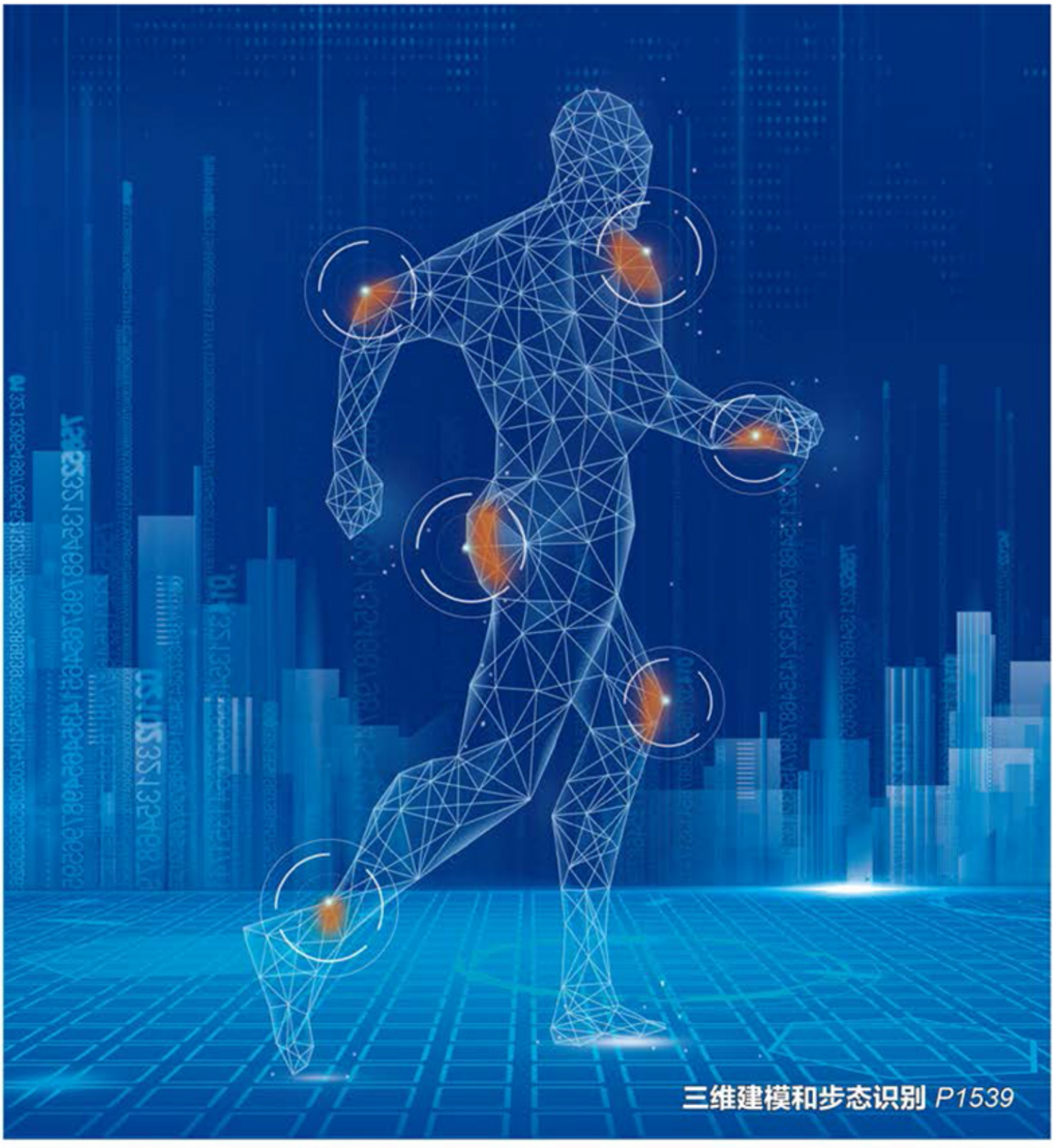
JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020
08
VOL.25

ISSN1006-8961
CN11-3758/TB



三维建模和步态识别 *P1539*

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第25卷第8期（总第292期）
2020年8月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

主管单位 中国科学院
主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主　　编 顾行发
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮　　编 100190
电子信箱 jig@radi.ac.cn
电　　话 010-58887035
网　　址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订　　购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
（邮政信箱：北京399信箱　邮编：100048）
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@radi.ac.cn
Telephone 010-58887035
Website www.cjig.cn

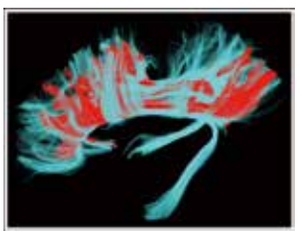
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

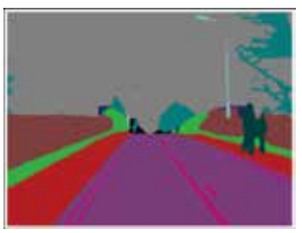
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2020年8月 第25卷 第8期（总第292期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

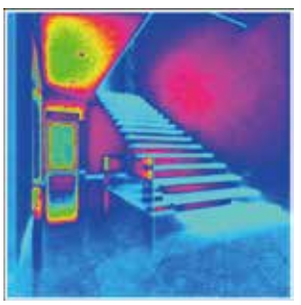
目次



神经纤维跟踪算法研究进展(第1513页)



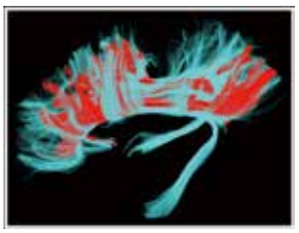
结合注意力机制的双路径语义分割(第1627页)



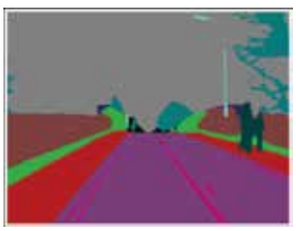
突变策略下多通路Metropolis光照传播(第1658页)

Volume 25, Number 8
Published August 16, 2020

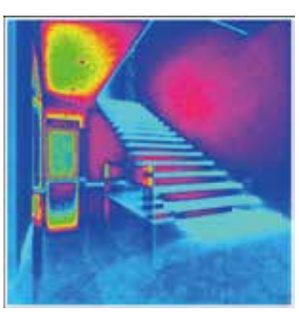
CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Research progress of neural fiber tracking(P1513)



Two-path semantic segmentation algorithm combining attention mechanism(P1627)



Improved multiplexed Metropolis light transport based on fusion mutation strategy(P1658)

中图法分类号:TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2020)08-1539-12

论文引用格式:Luo J, Li M X and Luo S G. 2020. Parametric 3D body modeling and view-invariant abnormal gait recognition. Journal of Image and Graphics,25(08):1539-1550(罗坚,黎梦霞,罗诗光. 2020. 异常步态3维人体建模和可变视角识别. 中国图象图形学报,25(08):1539-1550)[DOI:10.11834/jig.190497]

异常步态3维人体建模和可变视角识别

罗坚,黎梦霞,罗诗光

湖南师范大学信息科学与工程学院,长沙　410000

摘　要:**目的**　运用视觉和机器学习方法对步态进行研究已成为当前热点,但多集中在身份识别领域。本文从不同的视角对其进行研究,探讨一种基于点云数据和人体语义特征模型的异常步态3维人体建模和可变视角识别方法。**方法**　运用非刚性变形和蒙皮方法,构建基于形体和姿态语义特征的参数化3维人体模型;以红外结构光传感器获取的人体异常步态点云数据为观测目标,构建其对应形体和姿态特征的3维人体模型。通过ConvGRU(convolution gated necurrent unit)卷积循环神经网络来提取其投影深度图像的时空特征,并将样本划分为正样本、负样本和自身样本三元组,对异常步态分类器进行训练,以提高分类器对细小差异的鉴别能力。同时对异常步态数据获取难度大和训练视角少的问题,提出了一种基于形体、姿态和视角变换的训练样本扩充方法,以提高模型在面对视角变化时的泛化能力。**结果**　使用CSU(Central South University)3维异常步态数据库和DHA(depth-included human action video)深度人体行为数据库进行实验,并对比了不同异常步态或行为识别方法的效果。结果表明,本文方法在CSU异常步态库实验中,0°、45°和90°视角下对异常步态的综合检测识别率达到了96.6%,特别是在90°到0°交叉和变换视角实验中,比使用DMHI(difference motion history image)和DMM-CNN(depth motion map-convolutional neural network)等步态动作特征要高出25%以上。在DHA深度人体运动数据库实验中,本文方法识别率接近98%,比DMM等相关算法高出2%～3%。**结论**　提出的3维异常步态识别方法综合了3维人体先验知识、循环卷积网络的时空特性和虚拟视角样本合成方法的优点,不仅能提高异常步态在面对视角变换时的识别准确性,同时也为3维异常步态检测和识别提供一种新思路。

