分　类　号：TN391　　单位代码：10190
研究生学号：201505004　　密　　级：公 开

長春工業大學

硕士学位论文

李延林

2018年**6**月

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

基于深度运动图的人体行为识别研究

Human Action Recognition Based on Depth Motion Maps

硕士研究生： 李延林

导　　师： 史东承教授

申请学位： 工学硕士

学　　科： 信息与通信工程

所在单位： 计算机科学与工程学院

答辩日期： 2018 年 6 月

授予学位单位： 长春工业大学

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

长春工业大学研究生学位论文原创性声明

本人郑重声明：所呈交的研究生学位论文《基于深度运动图的人体行为识别研究》，是本人在指导教师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作 者 签 名：李延林

日　　期：2018年 6月 11日

长春工业大学硕士学位论文版权使用授权书

本学位论文作者及指导教师完全了解“长春工业大学研究生学位论文版权使用规定”，同意长春工业大学保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权长春工业大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，也可采用影印、缩印或扫描等复制手段保存和汇编学位论文。

作 者 签 名：李延林

日　　期：2018年 6月 11日

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

摘 要

随着计算机视觉领域的快速发展，人体行为识别作为其重要的分支，已经成为各大高校和科技公司研究的重中之重。由于人体行为在公共安全领域有着极其重要的应用前景，我国近些年也开始投入大量精力进行研究。由于应用范围不断扩大，如何提高人体行为的识别准确率和算法的鲁棒性，逐渐成为每位研究者的研究重点。

本文的侧重点在于研究人体行为的特征描述和人体行为分类这两个方面。由于传统的人体行为的识别方法大多依赖普通的 RGB 摄像机所获取的人体行为视频。其所记录的人体行为易受到光照、视角等因素的影响而造成识别率的下降。这十分不利于人体行为分析的商业化推广。为此，本文尝试利用图像景深作为突破点进行人体描述，来避免因光照和视角等因素造成的识别率下降。

首先，利用 Microsoft 所开发的 Kinect 进行人体行为的数据采集。获取我们所需要的人体行为的深度数据，利用深度数据在景深和纹理上的描述优势，进行三维投影以获取人体行为特征描述并将其转换成所需的深度运动图特征向量，并使用 LBP 算子对深度运动图再次进行特征提取，形成最终所需的人体行为描述子。

其次，利用获得的深度运动图特征向量进行 PCA 降维，以去除冗余特征向量并提升计算效率。随后，使用 L2 范式协同表示对人体行为进行分类处理。在应用协同表示时，我们将引用 Tikhonov 正则化，其目在于调整测试样本与测试样本中相似动作增加权重系数，对不相似动作减少权重系数，以此提高分类效果。

最后，本文将使用 Matlab 进行仿真实验，并使用 MSRAction3D 数据库作为训练和测试的标准数据库。在文中， MSRAction3D 数据库将按实验一和实验二两种实验条件进行对比试验。在实验一中，本文算法将与 Bag of 3D Points、DMM-HOG、HOJ3D 以及 Space-Time Occupancy Patterns 方法进行对比，并取得了理想的实验结果。在实验二中，本文算法与 DMM-HOG、Random Occupancy Patterns 和 Actionlet Ensemble 进行相比实验，识别率依然优于其他三种方法。

关键词：人体行为识别 深度运动图 局部二值模式 L2 范式正则化协同表示

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Abstract

With the rapid development of computer vision, human action recognition is an important branch. Many universities and companies have begun to focus on it. Because human action recognition has a very important application prospect in the field of public safety, China has invested a lot of money in recent years. As the application scope expanded, how to improve the recognition accuracy of human action and the robustness of the algorithm has gradually become the focuses of the researchers.

The emphasis of this paper is to study the characteristics of human action description and human action classification of these two aspects. Because the traditional human action recognition methods is mostly dependent on the ordinary RGB cameras. The recorded human action is subject to the influence of lights, visual angles and other factors. And these can be greatly influence the recognition rate. This is disadvantageous to the commercialization of the human action application. Therefore, this paper attempts to describe the human body by using the depth maps. It can avoid the influences caused by illumination and visual angles.

Firstly, we use the Kinect which was developed by Microsoft to collect the human action data. We need the depth maps of human action, and we will use the description advantage of depth maps in depth and texture. Then we use the depth maps to project to three perpendicular of the Descartes plane and converts the depth maps into the desired feature vectors of depth motion maps. Then we use the LBP to extract the features from the depth motion maps again. The formation of the final descriptors of human action is what we required.

Secondly, the feature vectors of depth motion maps is reduced the data dimension by using PCA, so as to remove the redundant feature vectors and improve the computational efficiency. Then the L2-regularized collaborative representation is used to represent the human behavior. In the application of collaborative representation, we will use the Tikhonov regularization, the purpose is to adjust the test samples and generated dictionary's weight coefficient. The similar actions will increase the weight coefficient, and the dissimilar actions will reduce the weight coefficient. It is aim to improve the classification effects.

Finally, this paper will use Matlab to carry on the simulation experiment, and use the MSRAction3D database as the standard database of training and testing. In this paper, the MSRAction3D database will carry on in the condition of Experimental One

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

and Experimental Two. In Experiment One, our algorithm will be compared with other methods, such as Bag of 3D Points, DMM-HOG, HOJ3D and Space-Time Occupancy Patterns, and the results of the experiment are very good. In Experiment Two, our algorithm will be compared with DMM-HOG, Random Occupancy Patterns and Actionlet Ensemble, and the recognition rate of our algorithm is still better than the other methods.

Key words: Human action recognition Depth motion maps Local Binary Pattern(LBP) L2-regularized collaborative representation

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

目 录

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

第 1 章 绪 论

1.1 课题的研究背景及意义

人体行为识别是计算机视觉(Computer Vision，CV)领域极其热门研究领域之一。它有着非常广阔的应用前景和现实价值，比如在视频监控、身份鉴别、人机交互、视频分析、自动驾驶等各种图像识别领域。随着世界范围内的公共安全问题日益凸显，各类极端恐怖主义也给各国社会的生产和生活带来了极大的威胁，如何积极应对和解决各类极端恐怖主义威胁成为各国政府亟待解决的难题。人体行为识别正因此种现实因素而逐步发展，如今各个国家都积极投入资金并搭建成了所谓的城市 360 度无死角视频监控。例如，美国在 911 事件后，在各个重大公共场所和要塞都加大了对于非正常人体行为的视频监控力度。也有报道指出，英国已经建立了非常完善的监控系统，这个系统使得每个在英国生活和工作的人都会被监控所记录，并据不完全统计，大约每个英国人每天会出现在三百多个不同的监控记录之中。但是，由于这种视频记录模式仅仅是记录视频内容，所以这种方式是极其粗放的，也给视频存储成本带来了很大的挑战。鉴于上述两种问题，全球各大高校、科技公司均看到了人体行为识别一旦商用，必然会极大的解决这两大难题，其潜在商业价值也必然是不可估量的，所以，各研究机构和科技公司对人体行为识别的科研经费投入也大幅提升。

人体行为一般按复杂度可分为四类，如一般姿态类、单人行为类、交互行为类以及群体行为类[1]。这四类虽然复杂度依次提高，但是，它们的侧重点依然各有不同。例如，姿态类和单人行为类一般侧重于人体骨架形态、人体相对于周遭环境的位置等，作为此类研究的主要课题。而交互行为类和群体行为类则侧重于研究人与人、人与物之间的时空特性以及逻辑关系等方面。

