

基于 Kinect 深度图像的指尖识别及手势判定

袁方剑¹, 王毅轩², 王毅刚¹, 杨道谈¹

(1. 杭州电子科技大学 计算机学院, 浙江 杭州 310018; 2. 伊川县交通局, 河南 伊川 471300)

摘要 在基于微软 Kinect 体感设备开发的交互应用系统中, 使用传统的鼠标、键盘等交互设备难以达到理想的效果。针对这种情况, 提出一种基于指尖识别的手势判定方法。采用 Kinect 传感器获取图像的深度信息, 通过 OpenNI 的内置模块获取手心的位置信息, 使用最近邻法实现手部的分割并对手形进行提取; 并采用射线求交法优化 Graham Scan 算法获取凸包点集合, 利用轮廓分析法从凸包点中识别出指尖。在此基础上, 结合指尖数目和面积比例实现对“抓取”手势的判定。实验结果表明, 该方法能有效地识别抓取动作的手势, 且具有良好的鲁棒性。

关键词 指尖识别; Kinect; OpenNI; 手势识别; 人机交互

中图分类号 TP391.41 **文献标识码** A **文章编号** 1007-7820(2014)08-006-05

Fingertip Detection and Gesture Identification Based on Kinect Depth Image

YUAN Fangjian¹, WANG Yixuan², WANG Yigang¹, YANG Daotan¹

(1. College of Computer Science, Hanzhou Electronic Science and Technology University, Hangzhou 310018, China;

2. Bureau of Communications of Yichuan County, Yichuan 471300, China)

Abstract In the development of interactive application systems based on the Microsoft Kinect somatosensory equipment, it is difficult to achieve desired results using traditional human computer interaction devices like mouse and keyboard. A gesture identification method based on fingertip detection is presented to address this problem. The depth information of the image is obtained using the Kinect sensor, and the position information of the palm center using the built-in module of OpenNI SDK. Then the nearest neighbor method is used to achieve hand segmentation and extract the hand shape. After that, the convex hull is obtained using the Graham Scan algorithm optimized with a ray intersection method. Finally, the fingertips are detected using the profile analysis method from the convex hull. On this basis, the recognition of the "grab" gestures is realized using the number of fingertips and the area ratio. Experiments results show that this method can effectively identify the grasp gesture and has good robustness.

Keywords fingertip detection; Kinect; OpenNI; gesture identification; human-computer interaction

在研究和工业领域, 自然人机交互(N-HCI)一直是重要的研究课题^[1], 而手势具有直观性、自然性等特点。近几年, 针对基于 Kinect 深度图像的指尖、手势识别技术这个问题已经有许多相关的研究^[2]。例如: 陈艳艳等人^[3]提出在形态学操作下, 基于形状对手掌部分进行剔除, 并结合手心位置从手指部分识别指尖的方法。这种方法的误差较大只能处理手掌展开且正对摄像头的情况, 鉴于手掌的自遮挡概率较高以及不同人手掌尺寸的差异性较大, 根据某一特定比例剔除手掌来分离手指的效果并不理想。邓瑞等人^[4]提出使用

支持向量机(SVM)进行特征训练, 并识别手势的方法。但该方法只能识别出经过训练的特定手势, 无法识别出指尖的具体位置。Feng 等人^[5]提出基于 K 均值聚类的方法将整个前臂的深度图分割成手和手臂两个区域, 分析手臂部分的方向识别出手的区域, 并根据该区域中的点到其他簇中心的距离来检测指尖的方法。这种方法可以得出指尖的位置信息, 但只能处理仅有一根手指伸出的情况, 且聚类的结果一般不够理想。

本文提出一种借助计算手形的凸包点来分析并识别指尖, 来判断手势的方法。通过 Kinect 传感器获取深度图像信息, 利用 OpenNI 手部分析模块获取手心的位置, 并结合最近邻法从深度图中分割出手的部分。提取手形轮廓, 采用射线求交法优化传统的 Graham Scan 算法, 并计算出手形的凸包点集合, 然后利用轮廓分析法对凸包点进行指尖检测。本文最后基于检测得到的指尖数目, 结合手形面积和凸包面积的比例等因素, 实现了对“抓取”动作的判定, 实验结果表明, 该

