基于 Kinect 的实时人体姿势识别

刘开余，夏 斌
（上海海事大学 上海 201306）

摘要： Kinect 的实时骨骼跟踪技术获取身体关节点的三维位置信息，为进行人体姿势识别提供了丰富的信息，拟在三维关节点位置信息的基础上，建立一种实时的人体姿势识别系统。选择关节角度作为姿势特征，结合逻辑回归（logistic regression，LR）分类算法对 54 种姿势进行识别研究。实验结果表明，该姿势识别系统可以准确实时地识别人体姿势。

关键词： 人机交互；姿势识别；Kinect；逻辑回归
中图分类号： TN02 **文献标识码：** A **文章编号：** 1674-6236(2014)19-0031-04

Real-time human posture recognition based on Kinect

LIU Kai-yu，XIA Bin
（*Shanghai Maritime University*，*Shanghai* 201306，*China*）

Abstract: Skeletal tracking provided by Kinect can acquire three-dimensional （3D） position of body joints，which provided a wealth of information for human posture recognition. A real-time human posture recognition system was built based on the 3D position of body joints. Joint-angles were selected as features and logistic regression （LR） algorithm was used to recognize 54 categories of postures. Experimental results showed that the human posture recognition system can recognize human postures in real time accurately.

Key words: Human Computer Interaction （HCI）；posture recognition；Kinect；logistic regression

姿势识别是机器视觉领域的研究热点，被广泛应用在人机交互、行为分析、多媒体应用和运动科学等领域。

姿势识别主要有两种方法。第一种是利用可穿戴传感器，比如戴在身体上的加速度计[1]或装在衣服上的张力传感器[2]。可穿戴传感器具有精确直接的特点，但会对肢体运动造成束缚，会给用户带来额外的负担。第二种是利用视觉捕捉技术[3]，例如视频或者静态图像，通过对视觉数据的处理来判断用户的动作。基于视觉捕捉技术在特征表达方面，起初是采用人体轮廓作为姿势特征表达[4-5]。但是轮廓特征从整体角度描述姿势，忽略了身体各部位的细节，不能精确地表示丰富多彩的人体姿势。有研究[6-7]采用基于身体部位的姿势表达，即把人体轮廓分成若干个身体部位，例如颈部、躯干和腿。由于这些姿势特征都是从二维彩色图像中抽取而来，需要处理人体定位、肢体被遮挡、不同光照条件等问题。

近年来，Kinect 等深度传感器不仅提供彩色图像数据，而且提供了三维深度图像信息。三维深度图像记录了物体与体感器之间的距离，使得获取的信息更加丰富。利用 Kinect 的实时骨骼跟踪技术和支持向量机 （support vector machine，SVM）识别 4 种姿势（站，躺，坐和弯腰）[8]。本文采用逻辑回归算法对 54 种姿势进行识别研究，设计开发实时的人体姿势识别系统。

1 方 法

1.1 特征提取

人体姿势可定义为某一时刻身体关节点之间的相对位置。如果得到关节点的三维位置信息，那么关节点之间的相对位置就确定。但由于不同人的体型存在差异，原始坐标数据过于粗糙，所以采用关节角度描述姿势特征。微软公司提供的 Kinect 体感器主要由红外发射器、RGB 摄像头、红外深度图像摄像头、传动马达和麦克风阵列组成，如图 1 所示。红外发射器和红外深度图像摄像头组合起来获取深度图像。RGB 摄像头获取彩色图像。传动马达用于调整 Kinect 设备的俯仰角。麦克风阵列可以捕获声音和定位声源。

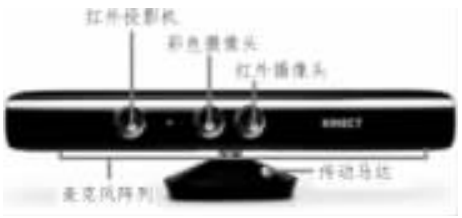

图 1 Kinect 传感器
Fig. 1 Kinect sensor

收稿日期：2014-04-26 **稿件编号：**201404245
基金项目：上海市科委项目（20130004）
作者简介：刘开余（1988—），男，浙江兰溪人，硕士研究生。研究方向：体感互动技术。