依据当前论文的表述情况，全球研究人员大多将研究方向侧重于单人行为类识别，许多国际知名的人体行为数据库也大都以单人行为类作为主要的识别对象。可见，人体行为识别的发展依然处于较为初步的状态，其仍然有着巨大的研究空间，将此课题作为研究内容必然拥有实际的理论意义和应用价值。

一般人体行为的研究方式按以下两种方式进行。第一种方式侧重于研究人体行为的特征描述，其目的是尽可能准确完整的将人体行为进行数学化表达。第二种方式侧重于识别人体行为的类别属性，第二种方法一般是在第一

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

种方法的基础上进行人体行为的归类。本文的主体结构也是从这两方面分别进行阐述。

随着人体行为识别研究的不断深入和发展，人体行为的研究方式逐步呈现多样化的趋势，同时人体行为的识别能力也得到了极大的提升。但至今尚未有可以在复杂的公共场所中可广泛应用的人体行为分析技术。因此，对于人体行为识别的发展仍需要投入大量的成本以积极应对市场的广阔需求，这对全社会的生产和生活活动都有着非同寻常的重大意义。

1.2 国内外研究现状

1.2.1 国外研究现状

目前，国外高校和高科技公司都投入了大量资金进行人体行为分析。如法国国家信息与自动化研究所(Institute of National Research on Information and Automation，INRIA)在电影视频的动作识别中取得了重大进展，他们利用好莱坞电影的台词与视频内容相结合，对视频内的人体动作进行识别。这种利用文本辅助信息进行人体行为识别的新方法值得进一步深入研究。在人群分析的研究中，中央佛罗里达大学(University of Central Florida，UCF)取得了相当理想的研究成绩。英国牛津大学的 VGG(Visual Geometry Group)实验室同样成果斐然，他们开发了基于 Matlab 的计算机视觉工具包 VLFeat 和 MatConvNet，极大的方便了全球其他学者在其成果上的二次开发，极大刺激了人体行为识别的发展速度。近些年，伴随着计算机硬件水平的不断提高，尤其是显卡运行速度和处理能力的突飞猛进，使得卷积神经网络逐渐走进研究人员的视野当中。如斯坦福大学的 Andrew Ng 和纽约大学的 Yann LeCun 在卷积神经网络的研究中走在了世界的前列。他们提出的算法被人体行为研究者广泛使用，并取得了非常不错的识别效果。美国的 Microsoft 公司在研发 Xbox 系列游戏时，为增强游戏的交互能力，Microsoft 公司推出了 Kinect 传感器，Kinect 不仅拥有普通摄像机所拥有的彩色视频的记录功能，它还拥有记录视频中物体景深的功能，这就将人体行为数据扩大了一个维度，自然也就给人体行为分析带来了更多处理手段，如今已经有一批学者利用 Kinect 所采集的视频图像进行人体行为识别并取得了与普通 RGB 摄像机所得视频相同甚至更高的识别效果，所以利用 Kinect 所得的深度信息正逐步受到人体行为识别学者的重视。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

1.2.2 国内研究现状

国内人体行为的研究相较国外稍晚，但每年国内相关论文的出版数量逐步增长，发展也极其迅速，可见人体行为识别也逐步成为国内学者的主攻方向之一。尤其是进入新千年，各大高校和高新技术企业都将大量的科研经费投入到其中。而其中走在国内最前列的要数中科院自动化研究所，早在 2008 年北京奥运会期间，投入到场馆运行的监控系统就是由此研究所研制开发的，为北京奥运会的顺利召开提供了强有力的安全保障。其他高校如清华大学、北京大学、北京邮电大学、电子科技大学、中山大学、西安交通大学、上海交通大学等等，也都成立了自己的研究实验室，并与社会上的一些企业合作研发人体行为分析的产品，并积极投入实际运营中。在一些高科技企业中，如百度，京东等积极致力于无人驾驶技术和无人机的研发工作，而无人驾驶技术和无人机的一个核心技术便是人体行为识别技术，因为让自动驾驶的小汽车和无人机可以成功识别和避让行人和车辆，其算法的基础便是利用人体行为识别技术进行车辆前方的行人识别。为此，百度更是从美国请来 Andrew Ng 作为技术指导，意图引领中国人体行为技术的发展潮流。

每年计算机视觉领域都会有许多会议举行和期刊发表，来分享全球各领域专家的研究成果。近些年来，关于计算机视觉的相关论文一直都是各个邻域学者关注的重点，其中人体行为识别的论文发表数量更是呈现井喷式增长，可见人体行为识别已经成为计算机视觉领域极其热门的方向之一。但从整体的论文发表内容看，现在的研究阶段依然处于底层视觉的研究阶段，对于高层视觉的研究依然处于较为空白的状态。由此我们可以断定，人体行为识别还有很大的发展空间等待研究人员进行深入的挖掘和探索。

1.3 研究难点

人体行为的发展虽然呈现着井喷式发展，但是始终都是延续各自的研究方式，没有一个可以相互统一的方法进行人体行为识别。这就造成了如今，人体行为识别方法虽然多如牛毛，但各种方法虽然针对特定的数据库往往有着不错的识别效果，一旦应用到其他数据库或者现实场景中，往往效果就不会像实验室效果一样理想。因此，如何致力于在现实场景中提升人体行为识别效果是研究者应该积极着力追求的。接下来，本文的后续章节将会按照下文所体现的，人体行为识别中的研究难点问题进行展开表述。

第一，在复杂场景中物体对人体动作的遮挡，人与人之间的身体遮挡以及

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

光照对人体动作的影响都很影响识别算法的准确率。其次，摄像机抖动对拍摄物体造成的模糊。或者由于人体所穿衣着颜色与周遭环境颜色相近时，造成很多识别算法会往往会将人体视为背景，而非人体行为。鉴于此种问题，利用 Kinect 采集深度信息，可以避免因光照和颜色等问题造成行为与背景的误判。

第二，对于现实场景中，对于相似动作的识别始终是人体行为识别中极其令人头痛的症结。例如，打高尔夫球的挥杆动作与打羽毛球的挥拍动作，描述这两种动作的特征矩阵结构和矩阵中元素的数值往往十分近似，难以判断。另外，相似体型的人在做拾起与放下的动作时，同样也会产生很大的相似性。如何创造出比较有区别性的描述子对动作进行描述，是提升识别效果的主要方式之一。

第三，充分利用高层语义进行人体行为识别是未来的发展方向，也是最能应用到现实生活中的主要技术。因为，现实生活中人们需要识别的往往都是具有高度复杂性的多重动作，这些动作一般是由一些连续的简单动作相互连接在一起的集合。不同的简单动作进行不同的排列组合往往也意味着不同的动作语义。另一个重要问题是，在一些重大公共场景中，对于人体行为识别的实时性要求有着非常高的需求，所以，千方百计提升和优化行为算法识别的速度。同时，还不会不降低识别准确率也就成为人体行为识别的未来难点之一。

1.4 本文主要研究内容及章节安排

本文主要侧重点放在研究 Kinect 传感器所获取的人体行为深度数据，该数据库是现在普遍认可的人体行为深度数据库，即 MSRAction3D 人体行为数据库。在人体行为特征提取方面，本文提出将投影算法和 LBP 特征提取算法相融合，以提高深度信息的特征提取精准性和鲁棒性，所用数据库为 MSRAction3D，该数据库是深度信息领域内十分重要和流行的标准数据库之一，很多算法都已将该数据库作为验证算法的主要标准之一。文中，我们将展示本文算法对深度运动图特征提取方面的实验结果，从实验图中可知，本文的特征提取算法表现较为良好。非常有利于下一步的人体行为分类器对所提取的深度特征进行分类。在动作识别方面，本文首先将特征提取的部分进行 PCA 降维和去除冗余信息，然后在利用 Tikhonov 矩阵计算训练数据库的各类动作的系数向量。最后，本文提出了利用 L2 范式正则化协同表示对人体行为进行识别。

