

# 基于 Kinect 的姿势识别与应用研究

谢 亮,廖宏建,杨玉宝

(广州大学,广东 广州 510006)

**摘 要:**姿势识别在三维虚拟实验中有着重重要的作用,其作用体现在通过特定的姿势,控制程序准确地响应某一个操作来完成所对应的实验中的特定功能。为了便于使用,同时使得姿势识别的准确率提高,文中提出了利用 Kinect 传感器得到的二十多个关节点信息,通过确定每个关节点之间的欧氏距离与角度来判别特定姿势的方法。结果表明,使用此方法来识别姿势,识别率较高,且可随时扩展动作库来满足判定不同姿势的需求。本算法计算较为简单,效率高,能满足日常的虚拟实验要求。其便利的姿势库扩展性,有利于软件复用,提高虚拟实验的使用效率。

**关键词:**Kinect;姿势识别;人机交互;骨骼追踪

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2013)05-0258-03

doi:10.3969/j.issn.1673-629X.2013.05.067

## Recognition and Application Research of Kinect-based Gesture

XIE Liang, LIAO Hong-jian, YANG Yu-bao

(Guangzhou University, Guangzhou 510006, China)

**Abstract:** Gesture recognition plays an important role in the three-dimensional virtual experiments, its role is to control program in responding to an action by a particular posture. For ease of use and making the gesture recognition more accurate, propose the use of Kinect sensor to get over twenty articulation points information, and determine the Euclidean distance and angle of the articular point to distinguish a particular posture. The results show that using this method can always expand action library to meet the needs of different postures of the decision. The algorithm is relatively simple to calculate and high efficiency, they are able to meet the day-to-day requirements of the experiment. It's helpful to software reuse, and improves the efficiency in the use of virtual experiment.

**Key words:** Kinect; gesture recognition; HCI; skeleton track

## 0 引言

传统的人机交互模式通过窗口、图标、菜单、指示装置等(WIMP)为大家所熟知,但是他们在操作上的限制随着技术的发展愈来愈明显。人们呼唤更自然的人机交互方式,将人与人的交互方式引入到人机交互中,被称为“自然”交互方式。这其中就包括了一系列对人体、手臂以及手势进行识别的技术<sup>[1]</sup>。在这些自然的交互方式里,又有姿势(pose)、手势(gesture)、语音等方式。姿势是人和其他物体重要区别<sup>[2]</sup>。人们通过姿势来表达一定的信息和情感,比如在体育比赛中,裁判员使用各种姿势来传递信息。姿势和手势通常会混淆,但它们是两个不同的概念。当一个人摆出一个姿势的时候,他会将身体的位置和样子保持一段

时间。但是手势包含有动作,例如用户通过手势在触摸屏上放大图片等操作<sup>[3]</sup>。

## 1 姿势识别

姿势是身体的某一部分和其他部分之间的一种静态关系。身体及各个关节点的位置定义了一个姿势,即关节点之间的相对位置就能确定一个姿势,姿势的类型和复杂度与识别算法的复杂度成正比<sup>[4]</sup>。可以通过关节点位置的重叠交叉或者关节点之间的角度进行姿势识别。

### 1.1 骨骼捕获

文中利用 Kinect 传感器来捕获用户的骨骼和姿势。Kinect 感应器有三个镜头,中间一个 RGB 彩色摄影机,左右两边镜头分别为红外线发射器和红外线 CMOS 摄影机所构成的 3D 结构光深度感应器。该传感器通过黑白光谱的方式来感知环境:纯黑代表无穷远,纯白代表无穷近。黑白间的灰色地带对应物体到传感器的物理距离。它收集视野范围内的每一点,并形成一幅代表周围环境的景深图像。接着, Kinect 会

收稿日期:2012-07-15;修回日期:2012-10-20

基金项目:广州市科信科技攻关项目(20121208001A);广州市2012年高校科研计划项目(2012A095)