骨骼跟踪是在深度图像的基础上，利用机器学习方法逐步实现[9]。第一步是人体轮廓分割，判断深度图像上的每个像素是否属于某一个用户，过滤背景像素。第二步是人体部位识别，从人体轮廓中识别出不同部位，例如头部、躯干、四肢等肢体。第三步是关节定位，从人体部位中定位 20 个关节点。

Kinect 的骨骼跟踪技术可以主动跟踪 2 个用户，被动跟踪 4 个用户。主动跟踪时，捕获用户身体 20 个关节点的三维位置信息，如图 2 所示，关节点名称详见表 1。被动跟踪时，只捕获用户的脊柱中心位置。骨骼坐标系以红外深度图像摄像头为原点，X 轴指向体感器的左边，Y 轴指向体感器的上边，Z 轴指向视野中的用户。

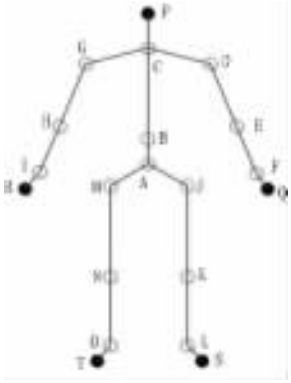

图 2　20 个人体关节点
Fig. 2　Twenty body joints

通过观察发现 15 个身体关节与姿势的关联度比较大，分别标记为“A”到“O”。另外 5 个标为黑色的关节由于和相邻的关节点距离太近容易产生抖动现象而未被使用。从 15 个关节点中提取可能与姿势有关联的 25 个关节角度特征，如表 2 所示。其中，角 1–4 和角 16–25 是两个向量之间的夹角，角 5–13 是一个向量和 Y 轴的夹角，角 14–15 是一个向量和 X 轴的夹角。所有角度的范围在(0°,180°)。

表 1　20 个关节点命名
Tab. 1　Names of twenty body joints

编号	命名	编号	命名	编号	命名	编号	命名
A	臀部中心	E	左肘	I	右腕	S	左脚
B	脊柱中心	F	左腕	R	右手	M	右臀部
C	肩膀中心	Q	左手	J	左臀部	N	右膝盖
P	头部	G	右肩	K	左膝盖	O	右踝
D	左肩	H	右肘	L	左踝	T	右脚

1.2　逻辑回归

逻辑回归是经典的分类算法，应用十分广泛。逻辑回归的原理是使用梯度下降方法进行多次迭代使得代价函数逐渐减少。当代价函数满足要求时，记录模型的参数。多个关节角度特征提取完成后，利用逻辑回归分类算法对姿势进行分类。假设 N 维特征向量 $x=[x_0,x_1,\cdots,x_{N-1}]^{\mathrm{T}}$，参数向量 $\theta=[\theta_0,\theta_1,\cdots,\theta_{N-1}]^{\mathrm{T}}$，函数模型为

$$h_\theta(x)=g(\theta^{\mathrm{T}}x) \tag{1}$$

其中 g 是内核函数，x_0=1。为了使模型输出范围在 0 到 1 之间，定义内核函数

$$g(z)=\frac{1}{1+\exp(-z)} \tag{2}$$

当 z 取较大正值时，$g(z)$接近 1，当 z 取较小负值时，$g(z)$接近 0。对于内核函数有两种理解：1)定义阈值 threshold，当模型输出大于 threshold 时，判断为 1，否则判断为 0；2)假如模型的输出为 0.8，表示为 1 的可能性是 0.8，为 0 的可能性是 0.2。

在一对多的逻辑回归分类中，每一类都要训练一个模型 $h_\theta^{(i)}(x)$。在进行预测时，选择 $h_\theta^{(i)}(x)$值最大的一类作为分类结果。假设训练样本为

$$S=\{(x^{(i)},y^{(i)})\}_{i=1}^{M}\subseteq(X\times Y)^M \tag{3}$$

其中 $x^{(i)}\in X\subseteq R^N$，$y^{(i)}\in Y=\{0,1\}$分别表示输入向量和标签。参数矩阵为

$$\theta_{N\times K}=[\theta^{(1)},\theta^{(2)},\cdots,\theta^{(p)},\cdots,\theta^{(k)}] \tag{4}$$

