M_{00} 和一阶矩 M_{01}, M_{10} 分别如下:

$$\begin{aligned} M_{00} &= \sum_x \sum_y I(x, y) \\ M_{10} &= \sum_x \sum_y xI(x, y) \\ M_{01} &= \sum_x \sum_y yI(x, y) \end{aligned} \quad (1)$$

由此, 可以计算出搜索窗口的质心:

$$(x_c, y_c) = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right) \quad (2)$$

- 5) 调整搜索窗口的中心到质心位置;

6) 判断是否收敛(窗口中心移动到质心的距离小于预设的固定阈值, 或者达到设定的迭代次数). 如果不满足收敛条件, 则重复步骤 4) 直到收敛为止; 如果满足收敛条件, 则返回搜索结果并利用当前所得的中心位置和区域大小作为新的搜索窗口进行下一帧的搜索.

2 运动目标三维坐标提取

2.1 Kinect 介绍

Kinect 是 2010 年 11 月 4 日微软在北美正式发布的 Xbox 360 游戏机的一款廉价外设. 实际上它就是一个 3D 摄像头, 由一个多阵列麦克风、一个 RGB 摄像头、一个单色的 CMOS 摄像头和一个红外发射器组成(图 1), 其中彩色 CMOS 产生彩色图像, 红外发射器和红外 CMOS 产生深度图像.

基于 Kinect Xbox 的开发平台主要有微软的 Kinect for Windows SDK Beta 和 PrimeSense 公司的 OpenNI, 本文采用的开发平台是 OpenNI.

2.2 三维坐标获取

首先根据 CamShift 算法获得了 RGB 图像中运动目标的质心位置, 然后再将 Kinect 获得的 RGB 图像和红外深度图像进行视角修正, 使得获得的质心位置在 RGB 图像中与在红外深度图像中的像素位置相对应, 得到该质心在深度图像中的索引位置, 从而获得该点在深度图像中的像素值, 也即该点的深度值.

OpenNI 求取深度值的方法是通过深度产生器节点得到深度映像, 然后利用 depth generator 的 ConvertProjectToRealWorld 函数将投影坐标系的点转换到真实世界坐标系中^[9], 该函数的具体实现方法如下^[10]:

- 1) 获得视角值, 也即 Kinect 在水平方向和垂直方向的视角;

2) 计算深度映像图的宽度值 W 、高度值 H 与距离 d 的比值, 如图 2 所示为真实世界坐标系, 坐标原点为红外 CMOS 的中心, $X-Y$ 平面即 Kinect 所在平面, α, β 分别表示水平和垂直视角的角度值, 由此得到映



图 1 Kinect 设备

像图中一点 (x, y) 与距离 d 的单位换算比例为:

$$xTz = 2\tan(\frac{\alpha}{2}) \quad yTz = 2\tan(\frac{\beta}{2}) \quad (3)$$

3) 从投影坐标系转换到真实世界坐标系,实际上只是对 x 和 y 的值进行了转换, z 的值不会改变, x 和 y 的计算公式如下((X, Y, Z) 表示真实世界坐标系中的一点):

$$\begin{aligned} X &= (\frac{x}{outputMode.nXRes} - 0.5) * z * xTz \\ Y &= (0.5 - \frac{y}{outputMode.nYRes}) * z * yTz \\ Z &= z \end{aligned} \quad (4)$$

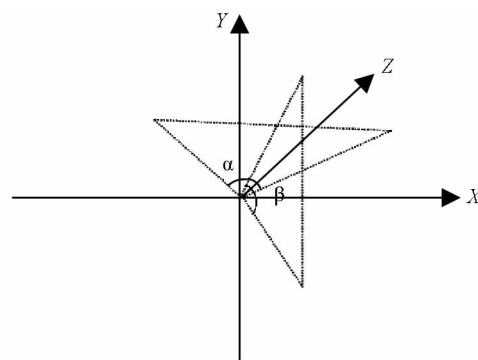


图2 真实世界坐标系

3 实验结果与分析

本文利用了 VS2010 开发平台和 OpenCV 库进行了实验验证,实验选择的跟踪目标是一个小盒子,盒子移动范围是 400 mm 左右至 2 000 mm 左右。实验结果表明,当移动目标离 Kinect 的距离小于 500 mm 时测量的数据误差非常大,而在 500 mm 至 2 000 mm 左右时测得的深度值与实际测量的距离相比,误差很小,因此,物体离 Kinect 的距离小于 500 mm 时的深度值无法测量。由于空间限制,该实验没有测量 Kinect 能测得的最大深度值,而在 Kinect 体感游戏开发网站上看到一位网友测得的最小深度值是 500 mm^[11],与本实验所测得的最小深度值相同,而他测得的最大深度值是 9 000 mm。