本论文总共包含六个章节，每章的具体安排如下：

第一章：绪论，本章主要介绍人体行为识别课题的研究背景及其意义，并介绍了国内外目前的发展现状。同时本文也提出了人体行为识别的研究难

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

点和热点，最后提出本章的总体内容和结构。

第二章：人体行为识别的常用关键技术，本章主要介绍人体行为识别的一般方式和流程，并提出人体行为的两方面重要技术，即特征提取算法和行为识别算法。本章将会详细介绍现阶段主流的特征提取和行为识别的发展情况。

第三章：基于深度映射图的人体行为特征提取算法，本章主要介绍利用 Kinect 传感器可获取的人体骨架数据和深度数据，本文算法将会利用深度数据作为人体行为识别的主要数据，即深度映射图(Depth maps)，通过对深度映射图三维投影以生成人体行为的深度运动图特征向量，并使用 LBP 算子对深度运动图再次进行特征提取。

第四章：基于深度运动图的人体行为识别算法，本章主要介绍利用深度运动图特征描述子，生成本文所需的训练样本集合，随后将 Tikhonov 正则化应用到 L2 范式协同表示中，对测试样样本进行分类，使用 Tikhonov 正则化的目的在于将生成的训练样本集合中与测试样本相似的动作给予较高的权重系数，对不相似的动作给予较低的权重系数，以期提升识别效果。最后，计算测试样本与各类训练样本的重构残差，以确定测试样本所属类别。

第五章：实验过程、结果与分析，本章主要介绍本文算法在特征提取和动作识别中的表现情况。并使用 MSRAction3D 数据库作为算法识别率的主要评价标准。实验中，本文将 MSRAction3D 按实验一与实验二两种条件分别进行对比试验，在实验一中，本文算法与 Bag of 3D Points，DMM-HOG，HOJ3D 以及 Space-Time Occupancy Patterns 进行准确率比对，表明本算法在识别准确率上有了明显提升。在实验二中，本文算法将于 DMM-HOG，Random Occupancy Patterns 以及 Actionlet Ensemble 再次进行实验对比，本文算法依然优于其他三种算法。

第六章：总结与展望，本章主要总结本文所做的主要工作，评价算法的优缺点，并问题未来的研究方向和内容作了进一步的展望。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

第 2 章 人体行为识别的常用关键技术

2.1 人体行为识别的一般流程

对于人体行为识别的研究其实并不短，它已经经过近半个世纪的研究和发展，已经逐渐形成了自己独特的行为识别流程。按照传统识别方式一般分为三部分，第一部分多以图像或视频的预处理为主，主要目的是去噪方便下一步的特征提取。第二部分则多以特征提取为主，将人体行为的主要特征进行数学化表达。比如，密集轨迹提取[2]、光流直方图[3]、运动历史图[4]等各种形式的特征提取方法，其目的不外乎形成可靠，快速的人体行为特征描述子，便于第三步的分类与识别。第三部分则多侧重于对前一步的特征描述子进行识别和分类，利用训练集内的特征数据先进行识别器的训练，使识别器的内部参数、结构可以调节到适当的形式。最后，输入测试集对识别器进行评估。图 2-1 为人体行为的一般流程图。

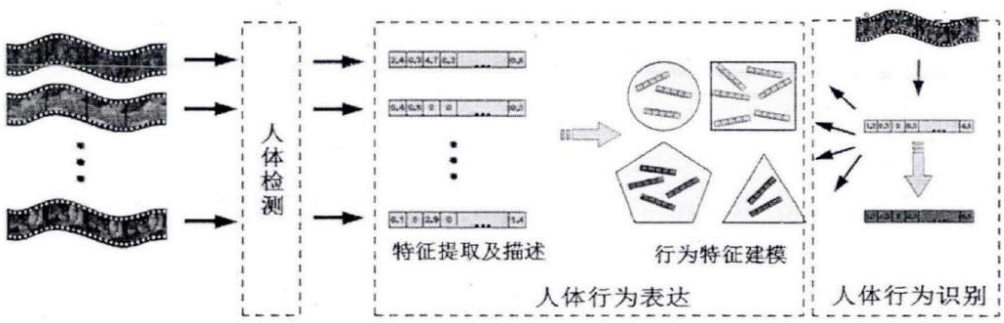

图 2-1 人体行为的一般流程图

近些年，由于计算机运行速度的快速提升，尤其是显卡计算能力的大幅提高，许多依赖于运行速度的算法逐渐受到学者们的重视。卷积神经网络便是近几年极为流行的深度学习算法之一。2006 年，Hinton 在 Science 上发表了一篇名为“Reducing the Dimensionality of Data with Neural Networks”[5]的文章，这是一篇具有划时代意义的论文，正是这篇文章掀起了神经网络发展的新高潮。由于卷积神经网络在特征提取与识别的结构上完全不同于传统方法，且其结构清晰明了，只需要将关键性参数设定完毕后，便可以将整个网络的训练交由计算机来完成，大大简化了整个人体行为识别系统的复杂构架。但其缺点也如其优点一样，对于卷积核的确定，网络层数的构架，全连阶层后的分类器的设定，这些繁琐的内容都需要人为去设定，无法做到自动提升与优化，而且现在还没有具体的数学理论或者其他理论算法可以指导使用。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

同时，卷积神经网络对于计算机硬件水平要求极高，为了实现其应用的便捷性和实时性，往往需要强大的服务器作为支撑，这一点也造成其使用成本极其高昂。

2.2 人体行为的常用特征提取算法

人体行为识别常用的特征提取算法大致分为全局特征法、局部特征法、全局与局部特征融合法。这些方法的基本思路是利用人体行为的连贯性和独特性，分别从整体思路和局部思路来分析人体行为。全局法的思路是从大至小，以整体特征为主。而局部法的思路恰恰相反，它是从小到大，以局部特征为主。全局与局部融合法，则是意图将整体与局部的各自优势相结合，更加丰富人体行为的描述精准性和完整度。

2.2.1 全局特征

全局特征，顾名思义就是对人体行为的整体性和总体综合性进行描述的一种方式，一般的常规使用方法有剪影法、光流法、模板法、纹理描述法等。全局特征一般需要先将视频或图像中的人体轮廓进行定位并提取出来，因此全局特征法所包含的人体信息非常丰富，但是它对光照、遮挡和噪声等又十分敏感，这也是全局特征的致命缺点之一。

剪影法是利用人体形态的一种方法，一般的获取途径是利用人体动作视频序列的差分。Wang 等人[6]基于此提出了 AME(Average Motion Energy)和 MMS(Mean Motion Shape)。其中 AME 是将一个完整视频序列的人体动作的剪影图像序列累积和的平均值。MMS 则是利用人体运动的形状，先求出人体形状的质心，然后利用人体边界点对人体行为进行描述。

光流法[7]是利用视频帧相邻的两帧之间的像素值的变化作为确定人体行为信息的一种方式。光流法使用的前提是基于以下三种假设，第一：相邻帧之间亮度保持不变；第二：相邻帧之间的运动物体在像素级别的变化较为“微小”；第三：被划归为同一个子图像的像素点都具有相同的运动。由于所需识别的人体动作大多背景不变，因此利用光流法可以不必考虑将人体轮廓提取出来，但光流法对噪声十分敏感，拍摄相机的抖动等，都会极大的影响光流法的应用效果。图 2-2 为利用光流法所得的运动轨迹。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图 2-2 光流法