其中 k 表示姿势的种类数。

对于每一种姿势，训练一个一对多的分类器 $\theta^{(p)}=[\theta_0^{(p)},\theta_1^{(p)},\cdots,\theta_{N-1}^{(p)}]^{\mathrm{T}}$。如果进来一个新的样本 $x^{(i)}$，计算概率向量 $p_{1\times K}=g(x^{(i)^{\mathrm{T}}}\theta)$，则 $p_{1\times k}$ 中值最大的元素下标就是识别出来的姿势编号。

表 2　25 个关节角度
Tab. 2　Twenty-five joint-angles

角度	描述	角度	描述		
1	(ED,EF)	10	(IH, +Y 轴)	19	(MN,MA)
2	(HG,HI)	11	(LK, +Y 轴)	20	(DB,DE)
3	(KJ,KL)	12	(ON, +Y 轴)	21	(GB,GH)
4	(NM,NO)	13	(BC, +Y 轴)	22	(AN,AK)
5	(ED,+Y 轴)	14	(DG, +X 轴)	23	(AO,AL)
6	(HG, +Y 轴)	15	(JM, +X 轴)	24	(JD,JK)
7	(KJ, +Y 轴)	16	(DC,DE)	25	(MG,MN)
8	(NM, +Y 轴)	17	(GC,GH)		
9	(FE, +Y 轴)	18	(JK,JA)		

2　实　验

在相关研究[5,10]提出的姿势基础上进行扩展，建立包含 54 种全身姿势的数据库。图 3 以镜像小图标的形式展示 54 种姿势。

2.1　离线实验

招募了 5 名（3 名男性 2 名女性）被试进行离线实验。Kinect 设备水平放置，距离地面 48 cm。传动马达的角度是正 10 度。背景是一面白墙。被试面对着 Kinect 设备，全身处在视野范围内，距离其 240 cm 的位置，按顺序做完 54 种姿势。对于被试 2~5，每种姿势采集 109 个样本，分别有 5 886 个样本。对于被试 1，每种姿势采集 218 个样本，共 11 772 个样本，详见表 3。被试 1 的姿势数据的 50%用于训练，另外 50%用于测试。被试 2~5 的姿势数据全部用于测试。

为了得到准确率最高的关节角度特征数量，从 5 个到 25 个逐步增加关节角度特征数量。当得到最优的关节角度特征数量时，从 2 类到 54 类逐步增加分析逻辑回归分类方法的

图 3　姿势数据库
Fig. 3　Posture dataset

表 3　每位被试的样本数量
Tab. 3　Number of examples for each subject

被试	1	2	3	4	5
样本	11 772	5 886	5 886	5 886	5 886

准确率。

2.2　在线实验

利用离线实验的最优关节角度特征数量和逻辑回归分类算法建立实时的姿势识别系统。Kinect 体感器以 30 帧每秒的速度捕获 20 个关节点的坐标数据。在实时的姿势识别中,连续采集 60 帧坐标数据,然后从每一帧中提取 21 个关节角度,并送入分类模型对每一帧进行识别。统计 60 帧中每种姿势出现的次数,出现次数最多的姿势认为是识别的姿势。实时姿势识别系统界面如图 4 所示。

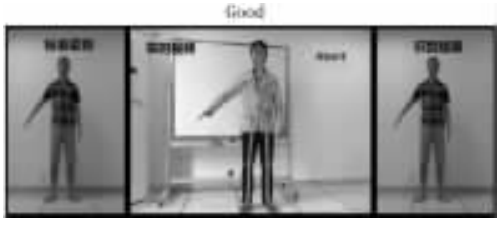

图 4　实时人体姿势识别系统界面
Fig. 4　Interface of real-time human posture recognition system

界面中间是实时的彩色视频画面,当被试把手移动到开始按钮上时,在界面左侧显示标准姿势,引导被试在 3 秒内模仿标准姿势并保持 2 秒钟。之后界面右侧显示姿势识别的结果。如果被试的姿势和标准姿势相近,则提示“Good”,否则提示“Error”。