收稿日期: 2014-01-08

基金项目: 国防基础科研重点基金资助项目

作者简介: 袁方剑(1989—), 男, 硕士研究生。研究方向: 计算机图形学, 虚拟现实。E-mail: space3344@126.com。王毅刚(1973—), 男, 博士后, 博士生导师。研究方向: 计算机图形学, 虚拟现实。杨道谈(1989—), 男, 硕士研究生。研究方向: 计算机图形学, 虚拟现实。

方法具有良好的识别效果及鲁棒性。

1 Kinect 设备获取手部信息

1.1 Kinect 简介

2010 年 11 月,微软发布了一款针对 XBOX360 游戏机开发的体感外设,其核心是三维体感摄像机,能够实现即时动作捕捉、影像辨识、语音输入等功能,图 1 为 Kinect 外观图。

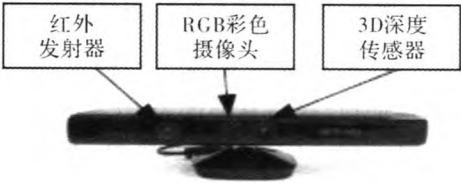


图 1 Kinect 体感设备

如图 1 所示,Kinect 主要是由红外线发射器、RGB 彩色摄像头和红外线 CMOS 摄像头所构成的 3D 深度传感器所组成^[6]。与普通摄像头相比,Kinect 的主要特点是具有 CMOS 红外传感器,通过这个传感器感知周边环境,而且不受周围光照环境的影响。红外传感器通过黑白光谱的方式来标度被测物体的深度值,即其距离传感器的距离:越黑表示距离越远,越白表示距离越近^[7]。它收集感知范围中每一点的深度信息并保存成一幅深度图像。传感器以每秒 30 帧的速度生成深度图像流,以达到对周围环境的实时监测。

1.2 OpenNI 获取手心位置

OpenNI 是一个多语言、跨平台的框架。该框架定义了编写自然交互程序所需要的 API 接口,以方便利用视觉、声音等设备进行开发。目前 OpenNI 支持的硬件包括:3D Sensor、RGB Camera、IR Camera 和 Audio Device4 类^[7]。

目前 PrimeSense 已经提供了一套 NITE 作为最主要的中间件,并提供上述 4 种功能。利用第 2 个手部分析的功能即可得到手心的位置信息,图 2 为利用 NITE 识别出的掌心位置。



图 2 NITE 识别掌心位置

识别手心需要通过挥手的动作激活。将手掌伸开,指尖朝上放在离 Kinect 一定距离(20 cm ~ 1 m)的

范围内,挥动手掌,不能过快或过慢,且幅度不能太小。NITE 识别出手掌以后,用灰色点来标识出掌心的位置,如图 2 所示。

2 指尖识别及手势判定

2.1 Kinect 坐标系

设 Kinect 捕获的深度图像尺寸为 $w \times h$ (默认 640×480 像素),则其坐标系如图 3 所示。

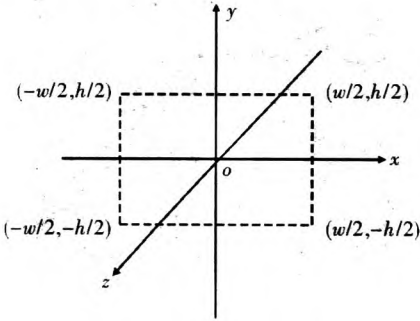


图 3 Kinect 坐标系

其中,原点为摄像头的位置; z 轴为摄像头到物体的方向。

2.2 最近邻法分割手部

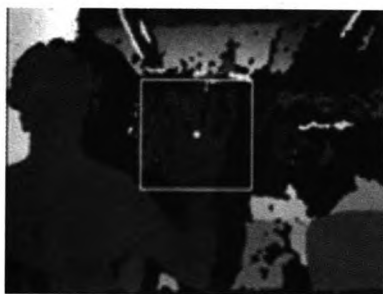
为了实现指尖识别的功能,首先需要从深度图中分割出手的部分。手掌具有一定厚度,且手心的位置之前已经利用 PrimeSense 的 NITE 提取出来,因此可以利用最近邻法从深度图像中分离出手部。

