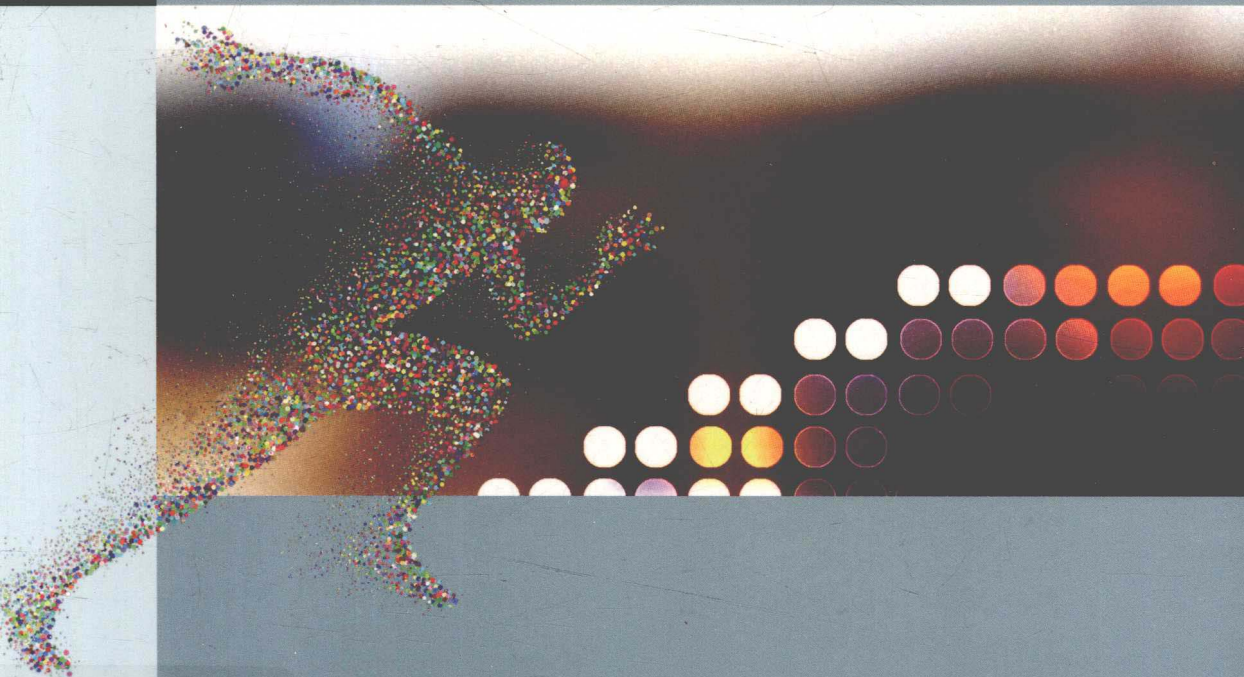


J
JIYU TUXIANG SHENDU XINXI DE
RENTI DONGZUO SHIBIE YANJIU

基于图像深度信息的人体动作识别研究

李建军 著



重庆大学出版社

J

JYU TUXIANG SHENDU XINXI DE
RENTI DONGZUO SHIBIE YANJIU

基于图像深度信息的人体动作识别研究

李建军 著



重庆大学出版社

内容提要

本书从人体动作识别的研究现状与挑战以及人体动作数据采集的发展现状入手,并以图像深度信息为研究对象进行人体动作识别研究,循序渐进地进行原理和设计方法的介绍。全书主要内容包括人体三维骨架关节点数据的获取原理,根据数据特点如何建立张量模型,并提出空间曲度概念进行数据集关键帧提取,以及基于3D深度卷积神经网络建立人体动作识别模型等。

本书可作为从事计算机视觉、模式识别研究人员的参考书籍。

图书在版编目(CIP)数据

基于图像深度信息的人体动作识别研究 / 李建军

著. -- 重庆:重庆大学出版社, 2018. 12

ISBN 978-7-5689-1427-7

I. ①基… II. ①李… III. ①人体—运动—图像识别—研究 IV. ①TP391.413

中国版本图书馆CIP数据核字(2018)第284573号

基于图像深度信息的人体动作识别研究

李建军 著

策划编辑:鲁黎

责任编辑:李定群 版式设计:鲁黎

责任校对:杨育彪 责任印制:张策

*

重庆大学出版社出版发行

出版人:易树平

社址:重庆市沙坪坝区大学城西路21号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn(营销中心)