关键词:机器视觉;人体识别;异常步态3维建模;虚拟样本合成;卷积循环神经网络

Parametric 3D body modeling and view-invariant abnormal gait recognition

Luo Jian, Li Mengxia, Luo Shiguang

College of Information Science and Engineering, Hunan Normal University, Changsha　410000, China

Abstract: **Objective** Gait has become a popular research topic that is currently investigated by using visual and machine learning methods. However, most of these studies are concentrated in the field of human identification and use 2D RGB images. In contrast to these studies, this paper investigates abnormal gait recognition by using 3D data. A method based on 3D point cloud data and the semantic body model is then proposed for view-invariant abnormal gait recognition. Compared with traditional 2D abnormal gait recognition approaches, the proposed 3D-based method can easily deal with many obsta-

收稿日期:2019-09-29;**修回日期**:2020-02-19;**预印本日期**:2020-02-26

基金项目:国家自然科学基金项目(61701179,41604117);国家留学基金项目(201808430285);湖南省自然科学基金项目(2019JJ50363)

Supported by: National Natural Science Foundation of China (61701179, 41604117); China Scholarship Council (201808430285); Natural Science Foundation of Hunan Province, China (2019JJ50363)

cles in abnormal gait modelling and recognition processes, including view-invariant problems and interference from external items. **Method** The point cloud data of human gait are obtained by using an infrared structured light sensor, which is a 3D depth camera that uses a structure projector and reflecting light receiver to gain the depth information of an object and calculate its point cloud data. Although the point cloud data of the human body are also in 3D, they are generally unstructured, thereby influencing the 3D representation of the human body and posture. To deal with this problem, a 3D parametric human body learned from the 3D body dataset by using a statistic method is introduced in this paper. The parameterized human body model refers to the description and construction of the corresponding visual human body mesh through abstract high-order semantic features, such as height, weight, age, gender, and skeletal joints. The parameters are determined by using statistical learning methods. The human body is embedded into the model, and the 3D parametric model can be deformed both in shapes and poses. Unlike traditional methods that directly model the 3D body from point cloud data via the point cloud reduction algorithm and triangle mesh grid method, the related 3D parameterized body model is deformed to fit the point cloud data in both shape and posture. The standard 3D human model proposed in this paper is constructed based on the body shape PCA (principal component analysis) analysis and skin method. An observation function that measures the similarity of the deformed 3D model with the raw point cloud data of the human body is also introduced. An accurate deformation of the 3D body is ensured by iteratively minimizing the observation function. After the 3D model estimation process, the features of the raw point cloud data of the human body are converted into a high-level structured representation of the human body. This process not only abstracts the unstructured data to a high-order semantic description but also effectively reduces the dimensionality of the original data. After 3D modelling and structured feature representation, a convolution gated recurrent unit (ConvGRU) recurrent neural network is applied to extract the temporal-spatial features of the projected depth gait images. ConvGRU has the advantages of both convolutional and recurrent neural networks, the latter of which is based on the gate structure. The tow gates (i. e., reset and update gates) help the model memorize useful information and forget useless data. In the final classification process, the samples are divided into positive, negative, and anchor samples. The anchor sample is the sample itself, the positive samples are same-category samples that belong to different objects, and the negative samples are those that belong to opposite categories. Training the classifier by using the triples elements strategy can improve its ability to discriminate small feature differences of different categories. At the same time, a virtual 3D sample synthesizing method based on body, pose, and view deformation is proposed to deal with the data shortage problem of abnormal gait. Compared with normal gait datasets, abnormal gait data, especially 3D abnormal datasets, are rare and difficult to obtain. Moreover, given the limited amount of ground truth data, most of the abnormal data are imitated by the experimental participates. As a result, the virtual synthesizing method can help extend the training data and improve the generalization ability of the abnormal gait classification model. **Result** Experiments were performed by using the CSU (Central South University) abnormal 3D gait database and the depth-included human action video (DHA) dataset, and different abnormal gait or action recognition methods were compared with the proposed approach. In the CSU abnormal gait database, the rank-1 mean detection and recognition rate of abnormal gait is 96.6% at the 0°, 45°, and 90° views. In the 90°-0° cross view recognition experiment, the proposed method outperforms the other approaches that use DMHI (difference motion history image) or DMM-CNN (depth motion map-convolutional neural network) as feature representation by at least 25%. Meanwhile, in the DHA dataset, the proposed method result has a rank-1 mean detection and recognition rate of near 98%, which is 2% to 3% higher than that of novel approaches, including DMM based methods. **Conclusion** Based on the feature extraction method of the 3D parameterized human body model, abnormal gait image data can be abstracted to high-order descriptions and effectively complete the feature extraction and dimensionality reduction of the original data. ConvGRU can extract the spatial and temporal features of the abnormal gait data well. The virtual sample synthesis and triple classification methods can be combined to classify and recognize abnormal gait data from different views. The proposed method not only improves the recognition accuracy of abnormal gait under various view angles but also provides a new approach for the detection and recognition of abnormal gait.

Key words: machine vision; human recognition; 3D abnormal gait modeling; virtual sample generation; convolutional recurrent neural network

0 引　言

异常步态通常指人体行走时的各种非正常的动作和姿态，具有明显的动态特性。作为一种重要的生物特征，它可以用来实现安防领域中的异常行为监测，自动驾驶环境下的行人异常分析，医学中的步态症状检测和分析等（Pogorelc 等，2012）。

随着各种监控摄像头的布置以及微型传感器的发展，人们使用监控视频或穿戴传感器来进行各种异常步态行为的检测，包括变电站中的意外跌倒检测、独自居住老年人的摔倒检测、人体行为分析和公共场合下的异常步态行为监测等（王磊 等，2017；Elmadany 等，2018）。其中，运用图像来进行异常步态行为的检测方法，多使用 2 维彩色摄像机来获取人体的步态运动数据，孙朋等人（2017）使用改进混合高斯模型来去除背景，检测出异常跌倒人体，并根据人体质心所在的相对位置来判断跌倒与否。Bauckhage 等人（2005）使用二值步态轮廓来进行异常步态行为的检测，通过将步态轮廓进行栅格化处理，并提取栅格内部特征数据，使用支持向量机来进行摇摆、摔倒等异常步态行为的检测。Wang（2006）使用步态轮廓光流运动图来进行异常步态特征的表述和分类识别，光流图可以较好地反映前后运动帧的特征变换，具有一定的动态特征。然而能量图将一个步态周期内的图像根据统计方法压缩到了一幅图像上，必然会引起步态时间特征的丢失，从而影响其应用效果。由此可见，使用 2 维图像来进行异常步态行为检测的方法较为直观且实现方便，但局限性也比较明显。因为人类的视觉是 3 维的，2 维步态图像由于缺少了深度信息，在面对视角大幅变化时，检测模型的鲁棒性将会受到限制。因此，Yang 和 Tian（2017）尝试使用深度摄像机来进行人体行为识别，提出了一种使用超法向量来提取人体运动能量特征的方法，以实现分类和识别。Xia 和 Aggarwal（2013）使用一种滤波方法来获取深度视频中的局部时空特征点，提出一种深度长方体相似性特征描述和度量方法，并将其应用于行为检测算法中。

除了使用视频和图像来研究异常步态的特征提取和识别方法外，可穿戴移动电子设备也越来越多地用在异常步态数据的采集和分析上。通过运动传感器所采集的数据，其精度和重复性同样受到传感器的穿戴位置、角度和方式等影响。为解决此问题，Ngo 等人（2015）使用三组运动传感器，通过固定在人体腰部位置来采集步态数据，同时利用陀螺仪的数据来进行坐标修正。但实验结果表明，上下坡相似的步态仍存在一定的混淆。Li 等人（2018）通过将传感器安装在鞋上进行异常步态数据的采集，同时加入了足底压力传感器，因此，模型参数相对较多。总之，无论是哪种运动传感器、安装位置如何，它们都需要被监测对象的主动配合，而且对于传感器的穿戴方式有着明显要求，这使其应用仍受到一定限制。