首先对 Kinect 深度图进行高斯滤波处理,其作用是对整幅图像进行加权平均,使其中每个像素的值都由周围邻域内像素点的值加权平均得到,从而减少噪声的干扰。其次分离手部深度图像,并对其二值化。Kinect 捕捉的深度图用灰度的方式表示物体离摄像头的距离。约定识别时手部伸到身体前侧,这样手的部分在深度图中的灰度值会比后方身体及周围环境的灰度小,然后根据事先获取的手心的深度值,在 z 轴方向上设定一个阈值 T ,将在手心深度加、减 T 的范围内的点按照式(1)提取并进行二值化。

$$\text{hand}(x,y)=\begin{cases}0, \text{depth}(x,y) \leq d-T \\ 0, \text{depth}(x,y) \geq d+T \\ 255, \text{otherwise}\end{cases} \quad (1)$$

式(1)中, d 为手心的深度值; T 为手掌厚度阈值; $\text{depth}(x,y)$ 为深度图像中坐标为 (x,y) 的点的深度值; $\text{hand}(x,y)$ 即为分离出的手部深度图像中坐标为 (x,y) 的点经过二值化处理后的值。

图 4 为手部分割的效果,其中图 4(a)为 Kinect 深度图,被白色矩形圈出来的为手部,图 4(b)为用最近邻法分离出来,并经过二值化的手部图像。



(a) 深度图中的手部



(b) 分割出的手部图像

图4 手部分割效果

2.3 手形的提取

从深度图中分离出手部的二值化图像以后,需要接着将手形轮廓从二值化图像中提取出来,以方便后面对凸包的计算。

本文采用了一种比较简便的方法获取手形轮廓:逐行遍历二值化图像,用一个变量 preColor 来标记前一个像素点的颜色值,每次与当前点的颜色值进行比较,如果颜色值不同,即说明当前点为轮廓上的点,将该点保存到轮廓数组中。遍历完整幅图像后,得到一个保存了所有轮廓点的集合,作为下一步计算凸包的输入。

2.4 优化 Graham Scan 算法计算凸包

本文提出的指尖识别算法基于计算凸包点实现。手掌在处于伸开状态时,通常情况下 5 个指尖均是凸包点之一,因此在识别指尖之前,需要计算出手形轮廓点集的凸包。

平面点集 Q 的凸包(Convex Hull)是一个最小的凸多边形 P ,满足 Q 中的每个点或者在 P 的边界上,或者在 P 的内部,本文中凸包用 $CH(Q)$ 来表示。对于平面中点集 Q 的 N 个点 $p_0, p_1, p_2, \dots, p_N$,其凸包 $CH(Q)$ 可以表示为

$$CH(Q) = \left\{ \sum_{i=0}^N \lambda_i p_i \mid \lambda_i \geq 0, \sum_{i=0}^N \lambda_i = 1 \right\} \quad (2)$$

式(2)中, $CH(Q)$ 为凸包; p_i 为凸包内第 i 个点; N 为凸包内点的数量; λ_i 为 p_i 的权值。

如图 5 所示,点集 $Q = \{p_1, p_2, \dots, p_{12}\}$ 的凸包为由 $p_0, p_1, p_3, p_{10}, p_{12}$ 围成的凸多边形,其中 p_0, p_1, p_3, p_{10} ,

p_{12} 为凸包点。

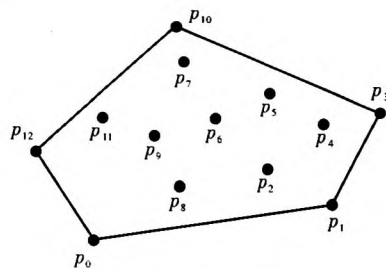


图5 点集 Q 及其凸包 $CH(Q)$

本文采用 Graham Scan 算法^[8]来计算凸包。Graham Scan 算法实现简单,时间复杂度为 $O(N \log N)$,其中, N 代表参与计算凸包点的候选点的个数,亦即在上一步提取出的手形轮廓点。由于 N 的值比较大,因此带来的计算量也会增大,考虑到本文的指尖识别需要应用到实时 3D 虚拟试衣系统中,过大的计算量可能导致系统无法达到实时的交互效果,因此在计算凸包点之前,需要对候选点集合进行优化。