全国新华书店经销

重庆市正前方彩色印刷有限公司印刷

*

开本:787mm×1092mm 1/16 印张:7.75 字数:112千

2018年12月第1版 2018年12月第1次印刷

ISBN 978-7-5689-1427-7 定价:48.00元

本书如有印刷、装订等质量问题,本社负责调换

版权所有,请勿擅自翻印和用本书

制作各类出版物及配套用书,违者必究

「 前 言 」

人体动作识别是计算机视觉领域内的一个研究热点。它通过分析视频图像序列,提取有效动作特征信息,学习和理解人体动作行为,在人机交互、智能视频监控、虚拟现实以及社会公共安全等领域有着广阔的应用前景。早期的人体动作识别研究主要以 RGB 视频图像序列为研究对象,易受到复杂背景、视角变换和光照变化等因素的影响,动作识别率较低,算法鲁棒性较差。深度传感器能获取更具判定性的人体动作深度信息,基于 RGB-D 图像序列的人体动作识别研究受到越来越多学者的关注。

本书在习近平新时代中国特色社会主义思想指导下,落实“新工科”建设新要求,基于 Kinect 传感器采集的人体三维骨架序列和深度图像进行人体动作识别研究,主要探讨如何高效、准确地进行人体动作的特征描述、行为建模和特征融合等关键技术,克服环境变化等因素影响,弥补现有研究不足。本书共 6 章,主要包括:

第 1 章绪论,介绍目前人体动作识别的研究现状与挑战以及人体动作数据集的发展现状。

第 2 章基于三维骨架关节点信息的动作识别,内容包括获取人体三维骨架关节点数据的原理,以及利用张量概念,根据骨骼模型数据特性,将骨骼点序号、点上下文序号、帧数作为张量的 3 个模,建立张量模型,提出张量形状描述子。实验

表明,该算法在提高人体动作识别率的基础上,有效降低了空间运算复杂度,同时具有旋转、平移和缩放不变性。

第3章基于动作捕获数据的关键帧提取,介绍人体动作数据集关键帧提取方法的原则以及研究现状,并基于动作序列轨迹点和其近邻域点构造的空间曲面,提出空间曲度概念。通过阈值判决法完成动作序列的关键帧提取。实验表明,该算法能去除大量冗余数据,减少帧和帧之间的干扰。

第4章基于深度图像多特征融合的交互动作识别,介绍了底层视觉特征的提取方法,并使用加权融合策略进行多特征融合,提出的算法避开了图像分割,更稳定地提取了图像特征,具有旋转、平移和缩放不变性。实验表明,该融合特征比单一特征具有更高的识别率和更好的稳定性,有效解决了单一特征表达精度不高、鲁棒性差的问题。

第5章基于深度图像和卷积神经网络的动作识别,介绍目前较流行的深度学习中典型的3D卷积神经网络,提出了基于3D深度卷积神经网络的人体动作识别模型,并通过迁移学习提高网络训练的收敛速度,很好地解决少量样本数据的学习问题,进一步提高动作识别率。实验表明,该模型具有较强的特征表达和泛化学习能力。

第6章总结与展望,对本书的内容进行总结,并提出了在该方向的后续研究内容。

本书由李建军撰写,受到国家自然科学基金项目(项目编号51565046)、内蒙古自治区科技计划项目(协整分析下基于大数据的风电机组齿轮箱故障诊断与性能预测研究)和内蒙古自然科学基金项目(项目编号2018MS06018,基于时空上下文的动作行为类内差异性研究)的资助,感谢张超、赵建峰老师提供的大力帮助。

本书可作为图像处理、模式识别研究者的参考资料。

由于作者水平有限,书中难免存在疏漏和不足之处,恳请读者给我们以这方面的反馈信息。

李建军

2018年5月

目 录

第1章 绪 论	1
1.1 人体动作识别的研究背景与意义	1
1.2 人体动作识别研究现状与挑战	3
1.2.1 基于骨架关节点的动作识别研究	4
1.2.2 基于深度图像的动作识别研究	8
1.2.3 基于深度学习的动作识别研究	11
1.3 人体动作数据集的发展现状	12
1.4 本章小结	17
第2章 基于三维骨架关节点信息的动作识别	18
2.1 概 述	18
2.2 获取人体三维骨架关节点数据的原理	20
2.3 基于张量形状描述子的动作识别	21
2.3.1 张量形状描述子	22
2.3.2 多线性主成分规整	29