模板法[8]大多以运动能量图(Motion Energy Images,MEI)和运动历史图(Motion History Images,MHI)作为常用方法。其思想核心是将人体动作的整体作为一个完整的模板，运动能量图记录的是某一人体动作中产生运动的部分，而运动历史图则记录了某一人体动作的按照时间发生的顺序序列，学术界往往都是将这两种人体行为的特征描述子作为人体行为的特征向量。然后利用所需的分类器计算不同的人体行为特征之间的距离，当两个特征距离越近说明动作相似度越高，如果距离小于所设定的阈值时，便可说明两者之间的所做的动作是一样的。图 2-3 为利用 MHI 处理图像的情况。

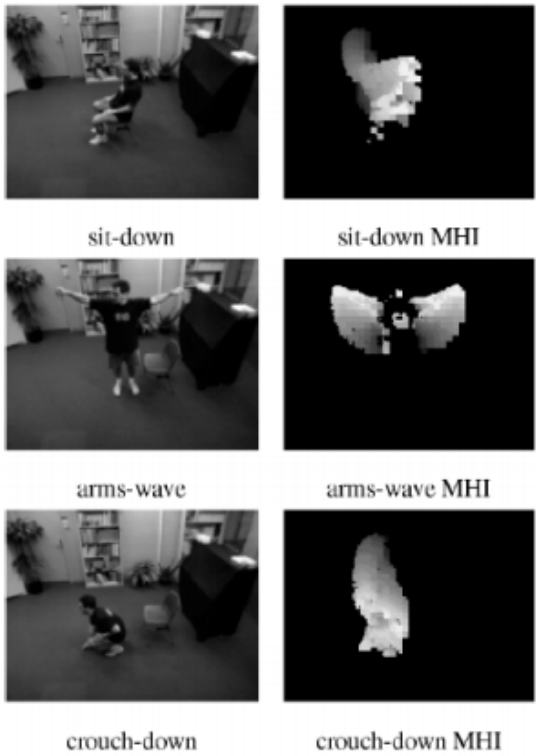

图 2-3 利用 MHI 处理图像的情况

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

图像纹理信息[9]的出现是通过仿照人类眼睛处理具有纹理类属性图像的一般方式方法所提出的，所以一直都是计算机视觉领域里学者研究的重要方向之一。依据人类处理图像信息的一般过程，我们可知图像的纹理信息是人类做出识别的重要依据之一。但是，到目前为止，在脑科学领域和计算机视觉领域中一直没有明确的数学定义来对图像纹理进行定量的描述。直到 Timo Ojala 在 2000 年首先提出利用局部二值模式(Local Binary Pattern, LBP)作为一种图像纹理特征提取算法，LBP 算法可以实现对于图像表面材质、光照、旋转等进行十分详细的特征描述，极大丰富了图像的全局特征的描述能力。本文将使用 LBP 算子作为全局特征描述的方法之一，并将在第三章中进行详细介绍。

2.2.2 局部特征

局部特征的思想方法恰好与全局特征相反，其目标是利用视频帧之间运动物体细节像素的变化作为人体行为的主要描述特征。其优点是对光照、噪声和遮挡等不是十分敏感，可以提升人体行为描述的鲁棒性。

局部特征多以时空兴趣点(Space and Time Interested Point,STIP)[10]作为主要描述方法。其思想是寻找视频序列中运动变化较大的像素点作为关键点，然后在视频帧中的规定区域内搜寻这些关键点的变化情况，将关键点周围发生运动的特征区域提取出来作为特征描述内容，并以此作为识别基础。

HOG(Histogram of Oriented Gradient)[11]，即方向梯度直方图。它是计算机视觉领域十分常见和流行的局部描述子之一。其核心思想是将所要计算的图像先进行分块，然后统计各个分块区域内梯度直方图，原因是图像梯度主要表现在图像纹理中的边缘，所以边缘方向分布状况以及梯度统计情况可以进行较为准确的描述。Navneet Dalal 在 2005 年成功利用 HOG 算子应用到人体检测，并得到了非常好的实验效果。

Harris Corner[12]是典型的局部特征描述子，Harris Corner 一般称为角点检测描述子，它一般是选取人体动作的遮挡位置边沿，边缘交界处或者图像纹理信息丰富的位置，这些位置往往有着较高的稳定性和重复率，可以较为鲁棒的描述人体行为信息，比如匀速行走的人，其腿部的变化是具有一定周期规律的，利用 Harris Corner 进行特征描述可以获得非常良好的效果。

SIFT(Scale Invariant Feature Transform)[13]是现阶段引用十分广泛的局部特征描述方法之一，它是由 D.Lowe 首先研发出来，并逐步得到完善。利用 SIFT 提取图像特征可以实现特征描述子的尺度、旋转以及光照的不变性，非

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

常适合描述各种人体行为特征，使其特征具有足够的鲁棒性和可识别性。

SURF(Speeded Up Robust Features)[14]是 SIFT 算法的一种快速算法，它使用 Haar 小波取代 SIFT 中的梯度算法，并使用积分图法共同提升 SURF 的计算速度，使得 SURF 的计算速度是 SIFT 的 3 到 7 倍，而且多数使用情况下实现效果与 SIFT 不相上下，非常适合于对实时性要求较高的视频监控领域和自动驾驶领域。DAISY[15]在思想本质上与 SIFT 差不多，两者都是将图像分块，然后统计各个分块中梯度方向直方图。不过，它们的不同之处在于 DAISY 在图像分块的方式上与 SIFT 不同，DAISY 利用高斯卷积实现梯度方向直方图的分块汇聚，因为 SIFT 方法中使用的是加权梯度直方图，当所要计算的像素位置发生平移或者旋转时，其梯度直方图也会随之发生改变，自然，其梯度直方图也就需要再次重新计算。而 DAISY 方法由于使用高斯卷积对梯度直方图进行加权运算，加之高斯核具有各向同性的优点，因此，当所计算图像的像素发生旋转时，只是直方图 bin 的顺序发生变化，其它则没有，不需要重复计算，所以十分便于提升计算效率。图 2-4 为 DAISY 的计算方式。

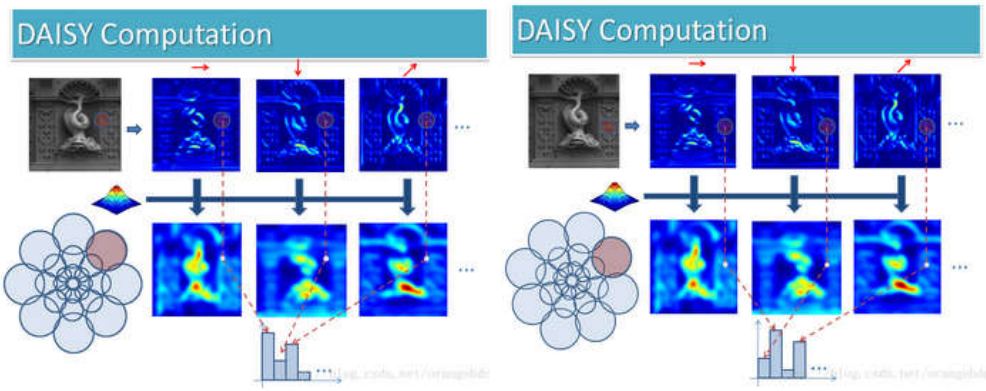

图 2-4 DAISY 计算方式

Klaser 将 HOG 描述子扩展成三维，形成 3D-HOG，并允许对 cuboid 快速密集采样。Scovanner 等人[16]将 SIFT 算法进行三维扩展，形成 3D-SIFT 描述子，并取得了比较高的识别率。同样的，Willems 将 SURF 描述子也进行了三维扩展，并将其率先应用到视频序列中。

