**南 京 师 范 大 学**

**毕 业 设 计（论 文）**

**（2019届）**



|  |  |  |  |
| --- | --- | --- | --- |
| **题 目：** | 基于Kinect体感信息的动作  及行为识别技术研究 | | |
| **学 院：** | 计算机科学与技术学院 | | |
| **专 业：** | 计算机科学与技术 | | |
| **姓 名：** | 仇思宇 | | |
| **学 号：** |  | 21150611 |  |
| **指导教师：** | **宋凤义** | | |

**南京师范大学教务处 制**

# 

# 摘 要

**关键词：**人脸识别；在线学习系统；……

# Abstract

With the development of information technology,

……

**Key words:** face recognition,online learning system,……

**目 录**

[摘 要 1](#_Toc2026071)

[Abstract 2](#_Toc2026072)

[第 1 章 绪论 0](#_Toc2026073)

[1.1 本课题的目的及研究意义 0](#_Toc2026074)

[1.2 国内外研究现状 0](#_Toc2026075)

[1.2.1 RGB图像特点及其劣势 1](#_Toc2026076)

[1.2.2 深度图像特点及其优势 2](#_Toc2026077)

[1.2.3 利用骨架的动作表示与识别方法 2](#_Toc2026078)

[1.2.4 利用三维模型特征的动作表示与识别方法 2](#_Toc2026079)

[1.2.5 利用空-时特征的动作表示与识别方法 3](#_Toc2026080)

[1.3 本文的目的及研究意义 3](#_Toc2026081)

[第 2 章 动作识别方法评价体系 0](#_Toc2026082)

[第 3 章 基于骨架关节点的动作识别方法 0](#_Toc2026083)

[3.1 身体部位判断和关节点获取 0](#_Toc2026084)

[3.1.1 关节特征获取 0](#_Toc2026085)

[3.1.2 关节位置预测 0](#_Toc2026086)

[3.2 动作表示的方法 1](#_Toc2026087)

[3.3 动作的识别 2](#_Toc2026088)

[第 4 章 基于三维模型特征的动作识别方法 0](#_Toc2026089)

[4.1 局部占用模式的方法 0](#_Toc2026090)

[4.2 曲面法线的方法 0](#_Toc2026091)

[第 5 章 基于空-时特征的动作识别方法 0](#_Toc2026092)

[5.1 骨架的时间序列方法 0](#_Toc2026093)

[5.2 空-时体积方法 0](#_Toc2026094)

[第 6 章 基于学习特征的动作识别方法 0](#_Toc2026095)

[参考文献 0](#_Toc2026096)

[致 谢 0](#_Toc2026097)

[本科期间主要研究成果 1](#_Toc2026098)

# 绪论

目前，体感识别技术的课题主要是研究人体姿态和手势信息提取与识别等相关技术，如基于Kinect传感器深度信息的手势检测和识别技术[1]，为人机交互提供了新的方法和思考。在识别简单手势和动作识别技术逐渐成熟并广泛运用于人们日常生活中后，基于Kinect传感器的人体动作识别技术开始出现[2][3]。与此同时，识别和分析生物行为信息的技术也开始逐渐发展，如：针对小型动物的行为识别和分析系统[4]；利用Kinect深度传感器得到的深度图像，对猪群的攻击行为进行检测和辨别[5]；以及对老年人日常生活的深度图像进行分析，从而发现他们身体功能恶化的早期迹象，从而对可能产生的疾病进行预测[6]。

## 本课题的目的及研究意义

人类行为识别研究在过去十年取得了重大进展，并在各种学科中得到越来越多的关注。从诸如RGB相机，深度相机，距离传感器，可穿戴惯性传感器或其他类型传感器中获取相关数据[8]，进而利用这些数据进行人体动作和行为识别和分析。而由于从不同类型传感器中获取的数据处理方法不同、获取并利用的信息不同、使用的任务范围也不尽相同。从行为监视，视频分析，人机交互[7]，人类的动作和行为识别技术已经被广泛应用于日常生活和各个应用领域，同时，辅助生活，健康监控，危险行为预警等相关技术也应运而生[6]。对应于不同的传感器类型，用于识别人体动作的主要有基于视觉的动作识别和基于惯性的动作识别这两种主要的技术。