2.4	实验结果与分析	35
2.4.1	不同输入模型的对比实验	35
2.4.2	动作识别对比实验	37
2.5	本章小结	39
第3章	基于动作捕获数据的关键帧提取	41
3.1	概述	41
3.2	人体动作数据集的关键帧提取	42
3.2.1	关键帧提取的目的和原则	42
3.2.2	关键帧提取方法的研究现状	43
3.3	基于空间曲度的关键帧提取	46
3.3.1	空间曲度	47
3.3.2	关键帧提取	50
3.4	实验结果与分析	53
3.4.1	关键帧提取实验	53
3.4.2	基于关键帧的动作识别实验	55
3.5	本章小结	58
第4章	基于深度图像多特征融合的交互动作识别	59
4.1	概述	59
4.2	底层视觉特征提取	62
4.2.1	边缘特征提取	63
4.2.2	融合边缘与光流特征	65
4.3	实验结果与分析	70
4.3.1	双人交互动作识别实验	72
4.3.2	人物交互动作识别实验	74
4.4	本章小结	76

第5章 基于深度图像和卷积神经网络的动作识别	77
5.1 概 述	77
5.2 3D 卷积神经网络	78
5.2.1 卷积神经网络在动作识别应用中的优点	79
5.2.2 卷积神经网络基本结构	80
5.2.3 构建3D 卷积神经网络	83
5.3 实验结果与分析	86
5.3.1 基于3D CNNs 的动作识别实验	86
5.3.2 基于迁移学习的动作识别实验	88
5.4 本章小结	89
第6章 总结与展望	91
参考文献	95

第1章 绪论

1.1 人体动作识别的研究背景与意义

人体动作识别能自动分析、理解人在环境中发生的动作变化,从而做出相应的决策,是视频图像语义分析方向的一个新兴研究课题。该研究融合了人工智能、模式识别、图像处理、计算机视觉以及认知科学等多学科知识,在智能视频监控、人机交互、运动分析、虚拟现实等领域有着广阔的应用前景,对推动经济和社会发展具有重要意义。

(1) 智能视频监控

通过人体动作识别技术自动对监控环境中人物的行为举止进行分析、预测,并对监控环境中任务发生的变化进行识别、定位和跟踪,及时发现可疑人员。

(2) 人机交互

人体动作识别是计算机感知视觉信息的基础。识别技术是智能人机交互系统的关键技术,与其他常见交互方式相比,人体动作包含更多的信息,可极大地提升计算机对外部视觉信息的感知能力,实现方便、快捷的人机交互。

(3) 运动分析

对视频中的运动员情况进行人体动作识别,通过评价,完善训练系统,从而提高运动员成绩。也可通过步态分析,为腿部受伤运动员的康复训练治疗方案提供一个科学的技术支持。

(4) 虚拟现实

人体动作识别是虚拟现实技术的重要组成部分,通过对人体动作进行建模分析,可更好地对场景进行仿真模拟。例如,飞行员的模拟驾驶、虚拟现实游戏等,带给使用者更加优越的体验感受。

最初,人体动作识别主要是以 RGB 视频图像为研究对象,经过多年的努力,人体动作识别技术已取得了较快的发展。但是,由于环境光线、遮挡和动态背景等因素造成的影响,以及从三维空间到二维图像平面投影导致深度信息丢失等原因,造成了基于 RGB 视频图像的人体动作识别准确率较低、算法鲁棒性较差。因此,人体动作识别还存在极大的发展空间。

随着科学技术的发展,传感器工艺和性能不断更新,打破了传统的视频图像采集方式。2010 年问世的深度传感器 Kinect 不仅能采集 RGB 视频图像和深度图像,而且还可实时地对人体骨架进行跟踪,采集人体骨架关节点的坐标信息。深度图像能方便地实现背景剔除、前景人体检测和提取,弥补空间投影过程中深度信息丢失的问题。Kinect 传感器采集的深度信息为人体动作识别研究注入了新的活力,受到越来越多学者的关注。

