

DOI:10.3969/j.issn.1673-4785.201511050

网络出版地址: http://www.cnki.net/kcms/detail/23.1538.TP.20150113.1131.009.html

面向灵巧操作的视觉目标识别

雷丽充^{1,2}, 刘华平², 孙富春², 高蒙³, 张军惺³

(1. 中铁十七局集团有限公司 勘察设计院, 山西 太原 030032; 2. 清华大学计算机科学与技术系 智能技术与系统国家重点实验室, 北京 100084; 3. 石家庄铁道大学 电气与工程学院, 河北 石家庄 050043)

摘要: 为了能够实现灵巧手对目标物体进行精准操作, 研究了一种利用 Kinect 检测出目标物体, 在帧差法的基础上对获取的深度进行背景相减, 获取出目标物体的运动点, 在此基础上利用获取的目标物体的特征采用 T-S 模糊逻辑判断出目标物体的方法, 以 BH8-280 对目标物体进行抓取实验为例, 在实验中, Kinect 在帧差法的基础上检测出目标物体的位姿, 大小, 形状, 以此为基础建立起 T-S 模糊逻辑系统, 判断出目标物体的属性和类别, 通过实验结果进一步说明了利用本文研究的方法显著地提高了判断物体的准确率和效率, 为灵巧手的精细控制抓取奠定了基础。

关键词: Kinect; 视觉; 背景相减法; 帧差法; T-S 模糊逻辑; 目标识别; 精细控制; 灵巧操作

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2015)01-0037-06

中文引用格式: 雷丽充, 刘华平, 孙富春, 等. 面向灵巧操作的视觉目标识别[J]. 智能系统学报, 2014, 9(2): 37-42.

英文引用格式: LEI Lichong, LIU Huaping, SUN Fuchun, et al. Visual object recognition for smart manipulation[J]. CAAI Transactions on Intelligent Systems, 2014, 9(2): 37-42.

Visual object recognition for smart manipulation

LEI Lichong^{1,2}, LIU Huaping², SUN Fuchun², GAO Meng³, ZHANG Junxing³

(1. China Railway 17th Bureau Co., Ltd. Survey and Design Institute, Taiyuan 030032, China; 2. State Key Lab of Intelligent Technology and Systems, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; 3. School of Electrical and Electronics Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China)

Abstract: In order to achieve fine operation of dexterous hand, a new method is proposed. The method consists of detecting a targeted object by Kinect, conducting background subtraction for the obtained depth on the basis of frame difference method to acquire the motion point of the targeted object and judging the targeted object by the T-S fuzzy logic along with the features of the obtained targeted object. This paper uses BH8-280 for an object grasping experiment to verify the accuracy and rapidity of the method. In this experiment the pose, size and shape of the objects can be determined. The property categories of objects are judged by the T-S fuzzy logic system. It is seen that the dexterous hand can grasp the object accurately. The research of this method builds a foundation for fine operation of a dexterous hand.

Keywords: Kinect; visual; background subtraction; the frame difference; T-S fuzzy logical; object recognition; fine control; smart manipulation

目前, 随着灵巧手在各个方面的应用越来越广

泛, 对灵巧手的研究越来越深入, 尤其对控制系统投入了大量的研究, 但是对控制的研究也是以鉴别出目标物体为前提的。因此, 对目标识别的智能识别研究也成为了研究热点之一。

目标识别的智能识别研究就是在计算机领域内

收稿日期: 2013-11-20. 网络出版日期: 2015-01-13.

基金项目: 国家自然科学基金重大国际合作研究资助项目(61210013);
国家自然科学基金资助项目(60621062, 90716021).

通信作者: 雷丽充. E-mail: lei.lichong@163.com.