传统RGB相机捕获的图像序列信息的方法用于动作或手势识别的主要限制在于处理RGB图像时的高计算需求以及对图像质量敏感等相关挑战。本课题将从适用性、可靠性、效率等角度对比现有的基于视觉的动作识别技术和基于惯性的动作识别技术，并针对现有人体动作识别技术存在的问题和局限进行相关算法的改进。

## 国内外研究现状

……

### RGB图像特点及其劣势

在1.1中提到的基于视觉的动作识别方法主要有：时空体积、时空特征和轨迹，它们被广泛用于传统RGB传感器捕获的视频序列中的人体动作识别。如[10]中，局部特征与SVM分类器的结合，证明了可以通过度量局部特征[[1]](#footnote-1)实现动作识别。在[11]中提供了一种对噪声和姿势变化具有更强鲁棒性的算法，这种算法使用空时空特征[[2]](#footnote-2)点（单张图像上的局部特征）来表征行为。为了降低动作分类结果对背景杂乱，遮挡和比例变化的敏感度，[12]中介绍了直接运动识别方法：使用时空特征包(BoF)[[3]](#footnote-3)，判断人体运动特征（判断局部图像块的运动如何进行），而不是通过恢复人的身体二维模型或三位模型，以其局部结构特征实现动作分类。动态能量图像(MEI)[[4]](#footnote-4)和运动历史图像(MHI)[[5]](#footnote-5)在[13]中作为运动模板被引入，以模拟已知的视频中人类行为的空间和时间特征，从而进行动作匹配。这些方法都基于强度或基于颜色，因此也具有相同的缺点，即：识别结果对照明变化的敏感性，限制了识别稳健性 。

虽然基于视觉的人类动作识别技术作为模式识别和计算机视觉研究的重要组成部分仍在持续发展，但识别性能正在受到各种挑战。除去上一段中所介绍的，动作识别面临的挑战还有例如遮挡，摄像机位置，执行动作中的主体变化，背景杂乱等[8]因素影响识别结果。实际上，除此之外，使用者或研究者还需要拥有大量的硬件资源才能运行计算密集型图像处理和计算机视觉算法，并且还需要处理传统图像中缺少3D动作数据的问题。

### 深度图像特点及其优势

近年来，低成本深度传感器的出现，使它们大量被应用于人体动作识别及其相关领域。利用深度传感器提取的深度图像，可以解决传统RGB图像中缺失的3D动作数据，也因此具备可以更加精确识别人体动作的潜能。

与由摄像机捕获的传统RGB图像相比，深度相机生成的深度图像显示出对照明变化不敏感并且在人类动作识别中具有高性能、实时性强等特点。同时，人体骨骼信息也可以从深度图像中获得[9]。微软的Kinect设备的原理就是利用了深度或距离传感器，进行人体骨骼和动作的识别。

### 利用骨架的动作表示与识别方法

利用在[14]中的从单个深度图像快速准确地预测身体关节的空间位置的方法，提取出由关节点构成的人体骨架，并利用以关节位置差异作为特征提取出姿态信息、姿态运动信息，然后使用朴素贝叶斯最邻近分类器（NBNN）[15]进行动作分类是最为简单的方法。在结合两个人之间的距离和相对位置后，利用动作森林模型（AF）[2]，可以识别两个人的交互行为特征，并且具有更高的整体识别效率和自由度。由于骨架估计的不准确性，这种基于骨架的方法具有局限性。并且，骨架信息在许多应用场合中并不总是可用。在[16]中，使用3D关节位置直方图（HOJ3D）[[6]](#footnote-6)表示姿势，通过姿态的深度图像序列计算HOJ3D并使用LDA[[7]](#footnote-7)重新投影，然后聚类成若干个姿势视觉词。人体的静态姿势便由这些姿势视觉词序列构成。由离散隐马尔可夫模型（HMM）建模分析这些视觉词的时间序列，将其分类为若干已知动作。