本文采用基于射线求交的扫描方法对 Graham Scan 算法进行改进。基本思路为:在执行扫描算法之前,首先遍历轮廓点集合 Contour,找到最顶、最底、最左、最右的 4 个点 PT、PB、PL、PR,如图 6(a)所示。在计算这 4 个点的过程中,对于顶端和底端的点,如果有多个点有相同的纵坐标,则取横坐标最小的点;同样对于最左和最右的点作为 PT(或 PB),如果有多个点拥有相同的横坐标,则取纵坐标最小的点作为 PL(或 PR)。找到 4 个端点以后,由这 4 个点构成一个四边形 LTRB,见图 6(a)中的四边形。可以证明在此四边形内部的轮廓点均不是凸包上的点,可以将其剔除。实验表明,剔除率平均水平高达 60%,这样可以大幅提高扫描算法的效率,保障了系统的实时性。



(a) LBRT四点



(b) 射线求交

图6 利用射线扫描方法优化 Graham Scan 算法

射线扫描法的关键问题在于如何快速地判断候选点在四边形的内部还是外部。具体的思想如下:首先,设4个顶点的坐标分别为 $PT(x_1, y_1)$ 、 $PL(x_2, y_2)$ 、 $PR(x_3, y_3)$ 、 $PB(x_4, y_4)$ 。根据坐标分别写出4条直线的方程,以直线 LB 为例,其方程为

$$\frac{x - x_{PL}}{x_{PB} - x_{PL}} = \frac{y - y_{PL}}{y_{PB} - y_{PL}} \tag{3}$$

然后,对于某个候选点 $P(x, y)$,沿着平行于 X 轴的方向,向某一方向做射线,并分别求出与 LB 、 BR 、 RT 、 TL 四条直线的交点。可以证明,倘若交点个数为奇数,则说明点 P 在四边形内部;反之若为偶数,说明点 P 在四边形外部。如图 5(b) 所示,分别由点 P_1 、 P_2 水平向右做射线,求出交点个数: P_1 为奇数,在四边形内部; P_2 为偶数,在四边形外部。利用射线扫描法对候选点进行优化后,就可以运用经典的 Graham Scan 算法计算出凸包点。

2.5 轮廓分析法识别指尖

从凸包点中识别出指尖的方法主要有模板匹配和轮廓分析两种^[6]。模板匹配的效率不高,因此本文采用基于曲率特征的轮廓分析方法对指尖进行识别。

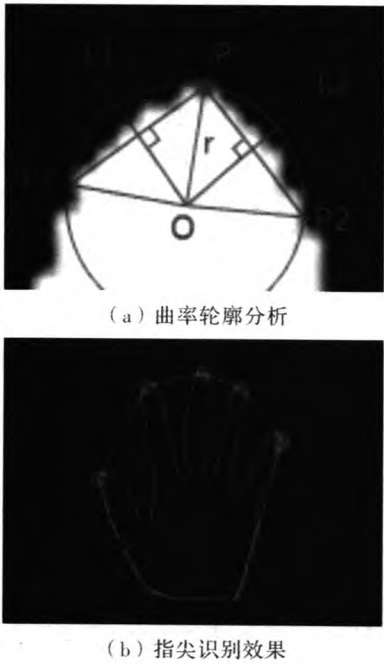


图7 指尖识别原理及效果

对每个候选凸包点,如图 7(a) 中点 $P(x, y)$,从轮廓数组 $Contour$ 中分别找出位于 P 左右两侧,且到 P 的距离均为特定长度 d 的点 P_1 和 P_2 。分别做线段 PP_1 和 PP_2 的中垂线 L_1 和 L_2 ,求出交点 O ,则线段 OP 即为过 P 、 P_1 和 P_2 三点的圆的半径 r 。以 r 来表示点 P 处的曲率特征,设定阈值 T ,如式(4)所示,当满足 r 在阈值范围内时,即判定凸包点 P 为指尖