模拟生物的视觉系统功能对目标物体进行辨认。目标识别是利用摄像机对场景的记录,获取到环境状态信息,通过信息、数据的变化采集到运动区域,通过对数据的分析寻找到感兴趣的范围。近年来,学者们一直在努力探索,在文献[1-3]中阐述了目标识别近年来在军事上和民间的发展和应用。文献[4]中采用了双目系统对目标物体进行识别,文献[5]采用了多传感器对目标物体的识别,文献[6]主要提出了有效的地标视觉提取和识别的方法,这些研究极大地促进了目标识别的发展,为后续任务打造了坚实的基础。

本文提出了将背景相减法和帧差法相结合^[7]的方法将目标物体识别出来。然后利用 T-S 模糊逻辑方法^[8]确定出目标物体的属性,进而,灵巧手根据不同物体对应不同力值的方法对目标物体施加恰当的力值,使灵巧手能够对目标物体进行控制。

1 目标识别

目标识别实际上是将一个目标物体从一个特定的环境中提取出来的过程。从而判定目标物体的类别和类型^[9]。针对固定摄像机的拍摄,可以采取很多方法检测识别出目标物体。对于颜色较明显的物体可以采用抓取颜色处理的办法进行处理,对于其他的没有颜色特征的物体,常用的目标检测方法有 3 种:背景相减法、帧间差分法、光流法^[10]。

在摄像机^[11]是固定的,目标物体运动到特定环境中静止的情况下,背景相减法只能大体上确定出目标物体的运动范围,不能判定出目标物体是否已经停止。帧差法^[12]只能确定相邻图像之间的区别,对动态环境有很好的适应性,但是位置不准确,不能完全提取出运动目标的所有相关点。

面对上述传统的目标检测方法的不足,文中做如图 1 算法改进。

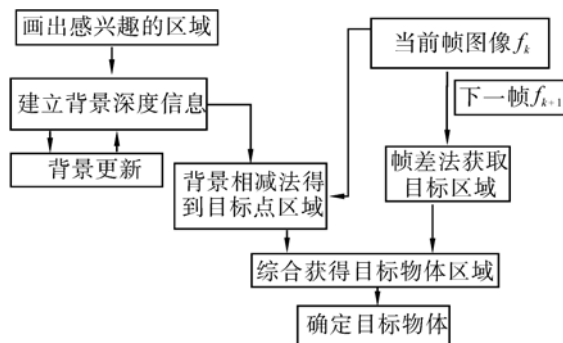


图 1 算法流程

Fig.1 Algorithm flow

图 1 中将背景相减法和帧差法相结合并且加入深度信息,首先利用帧差法确保提取出来目标物体

为运动的目标,最后根据目标物体的深度信息判定出目标物体的类别和类型。

1.1 前景提取

本文将背景相减和帧差法相结合的方法避免了两者方法的缺点,将优点结合在一起获得了极好的效果。背景相减法可以快速地检测到运动物体,并且容易实现,效果也较好。

背景相减法的研究内容包括背景初始化、背景更新、目标物体相关点的检测、噪音等虚假影响等等。“背景初始化”研究的是关于背景模型初始参数获取问题;“目标物体相关点的检测”研究的是如何在相关范围内检测出目标物体运动的序列,从而检测出运动的对象;“噪音等虚假影响”研究的是在检测出来的结果中剔除不是目标物体的像素,提高检测目标物体的准确率。

背景相减法可以分 4 部分。

背景初始化 检测环境中最原始的环境参数;背景更新:在检测目标物体时上一时刻的视频序列帧;目标物体:在检测的环境范围内感兴趣的目標物体;假象:由噪声等引起的图像变化,例如:边缘深度浮动等。

系统将背景初始化为没有放入物体的背景深度数值。根据式(1)确定目标物体的区域:

$$S_k(z) = |N_k(z) - B_k(z)| \quad (1)$$

式中: $B_k(z)$ 为背景图像上第 k 个点的深度值, $N_k(z)$ 为当前帧上第 k 个点的深度值,根据式(2)将得到的 $S_k(z)$ 的值与选取的阈值进行判断,从而改变图像的深度值:

$$G(k) = \begin{cases} 4000, & S_k(z) < \delta \\ 0, & S_k(z) \geq \delta \end{cases} \quad (2)$$

式中: δ 为设定的阈值,当 $G(k) = 4000$ 时表明没有目标物体,当 $G(k) = 0$ 表明有运动物体进入。

但是当人手将目标物体放入环境中时运动的物体有没有停止,人手有没有影响到检测目标物体的精确性等等,在这种情况下,背景相减法就不能够满足需要。因此,文中在此基础上添加了帧差法和深度的元素。

