Transactions on Computer Science and Technology
March 2014, Volume 3, Issue 1, PP.9-14

Fingertip Detection and Gesture Recognition Based on KinectDepth Data

Binjue Zheng#, Liaoyin Zhao, Yixuan Wang

Institute of Computer Application Technology, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China

#Email:zheng0211@gmail.com

Abstract

Vision-based gesture recognition is a key technique to achieve a new generation of human-computer interaction. This paper proposes a research program for fingertip detection and gesture recognition using depth information. The depth information of an image is captured using Microsoft access sensors, and threshold segmentation combined with the k-means clustering algorithm is used to obtain the hand image. The shape of the hand external convex hull detection combined with hand external contour detection is used to determine the fingertip. Vectors between each of the fingertips and the palm as well as the relative position of the relationship between the fingers are used for finger recognition. The finger gesture recognition of digital signal of 1~5 is conducted through the classification. The experimental results show that the research scheme can accurately detect finger, and hand gestures recognition rate is more than 95%.

Keywords: Hand Gesture Recognition; Fingertip Detection; Human-computer Interaction; Depth Data

基于 Kinect 深度信息的手指检测与手势识别*

郑斌珏，赵辽英，王毅轩

杭州电子科技大学 计算机应用研究所，浙江 杭州 310018

摘　要：基于视觉的手势识别是实现新一代人机交互的关键技术。本文提出了一种使用深度信息进行手指检测和手势识别的研究方案。利用微软 Kinect 传感器获取图像深度信息，通过阈值分割和 K-均值聚类算法获取手部图像。手型外部凸包检测结合手部外部轮廓检测实现指尖的确定，根据每个指尖到掌心的矢量和手指间的相对位置关系实现手指识别，通过手指的分类实现对数字手势 1～5 的手势识别。实验结果表明，该研究方案能准确检测手指，手势识别率超过 95%。

关键词：手势识别；指尖检测；人机交互；深度信息

引言

随着计算机技术的发展，人们对人机交互有了更高层次的要求。相比较鼠标，键盘等接触式的操作方式，自然而直观的手势更符合人的自然行为，是新一代人机交互的主要方法，在虚拟现实、远程控制和交通等领域逐渐得到应用[1]。自动手势识别成为近年来的一个研究热点[2-4]。

手势识别的初期研究主要集中在机器设备上，例如数据手套，即一个类似手套的传感器。计算机通过它可以获取手部各关节数据。该方法虽能精确的识别手势，但是设备昂贵且掩盖了手势自然的表达方式[5]。为此，学者们提出了基于视觉的手势识别方法。目前研究的手势识别系统主要分为基于 RGB 图像和基于深度图像。其中前者研究时间比较长，方法众多，但依然没有一个很好的方法解决问题。与 RGB 图像相比，深度图像具有物体三维特征信息，即深度信息。由于深度图像不受光源照射方向及物体表面的发射特性的影响，而且不存在阴影，不受颜色相近信号影响，所以可以更准确地识别手势。

*国家基金资助（61171152），浙江省自然科学基金项目资助（LY13F020044）

2010 年微软推出了用于 Xbox360 的体感外设 3D 摄像机 Kinect[6]，而且还提供了支持硬件的软件开发包和应用程序开发接口。Kinect 功能强大且相对廉价，研究者广泛采用它获取深度信息。

目前基于 Kinect 深度信息的手势识别方法主要有两类。一类是基于骨架识别获取手势位置[7]进而识别手势。这类方法使用 Kinect SDK 时必须要在摄像头全身可见的条件下才能进行较为准确的识别。另一类是深度信息转换为 2D 信息[8]，通过支持向量机训练后识别。这类方法需要构造训练样本，并且不能识别手指。郭康德等人[9]提出一种基于视觉技术的手指识别，这种方法能跟踪指尖但没有涉及手势识别。