本实验可以实时输出运动目标质心的三维坐标,图 3 显示了某一次实验所输出的三维坐标。同时在众多的实验数据中选择了其中一次实验的第 12,34,59,76,96,123,150,166,187 帧作为展示,如图 4(a),(b),(c),(d),(e),(f),(g),(h),(i) 所示。然后将这几帧测得的实验数据和实际测量的距离结果进行了比较(表 1),从这个比较结果也可以得出如下结论:在物体离 Kinect 的距离小于 500 mm 时不能测得深度值,而当距离大于 500 mm 直至 Kinect 能测得的最大深度值时,平均误差只有 3.1 mm 左右,精确度比较好,可以将该方法运用于对物体跟踪距离不是很近的工程中。

```
Hot keys:
ESC or s - quit tracking
To initialize tracking, select the object with mouse
帧_4的三维坐标为: (40.5703,37.4495,599)
帧_5的三维坐标为: (36.806,29.0301,597)
帧_6的三维坐标为: (37.3245,16.5886,597)
帧_7的三维坐标为: (31.6282,11.9969,628)
帧_8的三维坐标为: (33.0314,11.0105,634)
帧_9的三维坐标为: (34.0248,9.87815,632)
帧_10的三维坐标为: (32.9793,10.9931,633)
帧_11的三维坐标为: (32.9793,12.0924,633)
帧_12的三维坐标为: (32.9793,10.9931,633)
帧_13的三维坐标为: (34.4555,11.1147,640)
帧_14的三维坐标为: (34.3478,11.0799,638)
帧_15的三维坐标为: (34.4017,11.0973,639)
```

