

基于骨骼信息下的手势识别研究

杨和稳¹ 杨萍萍² 郭海晨²

¹(南京信息职业技术学院 江苏 南京 210000)

²(南京大学工程管理学院 江苏 南京 210000)

摘要 在人体行为动作中,基于手势动作的识别研究具有便捷、直观、互动性强、表达信息丰富等特点而成为众多研究者对人体行为分析的首选。传统的动作识别一般是建立在二维彩色图像上进行研究,而二维 RGB 图像容易受到背景扰动、光线、环境等因素对人体目标检测的干扰。对微软 Kinect 体感设备进行研究,利用 Kinect 捕获的深度图像,将其转换成骨骼图并提取骨骼关节点空间坐标信息,设计一个人体手势识别系统。借助动态时间规整算法将测试模板与训练样本进行相似度的匹配计算,达到姿势识别的目的。在 Unity 3D 里进行虚拟模型同步演示,实现了对人体动作手势指令的识别和对模型的控制。

关键词 姿势 识别 动态时间规整算法 Kinect

中图分类号 TP391.9

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2018.12.042

GESTURE RECOGNITION BASED ON SKELETON INFORMATION

Yang Hewen¹ Yang Pingping² Guo Haichen²

¹(Nanjing Vocational College of Information Technology, Nanjing 210000, Jiangsu, China)

²(School of Management and Engineering, Nanjing University, Nanjing 210000, Jiangsu, China)

Abstract In human behavior, the recognition research based on gesture has many characteristics such as convenience, intuition, strong interaction, and rich expression information, so it has become the first choice for many researchers to analyze human behavior. Traditional motion recognition is generally based on two-dimensional color image, and two-dimensional RGB images are easily interfered by background disturbance, light, environment and other factors when human are detected. In the paper, we studied Microsoft Kinect somatosensory device which was used to capture depth image, converted the depth image into bone image to get spatial coordinate information of bone joint point, and designed the human body gesture recognition system. The dynamic time warping algorithm was used to calculate the similarity between the test samples with the training samples, so as to achieve the purpose of gesture recognition. The virtual model synchronization demonstration was implemented in Unity 3D, which realized the recognition of human gesture instructions and the control of model.

Keywords Gesture Recognition Dynamic time warping algorithm Kinect

0 引言

随着科学技术的不断进步,人类和计算机的交互方式进入了一个新的发展阶段,越来越多的人开始关注研究人机交互技术,而人机交互方式也成为近些年研究者的一个重要的研究课题。作为人机交互的最新研究领域,手势作为更加便捷(相较于身体姿势、眼

神、视线)、表达力更加丰富的交互方式,越来越多的学者开始着重于手势识别的研究。

手势具有简单、直观、易于理解和操作等特点,已被广泛应用于机器人控制、虚拟键盘、体感游戏、虚拟现实医疗等诸多领域,科技的进步使得人机交互设备也在不断更新,并且识别精度会随着设备的更新而不断提高。文献[1]利用关节点的旋转原理,在关节点处贴上荧光标签,然后结合计算机来研究关节点的旋

转特性对人体动作进行分析观察,大量的实验数据表明人体骨骼节点的运动特性对人体的动作行为有很好的预测作用。文献[2]研发了一个可以识别手势的工具,用来训练以及识别手势,同时也支持 DTW 和 HMM 识别算法,只需将手势进行封装,将手势模板代码嵌入到应用程序中,从而可以减少管理的复杂性姿态数据和算法。文献[3]采用基于激光技术的深度相机获取深度图像,并对深度图像进行手部的三维建模从而实现对人体手势跟踪识别的功能。传统的数据采集装备虽然能在一定程度上采集人体的动作特征点,但是装备复杂、操作不便。美国 NaturalPoint 公司推出的 OptiTrack 运动捕捉系统,通过摄像机跟踪标记点完成人体动作捕捉。随后,Leap 公司研发了 Leap Motion 非接触式体感设备,依照三角测量法,基于双目摄像头的三维图像识别技术,可以定位出目标手势的准确位置。但由于光学式和非接触式动作识别易受外部环境的干扰,如遮挡、光线等问题的制约。因此,需要一个便捷的体感设备能够帮助我们完成对人体特征的获取以及识别,通过更方便的人机交互方式完成研究。

针对以上问题,微软经过多年研究于 2015 年推 Kinect 体感设备^[4],Kinect 综合了计算机视觉处理算法和空间定位性能,可以利用高分辨率对空间中的细微变动进行识别。且不受光照、背景等因素的影响,从而提高了手势轨迹别的稳定性和鲁棒性,最后实现手势动作识别。