作者简介:谢 亮(1983-),男,陕西岐山人,实验师,硕士,从事教育技术、计算机软件应用研究。

对景深图像进行像素级评估,来辨别人体的不同部位,采用分割策略将人体从背景环境中分离。再通过人工智能系统来处理大量的数据,识别手、脚及身体的其他部位,最后根据追踪到的 20 个关节点来生成一幅骨架系统。追踪到的 20 个关节点如图 1 所示。

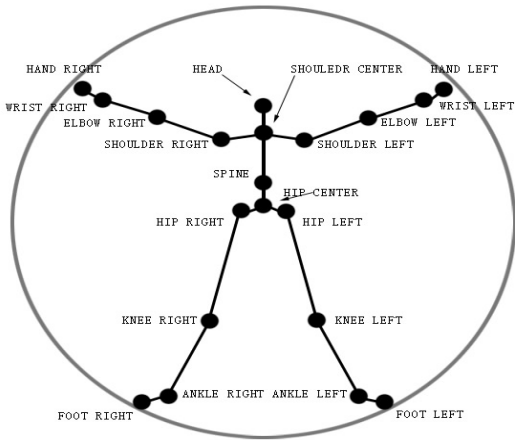


图 1 Kinect 识别到的 20 个骨骼点

1.2 算法实现

在得到了人物 20 个主要关节点以后,对姿势的识别就可以理解为对 20 个关节点的相对位置进行辨识的问题。关节点之间的相对位置可以用距离和角度来确认。如果姿势中出现关节点交叉,只要对关节点进行命中测试即可。例如双手叉腰动作,可以从骨骼追踪的数据获取左右髋关节和左右手的位置。然后计算左手与左髋关节的位置、右手与右髋关节的位置。如果它们的距离小于某一个阈值,就认为这两个点相交。

当然,并不是所有的节点交叉都需要三维坐标信息。一些简单的姿势可能只需要一个坐标平面的信息,比如立正姿势。在这个姿势中,手臂上的关节点和肩膀几乎在一个垂直坐标轴内。在这个姿势中,只需要测试手和肩部节点的 X 坐标的差值,如果在一定的阈值内就可以简单判断这些关节点处于一个平面内。为了保证识别的精确性,可能还需要判断手在 Y 坐标轴上应该低于肩部等其他信息,或者是判断所有的 20 个关节点的 Z 值是否处在近似相等的位置。

为了保证姿势识别的准确性,可以通过计算某些关节点连线之间的角度来进行特定姿势的识别。利用某两个关节点之间的欧氏距离和余弦定理,可得出两个关节点之间的夹角。假设有 点  $x$  与 点  $y$ ,则这两点之间的欧氏距离就为:

$$d(x,y)=\sqrt{\sum (x_i-y_i)^2} \quad i=(0,1,2,\cdots,n)$$

假如要判断举起右手的姿势,由于肩部和肘部在一条直线上并且和地面平行,手腕与肘部组成的连线和胳膊是垂直关系,在这个姿势中,很明显有一个直角(或者钝角)三角形。要让 Kinect 识别这个姿势,只要

判断右肩关节点 (JointType. ShoulderRight)、右肘关节点 (JointType. ElbowRight) 和右手腕关节点 (JointType. WristRight) 三个点之间的角度在一定的阈值范围内即可。这三个点组成的三角形与夹角的关系如图 2 所示。