同样，使用图像进行异常步态行为检测也存在不足，比如使用普通摄像机对老龄人的生活监测，将直接涉及个人隐私。与此同时，异常步态相比正常步态而言，数据量小，训练样本少，视角变换、遮挡和衣着变化等外在因素将直接影响到识别模型的泛化能力和鲁棒性。2 维步态图像由于缺少深度信息，无法构建 3 维空间模型，无法充分发挥 3 维机器视觉的特点，针对此问题，本文探讨一种使用 3 维结构光传感器，通过采集点云数据作为步态数据来源的异常步态识别方法。

基于结构光深度传感器的人体行为检测研究得到了广泛的关注。但结构光传感器所获取的人体点云数据往往是杂乱无章的，同时也会受到光线、背景和遮挡等外界干扰，数据存在噪声和缺失。而且 3 维点云人体模型，由于没有内嵌人体骨架结构，未进行形体参数的学习，无法像参数化人体模型一样进行形体和姿态变形，在进行数据存储时也需要保存所有点云数据，冗余数据多，灵活性不高，无法发挥 3 维人体模型的特点。而使用骨架数据来进行 3 维人体动作识别的算法，则需要准确的骨架关节提取算法，由于容易受到噪声、遮挡和人体形体的影响，在进行较少关节的相似动作识别时，其效果往往不高，比如对画三角形和画圆形状动作的区分度不高。

与此同时，异常人体步态数据获取相对困难，比如摔倒、昏厥和跛脚等数据采集难度大，真实数据较少，大多是通过正常人员或专业演员模拟得到的单一视角数据，因此相关多视角的公共数据集也较少。针对异常人体步态多视角训练样本少、点云人体模型和骨架关节模型存在的不足，本文通过引入 3 维参数化人体模型，利用采集的 3 维人体点云数据，来

估计结构化的人体参数模型,再通过对3维人体模型的视角、形体和姿态变换,来虚拟合成各视角下的数据,从而能有效扩充样本数据,提高识别模型面对视角变换时的鲁棒性。另外,由于人体步态的检测是基于图像序列的,本文使用具有时序特征学习机制和空间特征学习能力的ConvGRU(convolution gated recurrent unit)(Cho等,2014)卷积循环神经网络来学习和提取异常步态的时空特征。最终通过三元组分类器,来完成对异常步态的训练和识别。

1 3维结构光影像和参数化人体模型

1.1 人体光学影像

随着低成本的体感摄像机和3维结构光传感器的出现,人们获取3维数据的精度和效果都大大得到提升。由于3维结构光传感器获取的信号比传统摄像机的数据多了1维的深度信息,因此可以得到更多的原始运动数据,有利于更精细化的特征提取,更好地区分相似动作和行为,以及解决视角变化等问题。

如图1所示,3维结构光传感器和传统的2维摄像机在结构上存在明显不同,常见的红外结构光传感器,其内部除具有正常的彩色摄像机外,还包括红外结构光发射器和结构光接收装置。比起传统的双目3维摄像机,基于结构光和TOF(time of flight)技术的3维传感器显示更加精确,分辨率更高,受光照等外在因素影响更小。同时它们能实时计算所拍摄物体的深度值,相比双目摄像机的复杂算法,实时性能更好。

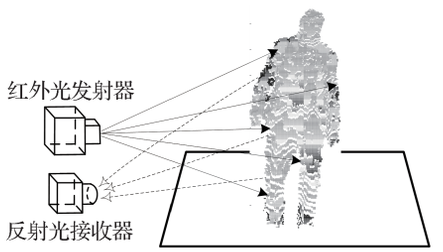


图1 红外结构光传感器采集点云数据示意图

Fig. 1 Diagram of infrared structured light sensor capturing point cloud data

常见的Kinect结构光传感器,可输出RGB彩色图像,以及包含距离信息的深度图像,令$(i,j) \in \mathbf{R}^2$为输出深度图像的像素坐标,像素(i,j)所对应的深度值为d,依据深度图像和结构光传感器的内参数,可计算出所有像素点在3D坐标系中坐标$(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})^{\mathrm{T}} \in \mathbf{R}^3$(Smisek等,2011),即

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\\ \boldsymbol{z}\end{bmatrix}=\frac{1}{c_1 d+c_0}Dis^{-1}\left(\boldsymbol{K}^{-1}\begin{bmatrix}i+u\\ j+v\\ 1\end{bmatrix},k\right) \tag{1}$$

式中,c_0和c_1为模型参数,$Dis(\ )$表示畸变函数,k为畸变参数,u和v为平移校正数值,d为像素点(i,j)对应的深度信息,$\boldsymbol{K}$为结构光传感器校正矩阵,模型参数的获取和校正见Smisek等人(2011)的方法。

1.2 人体语义参数化3维模型

针对3维点云人体模型过于粗糙的问题,以及关节骨架模型在特定应用场景下的不足,以人体点云数据为观测基础,通过对标准参数化人体模型的形体语义参数和姿态语义参数变换,来估计其所对应的参数化3维人体模型。以参数化人体模型的形体和姿态参数为特征,实现人体异常步态分类和识别。所谓的人体语义参数化模型,是指通过抽象的人体高阶语义特征(比如身高、体重、年龄、性别和骨架关节等)来描述和构建与之对应的可视化人体网络模型,其中所涉及的参数变形都是以统计学习方法为基础的。

由于参数化3维人体模型同时嵌入了人体骨架信息,因此通过对标准人体模型的骨架关节变换和形体变形,可使参数人体模型与点云数据所表述人体姿态和形体基本一致,达到对人体姿态和形体参数估计的目的,以及完成对异常步态的3维建模。表1为本文所使用的主要人体形体参数语义特征,令所有形体参数所对应的数值为$\boldsymbol{g}=[\boldsymbol{g}_1,\cdots,\boldsymbol{g}_L]$,$L$为最大形体数值。

参数化模型内嵌有3维CMU mocap(Carnegie Mellon University motion capture database)运动人体关节骨架信息(如图2所示)。骨架模型主要包括22处关节:头、颈、锁骨、肩、肘、手踝、胸骨、臀、膝、脚踝、脚趾和根关节等。所有运动信息通过人体骨架中各关节相对旋转角度进行表示,即姿态语义特征$\boldsymbol{r}=[\Delta\boldsymbol{r}_1,\cdots,\Delta\boldsymbol{r}_M]$,$M$为最大关节数值,$\Delta\boldsymbol{r}_M \in \mathbf{R}^3$。

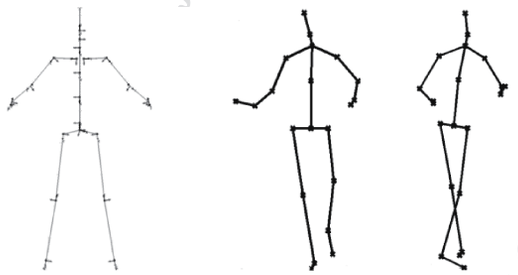
本文使用Makehuman生成的3维人体模型库,每一个3维人体模型都具有固定的顶点数量v_i和网格面数,同时所有模型都内嵌有CMU mocap人体骨架,但生成的人体模型没有衣着。针对此,利用3维

辅助设计软件，生成标准 T 姿态人体衣着一套，对所有训练模型依据 Liu 等人(2017)方法进行虚拟穿衣，衣着前后效果如图 3 所示。

表 1　人体形体语义特征参数

Table 1　Main semantic parameters of human body shape

类别	参数	类别	参数
总体特征	性别	头部	头部尺寸
	身高	颈部	颈长
	体重		颈粗
	肌肉	肩部	肩宽
躯体	躯干厚度	胸	胸部尺寸
	躯干长度	腿	大腿长度
手臂	上臂长度		大腿厚度
	上臂厚度		小腿长度
	前臂长度		小腿厚度
	前臂厚度	脚	脚长

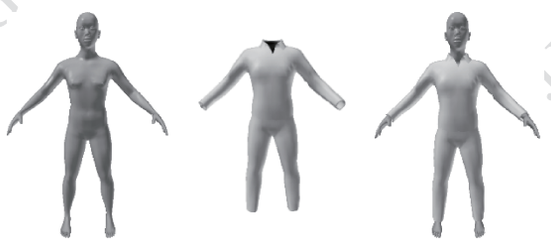


(a) CMU mocap骨架　　(b) 骨骼动画示意图

图 2　3 维人体运动骨架

Fig. 2　3D human motion skeleton

((a) CMU mocap skeleton; (b) skeleton animation diagram)



(a) 人体模型　　(b) 3D衣服　　(c) 虚拟衣着后效果

图 3　虚拟穿衣

Fig. 3　Virtual clothing

((a) body model; (b) 3D clothes; (c) virtual clothing effect)