2.2.3 全局与局部特征融合

全局与局部特征融合法其目的是充分结合全局特征的完整性和局部特征的鲁棒性，来提高人体行为的的识别效果。其大体可分为两类，第一类是原始数据的融合，第二类是特征描述的融合。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

原始数据的融合大多是从视频采集方面入手，例如将普通的 RGB 摄像机视频与远红外摄像机视频共同作为视频源，在记录普通 RGB 视频的同时也记录物体的景深信息，这样可以增加所记录视频数据的信息维度。Kinect 传感器便是如此，不仅可以拍摄普通的 RGB 视频，而且也可以记录视频得到景深信息，并利用光编码技术生成人体行为骨架结构。本文所使用的深度运动图就是基于 Kinect 所采集的数据而来。这些都可以视为原始数据方面的融合，它们将全局和局部信息全都保存下来，有利于学者对于所需特征的进一步选择。

特征描述融合的一般思想方法是将全局特征的人体行为信息和局部特征的人体行为信息分别提取出来后，然后，将二者有机的联系在一起。如 Thi[17]利用 AIFT 算法优化全局特征 MHI 算子，同时利用 SBFC 算法优化局部特征 STIP 中信噪比较小的特征，增加局部特征的鲁棒性。最后，将这两种特征输入到 ISM(Implicit Shape Model)分类器中，获取输出结果。图 2-5 为 Thi 所提方法流程图。

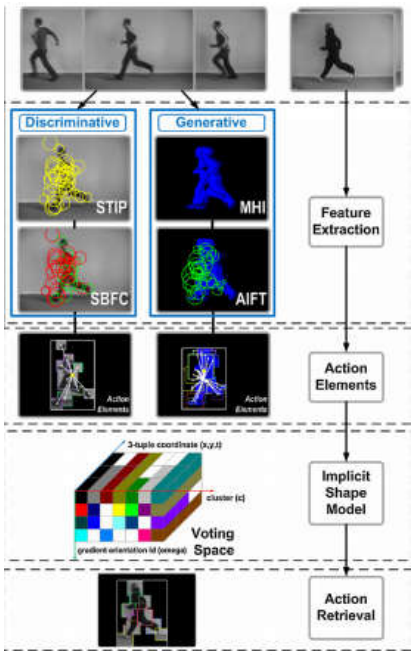

图 2-5 Thi 所提方法流程图

全局与局部特征融合虽然理论上可以丰富人体行为的表述能力，但如果过多的融合了大量特征，反而会使识别效果大幅下降。只有有选择有取舍的将特征间的优势与劣势互补才可以提升识别效果。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

2.3 人体行为识别的一般算法

前文已经将人体行为的一般的特征描述子进行了详细阐述。在一般的人体行为识别系统中，提取出人体行为特征后，一般要与适当的分类器或者说是识别器相结合，才能达到不错的识别效果。当前的分类器模型有很多种类，大体可分为三类，第一类是判别式模型，第二类产生式模型，第三类是深度学习模型。接下来，本章将相继进行介绍。

2.3.1 判别式模型

判别式模式的基本思想是将特征向量之间的差异通过类别间的最优平面进行划分。判别式模式的常用模型有支持向量机(Support Vector Machine,SVM)[18]、线性判别式分析 (Linear Discriminant Analysis,LDA)[19]等。

SVM 是一种典型的判别式模型，SVM 的基本思想是将那些在低维空间不可分割或者无法找到最有分割的数据转换到高维空间，使需要分割的数据在高维空间中变成线性可分的数据，从而找到满足条件的“超平面”。通常情况下，满足条件的超平面有很多，SVM 要找的超平面是两类数据之间的最大间隔超平面。LIBSVM[20]函数库是由林智仁教授研发出的 SVM 库，该函数库调用简单，输入参数少，运算速度快，而且是开源的，所以现在许多论文中应用的 SVM 大都以 LIBSVM 为基础。图 2-6 展示了 SVM 的分类方式。

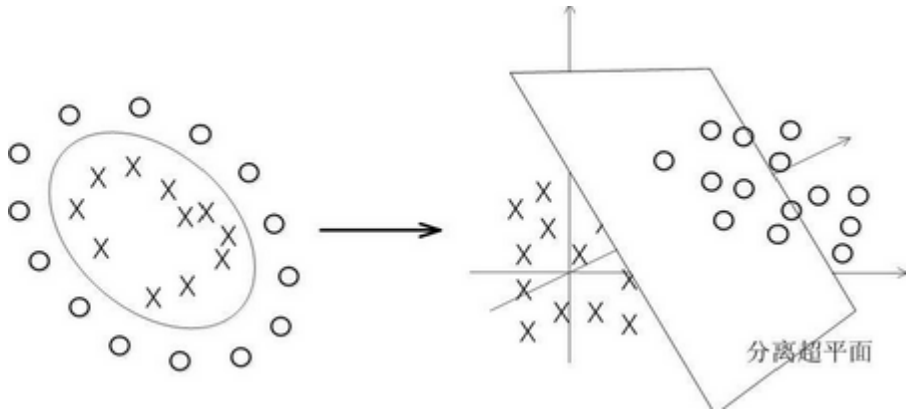

图 2-6 SVM 分类方式

LDA 亦可称为 Fisher 线性判别(Fisher Linear Discriminant,FLD)。其基本思想是将高维数据投影到最佳的矢量空间进行识别，可以达到抽取分类信息和压缩特征空间维数的作用，数据投影后应保证在新的矢量空间中类内距离最小、类间距离最大，也就是要达到数据在此矢量空间中有最佳的可分离

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

性。所以，使用 LDA 是进行识别的有效算法之一，它能够将人体行为特征经投影后形成类内散布矩阵最小，类间散布矩阵最大，从而达到较好的分类效果。

2.3.2 产生式模型

产生式模型是从统计学角度研究人体行为特征向量的分布情况，并利用后验概率进行建模，从而解决同类型特征向量的相似度问题。产生式模型一般包含有隐马尔可夫模型(Hidden Markov Model, HMM)、动态贝叶斯网络(Dynamic Bayesian Network,DBN)[21]等。

HMM 是由 L.E.Baum 等学者在统计学期刊论文里发表的。并在上个世纪逐渐应用到商业领域中，并成功应用于语音识别，自然语言处理和生物信息等尖端科技领域。进入新世纪，富有创造力的学者们逐渐开始利用 HMM 进行人体行为识别方面的研究。HMM 是将人体行为的各种动作状态按一定的逻辑纳入到其模型中，每个人体动作之间又有一定的状态转移概率，且各个人体动作之间转移是随机的。在 HMM 领域内进行人体行为识别往往都取得了不俗的成绩，可见应用该领域进行人体行为识别有着不错的发展空间。如图 2-7 展示了三层 HMM 模型。