帧差法是计算相邻 2 帧图像 $f_k(x_i, y_i)$ 和 $f_{k+1}(x_i, y_i)$ 的差值,通过差分图像来快速检测到目标物体。具体算法描述如下:

设当前图像序列为 $f_k(x_i, y_i)$, 其中 (x_i, y_i) 为图像中像素点的位置坐标。当下一帧图像出现时获取图像的序列 $f_{k+1}(x_i, y_i)$, 得到差值图像 $d_{\Delta k}(x_i, y_i)$, 将其二值化后得到二值化图像 $b_{\Delta k}(x_i, y_i)$:

$$d_{\Delta k}(x_i, y_i) = f_{k+1}(x_i, y_i) - f_k(x_i, y_i) \quad (3)$$

$$b_{\Delta k}(x_i, y_i) = \begin{cases} 0, d_{\Delta k}(x_i, y_i) < T \\ 1, d_{\Delta k}(x_i, y_i) \geq T \end{cases} \quad (4)$$

根据式(3)、(4),对于进入环境后静止的目标物体,当目标物体运动稳定后当前帧图像和上一帧图像同一位置的像素点应该一致,因此通过选取适当的阈值进行二值化后得到的差值图像持续为 0 时标明运动物体运动停止。

1.2 模糊逻辑

用背景相减和帧插法相结合的方法只能检测出目标物体的形状、大小、位姿,却不能具体说明此目标物体的属性。因此,可以利用目标物体的形状、大小、位姿建立 T-S 模糊逻辑,从而判断出目标物体的属性和类别。

T-S 模糊逻辑的基本原理是:若模糊规则为 R^l :如果 x_1 为 F_1^l, \dots , 且 x_n 为 F_n^l , 则 y_1 为 B_l ; $l=1, 2, \dots, M$, 则模糊系统的输出为 $y^l = c_0^l + c_1^l x_1 + \dots + c_n^l x_n$, 则 T-S 系统为

$$y(x) = \frac{\sum_{l=1}^M \sum_{j=0}^n c_j^l x_j \prod_{i=1}^n F_i^l(x_i)}{\sum_{l=1}^M \prod_{i=1}^n F_i^l(x_i)} \quad (5)$$

式中: F_n^l 代表物体要满足的元素之一。

T-S 模糊逻辑的基本框图如图 2 所示。

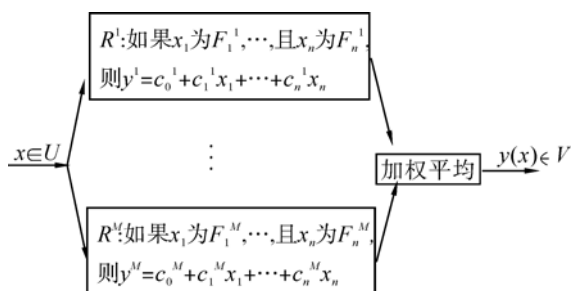


图 2 T-S 逻辑的基本框图

Fig.2 Basic block diagram of T-S logical

2 触觉建模

本文采用的是美国 BarrettHand 公司生产的 BH8 系列中的 BH8-280。灵巧手总共有 4 个触觉传感器^[13-15],每个传感器上有 24 个电容细胞。它的测量范围可以达到 10 N/m。为了能够抓取各种大小、形状的物体,本文选用饮料杯子、QQ 宠物、手机盒、小龙、圆口杯子作为代表性的物体进行建模。

灵巧手的触觉传感器上对目标物体最大的力值如表 1 所示。

3 实验结果

文中采用的摄像机是微软公司生产的 Ki-

nect^[16], Kinect 拥有分辨率为 320×240 , 16 位色深, 30FPS 的彩色摄像机和分辨率为 640×480 , 32 为色深, 30FPS 的深度摄像机。将 Kinect 放置到天花板上, 从上面俯视地面, 如图 3 所示, 显示了 Kinect 摄像的背景图, 图像的左半边是 Kinect 彩色摄像机显示的彩色图, 右半边是深度摄像机获取环境的深度信息。图 4 是背景初始化时的图像显示。当有目标物体出现时, 检测出来的图像如图 5 所示。