令所有衣着后的 3 维人体训练模型表示为 $\boldsymbol{S}=\{\boldsymbol{S}^1,\boldsymbol{S}^2,\cdots,\boldsymbol{S}^K\}$，$\boldsymbol{S}^K=\{\boldsymbol{V}^K,\boldsymbol{P}^K\}$，其中 K 为模型样本数，$\boldsymbol{V}=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_M\}$ 表示模型 M 个顶点信息，$\boldsymbol{v}_M\in\mathbf{R}^3$，$\boldsymbol{P}=\{\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_K\}$ 为模型网格面数据，$\boldsymbol{p}_K\in\mathbf{R}^t$，一个网格面含 t 个顶点。对所有模型顶点 $\boldsymbol{V}\in\mathbf{R}^{3M\times K}$ 进行主成分分析得

$$\boldsymbol{V}^k=\boldsymbol{U}\boldsymbol{p}^k+\overline{\boldsymbol{V}} \tag{2}$$

式中，$\boldsymbol{V}^k$ 表示样本 $k\in[1,\cdots,K]$ 的顶点信息，$\boldsymbol{U}\in\mathbf{R}^{3M\times K}$ 为特征矩阵，$\boldsymbol{p}^k\in\mathbf{R}^{K\times 1}$ 为个体形体差异系数，$\overline{\boldsymbol{V}}\in\mathbf{R}^{3M\times 1}$ 为 $\{\boldsymbol{V}^K\}$ 均值向量。PCA 可以将个体形体的差异变化进行表征，但是并不能直接构建出抽象的形体语义特征(如身高、体重和臂长等)与 3 维参数模型的重建关系。因此采用线性回归分析方法，来实现完成语义特征对 3 维人体模型的直接变形。见表 1 所示，令有 L 种形体语义参数，$\boldsymbol{g}_L^k$ 表示样本 k 对应的形体特征值，构建如下投影 $\boldsymbol{F}$ 矩阵

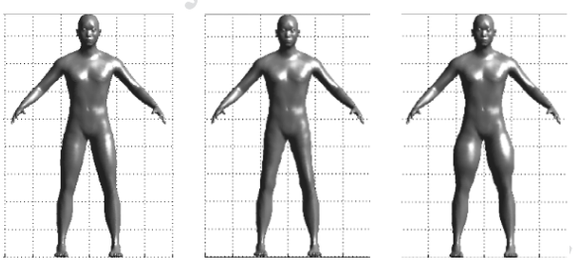
$$\boldsymbol{F}[\boldsymbol{g}_1^k,\cdots,\boldsymbol{g}_l^k,\cdots,\boldsymbol{g}_L^k,1]^{\mathrm{T}}=\boldsymbol{p}^k \tag{3}$$

联合所有训练样本，投影矩阵式(3)可表示为

$$\boldsymbol{F}=\boldsymbol{P}\boldsymbol{G}^- \tag{4}$$

式中，$\boldsymbol{P}=[\boldsymbol{p}^1,\cdots,\boldsymbol{p}^K]$，$\boldsymbol{G}=[\boldsymbol{g}^1,\cdots,\boldsymbol{g}^K,1]^{\mathrm{T}}$，$\boldsymbol{G}^-$ 为 $\boldsymbol{G}$ 的逆矩阵。为了更好地进行形体变化，采用增量变形方式，令 $\Delta\boldsymbol{g}=[\boldsymbol{g}_1^k,\cdots,\boldsymbol{g}_l^k,\cdots,\boldsymbol{g}_L^k,\boldsymbol{0}]^{\mathrm{T}}$，PCA 各个系数权值变化为 $\Delta\boldsymbol{p}=\boldsymbol{F}\Delta\boldsymbol{g}$，新生成的人体模型 $\boldsymbol{V}^{\mathrm{new}}=\boldsymbol{U}(\bar{\boldsymbol{p}}+\Delta\boldsymbol{p})^k+\overline{\boldsymbol{V}}$。

如图 4 所示，利用大腿相对厚度为 1 和 0.4 的参数模型进行训练，估计大腿厚度为 1.5 的 3 维参数化人体模型如图 4(c)所示。



(a) 大腿标准厚度 1　　(b) 相对厚度 0.4　　(c) 估计相对厚度 1.5的人体模型

图 4　按大腿厚度语义参数进行训练和估计的人体模型

Fig. 4　Human models trained and estimated according to the semantic parameters of thigh thickness((a) standard thigh thickness 1; (b) related thigh thickness 0.4; (c) estimated body model with 1.5 relative thigh thickness)

3 维人体模型的姿态变形采用骨骼蒙皮方法，将网格模型作为皮肤，绑定到人体的骨架上，即将网格顶点附着在不同的骨头上。在姿态变形时，先进

行关节骨架变化,然后更新对应网格顶点即可。对于每个关节,如图2(a)所示,都定义有局部坐标系,采用3个欧拉角的旋转变换来实现各个关节的相对运动。定义绕X轴、Y轴和Z轴旋转角度分别为α,β和γ,刚性模型联合变换矩阵为$\boldsymbol{R}_{XYZ}$,c表示cos运算,s表示sin运算,则

$$\boldsymbol{R}_{XYZ} = R_X(\alpha)\,R_Y(\beta)\,R_Z(\gamma) = \begin{bmatrix} c\beta c\gamma & -c\beta s\gamma & s\beta \\ s\alpha s\beta c\gamma + c\alpha s\gamma & -s\alpha s\beta s\gamma + c\alpha c\gamma & -s\alpha c\beta \\ -c\alpha s\beta c\gamma + s\alpha s\gamma & c\alpha s\beta s\gamma + s\alpha c\gamma & c\alpha c\beta \end{bmatrix} \tag{5}$$

因此,给定形体特征$\boldsymbol{g}$和姿态变化特征$\boldsymbol{r}$,通过形体$\boldsymbol{S}(\boldsymbol{g})$和姿态旋转变形$\boldsymbol{R}(\boldsymbol{r})$,就可以合成新的人体模型,令$\boldsymbol{X}_{\text{std}}$表示标准姿态和形体模型,$\hat{\boldsymbol{Y}}$为依据形体和姿态语义参数生成的新模型,则

$$\hat{\boldsymbol{Y}} = f(\boldsymbol{g},\boldsymbol{r}) = \boldsymbol{R}(\boldsymbol{r}) \cdot [\boldsymbol{S}(\boldsymbol{g}) \cdot \boldsymbol{X}_{\text{std}}] \tag{6}$$

1.3 人体异常步态3维模型估计

使用单个3维传感器采集到的人体点云数据存在自遮挡,只有前表面的数据,且原始的点云数据杂乱无章,数量不一,存在噪音干扰(见图5所示)。由于未经过网格剖面处理和嵌入骨架,人体点云数据不能进行姿态和形体变形。因此,需要以采集点云数据为观测对象,估计其对应的参数化人体模型。

(a) 90° (b) 45° (c) 0°

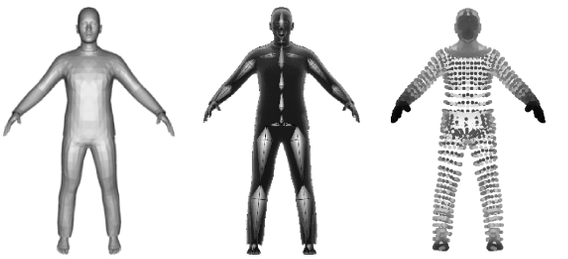
图5 90°,45°和0°异常步态点云深度图像投影示意图

Fig. 5 90°, 45° and 0° abnormal gait depth images of point cloud data((a) 90°; (b) 45°; (c) 0°)

令所采集人体步态点云数据视角α,经过归一化处理后,其对应的点云投影深度图像定义为$\boldsymbol{P}_\alpha$,如图5所示,不同灰度颜色表示不同的深度信息。令$\boldsymbol{r}_\eta$姿态和标准形体$\boldsymbol{g}_{\text{std}}$下的参数化人体模型及其点云深度图像投影如图6所示。其3维模型表示为

$$\hat{\boldsymbol{Y}}_\eta = \boldsymbol{R}(\boldsymbol{r}_\eta) \cdot \boldsymbol{S}(\boldsymbol{g}_{\text{std}}) \cdot \boldsymbol{X}_{\text{std}} \tag{7}$$

定义3维参数化人体模型在α视角下的点云投影深度图像为$\boldsymbol{Y}_\alpha(\boldsymbol{r},\boldsymbol{g})$(如图6(c)所示),为实现图5所采集点云数据到参数化人体模型的估计,定义以下