选择被试 1,再招募 3 名(3 名女性)被试进行在线实验。Kinect 设备水平放置,距离地面 48 cm。传动马达的角度是正 10 度。背景是一面白墙。被试面对着 Kinect 设备,全身处在视野范围内,距离其 240 cm 的位置,按顺序做完 54 种姿势。

3　结果

关节角度特征数量从 5 个到 25 个逐渐增加分析姿势识别准确率,结果如图 5 所示。

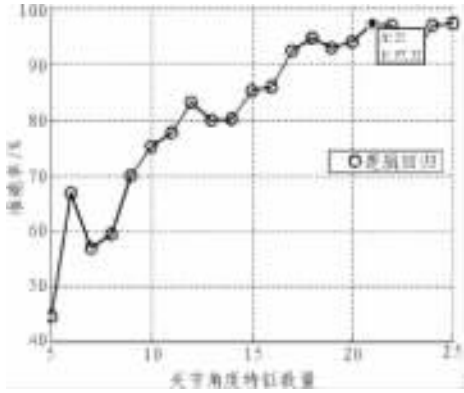

图 5　不同的特征数量和准确率
Fig. 5　Different number of features and accuracy

随着关节角度特征数量的增加,准确率总体呈上升趋势。当关节角度特征数量为 21 个时准确率最高,为 97.32%。当关节角度特征数量大于 21 个以后,准确率趋向平稳。

姿势种类数从 2 种到 54 种逐渐增加分析姿势识别准确率,结果如图 6 所示。

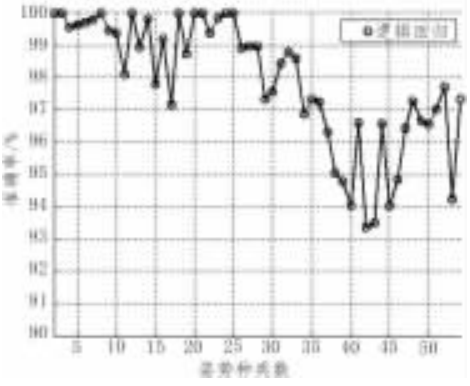

图 6　不同的姿势种类数和准确
Fig. 6　Different categories of postures and accuracy

随着姿势种类数的增加,姿势识别准确率总体呈下降趋势。当姿势种类数为 2、3、18 和 22 时,准确率达到了 100%。姿势种类数不超过 25 种时,准确率维持一个较高的水平,超出 25 种时,准确率开始下滑。

在线实验中,被试 1 的准确率高达 96.30%。三位新招募的被试准确率都在 85%以上,如表 4 所示。

表 4　在线姿势识别准确率
Tab. 4　Accuracy of real-time posture recognition

被试	1	6	7	8
准确率(%)	96.30	88.89	85.19	85.19

4　讨　论

利用 Kinect 的实时骨骼跟踪技术和逻辑回归分类算法实现了实时的人体姿势识别。Kinect 的实时骨骼跟踪技术可

以获取身体关节点的三维坐标数据。由于不同人的身高和体重不同,原始三维坐标数据存在维数高、泛化效果差等问题。关节角度特征可以较好地描述人体姿势。

离线实验中,利用 Kinect 设备捕获关节点三维坐标数据,抽取 25 个与姿势有关联的关节角度,采用逻辑回归算法训练分类模型。关节角度特征数量从 5 个到 25 个逐渐增加,分析相应的准确率,得出当关节角度特征数量为 21 个时准确率最高。姿势种数从 2 种到 54 种逐渐增加,得到平均准确率为 97.88%。逻辑回归可以较好地识别多种人体姿势。