利用深度图像不受肤色的影响的特点，本文提出了一种基于 Kinect 深度信息检测手指，不需要特征提取和训练，快速实现手势识别的方法。

1 手势提取

1.1 计算实际深度

Kinect 获得的深度数据以像素为单位记录。在进行手势识别时需要根据实际距离的大小进行距离判定，Kinect 返回的深度值不是实际距离，而是一个相对数据。故首先要把以像素为单位的数据转换成以实际距离为单位的数据。设 Kinect 返回的的某一点深度值为 d_{raw}，根据 Kinect 的校准过程，可求得相应的实际距离深度值为

$$d = K\tan(d_{raw}/2842.5 + 1.1863) - O \tag{1}$$

其中：K = 0.1236 m，O = 0.037 m。

根据实际距离深度值，可求得像素位置为(i，j，d)的实际位置(x，y，z) ，其转换式

$$\begin{cases} x = \left(i - \frac{w}{2}\right) \times (d - 10) \times s \times \frac{w}{h} \\ y = \left(j - \frac{h}{2}\right) \times (d - 10) \times s \\ z = d \end{cases} \tag{2}$$

其中：s = 0.0021。

1.2 基于阈值分割获取手部图像

获得实际深度数据后，遍历深度矩阵的每个值，获取距离最小值。根据一般做手势的时候，手部在整个身体最前面且与身体有一定距离。所以在最小值的基础上加上一个固定的值作为阈值。这种方法的好处是，距离范围不固定，当手的距离变化时，获取数据的范围也会动态变化。图 1 给出了基于阈值分割获取的手部图像。



图 1 基于阈值分割获取的手部图像

1.3 基于聚类算法区分左右手

基于阈值的分割技术只能把手势从背景中分割出来，不能区分两只手。k 均值算法根据像素在 2D 空间的位置，把所有像素划分为两个聚类中心，初始化 k 为 2。每个聚类区域代表一只手。然后计算各个样本到聚类中的距离，把样本归到离它最近的那个聚类中心所在的类。计算新形成的每一个聚类的数据对象的平均值来得到新的聚类中心，如果相邻两次的聚类中心没有任何变化，说明样本调整结束，聚类准则函数

已经收敛。

在每个手部的像素点划分好后，如果在两只手的中心距离小于提前定义的一个值，两个聚类将合并成一个。图 2 给出了经过 K 均值聚类算法判断的左右手，两只手用不同的颜色表示。

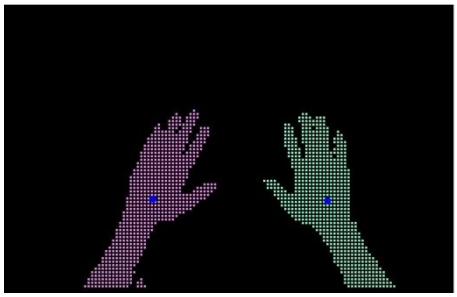

图 2 经过 K 均值聚类算法判断左右手结果

2 手指检测

2.1 凸包算法

在做手势的时候，手指经常远离手掌中心，而且指尖离的最远。根据这个情况建立数学模型。因为凸包包含指尖的点，基于 Graham's Scan 算法[10]找到手型外部的凸包。具体过程如下：

首先选取基点。在手部图像中选取 y 坐标最小的一点 H，当作基点。如果存在多个点的 y 坐标都为最小值，则选取 x 坐标最小的一点。排除坐标相同的点。如图 3 中，y 坐标最小点 H 为基点。

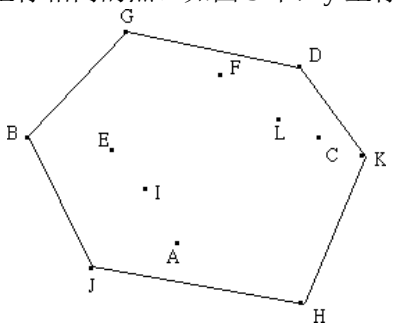

图 3 凸包算法中基点示意图

然后按照其它各点和基点构成的向量与 x 轴的夹角进行排序。当加入一点时，必须考虑到前面的线段是否会出现在凸包上。从基点开始，凸包上每条相邻的线段的旋转方向应该一致，并与扫描的方向相反。如果发现新加的点使得新线段与上线段的旋转方向发生变化，则可判定上一点必然不在凸包上。按照上述步骤进行扫描，直到点集中所有的点都遍历完成，即得到凸包。如图 4 所示，与$\overrightarrow{HC}$相邻的$\overrightarrow{HK}$和$\overrightarrow{HL}$旋转方向不一致，$\overrightarrow{HC}$被判定不在凸包上，与$\overrightarrow{HK}$相邻的线段旋转方向都一样，因此$\overrightarrow{HK}$在凸包上。