本文通过对虚拟现实技术以及人机交互技术进行充分研究,基于 Kinect 进行人体手势动作识别。利用 Kinect 捕获到的人体骨骼关节点坐标,结合动态时间规整算法进行动作识别。并在 Unity 3D 进行虚拟模型同步演示,将定义的手势应用到仿真系统里,代替传统的键盘交互形式。该项技术的创新能够更好地探讨虚拟现实交互系统里新的交互通道存在的应用价值。

1 Kinect 简介

Kinect 是微软公司自主研发的一款体感外设产品,2010 年正式推出,研发的初衷是为了让玩家用户拥有全新的人机交互体验^[5]。Kinect 由彩色(RGB)摄像头、深度(DEPTH)摄像头、红外摄像头以及麦克风等部分组成,如图 1 所示。



图 1 Kinect

本次实验采用的二代 Kinect 利用 Time of Flight^[6],也就是根据光脉冲的来回时间去捕获目标物体的深度图像。Kinect 最大的优势是其可以获得目标人物的骨骼图像从而进行人体行为动作的研究。而 Kinect 传感器将深度图像转换成骨骼图像的步骤主要有三部分:将人体轮廓从背景图像抽离出来、对人体关键部位的定位、骨骼节点的标识^[7]。

首先 Kinect 会对人体深度图像的像素点逐点扫描,将属于人体部位的像素点从深度图像中抽离出来;接着 Kinect 通过特征值将人体各个部位进行分类,比如躯干、头部、手臂等;最后 Kinect 根据机器学习的方法确定骨骼关节点。Kinect2.0 可以识别人体的 25 个骨骼关节点,包含肘、踝、脚、手指、肩等关节,相比一代 Kinect,识别的骨骼更多,以达到全身体感操作,Kinect 体系架构如图 2 所示。而本文对动作的分析则是基于人体的骨骼信息来进行研究的。

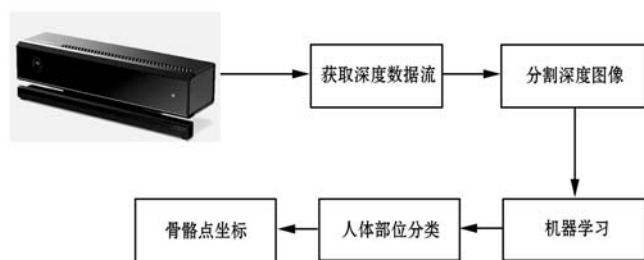


图 2 Kinect 骨骼追踪的体系架构

2 基于 Kinect 骨骼数据的姿势识别方法

对人体动作的识别从本质上来看是指对提取的动作特征进行分析,选取合适的识别算法对动作特征进行分析研究,从而实现动作识别。本文对动作特征提取的研究的出发点是基于人体骨骼关节点信息,Kinect 会捕获 25 个关节点数据,我们需要对这 25 个关节点的数据做进一步详细的分析才能提取出有用的姿态特征。如果直接利用 25 个骨骼节点的空间坐标数据作为人体的运动特征,会存在两个方面的缺点:一是 25 个关节点的维度较高,高纬度的训练样本会使得本次系统的计算量和复杂度增加,导致训练效率极其低下;二是不同的人的骨骼关节点空间坐标信息不同,需要找到统一的动作特征才能对人体动作进行识别。

2.1 动作特征的提取

在日常生活中见到的手臂弯曲、挥手、跳跃等行为动作其实是由不同的关节的旋转带动的,比如,跳跃时,胯关节带动大腿运动,而大腿则通过膝关节带动小腿运动。图 3 展示了人体四肢运动时,各个关节角度度数。



图3 人体关节角度图

从图中可以看出,运动特征其实可以对应于各个关节旋转的角度。反观不同的人的生理结构,关节之间的变化可以不受不同人体身体结构差异的影响,所以可以推测,关节的旋转特性决定人体四肢在空间中的位置,而旋转特性其实可以对应关节的角度。在人体运动过程中,关节之间的变化不受不同人体身体结构差异的影响。因此本文接下来研究在运动过程中骨骼关节角度的变化情况。