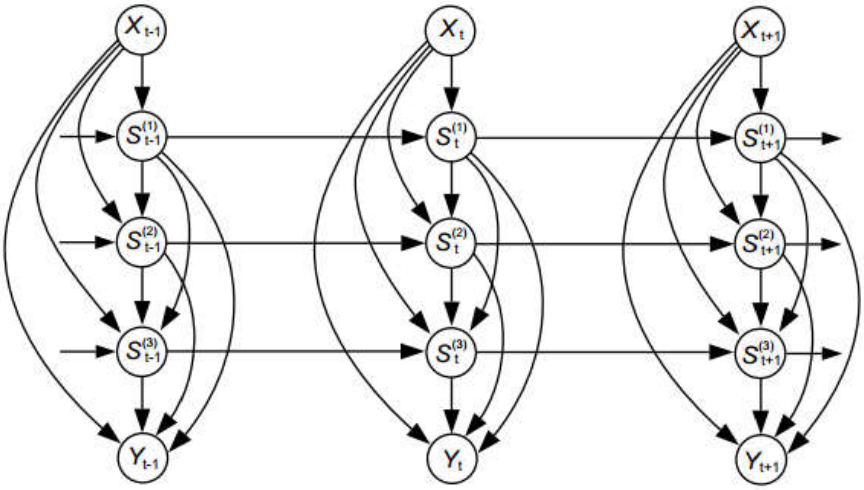

图 2-7 三层 HMM 模型

DBN 作为产生式模型，它与判别式模型中的神经网络相比较，DBN 处理方式是将观测数据与数据标签结合起来形成联合概率分布。在 2006 年，Hinton 在 Science 上发表的论文就是以 DBN 网络结构为主，如图 2-8。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

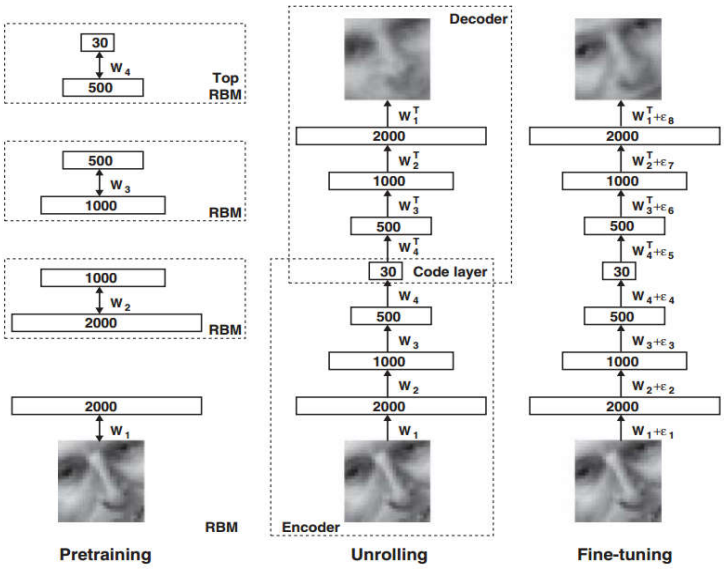

图 2-8 DBN 网络模型

论文[5]通过若干层 RBM(Restricted Boltzmann Machine)的级联，并在最后一层使用 BP(Error Back Propagation)算法，形成典型的 DBN 网络模型。各层之间相互连接，但同层之间没有连接，从输入层输入训练数据，在输出层利用 BP 算法对各层之间节点的权重进行更新，从而训练出满足人们需求的识别系统。DBN 网络是近些年刚刚兴起的新兴技术，随着现代计算机硬件水平的逐渐提高，将 DBN 网络应用到人体行为识别中必然会有宽阔的发展前景。

2.3.3 深度学习模型

深度学习是近几年比较热门的研究领域，它的第一个成功实现的网络是由日本学者 Fukushima 提出的基于感受野(Receptive Field)的神经认知机(Neocognition)[22]。神经认知机的处理方式是，第一步：将视觉模式分解成许多子模式；第二步：是进入分层递阶式相连的特征平面进行处理，通过视觉系统的模型化，对于物体位移或轻微变形的依然能完成识别。神经认知机利用位移恒定能力从激励模式中学习、识别各种模式的变化形式。

随着计算机硬件运算速度的不断提升，尤其是 GPU 运算能力幅度的大幅提升，卷积神经网络(Convolution Neural Network,CNN)[23] 这一曾经被人们搁置的方法，开始逐步进入学者们的研究视野。尤其是在图像领域，由于人体行为识别需要计算的数据是空前巨大的，传统方法在特征提取和特征识别上算法相对比较复杂，在噪声严重、光照变化大的情况下，对传统方法一直都是比较大的挑战。利用 CNN 进行人体行为识别可以克服上述难点。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

CNN 网络结构相对传统方法比较简单，它由卷积层、池化层交替联结，最后在联结由两个全连阶层，这便是 CNN 网络的一般结构。CNN 的发展势头虽然不错，但是网络内部的卷积核的设定，池化方式的设定以及全连阶层的设定都需要人工设计，并且没有强大的数学算法作为依托，只能凭借学者经验进行一一设定。所以，许多学者现在开始尝试利用 CNN 与传统方法相结合，也取得了不错的效果。图 2-9 展示的是 CNN 的一般实现流程图。

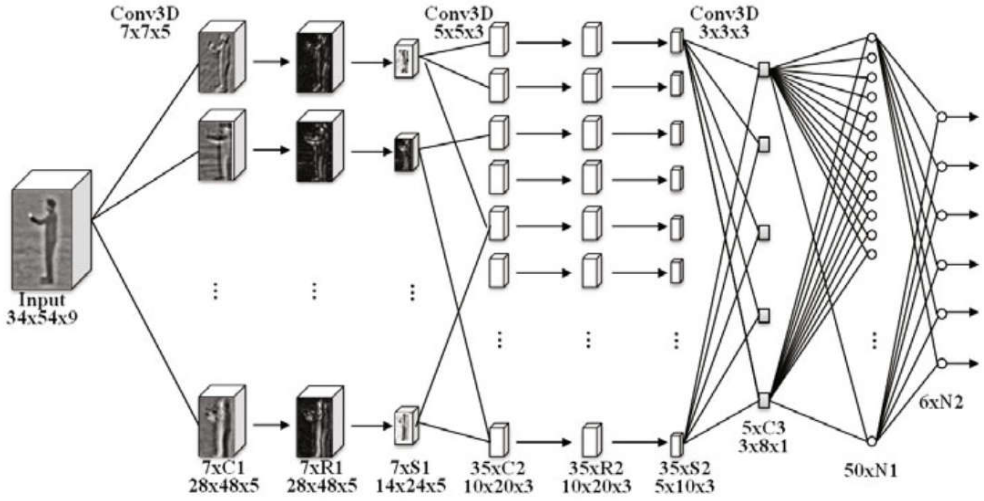

图 2-9 CNN 的一般流程图

2.4 本章小结

本章主要阐述了现阶段人体行为分析的一般特征提取与特征识别的常用算法。首先，介绍了人体行为识别的一般算法流程，其算法具有一定的一般性，也是主流算法的基本流程。紧接着，介绍了人体行为的一般特征提取算法，其中很多算法已经应用到商业领域，也是现在广大学者经常使用的特征提取方法。最后，介绍了现在主流的人体行为识别算法的大体分类情况，其中尤其以深度学习最为热门，现在很多高校和高科技公司皆积极投入其中，是值得研究的领域之一。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

第 3 章 基于深度映射图的人体行为特征提取算法

3.1 Kinect 获取的人体行为数据原理介绍

Kinect 是由 Microsoft 公司研发的一款体感游戏设备，它可以生成玩家的人体行为数据，生成的数据通过 Microsoft 的识别技术，使游戏玩家可以彻底摆脱鼠标键盘等游戏接入设备，解放人体操控游戏方式单一模式，提升游戏的体感真实性。Kinect 拥有三个镜头，中间的镜头是普通的 RGB 摄像头，可以记录周遭环境的 RGB 信息。Kinect 的另外两个装置分别是红外线发射器和红外线 CMOS 摄像机，利用红外线发射器发射红外线，另一个则是接收红外反射信号，从而形成人体行为的深度信息，即深度映射图。同时 Kinect 还拥有四个阵列式麦克风和底座转动电机，拥有去噪、定位和追踪的功能。实际的 Kinect，如图 3-1。