$$r \leq T \tag{4}$$

经过实验测试,定长 d 选择约在 30 像素,阈值 T 设定约在 35 像素时,效比较理想。

2.6 手势的判定

本文的手势判定算法主要针对 Kinect 体感三维虚拟试衣软件的交互情况进行研究。在三维虚拟试衣软件中,主要用到的手势即为“抓取”手势,用户通过该手势来表达“选中”的命令,因此本文主要对“抓取”动作进行识别。抓取动作分为两个状态:手掌张开和手掌握拳,对于单独每帧的状态,本文提出根据指尖的数目和面积比例进行判断的方法进行判定。

根据提出的指尖识别算法,可以计算出每帧图像中指尖的数目 N_{tips} 。利用 OpenCV 的 $ContourArea$ 函数分别计算出手形轮廓的面积 S_{hand} 和凸包的面积 S_{ctr} ,二者的比例 s

$$s = S_{hand} / S_{ctr} \tag{5}$$

设定阈值 T_1 和 T_2 (实验中 T_1 和 T_2 分别设置为 0.82 和 2),当指尖数目 $\leq T_1$,并且 $s > T_2$ 时,判定此时状态为握拳状态;而其余情况下,则判定当前为张开状态。

在单帧手势状态判定的基础上,本文采用大小为 N 的缓冲队列对“抓取”手势进行识别。程序运行时,统计队列中握起状态的数量,当大于阈值 M (实验中设定为 8) 时,判定发生“抓取”动作。算法流程图如图 8 所示。

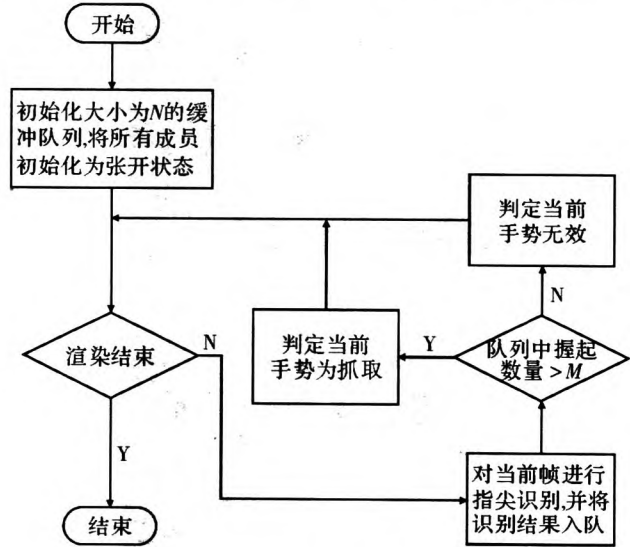


图8 手势识别算法流程图

在其他领域的研究及实际开发中,经常需要对其他类型的手势进行识别,而本文提出的基于指尖识别的手势判定方法具有一定的通用性和可扩展性,该算法可以扩展到除“抓取”手势之外的其他手势的识别上,如用手指表示罗马数字 1~5 的手势以及其他一些与指尖的数量相关的手势,使用指尖数目结合特定的判定条件即可实现对其他相关手势的判定。

3 实验结果与分析

3.1 手势判定实验

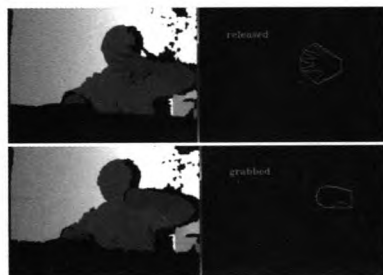
“抓取”动作发生时,手的朝向会有多种情况,图9所示的分别为在手指朝上、朝下、朝左、朝右的情况下“抓取”动作的识别。