(a) 3D人体模型 (b) 人体模型骨架 (c) 人体点云

图6 嵌入骨架的参数化人体模型和0°点云深度图像投影

Fig. 6 Skeleton embedded parametric body model and its point cloud projecting image in 0°((a) 3D body model; (b) skeletion of body model; (c) point cloud data of body)

基于深度点云轮廓和重要关键关节点匹配相似度函数,即

$$E = w_1 \sum_{i=1}^{I} \left\| \Gamma_i(\boldsymbol{Y}_\alpha(\boldsymbol{r},\boldsymbol{g})) - \Gamma_i(\boldsymbol{P}_\alpha) \right\|^2 + w_2 \sum_{n=1}^{N} \left\| Mark_n(\boldsymbol{Y}_\alpha(\boldsymbol{r},\boldsymbol{g})) - Mark_n(\boldsymbol{P}_\alpha) \right\|^2 \tag{8}$$

式中,w_1和w_2为权重值,$\Gamma_i(\cdot)$定义为点云深度图像中的人体轮廓提取函数,表示当前视角下人体边沿轮廓中的第i个离散点相对坐标r_i和深度值d_i,表示为(r_i,d_i)。人体边沿轮廓提取是以人体表面质心为3维坐标原点,按顺时针方向取I个点,表示为$r_i = x_i i + y_i j$。$\boldsymbol{P}_\alpha$表示采集的点云人体轮廓在视角α下的投影深度图像(见图5),同样以人体表面质心作为参考原点,求取其I个轮廓点信息$\Gamma_i(\boldsymbol{P}_\alpha)$。

$Mark_n(\boldsymbol{Y}_\alpha(\boldsymbol{r},\boldsymbol{g}))$表示视角$\alpha$下,参数化人体模型深度投影图像中的第$n$个重要关节点坐标和深度值,表示为$(r_n,d_n)$,$Mark_n(\boldsymbol{P}_\alpha)$表示点云人体投影图像$\boldsymbol{P}_\alpha$的第$n$个重要关节点对应坐标和深度值。本文取人体模型的头部关节点,左右手踝关节点和左右脚踝关节点作为5个人体重要关节点。重要关节点的确定参见Shotton等人(2013)方法中所提出的人体关节点提取算法。如式(8)所示,外围轮廓距离差度量函数保证了参数人体模型和采集点云人体模型的全局匹配,而重要关节点距离度量函数,保证估计参数人体模型的局部细节的匹配。

因此,以采集的3维人体点云轮廓和重要关节点为约束,通过求解投影轮廓相似度函数的极小值:argmin E,即可得到对应最优的形体和姿态估计语义特征值$\boldsymbol{r}_{\text{opt}}$和$\boldsymbol{g}_{\text{opt}}$,所对应的3维参数化人体模型

为 $\boldsymbol{Y}_{\text{opt}} = \boldsymbol{R}(\boldsymbol{r}_{\text{opt}}) \cdot \boldsymbol{S}(\boldsymbol{g}_{\text{opt}}) \cdot \boldsymbol{X}_{\text{std}}$，式(8)为非线性优化问题，若初值选取不合适，在求解过程中可能陷入局部最优。因此在求解的过程中，需要尽可能选取接近最优解的良好初值，同时控制迭代演化过程中参数的变化速度和限定范围，避免局部最优解的情况。参考前期研究成果(Luo 等，2016)，通过聚类分析方法来选取良好的初值，再利用改进鲍威尔共轭方向迭代法，先固定标准形体，对姿态参数 r 进行迭代，然后固定姿态参数，再对形体 $\boldsymbol{g}$ 进行优化求解，最后再得到联合最优解的形式来求解式(8)的最优解，从而估计出 3 维人体的最优姿态和形体参数。图 7 为跛脚和摔倒两个异常步态的 3D 人体估计模型示意图。

(a) 坡脚点云　(b) 坡脚3D模型　(c) 摔倒点云　(d) 摔倒3D模型

图 7　异常人体步态估计模型

Fig. 7　Estimated abnormal human gait model

((a) point cloud data of the lame body; (b) 3D model of the lame body; (c) point cloud data of the falling body; (d) 3D model of the falling body)

2　多视角 3 维异常步态样本合成方法

正常的步态数据可以通过固定场景下的摄像机来获取。但异常步态数据获取难度大，通常只能通过人为模拟方式来获得。如果没有足够的异常步态数据，只有小样本或单一视角的数据库，将极大地影响到异常步态检测模型的泛化能力和面对各种情境下的识别准确性。

针对此问题，通过对参数化 3 维人体模型的形体、姿态和视角参数进行适当变换来虚拟合成各视角下的新样本，从而达到扩充异常步态数据库的目的。令当前的标准异常步态模型的形体和姿态参数为 $\boldsymbol{g}_{\text{std}}$ 和 $\boldsymbol{r}_a$，其 3 维人体模型表示为 $\widehat{\boldsymbol{Y}} = \boldsymbol{R}(\boldsymbol{r}_a) \cdot \boldsymbol{S}(\boldsymbol{g}_{\text{std}}) \cdot \boldsymbol{X}_{\text{std}}$。首先，固定姿态参数，生成虚拟的形体参数集，表示为 $\boldsymbol{\Omega} = \{\bar{\boldsymbol{g}}\}$，式中 $\bar{\boldsymbol{g}}$ 服从多维高斯正态分布，即

$$N(\bar{\boldsymbol{g}} \mid \boldsymbol{g}_{\text{std}}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} (\bar{\boldsymbol{g}} - \boldsymbol{g}_{\text{std}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{g}} - \boldsymbol{g}_{\text{std}}) \right] \tag{9}$$

式中，$\boldsymbol{\Sigma}$ 表示协方差，$\boldsymbol{g}_{\text{std}}$ 为 D 维的标准形体参数向量。然后，对形体变换后的人体模型参考上述方法，变换姿态参数，生成虚拟姿态参数集，表示为 $\boldsymbol{\Phi} = \{\bar{\boldsymbol{r}}\}$，同样符合正态分布，但是其方差必须要限制在较小的范围之内，即如果姿态变换太大，则有可能成为另一种异常步态，如果只是小幅度变换，则可以归属于同一类别。形体和姿态参数虚拟合成完成之后，再变换视角，生成不同视角下的点云轮廓投影，定义视角集 $\boldsymbol{\Theta} = \{\alpha_0, \alpha_1, \cdots, \alpha_N\}$，共 N 个离散视角，视角 α 下点云投影深度图像集表示为

$$\boldsymbol{\Psi} = \{\boldsymbol{Y}_\alpha(\bar{\boldsymbol{r}}, \bar{\boldsymbol{g}}) \mid \alpha \in \boldsymbol{\Theta}, \bar{\boldsymbol{r}} \in \boldsymbol{\Phi}, \bar{\boldsymbol{g}} \in \boldsymbol{\Omega}\} \tag{10}$$

式中，集合 $\boldsymbol{\Psi}$ 即为虚拟合成的异常步态样本特征集，图 8 为虚拟异常步态样本示意图。

(a) 真实点云数据　(b) 估计参数化人体模型　(c) 虚拟形体变换　(d) 虚拟姿态变换　(e) 不同投影视角

图 8　虚拟异常步态样本示意图

Fig. 8　Diagram of synthetized virtual abnormal gait data

((a) real point cloud data; (b) estimated parameterized human model; (c) virtual shape transformation; (d) virtual posture transformation; (e) different projecting view)

3　基于卷积时空网络的异常步态识别

以 3 维参数化人体模型在 α 视角下的点云投影深度图像 $\boldsymbol{Y}_\alpha(\boldsymbol{r}, \boldsymbol{g})$(见图 6)作为异常步态时空特征提取的来源数据。对于 2 维图像，深度卷积神经网络(convolutional neural network，CNN)在图像分割、目标分类等图像处理问题上有很好的效果，因此本文拟将深度卷积网络引入到对步态深度图的特征提取上。但步态数据具有周期性的特点，即时空特征，因此，如果仅使用卷积神经网络，则只提取到空间特征，不利于对步态时空特征的提取，进而影响到分类和识别效果。因此，本文采用具有时空特征的 ConvGRU 时空卷积网络来进行步态时空特征提取。