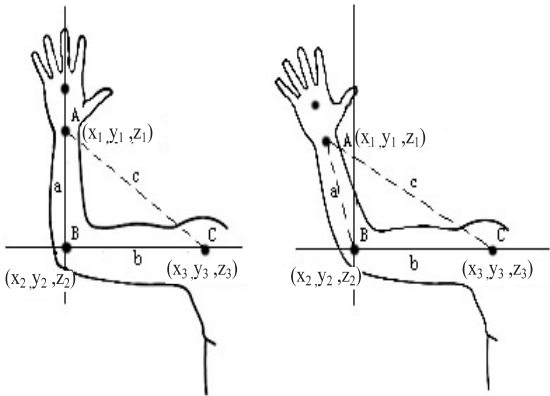


图 2 举起右手关节点各点之间关系

$$\begin{aligned} a &= \sqrt{((x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2)} \\ b &= \sqrt{((x_2-x_3)^2+(y_2-y_3)^2+(z_2-z_3)^2)} \\ c &= \sqrt{((x_1-x_3)^2+(y_1-y_3)^2+(z_1-z_3)^2)} \\ \angle ABC &= \cos^{-1}(a^2+b^2-c^2)-(2ab) \end{aligned}$$

1.3 实现 Kinect 姿势识别

为了实现姿势判断,笔者将每一个姿势的相关关节点取出,每次取两个关节点进行一次角度上的判断,如果同时满足需要的条件,即每两个关节点的角度都在阈值范围内,判断为有效姿势。判断规则为:以第一个点为中心来判断第二个点与其夹角是否在一定的范围内,将希望满足的角度记做 Angle, Threshold 为阈值,此次判断的权重记为 Weightfactor。则此条件可以简单记做:

$$\text{condition}=\{\text{point One, point Two, Angle, Threshold, Weightfactor}\}$$

举起右手这个姿势,可选择右肩点、右肘点、右手腕点三个关节点。当 Kinect 追踪到骨骼点以后,首先判断右肩点和右肘点。以右肩点为中心,判断右肘点和其夹角是否为 0 度,为了识别的准确,假设在 25 度的范围内就认为这个角度是可以接受的,25 就为阈值,因为举起右手的这个姿势重点是判断上述三个关节点,因此,这个判断的权重因子记为 1,否则记为 0,则上述条件记做:

$$\text{condition}=\{\text{Joint Type. ShoulderRight, Joint Type. ElbowRight, 0, 25, 1}\}$$

接着再判断右肘点和右手腕点是否能满足 90 度的要求,阈值仍然取 25 度。等到这两个条件都满足以后,可以说举起右手这个姿势已经基本判别完毕。为了使误识率降低,可以将左肩点、左肘点、左手腕点也

进行一次判断,判断其是否处于自然下垂状态。否则,程序会将举起双手也判断为举起右手,因此举起右手姿势要满足的四个条件就为:

```
condition1 = { Joint Type. ShoulderRight, Joint Type.
ElbowRight,0,25,1}
condition2 = { Joint Type. ElbowRight, Joint Type.
WristRight,90,25,1}
condition3 = { Joint Type. ShoulderLeft, Joint Type.
ElbowLeft,180,25,0}
condition4 = { Joint Type. ElbowLeft, Joint Type.
WristLeft,180,25,0}
```

为了尽可能多地识别不同的姿势和方便添加新的姿势,可以创设一个姿势库(存储条件的数组)用来方便存储和录入已有的姿势。将常用的姿势添加进姿势库(数组)中,在程序中调用即可<sup>[5,6]</sup>。

1.4 应用实践

笔者使用 Microsoft Visual Studio2010 与 Microsoft Kinect SDK 实现上述算法。采用加入计时器的方式,要求被测者要保持特定姿势 2 秒钟,方可认为此姿势识别正确。选取了 10 个人作为测试者,要求每个被试者面对 Kinect 感应器,距离感应器 1.5 米到 2 米之间,做已经添加到姿势库中的姿势来进行识别,实验结果如表 1 所示。

表 1 姿势识别统计

姿势	人次	识别次数	误识次数	拒识次数
举起左手	10	10	0	0
举起右手	10	10	0	0
伸平双臂做 T 字状	10	10	0	0
抬起双手	10	8	2	0