### 利用三维模型特征的动作表示与识别方法

在[17]中，将深度图像分别投影到三个坐标平面上，并利用投影图像计算相关的运动能量，组合为深度运动图（DMM）。从三个DMM中提取定向梯度柱状图（HOG）并将其连接为DMM-HOG表示动作。

与投影方法将三维图像转变为二维图像的思路不同，将空间划分为若干子空间，并计算落入子空间中的占有体积的特征被称为随机占用模式[18]（Random Occupancy Pattern，简称ROP）。在使用稀疏编码对该特征进行编码后，使用SVM对编码系数进行分类，从而实现动作识别。

在ROP特征的基础上，在文献[19]中提出一种新的人体动作特征和一种新的动作识别方法：局部占用模式（Local Occupancy Pattern，简称LOP）和动作类集合模型（Actionlet Ensemble Model），并明确了动作类是关节子集的特征的特定组合。新的动作模型对于特征中的错误更加健壮，并且可以更好地表征动作中的类内变化。

### 利用空-时特征的动作表示与识别方法

## 本文的目的及研究意义

……

# 动作识别方法评价体系

本章介绍……，分析……，给出……。

# 基于骨架关节点的动作识别方法

在20世纪70年代进行了一项着名的实验[20]，其中一个人穿着黑色且各关节附有光源。另一人在观察移动灯光的同时，可以识别正在执行的动作。这项工作激发了后来的研究人员，使人们认识到可以从主要的关节运动中感知人类行为。早期利用多个相机并配以关节点标记来进行可靠的关节点位置估计，从而实现人体的动作识别，且具有较高的精确性。但是，在仅仅使用单一相机的情况下会出现自身遮挡和互相遮挡。而且，基于标记点的动作识别只能在特殊的室内环境使用，且价格昂贵。

## 身体部位判断和关节点获取

应用随机决策森林进行身体部位识别，并使用模式发现算法生成关节位置。

### 关节特征获取

对任意关节点的平面位置，表示该关节点的深度特征。

其中是图像中平面位置处的像素深度，参数中包含两个包含随机偏移位置的向量和。通过使偏移规范化，确保特征不随深度相机的位置变化而变化。因此，这些特征具有平移不变的特性。如果偏移像素位于背景上或图像边界之外，则被给予大的正数常量值。