本次实验是针对人体手势动作进行研究识别的,为避免不同人体手臂长度或宽度对动作识别的干扰,除了跟踪某些关节在空间唯一的轨迹和线速度之外,还要对上肢关节结构向量间的夹角进行研究。综合考虑简单动作和复杂动作,本文选取的动作特征参数包括:双手手臂关节角度、手臂相对于肩部、头部变化的距离和角度。因此本文只针对上半身的几个关键骨骼点进行分析,着重研究肩关节、肘关节、头部、腕关节。图4解释了手势动作关节的计算原理。

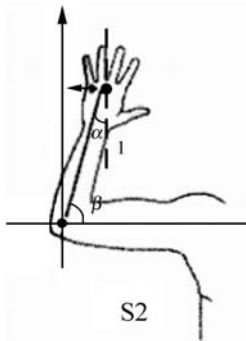


图4 夹角示意图

$$\alpha_j = \arctan \frac{|x_{t+1} - x_t|}{|y_{t+1} - y_t|} \quad (1)$$

$$\beta = 90^\circ - \alpha_j \quad (2)$$

具体操作是:把人体的运动看作是在 x, y 轴所在平面内的二维运动,根据式(1)计算出每个关节与 y 轴的夹角,而手肘关节的角度 β 由式(2)可以计算出。

将以上获得的夹角组成一组关节角度向量 $\mathbf{A} = \{\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2\}^T$,通过连续地捕捉每一帧的关节角度,就可以获得关节角度关于时间变化的函数,它可以充分地描述人体手部动作的运动特征。图5显示实验

者挥动左手时左肘关节角度关于时间变化的曲线图。

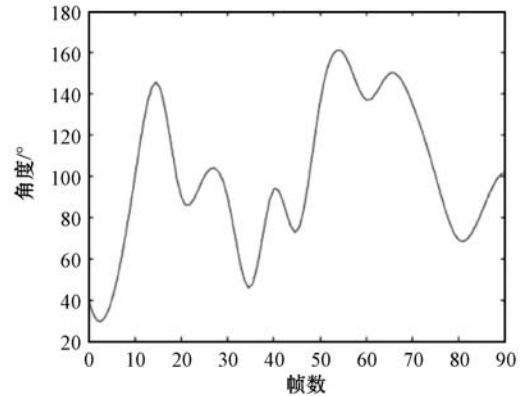


图5 左肘关节角度的变化图

从图中可以看出,在挥手这个动作过程当中,左肘关节角度是随时间的变化而不断变化的,这条基于时间序列的关节变化曲线图可以反映出挥手这个动作的运动特性。基于此,本文采用基于时间变化的骨骼关节角度的变化特性作为人体运动的动态特征。

2.2 动态时间规整(DTW)算法描述

通过实验对比分析可以发现,不同的测试者在做同一个挥手动作时,他们的肘关节角度曲线图在波形和幅值虽然有一定的偏差,但是整体的波形很相似,即将关节角度变化曲线图的相似度作为分类器的切入点。欧式距离是最常用的计算不同向量之间的距离,但是欧式距离是基于相同维度的向量进行计算的,而实际中关节角度变化曲线图的维度不同,这就要求我们需要借助动态规划的思想来解决。动态时间规整算法 DTW (Dynamic Time Wrapping)^[8]主要是通过时间规整的原理对两条不一样的时间序列进行相似度的匹配,寻找这两条曲线之间距离最小的相似度路径。时间规整是指将待测信号的长度进行压缩或拉长,从而与标准信号的模板里的长度一样。

假设有两个动作序列,长度分别为 m 和 n ,即:

$$T = \{t_1, t_2, \dots, t_m\}$$

$$G = \{g_1, g_2, \dots, g_n\}$$

为了得出这两条时间序列的相似度,我们引用一个 $m = n$ 的矩阵 \mathbf{M} ,该矩阵里的每一个元素 (i, j) 表示参考序列 t_i 和测试序列 g_j 两个点之间的距离 $d(t_i, g_j)$ ^[9]:

$$d(t_i, g_j) = (t_i - g_j)^2 \quad (3)$$

欧式距离是计算机领域中最常见的计算距离的方法,如式(3)所示。简而言之,该算法就是寻找出一条最优路径,且这条路径包含了这两个动作序列之间相距最短的点,即规整路径(warping path),如图6所示。

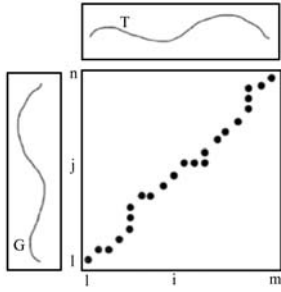


图 6 动态时间规整路径

规整路径 W 可表示为:

$$W = \{w_1, w_2, \dots, w_T\} \max(m, n) \leq T \leq m + n + 1$$

另外,规整路径不可以任意选择,它需要满足 3 个约束条件:

- (1) 边界性:规整路径 W 的起点必须是左下角,即 $w_1 = (1, 1)$, 路径的终点是右上角,即 $w_T = (m, n)$ 。
- (2) 连续性:规整路径 W 上的点必需与相邻的几个点对齐,也就是说当前点的下一个对应点只能是相邻的点,即 $w_T = (t, g), w_{T'} = (t', g'), t - t' \leq 1, g - g' \leq 1$ 。
- (3) 单调性:规整路径 W 上的点在时间轴上必须是单调递增的,即 $w_T = (t, g), w_{T'} = (t', g')$, 那么 $t - t' \geq 0, g - g' \geq 0$ 。

为了对两个时间序列进行匹配,这里我们定义累积路径值为 $\gamma(i, j)$, 对 T 和 G 这两个时间序列从起点 $(1, 1)$ 点进行计算匹配,每经过一个点,累计点 t_i 与点 g_j 之间距离再加上其四周与这个点距离最小的那个点之间的距离,如式(4)所示。当匹配到终点 (m, n) 的时候,最后得到的距离值代表了这两个时间序列的相似度。

$$\gamma(i, j) = d(t_i, g_j) + \min \left\{ \begin{array}{l} \gamma(i-1, j) \\ \gamma(i, j-1) \\ \gamma(i-1, j-1) \end{array} \right\} \quad (4)$$

直到匹配到终点 (m, n) 为止,最终所获得的路径累积值就是两天时间序列 T 和 G 是否匹配的依据。累计距离和相似度成反比,如果累计距离 $\gamma(i, j)$ 越小,这两个动作之间的相似度就越高,反之亦然。

对人体手势动作进行分析研究时,首先创建手势动作模板库,然后将测试动作与模板库中的动作进行相似度的匹配,根据 DTW 识别算法计算两个模板之间的最小匹配距离。找出匹配距离值最小的那个模板即为识别的动作模板。

3 实验分析

3.1 实验与分析

前面已经完成了对人体动作特征的提取和识别算

法的选取,下面通过实验来验证本文方法的有效性。

本文定义了基于交互的 5 种简单的手势动作,分别为上下举、左右挥、停止等,如表 1 所示。在人机交互中,可以通过这几种手势命令与计算机进行交互。

表 1 自定义的交互手势集

标识	语义	手势
1	上举	手臂上举,手指越过头顶
2	下举	手臂下举,手指从头顶到达胸部
3	左挥	手臂左挥
4	右挥	手臂右挥
5	握拳	握住手掌

手势 1 到 5 选择 5 个实验者进行 50 次实验,则每种手势有 250 组数据。本次实验由同一个人对这几种动作进行训练,并建立动作模板库。训练的模板长度为 90 帧, Kinect 记录的数据信息包括关节角度变化,上肢关节的坐标信息。这里以手臂上举,手指越过头顶为例进行分析。通过 Kinect 捕获其骨骼关节角度的变化作为测试序列,与标准的序列模板进行对比来检测系统识别效果。关节角度变化如图 7 所示,其中, Sample 作为标准模板, Test sample 作为测试序列模板。

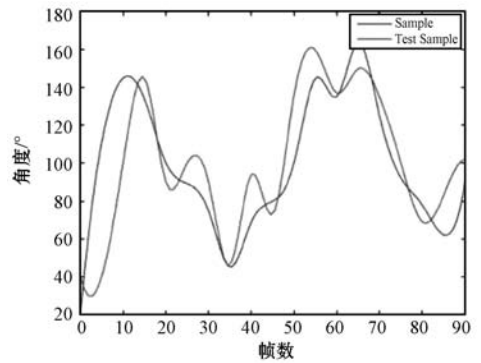


图 7 测试数据与模板数据对比

Kinect 对手势识别的结果是通过 Visual Gesture Builder 里的波形来表示,波形出现的范围越广,说明识别的相似度就越高。如果没有出现波形,则表明该动作没有被我们识别出来。识别的效果如图 8 所示。左边为人物的动作,右边为识别效果,波形代表相似程度,波峰越高代表与模板的匹配度越高,成功识别的判定标准为:波峰超过所在窗口的一般即认为成果识别。