图 3-1 Kinect 设备结构图

3.1.1 Kinect 的深度图像成像原理

Kinect[24]的深度图像成像原理相对比较简单，红外线发射器和红外线 CMOS 传感器是 Kinect 的主要核心元件，是生成深度信息的唯一来源。红外发射器通过透明散射体(如毛玻璃等)会在散射空间形成不规则分布的明暗光斑，这些斑点我们称之为激光散斑(Laser Speckles)。鉴于激光具有高度的相干性，且激光散斑在同一空间中任何两个不同位置的散斑图案都会不同，所

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

以激光散斑本身就会携带物体的位置信息。Kinect 就利用该原理在其红外线发射器所覆盖的范围内，每隔一定距离就会生成一个参考平面的散斑图像，当进行测量实物时就会拍摄实物的的散斑图像，并与参考平面的所有散斑图像做互相关运算，当空间中有所拍摄的实物时，做相关运算就会在相关度图像上产生一个峰值。将相关度图像上的峰值层层对叠在一起后，经插值运算便可以得到实物的三维图像信息以及整个场景的三维信息。最后，利用 Kinect 里的专属芯片便可计算出深度图像。

3.1.2 三维骨架数据

基于骨架数据[25]的人体行为识别最早是由 Johansson 提出来。起初，他在人体的各个关节点上安装发光点，并让佩戴者再一个暗室内做动作，然后让观察者观察佩戴者所做的动作。实验结果表明，在只看人体动作的关节点也可以识别出人体行为。由此利用人体骨架信息的人体行为识别逐步成为研究者研究的一个方向。

Kinect 作为一种可以记录图像深度信息的设备，其利用光编码技术(Light Coding)获取人体行为的深度映射图，然后利用深度映射图中每个像素所记录的三维信息，搜索深度映射图，将图像里的人体行为轮廓，将人体行为轮廓从背景中抽取出来。接下来，利用抽取出的人体行为轮廓识别头部、躯干和四肢，作为下一步标记人体骨架关节点的基础。最后，Kinect 会利用机器学习的方法标记头部、躯干和四肢的各个主要关节点。Kinect 会标记 20 个人体的关节点作为人体行为的主要数据。所记录的关节点如图 3-2 所示。

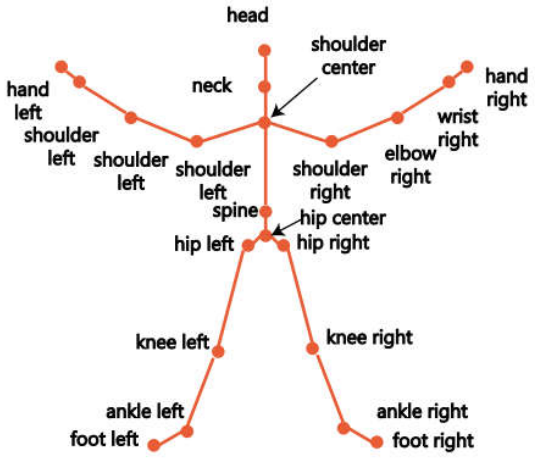

图 3-2　人体关节点示意图

利用 Kinect 生成的三维骨架数据在人体行为识别中有着比较广泛的应用。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

这是因为三维骨架数据的信息处理量较小，运算速度要比处理深度信息要快得多，但也正是由于三维骨架数据信息量相对较小，所以其对于相似动作识别率效果一般很难达到理想的效果。

3.1.3 深度图像

深度图像有时也被称为距离图像，它与一般的灰度图像和彩色图像不同，它所生成的深度图像对于人眼来说是一般是很难鉴别的，但是它记录的深度信息却为图像处理增加了一个维度。并且在深度图像中每个像素所记录的都是场景中的各点与 Kinect 的距离，这就是所谓的人体行为深度信息。深度图像的成像原理上文中已经提到，相对于三维骨架数据，深度图像数据量更大，可表示的人体行为信息更多，方便我们进一步提取人体行为特征。对于深度图像的处理一般分为时空体法、局部特征法以及投影法[26]。

时空体法和投影法一般归类于整体法之中，其中时空体法[27]一般是把人体行为某一动作序列的深度图像视为一个时空体，并在时空体中以不同的尺度采样不同位置，然后计算所采样的子时空体中各个像素的个数。随后，生成所谓的随机占有模式(Random Occupancy Patterns, ROP)特征，该特征对于噪声鲁棒，而且对遮挡不敏感。这就使得人体行为特征更加具有可识别性。然后将 ROP 特征进行稀疏编码，并利用 SVM 对人体行为进行分类。

投影法则是将时空体这一四维模型进行平面投影，如 Yang[28]提出将深度图像序列在笛卡尔平面坐标系中进行投影，并试图获取各个投影面中整个动作序列的运动能量图，然后利用 HOG 计算各个投影面的人体行为特征向量，最后利用 SVM 分类人体行为。投影法也将是本文特征提取的核心方法之一，具体内容将下一节中在详细介绍。

局部特征法一般是指提取深度图像序列的兴趣点，然后利用三维局部特征算法如 3D Harris Corner、Cuboid 等进行局部特征提取。Zhao[29]提出利用局部深度模式(Local Depth Pattern, LDP)作为对深度图像的特征提取方法，取得了十分理想的准确率。Lu[30]利用 STIP 方法从深度图像中获取 DSTIP 特征，并利用 DCSF(Depth Cuboid Similarity Feature)描述时空体中局部深度图像的 Cuboid，并将 DSTIP 与 DCSF 相融合，也得到了非常理想的识别效果。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

3.2 深度运动图的人体行为特征生成

3.2.1 深度映射图的投影方法

利用 Kinect 所获取的深度图像序列并参考文献[28]，在本文中将其统一称之为深度运动图(Depth Motion Map, DMM)，构成深度运动图的每一帧深度图像则称之为深度映射图。

由于深度映射图本身具有三维信息和结构，因此我们将深度映射图在三个互相垂直的笛卡尔平面坐标系中分别作二维投影，投影视角分别称为前视视角(front view)、侧视视角(side view)和俯视视角(top view)，可分别形成深度映射图的二维投影 map_f， map_s 和 map_t。为方便书写我们可以写成如下形式， $map_v(v=f,s,t)$。

设某一动作的深度运动图尺寸为 240*320，帧数为 44 帧。则每一帧深度映射图的三个二维投影 $map_v(v=f,s,t)$ 的求取方法为：

第一步，求取深度运动图中的最大像素值 $MAX=\max(DMM_{240\times320\times44})$。

第二步，设置三个投影映射图的投影尺寸，分别为 $size(map_f)=240\times320$ (map_f，即为深度映射图的当前帧，帧的尺寸和帧内像素的数值皆不发生变化)， $size(map_s)=240\times MAX$ 和 $size(map_t)=MAX\times320$。

第三步，则是利用前视视角 map_f 中非零像素值对 map_s 和 map_t 进行投影，假设 map_f 中第 i 行第 j 列的像素值不为零，即 $map_f(i,j)\neq0$。则该点在侧视视角的投影公式为 $map_s(i,map_f(i,j))=j$，俯视视角的投影公式为 $map_s(map_f(i,j),j)=i$。

3.2.2 深度运动图的生成

依据上文内容，不妨假设某一动作的深度运动图有 N 帧，则深度运动图的获取可由连续相邻帧的深度映射图的二维投影做绝对差分，并将这 N-1 个绝对差分进行累加，即可形成本文所需的深度运动图(DMM)，运算公式为

$$DMM_v=\sum_{i=2}^{N}\left|map_v^i-map_v^{i-1}\right| \tag{3-1}$$

如图 3-3 即为利用公式 3-1 所生成的深度运动图。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