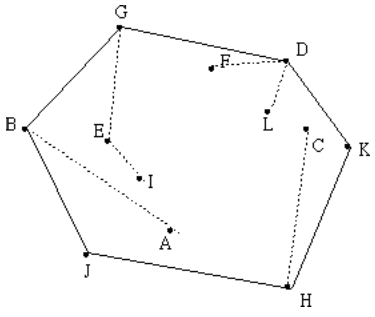

图 4 凸包算法中线段示意图

2.2 手部轮廓检测

手部轮廓检测采用 Moore neighborhood 算法[11]实现。定义 N(a)为像素 a 的八个领域像素。p 表示为当前

轮廓像素。C 表示检测到的轮廓点集合，初始化为空集合。轮廓检测算法描述如下：

1）从上面到下面，并从左到右，在荧幕上扫描所有像素，直到发现手部的像素 s 就判定为起始点。

2）设置当前轮廓像素 p 为 s，设置初始像素的领域。

3）把 p 插入 C 中，然后计算 N(p)的领域。

4）从 p 开始顺时针方向的遍历 N(p)的领域知道另外一只手的像素 r 被找到。

5）把 q 设置为 p，p 从新的轮廓像素 r 开始，然后重复步骤 3 直到遍历到起始点 s，或者检测像素超过最大值。

检测到手部的轮廓后，计算轮廓内接圆的中心坐标得到掌心。手掌中心被用来计算手指的位置。

2.3 识别手指

手指识别首先要确定指尖，通过计算那些既在凸包上又在轮廓上的点来找到指尖。算法描述如下：

1) 初始化 C 序列为既是凸包上的点又是轮廓上的点的集合。

2) 遍历 C 序列上每个点 p_0，取其左右 q_1 和 q_2 点，分别计算它们到 p_0 的矢量值为$\overrightarrow{p_0q_1}$和$\overrightarrow{p_0q_2}$。

3) 计算$\overrightarrow{p_0q_1}$和$\overrightarrow{p_0q_2}$的夹角为α_i。如果$\alpha_i < 45^o$，$i = 1 \ldots N$，则确定为指尖，否则返回 2）重新检查下一个候选点。

指尖检测示意图如图 5 所示，其中白线连接区域表示凸包，黄线连接区域表示手部轮廓，红色点表示检测得到的指尖。

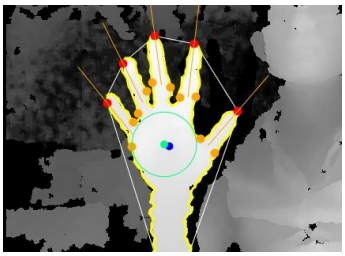

图 5 经过手指检测得到指尖的位置和手掌的中心

找到指尖点之后，可以与掌心做差值，得到从掌心到指尖的矢量。根据矢量长度和手指间的相对位置识别手指。两个手指之间距离相差最大的是大拇指和食指，小拇指离大拇指最远，中指离食指最近，剩下的是无名指。

图 6 所示的是手指识别结果，各个手指都被准确地识别出。



(a) 大拇指　　(b)中指

（c）食指　　（d）小拇指

图 6 手指识别结果

这种凸包结合轮廓的算法，比起 k-curvature 算法[12]，一开始可以排除了两个手指之间间隙这种情况。算法鲁棒性和效率更高。

3 手势识别

经过上面一系列的算法，现在我们获得了关于手指和手势的信息有手掌中心位置，手指个数，指尖的坐标和手指的向量。

这四种数据组成了手势的基本元素，通过这些数据可以识别大部分手势。可以通过计算伸开的手指的个数，实现数字手势的识别；还可以利用伸开手指的名字判断手势的意义；还可以根据伸开手指之间的角度来判断手势的意义。