### 关节位置预测

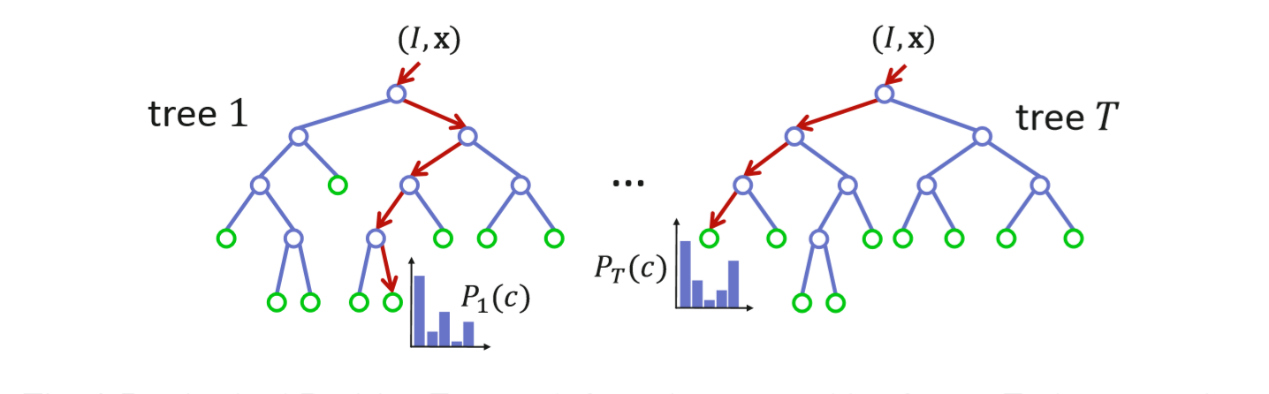


图 3‑1 随机决策森林原理示意图

分类器选择随机决策森林模型，该模型是决策树的集合。每棵树由中间节点（蓝色）和叶节点（绿色）组成。红色箭头表示不同树对特定输入可能采用的不同路径。

如图 3‑1所示，森林是个决策树的集合，每棵树由中间节点和叶节点组成。每个中间节点包含两个元素：特征和阈值。为了对图像的像素进行分类，从根处开始重复使用公式1并与阈值比较，进而向左或向右访问。在树t中到达的叶节点处，存储由身体部位标签得到分布，即各个偏移坐标所占的权重。对森林中的所有树的分布求平均值，以给出最终的分布。

我们采用基于加权高斯核的均值移位[]的局部模式发现方法，将每个身体部位的密度估计量定义为：

其中是为世界坐标，是图像像素的数量，是像素加权，是图像像素在给定深度下重投影到世界坐标空间中的位置，并且是学习的每部分带宽。像素加权为像素处所占的权重和像素深度平方的乘积：

估计量高于学习概率阈值的像素被用作部位c的起始点。给出最终的置信估计值作为到达每个模式的像素权重的总和。

## 动作表示的方法

当完成关节点的识别后，对于动作序列的每一帧，都包含个关节的空间坐标：。以关节的空间位置差异表示动作信息，需其包含三个特征：姿态特征，运动特征和偏移特征。

其中为当前帧中的成对关节差异，表征了当前帧的静态姿势信息。其计算方法为：

则为了捕获当前帧的运动属性，在当前帧和前一帧之间计算成对关节差异:

捕获了帧和初始帧之间的成对关节差异，表征当前帧中的偏移特征和整体位移：

将姿态特征，运动特征和偏移特征三者结合后得到每一帧的初步特征表示。但是对任意关节点，其三个值可能是不同坐标系中的坐标。例如为屏幕坐标，为深度坐标。为避免不同坐标系的噪声影响，需要进行标准化，即将关节点中的每个值缩放为，从而得到。