由于异常步态和正常步态的分析不同,正常步态具有明显的周期性,而异常步态周期不显示,因此,在进行异常步态分析时,不进行步态周期的估计,只提取固定 L_c 帧长度的步态投影深度图像进行特征提取。本文实验中选取 L_c = 20 帧步态数据,即将所有视频按 L_c 帧进行分割并附注异常步态类别标签。将 L_c 帧的 α 视角下的点云步态投影深度图,依据时间先后顺序,输入到 ConvGRU 深度卷积循环网络中,提取其时空特征,表示为 $\tilde{F} = convGRU(\boldsymbol{Y}_{k-L_c},\cdots,\boldsymbol{Y}_{k-l_c},\cdots,\boldsymbol{Y}_k), l_c \in [1, L_c]$,式中,$\boldsymbol{Y}_k$ 表示视频中的第 k 帧投影深度图像,取其相邻的 L_c 帧提取时空特征。

ConvGRU 卷积循环神经网络,是基于卷积网络和 GRU 循环神经网络两种特性的时空卷积网络,卷积网络对 2 维图像进行多尺度的特征提取效果很好,而循环神经网络则可充分记忆时序特征。基本的 GRU 单元是基于两种门的结构,一个为更新门 $\boldsymbol{z}_t$,另一个为重置门 $\boldsymbol{r}_t$,相比传统的长度时序记忆模型 LSTM(long short-term memory),GRU 少了 1 个门结构,同时移除了细胞状态单元,使得其结构更简单有效(Cho 等,2014)。ConvGRU 使用隐含状态来进行信息的传递,使用卷积结构来代替原来的全连接结构,其基本网络单元结构定义为

$$\begin{cases} \boldsymbol{r}_t = \sigma(\boldsymbol{x}_t * \boldsymbol{w}_r + \boldsymbol{h}_{t-1} * \boldsymbol{u}_r + \boldsymbol{b}_r) \\ \boldsymbol{z}_t = \sigma(\boldsymbol{x}_t * \boldsymbol{w}_z + \boldsymbol{h}_{t-1} * \boldsymbol{u}_z + \boldsymbol{b}_z) \\ \boldsymbol{h}'_t = f(\boldsymbol{x}_t * \boldsymbol{w}_h + \boldsymbol{r}_t \odot \boldsymbol{h}_{t-1} * \boldsymbol{u}_h + \boldsymbol{b}_h) \\ \boldsymbol{h}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{h}'_t + \boldsymbol{z}_t \odot \boldsymbol{h}_{t-1} \end{cases} \tag{11}$$

式中,$\boldsymbol{h}_t$ 表示更新的状态,由之前的隐含状态 $\boldsymbol{h}_{t-1}$ 和新的候选记忆状态 $\boldsymbol{h}'_t$ 共同决定。更新门 $z_t \subset [0,1]$,决定了之前隐含状态和新的候选状态在更新状态中的权重,而重置门 $\boldsymbol{r}_t$ 决定了之前的隐含状态 $\boldsymbol{h}_{t-1}$ 在当前候选记忆 $\boldsymbol{h}'_t$ 中的重要程度。$\odot$ 表示元素乘法,$\boldsymbol{w}$ 为要学习的模型参数,$\boldsymbol{b}$ 表示偏置参数,σ 表示 sigmoid 函数,f 为 tanh 激活函数。将所有步态投影深度图像以 L 帧长为单位,通过 ConvGRU 网络进行时空特征提取,令所有的异常步态时空特征集表示为

$$\boldsymbol{X} = \{\tilde{\boldsymbol{F}}^{\kappa}_{n,\alpha} \mid n \in [1, N]\} \tag{12}$$

式中,N 表示最大的时空特征样本数,α 为视角信息,κ 表示样本所属的异常步态类别标签。对所有的样本,按照三元组进行分类,分别为自身样本、正样本和负样本。其中,正样本是与自身样本属于同一类的样本,而负样本则是与自身样本不在同一类的样本。定义分类器 $C(x)$,x 为待分类的输入样本,$C(x) = Wx + b$ 为分类器,W 为分类器所要学习的权重参数,定义基于三元组的能量损失函数为

$$L_{\text{tri}} = \sum_{n}^{N} \left[\| \tilde{\boldsymbol{F}}^{\kappa_n}_n - \tilde{\boldsymbol{F}}^{\text{pos}}_n \|_2^2 - \| \tilde{\boldsymbol{F}}^{\kappa_n}_n - \tilde{\boldsymbol{F}}^{\text{neg}}_n \|_2^2 + \delta \right] \tag{13}$$

式中,δ 表示正样本和负样本的边界值,$\tilde{\boldsymbol{F}}^{\kappa_n}_n$ 表示第 n 个训练样本时空特征,$\tilde{\boldsymbol{F}}^{\text{pos}}_n$ 为其同一类的正样本,$\tilde{\boldsymbol{F}}^{\text{neg}}_n$ 表示其负样本,属于其他类别。通过最小化三元组能量损失函数,完成对三元组分类器的学习。基于三元组的分类器,可以将同类样本差异极小化,不同类样本之间的差异最大化,能很好地完成对测试样本的分类识别。

4 实验与结果分析

4.1 CSU 3 维异常步态数据库

CSU(Central South University)3 维人体异常步态数据库(见图 9)是由红外结构光传感器采集的(罗坚 等,2016)。该数据库共拍摄了 10 个样本的动作,每一个样本的步态序列包括 6 种状态:正常行走、坐下、异常向前摔倒、异常向后摔倒、左脚异常行走和右脚异常行走,每种状态采集 3 次。该数据库采集了 3 个角度的数据:0°,45° 和 90°,分辨率为 640×480 像素。CSU 人体点云数据经过背景去除,点云精简和归一化等操作。

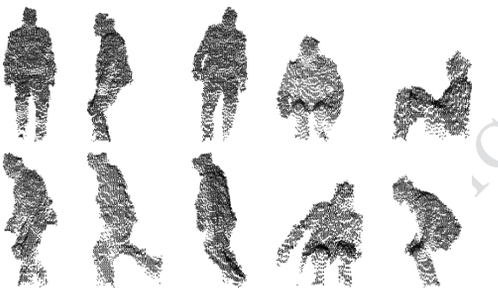


图 9 不同视角下的 3D 异常步态点云数据

Fig. 9 3D abnormal gait point cloud data under different views

使用该多视角异常步态库来验证本文 3 维异常步态识别方法在交叉视角时的效果,即训练和识别非同一个视角。首先,进行特征提取和分类模型的训练。步骤如下:

1)根据采集的不同目标人物,将异常动作分为2组,前1组为库中前1,3,5号目标人物数据,每个人物都采集了6种动作,3个视角,剩余的7个目标人物数据归属于第2组数据,实验设置训练样本数小于测试样本,以更好地验证虚拟合成样本方法的有效性。

2)对所有步态点云视频序列按照 $L = 20$ 帧进行分割,可以有重叠部分。按照本文所提出的基于点云轮廓的参数化3维步态估计方法,估计每帧点云所对应的3维参数化步态模型,并进行虚拟样本合成。针对虚拟样本合成,参照高斯分布,完成30种虚拟形体变换和30种虚拟姿态变换,共900个虚拟数据,将其与正常样本一起投影到 $\boldsymbol{\Theta} = \{0°, 45°, 90°\}$,共3个离散视角。然后再对各个视角 $\alpha \in \boldsymbol{\Theta}$ 投影深度图像提取时空特征,即

$$\widetilde{\boldsymbol{F}}^{\kappa}_{n,\alpha} = convGRU(\boldsymbol{Y}^{\alpha}_{n,1}, \boldsymbol{Y}^{\alpha}_{n,2}, \cdots, \boldsymbol{Y}^{\alpha}_{n,5}, \boldsymbol{Y}^{\alpha}_{n,6}) \quad l \in [1, L] \tag{14}$$

并标记其所属类别,以及它们的正样本和负样本。图10为3维异常步态行为识别方法在不同视角下的混淆矩阵。从图10中可以看到,下蹲和向后摔倒,以及正常行走和跛脚行走(跛左(脚)和跛右(脚))几个相似的步态动作易混淆。同时,对于左右脚异常行走,由于在45°和90°视角下存在自遮挡的问题,在一定程序上会影响到识别的结果。而0°视角下步态轮廓特征没有前面两个角度明显,因此参数化人体模型估计精度稍低,识别结果受到一定影响。