图3 三维坐标显示

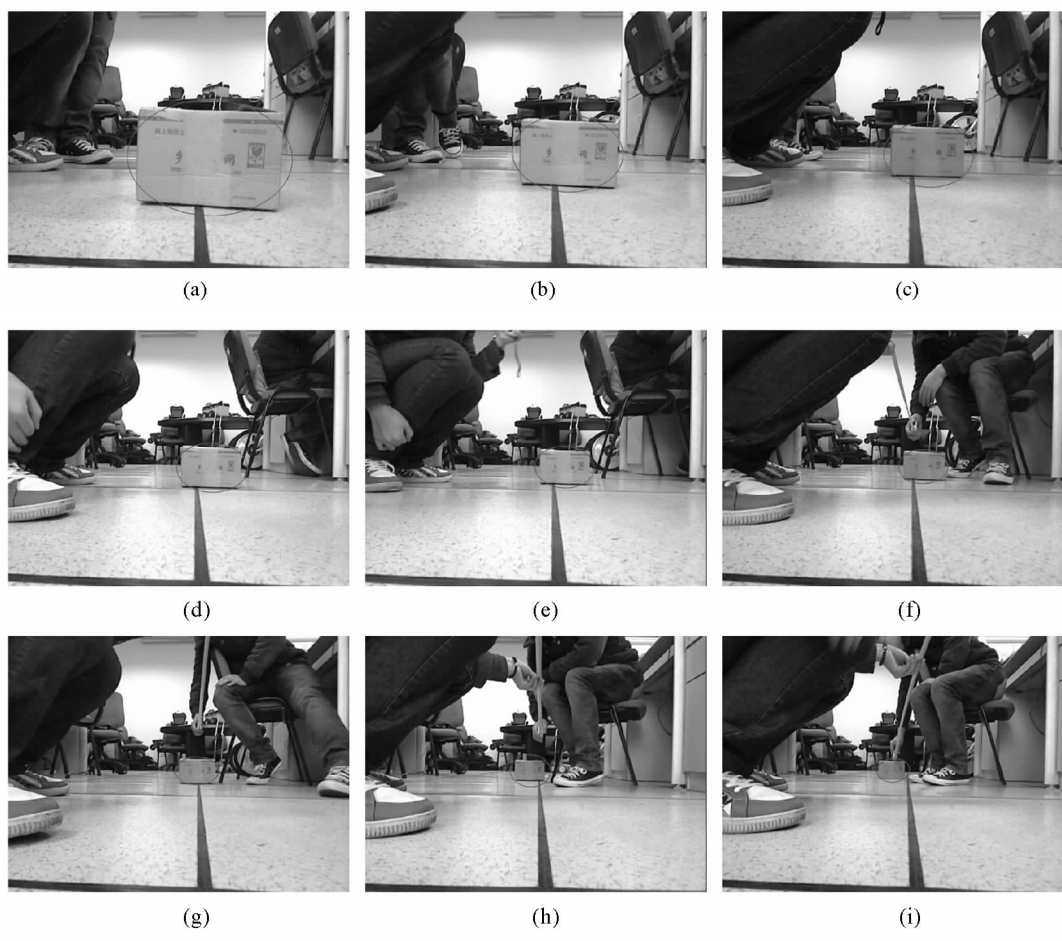


图 4 Kinect 序列图像跟踪结果

表 1 结果比较

帧序号	三维坐标/mm	实际深度/mm
12	(339.159, -82.157, 2426)	385
34	(52.639, -22.622, 501)	505
59	(107.507, -31.663, 668)	670
76	(88.896, -25.602, 819)	820
96	(91.332, -29.0212, 988)	996
123	(116.412, -66.521, 1197)	1 210
150	(79.536, -30.411, 1352)	1 360
166	(-30.069, -37.229, 1657)	1 660
187	(-1.635, -39.242, 1883)	1 885

4 结 论

本文利用了 CamShift 跟踪算法对移动目标进行跟踪,然后将 Kinect 采集到的 RGB 图像和深度图结合,利用 OpenNI 中的函数将投影坐标系下的坐标转换到真实世界坐标系中,从而得到移动物体的三维坐标。

本文方法的研究是直接选用的 CamShift 跟踪算法,因为此方法运行速度快,不受光照的影响,对本实验具有一定的优势,但是如果要将此方法运用到实际的工程中,则要根据实际的应用需求选择适当的跟踪算法。同时,本文只对三维坐标的获取进行了研究,下一步工作将是对算法进行优化,并将其应用到移动机器人的目标跟踪、避障等问题中。

参考文献:

- [1] 刘士荣,孙 凯,张波涛,等. 基于改进 CamShift 算法的移动机器人运动目标跟踪 [J]. 华中科技大学学报:自然科学版, 2011, 39(S II): 223—226.
- [2] 侯志强,韩崇昭. 视觉跟踪技术综述 [J]. 自动化学报, 2006, 32(4): 603—617.
- [3] 苏 虹,蔡自兴. 未知环境下移动机器人运动目标跟踪技术研究进展 [J]. 华中科技大学学报:自然科学版, 2004, 32(SI): 24—27.
- [4] 孙宇臣,葛宝臻,张以谟. 物体三维信息测量技术综述 [J]. 光电子·激光, 2004, 15(2): 248—254.
- [5] 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—589.
- [6] COMANICIU D, RAMESH V, MEER P. Real-Time Tracking of Non-Rigid Objects Using Mean Shift [C]//Proc IEEE Conf on Computer Vision and Pattern Recognition. New York: IEEE Press, 2000: 142—149.
- [7] BOYE M. The Effects of Capture Conditions on the CAMSHIFT Face Tracker [R]. Alberta: University of Calgary, 2001.
- [8] 邬大鹏,程卫平,于盛林. 基于帧间差分和运动估计的 CamShift 目标跟踪算法 [J]. 光电工程, 2010, 37(1): 55—60.
- [9] HERESY. Kinect + OpenNI 的深度值 [EB/OL]. (2011—12—27) [2012—09—27]. http://viml.nchc.org.tw/blog/paper_info.php?CLASS_ID=1&SUB_ID=1&PAPER_ID=278.
- [10] OPENSOURCE. Kinect 运用 OpenNI 产生点云 [EB/OL]. (2012—07—31)[2012—09—27]. <http://blog.csdn.net/opensource07/article/details/7804246>.
- [11] BEN Guo-hai. 实测 kinect 深度传感器参考数据 [EB/OL]. (2011—1—26) [2012—09—27]. <http://www.cnkinect.com/thread-508-1-1.html>.

On Moving Objects Tracking and 3D Measurement Based on Kinect

TAN Yan, WANG Yu-jun, LI Fei-long, Ge Geng-yu

School of Computer and Information Science, Southwest University, Chongqing 400715, China

Abstract: First of all, with the image sensor of the Microsoft's kinect obtains the RGB images, then Cam-shift algorithm tracking a selected object in time and obtain the object's center of mass has been studied in this paper. Finally, the center of mass is mapped to depth images and gets the three-dimensional coordinates of the movement target center of mass in time. In the color space of the target and background a certain deviation has been given. The experimental results show that the method can real-time tracking target object and get three-dimensional coordinates of the target object center of mass.

Key words: Kinect; moving objects tracking; Three-dimensional coordinate

责任编辑 张 桢