本书使用 Kinect 传感器采集人体动作的深度图像信息,探讨人体动作识别的相关理论和技术难点,着重研究如何消除可能对识别结果造成的干扰因

素,提高人体动作识别的鲁棒性与普适性。本书的主要目标是进一步分析人体动作行为的特性,提出鲁棒性更高的人体动作识别算法。

1.2 人体动作识别研究现状与挑战

由于人体动作的多样性和周边环境的复杂性,使人体动作识别充满了挑战。因此,主要集中在以下3个方面:

①环境影响。复杂背景、光照条件变化以及遮挡等外部环境影响。

②视角变化影响。不同视角将会导致环境背景不一致,同一动作不同的视觉效果将会带来人体重叠和遮挡等问题。

③动作多样性。不同人的动作习惯和同一动作个体表观的不同,将会导致动作类内差异性大和类间相似性高的问题。

人体动作识别最初的研究都是基于RGB视频图像序列,许多学者分别从静态特征、动态特征和时空特征3个方面对该领域的工作进行了总结^[1-4]。

静态特征多指原始的图像特征,如剪影轮廓^[5-8]、颜色直方图^[9]以及基于灰度纹理的特征描述子^[10]。

动态特征是分析视频图像中的运动信息,如光流场^[11-15]和运动轨迹^[16-18]等。光流场主要描述随时间变化的动作幅度和方向,但容易受噪声影响。Efros等^[11]针对低分辨率的动作主体提出了一种多通道光流场,使用校正方法提高了光流场对噪声的鲁棒性。张飞燕等^[15]通过数理统计方法,分析获取不同动作行为的梯度和光流直方图,分别提取梯度和光流的广义高斯分布模型参数,对人体动作进行特征描述,提出基于马氏距离的人体动作识别方法。Yu等^[17]通过尺度不变特征变换获取时空兴趣区域,引入权重因子,利用光流法提取区域内带权重轨迹的运动特征,然后使用光流词袋模型进行动作特征表示。

时空特征主要包括时空立方体、时空上下文等。其中,最为经典的是

Bobick 等^[19] 根据目标动作行为发生的区域和位置,提出的运动能量图 (Motion Energy Images, MEI) 算法,以及根据动作行为发生的时间顺序,提出的运动历史图 (Motion History Images, MHI) 算法。通过计算运动能量图和运动历史图的 Hu 矩特征,利用模板匹配方法进行动作识别。Weinland 等^[20] 基于运动历史图算法的思想,考虑了多视角问题,提出运动历史卷 (Motion History Volumes, MHV) 算法,将运动历史卷转换到柱状坐标系,使用 Fourier 变换进行动作特征提取。Laptev 等^[21] 提出时空兴趣点 (Space-time Interest Points, STIP) 概念,将提取 2D Harris 点的方法扩展到三维空间,获取时空兴趣点。Scovanner 等^[22] 将尺度不变特性 (Scale Invariant Feature Transform, SIFT)^[23] 扩展到三维空间,提出了同时具有抗光照变化、视角变化特点的 3D SIFT 描述子。由于尺度不变特征变换具有较高的运算复杂度,Bay 等^[24] 提出了能提高运算速度和实时性的加速鲁棒性特征 (Speeded-Up Robust Features, SURF) 描述子。Wang 等^[25] 通过观察多尺度时空上下文区域中每一个兴趣点的特征密度,获取上下文关系,提出了上下文特征描述子。

虽然基于 RGB 视频图像序列的动作识别研究取得了较好的成果,但因其复杂背景、相机视角、光照变化以及遮挡等问题,人体动作识别研究仍然存在诸多挑战。深度传感器的发布,使人们对三维世界的感知能力得到了进一步的提升,研究者可方便地提取更丰富的底层视觉信息。微软公司发布的 Kinect 深度传感器不仅能提供 RGB 视频图像,同时还能提供深度图像序列以及实时跟踪的人体骨架关节点位置信息,从而有效克服传统 RGB 相机导致的空间位置关系丢失、上下境建模精度不高的问题。Kinect 传感器提供的深度图像具有颜色无关性和纹理不变性,能较好地避免光照、阴影等因素造成的影响。因此,近几年,基于深度图像和骨骼模型的人体动作识别研究受到了学者们广泛的关注^[26-28],本书将从以下 3 个方面介绍人体动作识别研究现状:

1.2.1 基于骨架关节点的动作识别研究

人体结构可被视为由若干个骨架关节点链接成的刚性系统。因此,可通

过人体骨架关节点运动的位置状态信息进行动作描述,反映动作行为的语义信息。微软公司发布的 Kinect 深度传感器能主动跟踪两个人体的骨架关节点,将每个骨架关节点的位置状态信息存储为一个 X,Y,Z 的三维坐标。基于 Kinect 传感器跟踪的人体骨架 20 个关节点的平面示意图如图 1.1 所示。

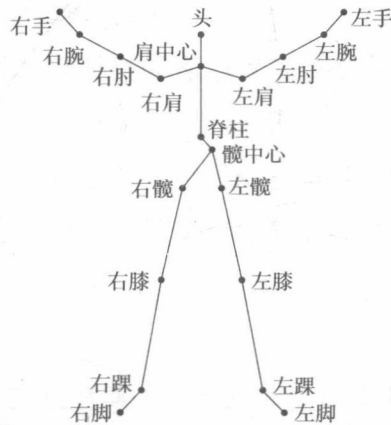


图 1.1 骨架 20 个关节点示意图

Kinect 深度传感器包含颜色和深度两个感应镜头,以及语音麦克风阵列和倾斜传感器机动调整部件。其可视范围为:水平视角,57°;垂视角,43°。Kinect 传感器机身转动范围为:±27°;传感深度范围为:1.2~3.5 m。传感器采集的数据流指标为:深度感应镜头,320×240,16 bit,30 fps;颜色感应镜头,640×480,32 bit,30 fps。

人体骨架 20 个关节点分别为头、肩中心、左(右)手、左(右)腕、左(右)肘、左(右)肩、脊柱、髋中心、左(右)髋、左(右)膝、左(右)踝、左(右)脚。人体骨骼模型数据不仅包含形状信息,还包含结构拓扑信息。

基于骨架关节点的人体动作识别研究最早出现在 1979 年,Johansson^[29]提出了经典的 MLD(Moving Light Display)实验,在人体骨架关节点处贴上亮点,获取黑色背景下亮点的运动轨迹信息,从而分析人体的运动行为。随着深度传感器的应用,基于骨骼模型的人体动作识别受到许多研究者的关注。

Li 等^[30]提出了一种词袋(Bag of Word, BoW)方法对人体动作行为进行建模,使用骨架关节点位置处的时空描述子构造视觉单词,用词典中单词出现的频率表示动作行为,计算每个动作的视觉单词直方图,使用支持向量机

(Support Vector Machines, SVM)分类器进行人体动作分类。Xia 等^[31]根据人体骨架关节点的位置状态,将 3D 空间进行特定方向的划分,并进行直方图计算,对骨架关节点的位置进行投影,构成关节点位置直方图(Histograms of 3D Joint Location, HOJ3D),采用线性判别式分析(Linear Discriminant Analysis, LDA)方法和 K-Means 聚类方法进行特征降维和聚类,使用马尔科夫模型进行动作分类。Li 等^[32]通过对骨架关节点运动特征相似性的判断,将人体的骨骼模型进行分割,分别对每个分割部分进行特征提取,采用逻辑回归方法计算每个特征的权值,用权值最大的类别进行人体动作识别。Ellis 等^[33]利用距离特征组合描述动作姿态,提取骨骼模型动作序列中包含运动信息的人体姿态,通过使用延迟感知学习训练逻辑回归分类器,从而解决人体动作识别时的高延迟问题。