表 1 灵巧手抓取不同目标物体的力值

Table 1 Force value of different targets grasped by the dexterous hand

目标物体	灵巧手种类	灵巧手状态	单个单元需要力的最大值/N
饮料瓶子	BH8-280	正常	70
QQ 宠物	BH8-280	正常	78
手机盒	BH8-280	正常	65
小龙	BH8-280	正常	500
圆口杯子	BH8-280	正常	60

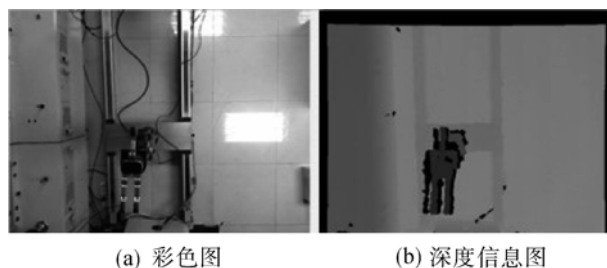


图 3 背景图

Fig.3 Background image



图 4 背景初始化

Fig.4 Background initialization



图 5 检测出来的目标物体

Fig.5 Detected objects

利用目标识别原理需要识别出备用的目标物体饮料杯子、QQ 宠物、手机盒、小龙、圆口杯子,如图 6。



图 6 备用目标物体

Fig.6 The target object

本文利用 Kinect 检测出的 5 种目标物体的效果图如图 7~11 所示。深度图像通过 IR 感应器采集到连续光,然后通过解码产生。因此,目标物体放置位置与 Kinect 的距离不同,检测出来的目标形状大小也不同。图 7~11 是以目标物体距离 Kinect 摄像机 107 cm 为例的效果图。