4 实验结果及分析



图 7 识别数字手势 1～5

利用 Kinect 深度摄像头及 OpenNI 框架 VisualStudio2010 开发环境上研究手指检测和手势识别算法。Kinect 输入的深度图尺寸为 640×480。手势提取中设置固定阈值为 500 毫米。图 7 所示的是识别数字手势 1～5 的结果。

为验证基于 Kinect 深度信息手指检测和手势识别的有效性，每种手势做 40 次。如表 1 所示。

表 1 手势识别结果

手势	输出 1	输出 2	输出 3	输出 4	输出 5	准确率(%)
手势 1	37	2	1			92.5
手势 2		38	1	1		95
手势 3			39	1		97.5
手势 4			1	39		97.5
手势 5				1	38	95

手势数字 1 由于干扰的可能性大，所以准确率最低。如果手指没有足够伸开，可能不被识别。大拇指的下面的点由于比较尖，也可能被识别为手指，所以手势数字 5 有一次被识别为数字 6。总体来看该算法手势数字识别速度快，准确率高。

5 结束语

在本文中，我们提出了基于 Kinect 深度图的手指检测和手势识别的算法。这个系统与基于 RGB 的手势识别不同，它可以在亮度低的情况下识别，并且对识别对象的肤色，服装和背景不敏感。我们不仅能找出手指，还能检测到手掌和手指的位置，并且识别不同的手指。通过这些信息，计算机能很方便的识别手势。

与现存的方法不同，我们的方法基于深度传感器，而且不需要数据手套，数据库和机器训练的过程。我们不仅能加测到手掌中心和指尖的位置，而且我们还能准确的找到每个手指，从而可以获得更多识别手势的信息。

由于前期的手指检测获取了丰富的信息，后续工作可以增加识别手势的种类，例如胜利的手势‘V’，可以通过计算手指矢量的夹角在 30 度到 60 度之间来判断。基于这个系统，可以很方便的开发出各种应用，比如手势控制鼠标，机器手臂，幻灯片播放等等。

REFERENCES

[1] SushmitaMitra, TinkuAcharya. Gesture Recognition: ASurvey[J]. IEEE Trans. on Systems, Man, and Cybernet-ics-Part C: Applications and Reviews, 2007, 37(3): 311-324

[2] 况鹰. 基于 Kinect 运动捕获的三维虚拟试衣[D], 广州: 华南理工大学, 2012

[3] 林海波, 梅为林, 张毅, 罗元. 基于 Kinect 骨骼信息的机械臂体感交互系统的设计与实现[J]. 计算机应用与软件, 2013,30(2):157-160

[4] 刘阳, 尚赵伟. 基于 Kinect 骨架信息的交通警察手势识别[J]. 计算机工程与应用, 2013,26(6):1578-1581

[5] 江立. 基于 CAS-GLOVE 数据手套的手势识别技术研究[D]. 北京: 北京交通大学, 2007

[6] Wikipedia. (2012, Jul.) Wikipedia. Kinect-wikipedia, the free encyclopedia.[Online].Available:http://en.wikipedia.org/wiki/Kinect

[7] 林海波, 等. 基于 Kinect 骨骼信息的机械臂体感交互系统的设计与实现[J]. 计算机应用与软件, 2013,30(2):157-160

[8] 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):1263-1274

[9] 郭康德, 张明敏, 孙超, 等. 基于视觉技术的三维指尖跟踪算法[J]. 计算机研究与发展, 2010,46(6):1013-1019

[10] R. Graham., “An efficient algorithm for determining the convex hull of a finite planar set,” Information Processing Letters 1, pp.132-133,1972

[11] J.Shotton, A. Fitzgibbon, M.Cook, T. Sharp, M.Finocchio, R.Moore,A.Kipman, and A.Blake, “Real-time human pose recognition in partsfrom single depth images,” in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, june2011, pp. 1297 –1304

[12] 陈静. 基于 Kinect 的手势识别技术及其在教学中[D]. 上海: 上海交通大学, 2013