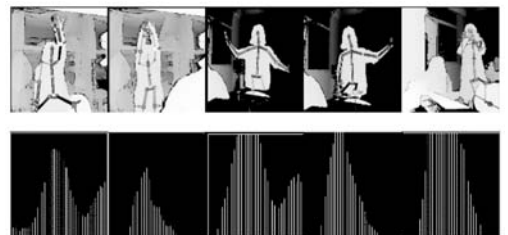


图 8 识别效果图

实验结果如表 2 所示。

表 2 识别结果

手势	测试次数	识别次数	识别率/%
上举	250	240	96
下举	250	230	92
左挥	250	238	95
右挥	250	235	94
握拳	250	230	92

从表中可以看出,基于上肢关节角度变化的动态时间规整算法的识别率普遍超过 90%。尽管对动作识别的精度不高,但是对于一些常见的动作已经可以实现对动作的识别。

3.2 鲁棒性验证

为检验光照环境变化对该实验鲁棒性的影响,做出另一组实验,在不同光照环境下每个手势由上一组实验者进行相同的实验,表 3 为实验者在光线充足与不充足的环境下的对比结果。

表 3 不同光照环境下每个手势的识别率

手势	光线充足			光线不充足		
	测试次数	识别次数	识别率/%	测试次数	识别次数	识别率/%
上举	250	240	96	250	238	95
下举	250	230	92	250	229	92
左挥	250	238	95	250	240	96
右挥	250	235	94	250	235	94
握拳	250	230	92	250	231	92

从对比结果可以看出,相同手势在不同光线下的识别率差距甚微,利用 DTW 识别算法以及 Kinect 传感器对光照变化的干扰具有较强的鲁棒性。

4 虚拟模型同步演示

本系统中的人机交互功能主要是使用人体手势动作控制虚拟模型的运动,而基于 Kinect 的虚拟仿真平台则是通过 Unity 3D 来实现对虚拟物体的三维展示功能。首先利用 Kinect 体感器获取现实场景中的深度数据,并获取人体骨骼数据,根据骨骼跟踪生成的关节点信息来识别当前用户的姿势,并通过约定的姿势来达到对展示的三维模型进行操作的目的是。

Unity 3D 是由 Unity technologies 开发的一款专业游戏引擎,它可以让用户轻松地创建各种三维游戏、实时动画、场景可视化等互动性较高的项目。本文将在

Unity 3D 中对虚拟物体的运动进行同步演示设计。如图 9 所示,在 Unity 3D 里创建一个虚拟小方块。小方块的每一面贴上不同的图纸,每个手势对应不同的界面。当手上举时,小方块就向上翻转一面。向左挥时,小方块向左旋转一面。当握拳时,小方块恢复成初始界面。

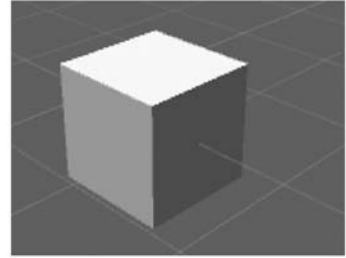


图 9 虚拟模型

虚拟物体模型是按照 Kinect 提供的骨骼节点设计的,这里只保留上肢节点。因为人体的运动是一个连续性的动态过程,所以需要时骨骼关节点的坐标位置进行实时地更新,这里我们建立一个 RefreshBodyObject() 函数用来实时地更新骨骼节点的坐标数据,并将这些数据复制给虚拟模型,从而使得虚拟模型可以同步地跟随用户运动。

最终的实验结果如图 10 所示,显示了用户上举、下举、左挥、右挥、停止动作时虚拟模型与其同步运动的情况。



(a) 上举

(b) 下举



(c) 左挥

(d) 右挥



(e) 握拳

图 10 同步运动结果

5 结语

本文利用搭建的 Kinect 的体感识别平台,通过 Kinect 捕获被测对象骨骼信息空间坐标数据,提出基于特征提取的模板匹配方法,分析关节角度与动作行为研究的相似性,提出将骨骼关节点的角度变化作为运

(下转第 292 页)