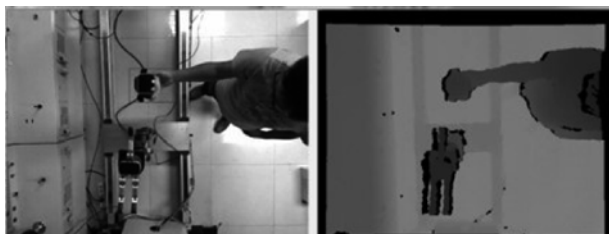
(a) 饮料瓶子的彩色图 (b) 饮料瓶子的深度图



(c) 结果图

图 7 检测饮料瓶子

Fig.7 Detection of the bottle



(a) QQ宠物彩色图 (b) QQ宠物深度图



(c) 结果图

图 8 检测 QQ 宠物

Fig.8 Detection of the QQ pet



(a) 手机盒彩色图 (b) 手机盒深度图



(c) 结果图

图 9 检测手机盒

Fig.9 Detection of the phone box



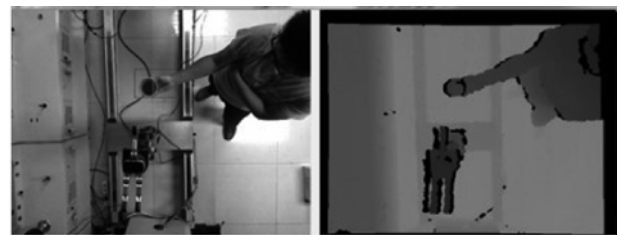
(a) 小龙盒彩色图 (b) 小龙深度图



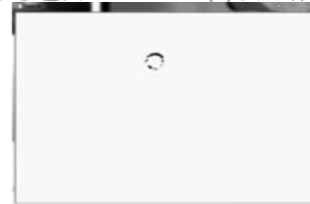
(c) 结果图

图 10 检测小龙

Fig.10 Detection of the dragon



(a) 圆口杯子彩色图 (b) 圆口杯子深度图



(c) 结果图

图 11 检测圆口杯子

Fig.11 Detection of the round cup

根据实际情况,目标物体放置位置距离传感器的距离为 127、117、107 cm。分析检测出来的

差值图像,将检测出来的目标物体画上最小外接矩阵。可以得出最小外接矩阵的周长和面积范围,如表 2 所示。

表 2 IR 传感器检测出来的目标物体状态

Table 2 Target object state detecting by IR sensor				
目标物体	检测出来的形状	目标物体离传感器的距离/cm	检测目标物体的周长/cm	检测目标物体的面积/cm ²
饮料瓶子	圆形	127	52.0	169.0
		117	64.4	259.2
		107	68.0	288.0
QQ 宠物	圆形	127	120.0	896.0
		117	142.0	1 258.0
		107	159.6	1 590.6
手机盒	长方形	127	152.0	1 363.0
		117	176.4	1 848.5
		107	190.0	2 146.0
小龙	不规则形状	127	89.1	495.997 2
		117	110.9	753
		107	113.6	807
圆口杯子	环形	127	108.5	735.2
		117	150.0	1 406.0
		107	172.3	1 855.3

因此,结合 T-S 模糊逻辑,有表 1 可以得出当检测的目标物体的周长为 52~68 cm,面积为 169~288 cm²,并且检测出来的形状为圆形时就认定此目标物体为饮料瓶子;周长为 120~159.6 cm,面积为 896~1590.6 cm²,检测出来的形状为圆形的目标物体认定为 QQ 宠物;周长为 152~190 cm,面积为 1 363~2 146 cm²,检测出来的形状为长方形的目标物体可以认定为手机盒;周长为 89~113.6 cm,面积为 495.997 2~807 cm²,检测出来的形状为不规则的形状,则认定目标物体为小龙;周长为 108.5~172.3 cm,面积为 735.2~1 855.3 cm²,检测出来的形状为环形,则认定目标物体为图 6 中的圆口杯子。

经过上述分析,可以以圆口杯子为例,当识别出来物体之后,采用美国 Barrett Technology 公司生产的 BH8-280 进行抓取,针对不同的物体,灵巧手使用不同的力值进行精细控制,参考文献[17]利用 PID 控制可以实现 BH8-280 对目标物体进行操作,其效果图(以圆口杯子为例)如图 12、13 所示。

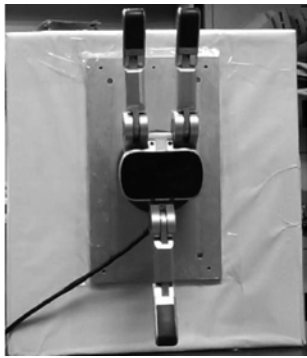


图 12 灵巧手初始化

Fig.12 Dexterous hand initialization



图 13 灵巧手抓取杯子

Fig.13 Dexterous hand is grasping the cup

4 结束语

为了能够实现灵巧手抓取目标物体对应的力值操作问题,本文研究了一种利用 Kinect 检测出目标物体,在帧差法的基础上对获取的深度值进行背景相减,获取出目标物体的运动点,在此基础上利用获取的目标物体的特征采用 T-S 模糊逻辑判断出目标物体的方法,使得 BH8-280 灵巧手抓取目标物体时能够更加智能和类人手化。利用 T-S 模糊逻辑为目标物体判断提供了理论基础,并且提高了目标物体的准确率。本文所提供的目标识别物体的方法更加智能化,扩充了目标识别的应用范围,为灵巧手更加智能化的发展提供了基础。

参考文献:

[1] 陈春玉. 目标识别技术的现状与发展[J]. 声学技术, 1999, 18(4): 185-188.
CHEN Chunyu. The present situation and developing trend of target discrimination techniques[J]. Technical Acoustics, 1999, 18(4): 185-188.
[2] 丛敏, 金善良, 罗翌. 自动目标识别技术的发展现状及其应用[J]. 飞航导弹, 1999(12): 1-9.
CONG Min, JIN Shanliang, LUO Yi. The present situation and developing trend of target automatic recognition techniques[J]. Winged Missiles Journal, 1999(12): 1-9.
[3] 舒海燕. 图像目标识别技术的研究与应用[D]. 西安: 西