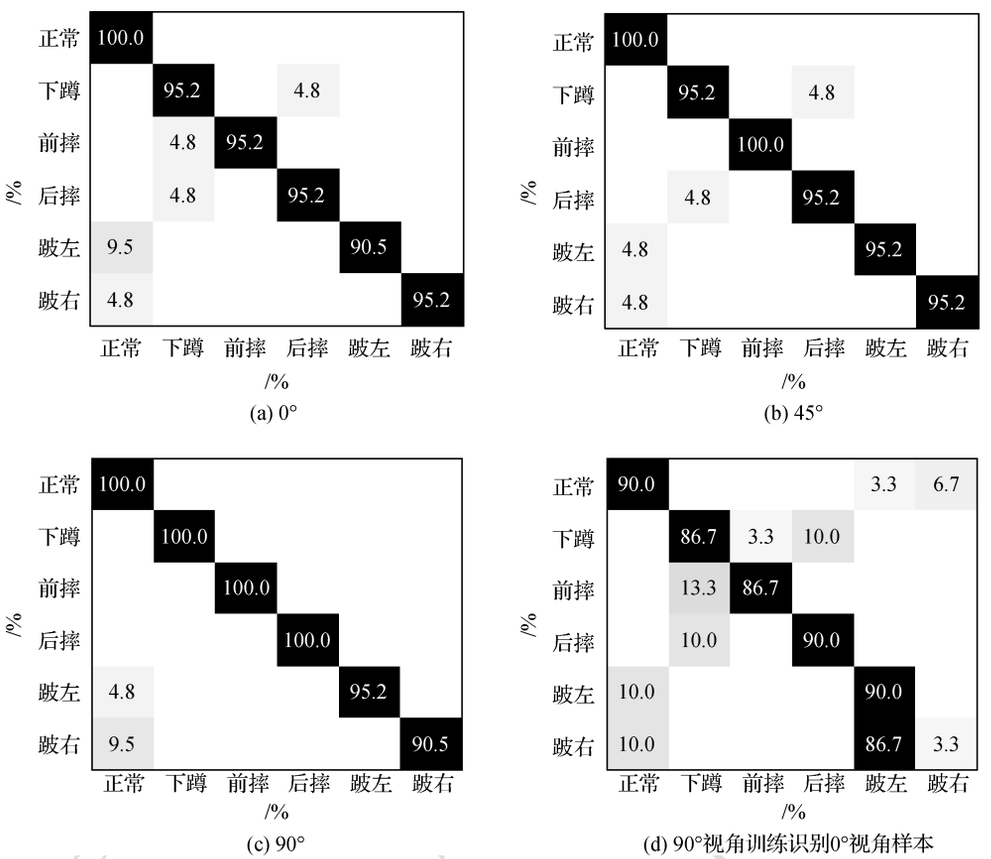


图10　不同视角下的识别混淆矩阵

Fig. 10　Recognition confusion matrix under different views((a) 0°; (b) 45°; (c) 90°; (d) 90° training and test 0°)

表2通过使用不同的步态特征来训练三元组分类器,并进行测试对比。主要特征包括:GEI(gait energy image)步态能量图(Han 和 Bhanu,2006),GFI(gait flow image)步态光流图(Lam 等,2011),差分深度运动历史图 DMHI(difference motion history image)(Gao 等,2015),基于 CNN 的深度运动图 DMM-CNN(depth motion map-CNN(Elmadany 等,2018)。

上述交叉视角实验中,仅使用90°视角数据进行训练,识别0°视角数据(训练和识别视角不相同)。从对比结果可以看出,本文提出的异常步态检测识别方法的效果要明显优于其他5种方法。其

表 2　不同步态特征的异常动作识别精度

Table 2　Abnormal gait recognition precision on different gait features

方法	正常/%	90°—0°交叉视角/%	平均/%
GEI	83.6	37.2	60.4
GFI	84.9	40.6	62.8
DMHI	92.6	53.9	73.3
DMM-CNN	93.1	61.1	77.1
本文	96.6	88.3	92.5

主要原理在于,基于 3 维参数化人体模型的特征提取方法,可以充分发挥 3 维光学步态影像的优点,不仅能提取 3 维特征,通过使用参数化人体模型,还可以虚拟合成不同形体、姿态和视角的样本,以扩充训练集,提高分类器的泛化能力,这在交叉视角实验中特别明显;同时,本文利用 ConvGRU 时空卷积网络对异常步态进行特征提取,可以将异常步态的时空特征充分挖掘出来。

前面两种特征,GEI 和 GFI 主要是针对二值步态轮廓图像进行特征表述的,在 2 维步态识别中应用较多,由于不能利用深度信息,其效果不明显,同时它们都需要一个完整周期的步态图像序列,才能更好地完成步态动作特征表示。DMHI 和 DMM-CNN 都是基于深度图像或点云图像的步态动作特征提取方法,效果要明显优于 GEI 和 GFI。但是它们都使用了压缩的能量图来进行特征表示,即将多个深度步态图通过统计平均方法在一张图上进行表示,会丢失有效的异常步态时序特征。比如,侧面视角的右脚异常由于自遮挡现象的存在,若干动作帧状态与正常步态行走相似(自遮挡时),如果不能有效地利用 3 维深度信息以及前后的时序特征,必然会影响到细小差异步态的分类效果。而本文所提出的基于时空卷积网络和三元组的分类方法,对此类问题有明显的针对性,效果更好。

4.2　DHA 深度人体行为数据库

DHA(depth-included human action video)人体行为数据集(Lin 等,2012)是一个包含深度数据的人体动作库。它是由 Microsoft Kinect 结构光传感器采集录制的,视频分辨率为 640×480 像素。数据集共包含 21 个人(12 名男生和 9 名女生),每个人采集 17 个动作,每个动作都有 RGB 图像、二值轮廓图和深度数据。17 个动作类别分别为:弯腰(Bend,Bd),Jack 动作(Jack,Jk),跳跃(侧面,Jump,Jp),招手(单手,One-handwave,Oh),跳跃(前向,Pjump,Pj),跑(Run,Rn),侧伸腿(Side,Sd),跳绳(Skip,Sk),招手(双手,Two-hand wave,Tw),步行(Walk,Wk),拍手(向前,Clap front,Cf),手臂摆动(Arm-swing,As),踢腿(Kick leg,Kl),投掷(Pitch,Pt),挥杆(Golf-Swing,GSw),拳击(Boxing,Bx)和太极(Tai-chi,Tc)。该数据集一共有 357 个视频序列,有些动作非常相似,如跑步、跳跃和跳绳等。

采用目标交叉方法进行实验,即前 10 人的视频数据作为训练使用,剩余 11 人的视频序列用于测试。由于 DAH 只录制了固定角度的视频,因此不存在交叉视角的实验,在实验时参考前面实验方法,只针对虚拟形体样本进行合成,并将合成的虚拟样本加入到训练数据中。表 3 对比了使用深度图进行动作识别的各种算法。从表 3 中可以看出,在使用深度图像进行行为分类识别时,本文所提出的 3 维异常步态识别方法,其效果要优于其他算法。

表 3　不同方法的动作识别结果

Table 3　Recognition rates of actions using different methods

方法	识别率/%
Lin 等人(Lin 等,2012)	87.0
DMHI-PHOG(Gao 等,2014)	90.6
D-DMHI-PHOG(Gao 等,2015)	92.4
DMMs-FV(Liu 等,2015)	95.4
cHCRF(Chen 等,2016)	95.9
本文	97.9

本文使用深度图像来构建 3 维参数化模型,而非直接从杂乱无章和存在噪音干扰的深度图像中提取特征进行分类识别。同时通过估计动作的参数化人体模型来虚拟合成不同形体的新的人体动作数据,从而有效扩充了样本数量,提高了系统识别精度和鲁棒性。

图 11 为本文方法对 DHA 人体行为数据库进行识别的混淆矩阵。从混淆矩阵图中可以看出,本文方法在对 17 类行为分类识别时,有 13 类的识别精确度都达到了 100%,实用性较强。

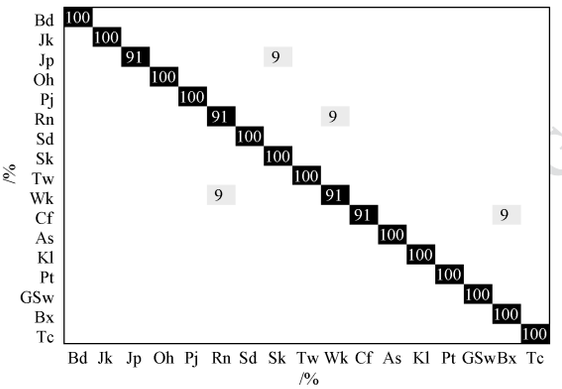


图 11　DHA 行为数据库的混淆矩阵

Fig. 11　Confusion matrix of DHA dataset

5　结　论

通过红外结构光传感器所获取的步态点云数据来估计其对应的参数化语义人体模型，基于参数化步态模型，提出了一种视角可变的异常步态样本虚拟合成的方法，以解决现实中异常样本不足的问题，达到提高异常步态识别模型泛化能力的目的。