- [3] Gooijer J G D, Hyndman R J. 25 years of time series forecasting [J]. *International Journal of Forecasting*, 2005, 22(3):443-473.
- [4] Guresen E, Kayakutlu G, Daim T U. Using artificial neural network models in stock market index prediction [J]. *Expert Systems with Applications*, 2011, 38(8):10389-10397.
- [5] Chen A S, Leung M T, Daouk H. Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index [J]. *Social Science Electronic Publishing*, 2003, 30(6):901-923.
- [6] Manahov V, Hudson R. A note on the relationship between market efficiency and adaptability—New evidence from artificial stock markets [J]. *Expert Systems with Applications*, 2014, 41(16):7436-7454.
- [7] Cao L, Tay F E H. Financial forecasting using support vector machines [J]. *Neural Computing & Applications*, 2001, 10(2):184-192.
- [8] Chourmouziadis K, Chatzoglou P D. An intelligent short term stock trading fuzzy system for assisting investors in portfolio management[M]. [S. l.]:Pergamon Press, Inc. 2016.
- [9] Kumar L, Pandey A, Srivastava S, et al. A hybrid machine learning system for stock market forecasting [J]. *Proceedings of World Academy of Science Engineering & Technology*, 2011, 20:315-318.
- [10] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. *Nature*, 2015, 521(7553):436-444.
- [11] Arévalo A, Niño J, Hernández G, et al. High-frequency trading strategy based on deep neural networks[M]//*Intelligent Computing Methodologies*. Springer International Publishing, 2016:424-436.
- [12] Chong E, Han C, Park F C. Deep learning networks for stock market analysis and prediction: methodology, data representations, and case studies [J]. *Expert Systems with Applications*, 2017, 83:187-205.
- [13] Zhong X, Enke D. Forecasting daily stock market return using dimensionality reduction[M]. [S. l.]:Pergamon Press, Inc. 2017.
- [14] Frost A J, Prechter R R. Elliott wave principle: key to market behavior[M]. [S. l.]:Elliott Wave International,1998.
- [15] Sanger T D. Optimal unsupervised learning in a single-layer linear feedforward neural network [J]. *Neural Networks*, 1989, 2(6):459-473.
- [16] Sontag E D. Some remarks on the backpropagation algorithm for neural net learning [R]. Rutgers Center for Systems and Control Technical Report, 1988: 88-07.
- [17] Zadeh L A. Fuzzy sets, information and control [J]. *Information & Control*, 1965, 8(3):338-353.
- [18] 张智星, 孙春在. 神经-模糊和软计算[M]. 西安:西安交

通大学出版社, 2000.

- [19] Møller M F. A scaled conjugate gradient algorithm for fast supervised learning[J]. *Neural Networks*, 1993, 6(4):525-533.
- [20] Mandelbrot B. How long is the coast of Britain? Statistical self-similarity and fractional dimension [J]. *Science*, 1967, 155(3775):636-638.

(上接第 232 页)

动特征。并借助于动态规划的思想来解决,即利用模板匹配的方法来进行人体手势动作的跟踪和识别。实验结果和对比结果表明,该方法成功实现了对人体手势特征的识别,并获得良好的姿势识别率、鲁棒性较强。同时将这一技术在虚拟现实同步演示,获得良好的演示效果,交互方式更直观、便捷。在未来的研究中,可以利用该方法来完成对真实物体模型进行观测的目的。

与此同时,本方法也存在一定的不足之处,对于手势识别,只是进行了主要动作的识别,应用相对有限,可以将动作识别的工作进一步深入,这些工作以后需要改进与完善。

参 考 文 献

- [1] 赵飞飞. 基于 Kinect 骨骼信息的手势识别及应用[D]. 太原:太原理工大学,2016.
- [2] Ibañez R, Soria ú, Teyseyre A, et al. Easy gesture recognition for Kinect [J]. *Advances in Engineering Software*, 2014, 76:171-180.
- [3] Ali S, Shah M. Human action recognition in videos using kinematic features and multiple instance learning [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2009, 32(2):288-303.
- [4] Biswas K K, Basu S K. Gesture recognition using Microsoft Kinect[C]//*The 5th International Conference on Automation, Robotics and Applications*. IEEE,2011:100-103.
- [5] 周慎. 基于虚拟现实的汽车驾驶模拟器建模技术研究[D]. 武汉:武汉理工大学,2005.
- [6] 吴国斌,李斌,阎骥洲. KINECT 人机交互开发实践[M]. 北京:人民邮电出版社,2013.
- [7] 董珂. 基于 Kinect 的人体行为识别研究[D]. 武汉:武汉科技大学,2015.
- [8] Sakoe H, Chiba S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition [J]. *IEEE Transactions on Acoustics Speech & Signal Processing*, 1978, 26(1):43-49.
- [9] 刘兵. 时间序列与聚类挖掘相关技术研究[D]. 上海:复旦大学,2006.