- 北工业大学, 2002: 1-65.
- SHU Haiyan. The research and application of image target recognition technology[D]//Xi'an: Northwestern Polytechnical University, 2002: 1-65.
- [4] 施雷, 胡文俊. 基于主动形状模型的双目识别系统的设计与实现[J]. 计算机应用与软件, 2009, 26(3): 223-225.
- SHI Lei, HU Wenjun. Design and implementation of binocular visual recognition system based on the active shape model[J]. Computer Applications and Software, 2009, 26(3): 223-225.
- [5] 孙涛, 张宏建. 目标识别中的信息融合技术[J]. 自动化仪表, 2001, 22(2): 1-4.
- SUN Tao, ZHANG Hongjian. The technology of information merging in object recognition[J]. Process Automation Instrumentation, 2001, 22(2): 1-4.
- [6] LEE L K, AN S Y, OH S. Efficient visual salient object landmark extraction and recognition[C]//IEEE/SMC International Conference on Systems, Man, and Cybernetics. Anchorage, Alaska, 2011: 1351-1357.
- [7] 方帅, 薛方正, 徐新和. 基于背景建模的动态目标检测算法的研究与仿真[J]. 系统仿真学报, 2005, 17(1): 159-165.
- FANG Shuai, XUE Fangzheng, XU Xinhe. Moving object detection based on self-adaptive background modeling[J]. Journal of System Simulation, 2005, 17(1): 159-165.
- [8] QUANG V D, Van BAN D, HA H C. A method of object identification based on fuzzy object functional dependencies in fuzzy object-oriented databases[C]//IEEE/KES Fourth International Conference on Knowledge and Systems Engineering. Danang, Vietnam, 2012: 46-53.
- [9] 孙红岩, 毛士艺. 多传感器目标识别的数据融合[J]. 电子学报, 1995, 23(10): 188-192.
- SUN Hongyan, MAO Shiyi. Multisensor data fusion for target identification[J]. Chinese Journal of Electronics, 1995, 23(10): 188-192.
- [10] 韩峥, 刘华平, 黄文炳, 等. 基于 Kinect 的机械臂目标抓取[J]. 智能系统学报, 2013, 8(2): 149-155.
- HAN Zheng, LIU Huaping, HUANG Wenbing, et al. Service robot desktop cleaning based on Kinect[J]. CAAI Transactions on Intelligent Systems, 2013, 8(2): 149-155.
- [11] AMER A. Voting-based simultaneous tracking of multiple video objects[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(11): 1448-1462.
- [12] 崔学超. 基于减背景和混合差分的运动目标检测算法[J]. 电子科技, 2010, 23(10): 85-88.
- CUI Xuechao. Moving objects detection algorithm based on background subtraction and hybrid difference methods[J]. Electronic Science & Technology, 2010, 23(10): 85-88.
- [13] BEKIROGLU Y, LAAKSONEN J, JORGENSEN J A, et al. Assessing grasp stability based on learning and haptic data[J]. IEEE Transactions on Robotics, 2011, 27(3): 616-629.
- [14] STEFFEN J, HASCHKE R, RITTER H. Experience-based and tactile driven dynamic grasp control[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA, 2007: 2938-2943.
- [15] BEKIROGLU Y, DETRY R, KRAGIC D. Learning tactile characterizations of object- and pose-specific grasps[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA, 2011: 1554-1560.
- [16] EL-LAITHY R A, HUANG J, YEH M. Study on the use of Microsoft Kinect for robotics applications[C]//IEEE/ION Position Location and Navigation Symposium. Myrtle Beach, USA, 2012: 1280-1288.
- [17] 刘金珉. 先进 PID 控制及其 MATLAB 仿真[M]. 北京: 电子工业出版社, 2003: 67-80.

作者简介:



雷丽充, 女, 1988 年生, 硕士研究生, 主要研究方向为智能控制、模式识别、图像处理。



刘华平, 男, 1976 年生, 副教授, 主要研究方向为智能控制及机器人、计算机视觉等。



孙富春, 男, 1964 年生, 教授, 博士生导师, 博士, IEEE 高级会员, 中国人工智能学会理事、智能控制与智能管理专业委员会副主任兼秘书长。主要研究方向为智能控制、机器人与飞行器的导航与控制、网络控制系统、人工认知系统的信息感知和处理等。发表学术论文 120 余篇, 其中被 SCI 检索 52 篇。