(a) 朝上的“抓取”动作识别



(b) 朝下的“抓取”动作识别



(c) 朝左的“抓取”动作识别



(d) 朝右的“抓取”动作识别

图9 不同朝向的手势识别实验

本文分别针对手指朝上、朝下、朝左、朝右4种情况进行了实验,分别采集了10为实验对象的手势,每位实验对象针对每种姿势做5次“抓取”动作,每种情况下最终得到50次抓取实验结果。

3.2 实验结果分析

实验中,每种情况下“抓取”动作的识别次数及准备率如表1所示。实验结果表明,在手的每种朝向下,“抓取”动作的识别率都在80%以上,对于识别失败的情况,主要原因有:首先是 Kinect 体感设备返回的深度图时会出现破碎的情况,导致凸包点的识别出现错误,最终致使识别结果出错;其次,在朝下和朝左两种情况下手腕会发生偏转,导致提取出的手形轮廓在手腕处容易出现尖角,且较难过滤,因此偶尔会被指尖识别算法误判为指尖,从而导致这两种情况下手势判断的准确率相对较低,在今后的工作中还将对指尖识别算法进一步改进。总体上,本文的手势识别算法具有较高的识别率及较强的鲁棒性。

表1 4种情况的抓取手势识别率

手的朝向	朝上	朝下	朝左	朝右
识别次数/次	49	41	43	45
准确率/%	98	82	86	90

4 结束语

本文利用射线求交法对 Graham Scan 算法进行了优化,改善了凸包的计算效率。另外,本文使用轮廓分析法识别指尖,相较于模板匹配法,该方法具有更高的准确性。最后,本文提出了一种基于指尖数目和面积比例判断“抓取”手势的方法,实验表明,该方法具有良好的识别率。在今后的工作中,将对轮廓分析法进行完善和改进,提高复杂情况下的识别准确率及鲁棒性,并增加对更多其他手势的识别。

参考文献

- [1] YEO H S, LEE B G, LIM H. Hand tracking and gesture recognition system for human - computer interaction using low - cost hardware [J]. Multimedia Tools and Applications, 2013 (6): 1 - 29.
- [2] REN Z, YUAN J, MENG J, et al. Robust part - based hand gesture recognition using kinect sensor [J]. IEEE Transactions on Multimedia, 2013, 15(5): 1110 - 1120.
- [3] 陈艳艳, 陈正鸣, 周小芹. 基于 Kinect 的手势识别及在虚拟装配技术中的应用[J]. 电子设计工程, 2013, 21(10): 4 - 7.
- [4] 邓瑞, 周玲玲, 应忍冬. 基于 Kinect 深度信息的手势提取与识别研究[J]. 计算机应用研究, 2013, 30(4): 1263 - 1265.
- [5] ZHANG X, YE Z, JIN L, et al. A new writing experience: finger writing in the air using a kinect sensor [J]. IEEE Transactions on MultiMedia, 2013, 20(4): 85 - 93.
- [6] LEE C K, LEE V Y. Fall detection system based on kinect sensor using novel detection and posture recognition algorithm [M]. Germany: Springer, 2013.
- [7] YANG K - P, VAN DELDEN S, BOND E. KinectFlix: a hand gesture recognition application for kinect to watch netflix [J]. World, 2014, 2(1): 6 - 9.
- [8] ADAMS W, BELOTTI P, SHEN R. Convex hull characterization of lexicographic orderings [J]. Writing, 2013, 23(6): 719 - 726.

基于Kinect深度图像的指尖识别及手势判定

作者:

[袁方剑](#), [王毅轩](#), [王毅刚](#), [杨道谈](#), [YUAN Fangjian](#), [WANG Yixuan](#), [WANG Yigang](#), [YANG Daotan](#)

作者单位:

[袁方剑, 王毅刚, 杨道谈, YUAN Fangjian, WANG Yigang, YANG Daotan \(杭州电子科技大学计算机学院, 浙江杭州, 310018\)](#), [王毅轩, WANG Yixuan \(伊川县交通局, 河南伊川, 471300\)](#)

刊名:

[电子科技](#)

英文刊名:

[Electronic Science and Technology](#)

年, 卷(期):

2014, 27 (8)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_dzkj201408002.aspx