具有时空分布特性的人体骨架关节点,其相互之间的位置关系能提供丰富的动作信息。Yang 等^[34-35]基于人体骨架关节点数据分别提取姿态特征、运动特征和偏移特征构成 EigenJoints 特征描述子,利用运动过程中当前帧和连续帧关节点间的位置差异表示动作行为,最后采用朴素贝叶斯最近邻(Naïve Bayes Nearest Neighbor, NBNN)分类器进行动作分类。Lu 等^[36]利用人体动作过程中关节点的偏移位置进行人体动作行为描述,采用 NBNN 分类器进行动作分类。Wang 等^[37]基于人体骨骼模型和点云信息,提取骨架关节点周围点云局部占有信息(Local Occupancy Pattern, LOP),每一个骨架关节点对应一个局部占有信息特征。该特征能很好地反映人体骨骼模型各个部分之间的关联特性,具有良好的平移不变性,同时还提出由不同骨架关节点特征构成的 Actionlet 模型。Yu 等^[38]主要是对骨架关节点数据的空间信息进行编码,通过对骨架关节点之间欧氏距离的大小排序,将该排序信息作为一个中层特征,称为“orderlet”特征。该特征是原始特征值之间的对比关系,比一般数值特征具有更强的抗噪能力。Vieira 等^[39]将人体骨架关节点之间的距离构成一个矩阵向量,通过矩阵不变特性进行动作特征表示,矩阵中的每一个元素表示人体任何两个骨架关节点之间的欧氏距离,刚体变换具有很好的距离不变性。因此,该算法的稳定性较好,鲁棒性较强。

骨骼模型数据不仅能提供人体动作姿态的位置信息,还能提供骨架关节点的运动信息。大量学者利用骨架关节点运动形成的轨迹进行人体动作识别研究。Sheikh 等^[40]利用骨架关节点投影到归一化 XYZ 空间的运动轨迹,构建动作行为轨迹包,通过不同样本轨迹之间的视角不变相似度进行人体动作识别。Xia 等^[41]利用骨架关节点的运动轨迹,匹配人体动作行为骨架图,使用混沌不变量和相关向量机原理进行空间轨迹重构,实现人体动作识别。Hammond 等^[42]使用小波变换频谱图,构建时空金字塔模型,从不同的时空尺度获取骨架关节点的运动轨迹信息,使用主成分分析(Principal Component Analysis, PCA)方法进行特征降维,利用 SVM 分类器进行动作分类。Devanne 等^[43]基于流形学习的概念提出一种人体动作识别方法,在 Riemannian 流形上计算骨架关节点运动轨迹的形状相似性,该方法能同时获取人体动作的形状信息和动态特性。Slama 等^[44]通过分析人体骨架关节点空间运动轨迹特性,提出了自回归移动平均模型(Auto Regressive and Moving Average Model, ARMA),将 ARMA 观测矩阵的列向量作为 Grassmann 流形上的一个点,通过流形学习进行人体动作识别。目前,还有很多基于人体动作识别的研究成果^[45-48]。

现有研究中,基于人体骨架关节点运动轨迹的描述大体上可分为两类:一类是将运动轨迹作为时间序列,采用序列匹配方式进行特征提取,如隐马尔科夫模型(Hidden Markov Model, HMM)^[49]、动态时间规整(Dynamic Time Warping, DTW)^[50]、条件随机场(Conditional Random Field, CRF)^[51]及连续动态时间规整(Continuous Dynamic Time Warping, CDTW)^[52]等方法,均有较高的准确率和实时性;另一类是将骨架关节点的运动轨迹作为三维曲线处理,使用形状描述子提取运动轨迹特征,其方法包括曲率特征空间^[53]、傅里叶描述子^[54]、小波描述子^[55]及曲线不变矩^[56]等。形状描述子具有维度低、稳定性强的优点,缺点是忽略了运动轨迹的时间信息,对形状相同的运动轨迹识别容易出现混淆现象。如何更好地结合动作序列和运动轨迹曲线特性进行人体动作识别研究,仍需要研究和探索。

1.2.2 基于深度图像的动作识别研究

深度传感器 Kinect 采集的深度图像与实时采集的骨骼模型深度信息相比,更能直观和完整地描述动作形状特征以及动态特征。深度图像与 RGB 视频图像相比(见图 1.2),具有以下优势:

- ①深度图像使用深度距离的大小表示图像像素值,对光照变化和环境变化不敏感,具有较好的稳定性。
 - ②深度图像能描述 3D 场景,反映几何特性和形状信息,比 RGB 视频图像的颜色和纹理信息具有更好的辨识性,可有效地克服遮挡和重叠等问题。
- 因此,越来越多的学者开始基于深度图像进行人体动作识别研究。