对于包含=20个关节点的深度图像，的维度为，需要进行数据降维。因此，在得到后需要进行主元素分析，从而实现关节点姿态的紧凑表示。

## 动作的识别

# 基于三维模型特征的动作识别方法

## 局部占用模式的方法

## 曲面法线的方法

# 基于空-时特征的动作识别方法

## 骨架的时间序列方法

## 空-时体积方法

# 基于学习特征的动作识别方法

# 参考文献

1. Vinh T Q , Tri N T . Hand gesture recognition based on depth image using kinect sensor[C]// Information & Computer Science. IEEE, 2015.
2. Chuan C H , Chen Y N , Fan K C . Human Action Recognition Based on Action Forests Model Using Kinect Camera[C]// 2016 30th International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE, 2016.
3. Fujino M , Zin T T . Action Recognition System with the Microsoft KinectV2 Using a Hidden Markov Model[C]// Third International Conference on Computing Measurement Control & Sensor Network. IEEE, 2017.
4. Wang Z, Mirbozorgi S A, Ghovanloo M. Towards a Kinect-based behavior recognition and analysis system for small animals[C]//Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE. IEEE, 2015: 1-4.
5. Jonguk L , Long J , Daihee P , et al. Automatic Recognition of Aggressive Behavior in Pigs Using a Kinect Depth Sensor[J]. Sensors, 2016, 16(5):631-.
6. Banerjee T, Yefimova M, Keller J M, et al. Exploratory analysis of older adults’ sedentary behavior in the primary living area using kinect depth data[J]. Journal of Ambient Intelligence and Smart Environments, 2017, 9(2): 163-179.
7. Dawar N, Kehtarnavaz N. Real-Time Continuous Detection and Recognition of Subject-Specific Smart TV Gestures via Fusion of Depth and Inertial Sensing[J]. IEEE Access, 2018:1-1.
8. Chen C, Jafari R, Kehtarnavaz N. A survey of depth and inertial sensor fusion for human action recognition[J]. Multimedia Tools and Applications, 2017, 76(3): 4405-4425.
9. Real-time human pose recognition in parts from single depth images[J]. Communications of the ACM, 2013, 56(1):116.
10. Schuldt C , Laptev I , Caputo B . Recognizing human actions: a local SVM approach[C]// Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. IEEE, 2004.
11. Dollar P , Rabaud V , Cottrell G , et al. Behavior recognition via sparse spatio-temporal features[C]// Joint IEEE International Workshop on Visual Surveillance & Performance Evaluation of Tracking & Surveillance. IEEE, 2006.
12. Laptev I , Marszalek M , Schmid C , et al. Learning realistic human actions from movies[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2008.
13. Bobick A F, Davis J W. The recognition of human movement using temporal templates[J]. IEEE Transactions on pattern analysis and machine intelligence, 2001, 23(3): 257-267.
14. Real-time human pose recognition in parts from single depth images[J]. Communications of the ACM, 2013, 56(1):116.
15. Yang X , Tian Y L . EigenJoints-based action recognition using Naïve-Bayes-Nearest-Neighbor[C]// 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). IEEE Computer Society, 2012.
16. Xia L , Chen C C , Aggarwal J K . View invariant human action recognition using histograms of 3D joints[C]// Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on. IEEE, 2012.
17. Yang X , Zhang C , Tian Y L . Recognizing actions using depth motion maps-based histograms of oriented gradients[C]// Acm International Conference on Multimedia. ACM, 2012.
18. Wang J , Liu Z , Chorowski J , et al. Robust 3D Action Recognition with Random Occupancy Patterns[M]// Computer Vision – ECCV 2012. 2012.
19. Wang J , Liu Z , Wu Y , et al. Learning Actionlet Ensemble for 3D Human Action Recognition[J]. IEEE Transactions on Software Engineering, 2013, 36(5):914-927.
20. Chen L , Wei H , Ferryman J . A survey of human motion analysis using depth imagery[J]. Pattern Recognition Letters, 2013, 34(15):1995-2006.

# 致 谢

四年的时间飞逝，……

# 本科期间主要研究成果

**发表论文：**

1. **……**

**完成项目：**

1. **高晨**，\*\*\*，\*\*\*. 大学生创新训练项目. 南京师范大，2016-2017，0.3万.

1. 局部特征：图像中的图案或不同结构，例如点，边缘或小图像块。 它们通常与图像贴片相关联，这些贴片在其纹理，颜色或强度上与其周围环境不同。 [↑](#footnote-ref-1)
2. 时空特征：短的局部视频序列，例如眼睛张开或膝盖弯曲，或者用于快速前后移动的爪子。 然后根据存在的特征点的类型和位置充分描述行为。 [↑](#footnote-ref-2)
3. 时空特征包(BoF)：一组时空特征的集合，反应了局部特征的运动特征如时空轨迹、周期性等。 [↑](#footnote-ref-3)
4. 动态能量图像（Motion-Energy Images，简称MEI）：表示图像序列中发生运动位置灰度图像。 [↑](#footnote-ref-4)
5. 运动历史图像（Motion-History Images，简称MHI）：标量值图像，其中每个像素的值是运动新近度的函数。 [↑](#footnote-ref-5)
6. 3D关节位置直方图（Histograms Of 3D Joint, 简称HOJ3D）：一种3D人体姿势的基于直方图的表示方法。 [↑](#footnote-ref-6)
7. 隐含狄利克雷分布（Latent Dirichlet Allocation，简称LDA），由Blei, David M.、Ng, Andrew Y.、Jordan于2003年提出，是一种主题模型，它可以将文档集中每篇文档的主题以概率分布的形式给出，从而通过分析一些文档抽取出它们的主题（分布）出来后，便可以根据主题（分布）进行主题聚类或文本分类。同时，它是一种典型的词袋模型，即一篇文档是由一组词构成，词与词之间没有先后顺序的关系。 [↑](#footnote-ref-7)