Kinect 侦测到特定的姿势后,应用程序如何处理这些信息来完成某个功能,是与姿势识别同等重要的。如果将 Kinect 看做一个输入设备,姿势识别的目的是触发一系列的操作,比如响应鼠标单击、键盘输入等。在传统的 WIMP 交互方式下,应用程序的开发者和使用者都已经熟悉按照鼠标点击、选择键盘输入等事件来触发程序的某个功能<sup>[7,8]</sup>。如果使用 Kinect 做为输入设备,应用程序必须知道什么时候用户的姿势表示要进行人机交互,什么时候仅仅只是用户的姿势而已。为了解决这个问题,通用的策略是加入一个计时器,当用户保持特有的姿势超过一定的时间时,触发对应的某个事件。或者是当监测到一系列的姿势时,触发某个事件。

将 Kinect 作为输入设备,其应用范围是比较广的。比如在使用 Power Point(以下简称 PPT)幻灯片播放时,当 Kinect 侦测到举起右手的姿势,就认为用户按下

了键盘上的向右键,这时 PPT 播放下一页;侦测到举起左手的姿势,就认为用户按下了键盘上的向左键,PPT 播放前一页。或者是在三维虚拟巡游系统中,使用原地踏步的姿势来触发摄像机的前进和停止运动,用左右手的抬起来控制摄像机左右移动视角等等。

2 结束语

文中利用 Kinect 感应器获取人物骨骼信息,计算不同骨骼关节之间的夹角来检测人物的特定姿势。为了使得 Kinect 辨认更多的姿势,创设了姿势库。通过与姿势库中的模板进行匹配的方式来检测某个特定的姿势。经过实验,发现这种识别方式具有较高的识别率。

当然,这种判断算法还存在着很多值得改进和加强的地方。首先,在 Kinect 姿势的识别上只考虑了静态的某个特定姿势,但实际生活中,人物的姿势往往是连续的。其次,所提出的这种通过关节点之间夹角关系来识别某个姿势的方法并不适用于手势识别。判断某个姿势的时候只使用了相关的关节点,存在着误识的可能性。为了识别的更为精确,可以判断 20 个关节点的信息,将 20 个关节点和角度按照一定的顺序编排可以方便地进行序列化,很容易实现姿势库的扩展。

实验只使用了 Kinect 感应器得到的人物骨骼信息。除此之外,Kinect 还提供了语音识别和深度识别,若同时采用多种识别方式,相信识别的准确性会大幅度提高,能为各种应用程序提供更好的自然交互方式。

参考文献:

[1] Pavlovic V I,Sharma R,Huang T S. Visual Interpretation of Hand Gestures for Human-computer Interaction; A Review [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1997,19(7):677-695.

[2] Webb J,Ashley J. Beginning Kinect Programming with the Microsoft Kinect SDK[M]. [s. l.]:Apress,2012.

[3] Besl P J,McKay H D. A method for registration of 3-D shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1992,14(2):239-256.

[4] 余龙华,王 宏,钟洪声. 基于隐马尔科夫模型的人脸识别 [J]. 计算机技术与发展,2012,22(2):25-28.

[5] 申鹏洋. 基于 3D 人体模型的单目图像运动姿势识别研究 [D]. 北京:北京理工大学,2011.

[6] 韩 磊. 图像序列中人的行为分析和识别方法[D]. 北京:北京理工大学,2009.

[7] 钱 锋,梅 雪,林锦国. 基于矩形骨架的人体动作识别 [J]. 微计算机信息,2012,28(4):100-102.

[8] 郑 韡,沈旭昆. 基于连续数据流的动态手势识别算法 [J]. 北京航空航天大学学报,2012,38(2):273-279.

# 基于Kinect的姿势识别与应用研究

作者: [谢亮](#), [廖宏建](#), [杨玉宝](#)  
作者单位: [广州大学, 广东 广州 510006](#)  
刊名: [计算机技术与发展](#)  
英文刊名: [Computer Technology and Development](#)  
年, 卷(期): 2013(5)

本文链接: [http://d.g.wanfangdata.com.cn/Periodical\\_wjtz201305069.aspx](http://d.g.wanfangdata.com.cn/Periodical_wjtz201305069.aspx)