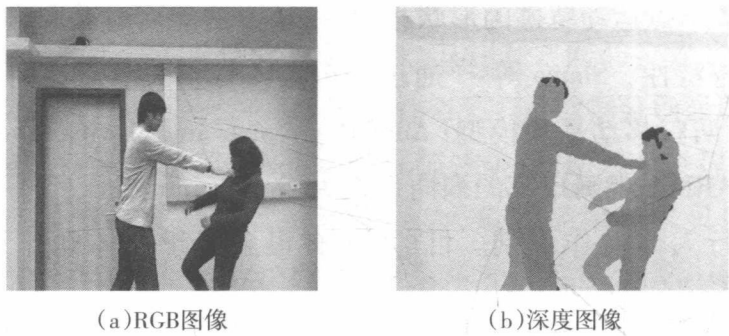


图 1.2 深度传感器 Kinect 采集的图像类型

基于深度图像提取的特征主要包括全局特征和局部特征。提取动作全局特征就是对人体动作行为整体特征进行描述。全局特征主要包括人体轮廓形状特征、时空形状特征、运动特征以及一些更高层次的特征等。Yang 等^[57]将深度图像序列投影到指定的二维平面视图中,利用投影过程的时空特征构造动作特征描述子,提出一种深度运动映射(Depth Motion Map, DMM)算法,即将深度图像分别投影到前视图、顶视图和侧视图 3 个正交平面,对每个平面上的连续图像帧进行差分计算,通过阈值判决法获取映射图像,并对该映射图像进行累积叠加,分别得到投影到每个平面的动作序列图像。Wang 等^[58]采用加权采样方案从深度图像序列中提取特征,首先对深度映射图像序列构成的四维时空体进行采样,获取不同时空位置和不同尺度的子时空体,

统计每个子时空体中的像素个数,即各子时空体在四维时空体中的占有信息,将这种特征描述子称为随机占有模式(Random Occupancy Pattern, ROP)特征描述子,经过稀疏编码后,使用SVM分类器进行动作分类。

基于深度图像的局部特征提取方法是人体动作识别领域中的一个研究热点。它不需要进行人体目标检测、前景和背景分割,也不需要精准定位和跟踪运动目标,只对动作显著性区域进行特征提取。其常用的局部特征描述子包括兴趣点邻域内像素分布特征、梯度大小及方向、形状特征和光流特征等。Gilbert等^[59]分别在 xOy , yOz , xOz 3个平面进行Harris兴趣点检测,通过数据挖掘技术筛选兴趣点,从而获取性能稳定的时空兴趣点。Willems等^[60]将二维SURF特征检测方法扩展到三维Hessian矩阵,提出显著度评价函数,利用非极大值抑制方法获取时空兴趣点。

近年来,人体动作识别采用了大量的局部特征提取方法,如词袋(Bag of Words, BoW)模型、稀疏编码(Sparse Coding, SC)、Fisher核、局部矢量聚合描述符(Vector of Locally Aggregated Descriptors, VLAD)、朴素贝叶斯最近邻分类器(Naïve Bayes Nearest Neighbor, NBNN)等。Zhen等^[61]基于以上常用时空局部特征提取方法进行全面研究,在统一的实验环境下对3个公开使用的动作数据集进行比较实验,并对该特征表示方法进行评价。Cheng等^[62]提出一种新颖的基于深度信息的描述子。该描述子用于描述动作的时空结构关系,首先使用Harris检测算法获取时空兴趣显著点,构建一个 $3 \times 3 \times 3$ 的时空立方体,然后计算立方体中心点到其他点的深度值,按照顺序进行编码,提出比较编码描述子(Comparative Coding Descriptor, CCD)。Xia等^[63]基于时空立方体特征相似性(Depth Cuboid Similarity Feature, DCSF),提出了时空特征描述子,通过比较动作内容的自相似性,获取局部区域的几何分布信息,从而描述图像序列的局部外观模式。Seo等^[64]提出一种新颖的时空局部回归核(Space-Time Local Regression Kernels, STLRK)特征表示方法,通过构建时空立方体进行动作匹配,完成人体动作识别。

在人体动作识别研究中,提取的时空兴趣点和底层视觉特征并不是独立存在的,而是具有一定的相关性,这种相关性称为视觉上下文关系。它主要