结合骨骼跟踪和逻辑回归的姿势识别方法在实时性、精确度等方面上都有良好的表现。骨骼跟踪的速率是 30 帧每秒,即每秒钟可以对姿势进行 30 次识别。当用户做出一个姿势,系统能够快速地识别出来并做出反应,达到友好交互的目的。在线实验中,利用最优的关节角度特征和逻辑回归算法设计开发实时的姿势识别系统。三位新的被试从未体验过 54 种姿势,她们模仿界面左边的标准姿势做完 54 种姿势。她们做了很多模棱两可的姿势,平均准确率达到 88.89%(见表 4)。如果三位被试不断体验姿势识别系统、熟悉 54 种姿势,则她们所做的姿势可以被正确地识别。姿势的种类可以继续扩展,对于一种新的姿势,给定一定量的训练样本,就可以训练出分类模型。

5 结束语

本文利用 Kinect 的实时骨骼跟踪技术获取身体关节点的三维位置信息,建立包含 54 种人体全身姿势的数据库。提取 25 个与姿势有关联的关节角度作为姿势特征,结合逻辑回归分类算法进行离线实验,得出当关节角度特征数量为 21 个时姿势识别的准确率最高。设计开发了实时的姿势识别系统并进行在线实验。实验证明,结合 Kinect 的骨骼跟踪和逻辑回归算法可以准确实时地识别人体姿势。

参考文献:

[1] Allen F R,Ambikairajah E,Lovell N H,et al. Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models[J]. Physiological Measurement,2006,27(10):935.

[2] Mattmann C,Clemens F,Tröster G. Sensor for measuring strain in textile[J]. Sensors,2008,8(6):3719-3732.

[3] Weinland D,Ronfard R,Boyer E. A survey of vision-based methods for action representation, segmentation and recognition[J]. Computer Vision and Image Understanding, 2011,115(2): 224-241.

[4] Boulay B,Brémond F,Thonnat M. Applying 3d human model in a posture recognition system [J]. Pattern Recognition Letters,2006,27(15):1788-1796.

[5] Cohen I,Li H. Inference of human postures by classification of 3D human body shape[C]//Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on. IEEE,2003:74-81.

[6] Mo H C,Leou J J,Lin C S. Human Behavior Analysis Using Multiple 2D Features and Multicategory Support Vector Machine[C]//MVA,2009:46-49.

[7] Souto H,Raupp Musse S. Automatic Detection of 2D Human Postures Based on Single Images[C]//Graphics, Patterns and Images (Sibgrapi), 2011 24th SIBGRAPI Conference on. IEEE,2011:48-55.

[8] Le T L,Nguyen M Q,Nguyen T T M. Human posture recognition using human skeleton provided by Kinect [C]// Computing, Management and Telecommunications (ComManTel), 2013 International Conference on. IEEE, 2013: 340-345.

[9] Shotton J,Sharp T,Kipman A,et al. Real-time human pose recognition in parts from single depth images [J]. Communications of the ACM,2013,56(1):116-124.

[10]Negin F,Özdemir F,Akgül C B, et al. A decision forest based feature selection framework for action recognition from rgb-depth cameras [C]//Image Analysis and Recognition. Springer Berlin Heidelberg,2013:648-657.

(上接第 30 页)

[3] Clerc M,Kennedy J. The particle swarm-explosion, stability and convergence in a multidimensional complex space [J]. IEEE Transactions on Evolutionary Computation,2002,6 (1):58-73.

[4] 徐甜,刘凌霞. Bezier 曲线的算法描述及其程序实现[J]. 安阳师范学院学报,2006:49-52.
XU Tian,LIU Ling-xia. Algorithm description and program implementation of Bezier curve[J]. Journal of Anyang Normal University, 2006:49-52.

[5] Clerc M,Kennedy J. The particle swarm-explosion, stability and convergence in a multidimensional complex space [J]. IEEE Transactions on Evolutionary Computation,2002,6(1): 58-73.

[6] 刘波. 粒子群优化算法及其工程应用[M]. 北京:电子工业出版社,2010.

[7] 焦鹏,王新政,谢鹏远. 基于粒子群优化LSSVM的模拟电路故障诊断方法[J]. 现代电子技术,2013(8):35-38.
JIAO Peng,WANG Xin-zheng,XIE Peng-yuan. Method of analog circuit fault diagnosis based on particle swarm optimization LSSVM[J]. Modern Electronics Technique,2013 (8):35-38.