为验证本文方法在可变视角下的识别效果，在 CSU 3 维异常步态数据库上，设计了视角大幅变化时的对比实验，即测试的 0°视角异常步态数据不参与训练，训练仅使用 90°视角数据。通过对 0°视角数据进行 3 维人体建模，以及形体、姿态和视角变化来虚拟合成 90°视角数据，以完成与训练库中 90°视角数据的对比和识别。识别结果显示，尽管在 90°—0°交叉视角实验中，整体的识别率比视角固定时要低 8% 左右，但基于 3 维人体模型的视角变化方法，为视角可变异常步态检测提供了一种新方法，比使用 DMHI 和 DMM-CNN 等 2 维步态运动特征要高出 25% 以上。因为后者基于 2 维的特征表达不能进行 3 维下的视角变化，只能将 0° 和 90° 具有明显视觉差异的数据直接进行对比，因此在交叉视角实验时，识别效果明显下降。

在 DHA 深度人体运动数据库实验中，本文方法识别率接近 98%，比 DMMs-FV/LBP 和 cHCRF/LBP 算法高出 2%—3%。DHA 不是一个多视角数据库，但本文方法综合了 3 维人体先验知识、循环卷积网络的时空特性、虚拟样本合成特点和三元组异常步态分类器的类内间区别能力强等优点，不仅能提高异常步态在面对视角变换时的识别准确性，同时能充分有效地提取异常步态的时空特征，有效区别异常步态中的细小差异，提高识别效果和提升算法在面对各种情境下的鲁棒性。

本文研究的视角可变异常步态识别方法是以参数化人体语义模型为基础，它具备 2 维人体模型没有的形体、姿态和视角等变换特性。但是杂乱的点云数据或深度图像在一定程序上影响了模型估计的精度。更加准确和快速地完成异常人体建模将是下一步研究的重点，比如通过事先构建一个较为完备的异常步态行为 3 维模型库，来加速模型的估计等。

参考文献(References)

Bauckhage C, Tsotsos J K and Bunn F E. 2005. Detecting abnormal gait//Proceedings of 2nd Canadian Conference on Computer and Robot Vision. Victoria: IEEE: 1-7 [DOI: 10.1109/CRV.2005.32]

Chen C, Liu M Y, Zhang B C, Han J G, Jiang J J and Liu H. 2016. 3D action recognition using multi-temporal depth motion maps and Fisher vector//Proceedings of the 25th International Joint Conference on Artificial Intelligence. Sony: AAAI Press: 3331-3337

Cho K, Van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H and Bengio Y. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation [EB/OL]. [2019-09-22] https://arxiv.org/pdf/1406.1078.pdf

Elmadany N E D, He Y F and Guan L. 2018. Information fusion for human action recognition via biset/multiset globality locality preserving canonical correlation analysis. IEEE Transactions on Image Processing, 27(11): 5275-5287 [DOI: 10.1109/TIP.2018.2855438]

Gao Z, Zhang H, Liu A, Xue Y B. 2014. Human action recognition using pyramid histograms of oriented gradients and collaborative multi-task learning. KSII Trans on Internet and Information Systems. 8(2): 483-503 [DOI: 10.3837/tiis.2014.02.009]

Gao Z, Zhang H, Xu G P and Xue Y B. 2015. Multi-perspective and multi-modality joint representation and recognition model for 3D action recognition. Neurocomputing, 151: 554-564 [DOI: 10.1016/j.neucom.2014.06.085]

Han J and Bhanu B. 2006. Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2): 316-322 [DOI: 10.1109/TPAMI.2006.38]

Lam T H W, Cheung K H and Liu J N K. 2011. Gait flow image: a silhouette-based gait representation for human identification. Pattern Recognition, 44(4): 973-987 [DOI: 10.1016/j.patcog.2010.10.011]

Li G Y, Liu T and Yi J G. 2018. Wearable sensor system for detecting gait parameters of abnormal gaits: a feasibility study. IEEE Sensors Journal, 18(10): 4234-4241 [DOI: 10.1109/JSEN.2018.

2814994]

Lin Y C, Hu M C, Cheng W H, Hsieh Y H and Chen H M. 2012. Human action recognition and retrieval using sole depth information//Proceedings of the 20th ACM International Conference on Multimedia. Nara City: ACM: 1-4 [DOI: 10.1145/2393347.2396381]

Liu A A, Nie W Z, Su Y T, Ma L Hao T and Yang Z X. 2015. Coupled hidden conditional random fields for RGB-D human action recognition. Signal Processing, 112: 74-82 [DOI: 10.1016/j.sigpro.2014.08.038]

Liu L, Su Z, Fu X D, Liu L J, Wang R M and Luo X N. 2017. A data-driven editing framework for automatic 3D garment modeling. Multimedia Tools and Applications, 76(10): 12597-12626 [DOI: 10.1007/s11042-016-3688-4]

Luo J, Tang J, Tjahjadi T and Xiao X M. 2016. Robust arbitrary view gait recognition based on parametric 3D human body reconstruction and virtual posture synthesis. Pattern Recognition, 60: 361-377 [DOI: 10.1016/j.patcog.2016.05.030]

Luo J, Tang J, Zhao P, Mao F and Wang P. 2016. Abnormal behavior detection for elderly based on 3D structure light sensor. Optical Technique, 42(2): 146-151 (罗坚, 唐琎, 赵鹏, 毛芳, 汪鹏. 2016. 基于3D结构光传感器的老龄人异常行为检测方法. 光学技术, 42(2): 146-151) [DOI: 10.13741/j.cnki.11-1879/o4.2016.02.011]

Ngo T T, Makihara Y, Nagahara H, Mukaigawa Y and Yagi Y. 2015. Similar gait action recognition using an inertial sensor. Pattern Recognition, 48(4): 1289-1301 [DOI: 10.1016/j.patcog.2014.10.012]

Pogorelc B, Bosnić Z and Gams M. 2012. Automatic recognition of gait-related health problems in the elderly using machine learning. Multimedia Tools and Applications, 58(2): 333-354 [DOI: 10.1007/s11042-011-0786-1]

Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Cook M and Moore R. 2013. Real-time human pose recognition in parts from single depth images. Communications of the ACM, 56(1): 116-124 [DOI: 10.1145/2398356.2398381]

Smisek J, Jancosek M and Pajdla T. 2011. 3D with Kinect//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Barcelona: IEEE: 1154-1160 [DOI: 10.1109/ICCVW.2011.6130380]

Sun P, Xia F, Zhang H, Peng D G, Ma X and Luo Z J. 2017. Research of human fall detection algorithm based on improved Gaussian mixture model. Computer Engineering and Applications, 53(20): 173-179 (孙朋, 夏飞, 张浩, 彭道刚, 马茜, 罗志疆. 2017. 改进混合高斯模型在人体跌倒检测中的应用. 计算机工程与应用, 53(20): 173-179) [DOI: 10.3778/j.issn.1002-8331.1604-0423]

Wang L. 2006. Abnormal walking gait analysis using silhouette-masked flow histograms//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China: IEEE: 1-4 [DOI: 10.1109/ICPR.2006.199]

Wang L, Jiang W J, Sun P and Xia F. 2017. Application of improved D-S evidence theory in human fall detection of transformer substation. Journal of Electronic Measurement and Instrumentation, 31(7): 1090-1098 (王磊, 江伟建, 孙朋, 夏飞. 2017. 改进D-S证据理论在变电站人体跌倒检测的应用. 电子测量与仪器学报, 31(7): 1090-1098) [DOI: 10.13382/j.jemi.2017.07.015]

Xia L and Aggarwal J K. 2013. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE: 2834-2841 [DOI: 10.1109/CVPR.2013.365]

Yang X D and Tian Y L. 2017. Super normal vector for human activity recognition with depth cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(5): 1028-1039 [DOI: 10.1109/TPAMI.2016.2565479]

作者简介

罗坚,1984年生,男,讲师,主要研究方向为模式识别和机器视觉。
E-mail:luojian@hunnu.edu.cn

黎梦霞,女,硕士研究生,主要研究方向为图像处理和生物特征识别。E-mail:1050706772@qq.com

罗诗光,男,本科生,主要研究方向为图像处理与模式识别。E-mail:1901675570@qq.com