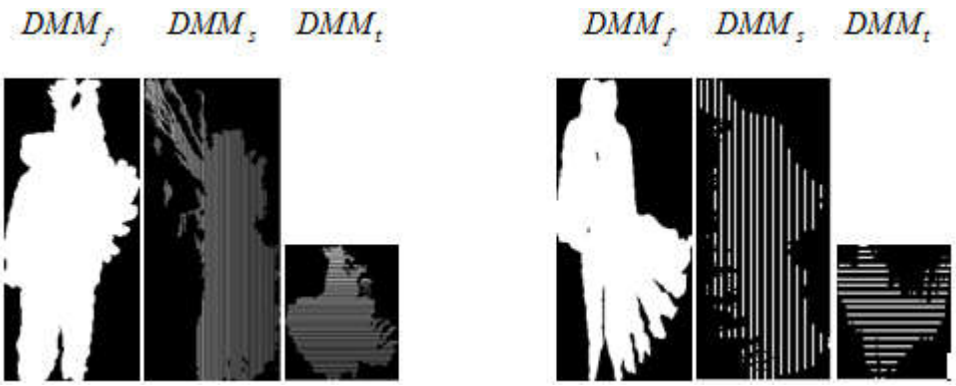

图 3-3 生成的深度运动图

在实验环境中，为增加运算速度并消除其他非人体行为造成的冗余信息，本文利用边界盒子来约束其他非人体行为对人体行为识别的干扰。在形成的深度运动图中，我们会统一将像素值为非零区域的位置视为人体运动区域，以此形成一个边界盒子。

如图 3-3，可以看出深度运动图的三个投影可以清晰的表达人体动作的类别。但是，同一个实验者所做的不同动作，或者是同一动作不同实验者所做的动作在三个投影面的投影尺寸可能大小不一，会造成特征向量的数量，维度等也不尽相同。所以，我们在求得深度运动图后，会利用双三次插值法将各自的动作的投影都形成统一的投影大小。一般我们可设每个投影的大小分别为 $m_f \times n_f$ (前视视角)、$m_s \times n_s$ (侧视视角)、$m_t \times n_t$ (俯视视角)。在实验中，我们规定三个投影面最后的大小分别为：前视视角为 102*54，侧视视角为 102*75，俯视视角为 75*54。由于本文所得的深度运动图是使用像素值作为特征值，为了减小大特征值在运算过程中，对小特征值的干扰，本文将会使所有特征值都进行归一化处理，以减少对于实验效果的影响。

3.3 深度运动图的 LBP 特征提取

局部二值模式(Local Binary Pattern, LBP)[31][32]是 Timo Ojala 最先提出的，LBP 算子是一种非常优秀的图像局部纹理描述算子，它具有灰度不变性和旋转不变性这两方面优点，对于提取深度运动图这种深度图有着较为明显的优势。下面，我们将详细介绍 LBP 算子及本文所要应用的 LBP 算子。

3.3.1 原始 LBP 算子

原始 LBP 算子的基本思想是在一个 3*3 的窗口内，以中心点的像素值为阈值，分别比较中心点与其周围的 8 个相邻点，比中心值大的点置为 1，反

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

之置为 0。围绕中心点的 8 个点便可以组成八位二进制数，这 8 位二进制数的顺序可由使用者自定。最后，将得到的八位二进制数转换成十进制数，使用这个十进制数表示这个中心点的纹理信息，即为 LBP 值。下图 3-4 即为 LBP 算子的计算方法。

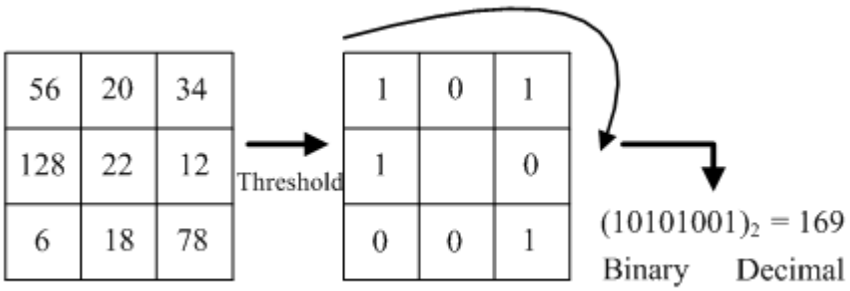

图 3-4 原始 LBP 算子计算方法

原始 LBP 算子只能在以 3*3 的固定窗口内进行纹理分析，无法进行扩展，这样就容易造成对不同频率和尺寸的图像纹理分析不够完善，造成许多纹理信息无法表达图像的真实情况。所以，将原始 LBP 算子进行扩展势在必行。

3.3.2 圆形 LBP 算子

原始 LBP 算子的局限性较大，其窗口大小固定，不适用于种类繁多的众多图像。因此，各位学者在此基础上又提出了许多改进和优化算法，以期提高 LBP 算子在不同尺度和纹理频率的图像上都可以获取出精准的纹理信息。Ojala 在其原有的方法上提出了扩展方法，不再局限于 3*3 的范围内，其使用半径为 r 的圆形窗口作为中心点的窗口半径，并可随意设置以 r 为半径的圆上等距分布的邻域点的数目 m。这极大的扩展了 LBP 算子描述图像纹理的能力。

设 g_c 为所要求得的 LBP 值，其坐标可设为 $(0,0)$，那么令 g_p 为以 g_c 为原点半径为 r 的 m 个相邻点中的一个邻域点，则 g_p 所生成的坐标为 $(-r\sin 2\pi p/m, r\cos 2\pi p/m)$，所示的如图 3-5。

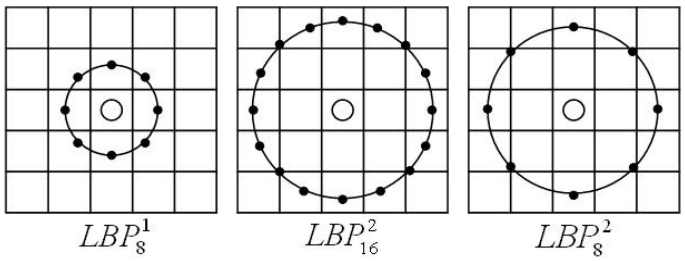

图 3-5 几种圆形 LBP 算子

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

从图 3-5 可知，有些邻域点并没有坐落在像素点的中心上，所以，所得的坐标并不都是整数。为了解决这一问题，我们将会使用双线性插值法来获取该位置的像素值。

3.3.3 灰度不变 LBP 算子

首先设 T 为图像纹理描述子，g_c为所要描述范围的中心点，其所要描述的范围为 r，g_c的邻域点个数为 m，则纹理描述子 T 可以使用在此范围内的像素灰度值的联合分布密度来表示，可由公式(3-2)描述

$$T = t(g_c, g_0, g_1, ..., g_{m-1}) \tag{3-2}$$

由于各像素点的值表达的是该像素点的亮度值，所以将g_c的 m 个领域点都减去g_c，可获得等式$T = t(g_c, g_0 - g_c, g_1 - g_c, ..., g_{m-1} - g_c)$，纹理描述本身描述的是像素之间的亮度关系，因此这并不会损失任何图像的纹理信息。接下来我们假设$g_m - g_c$与g_c是互相独立的，则公式 T 即可利用像素差值的联合分布来表示，即

$$T = t(g_c)t(g_0 - g_c, g_1 - g_c, ..., g_{m-1} - g_c) \tag{3-3}$$

在实际应用中，$g_m - g_c$与g_c并不是一定会确保相互独立的，因此公式 3-3 并不能保证一定相等。所以，只能用公式(3-4)大致表达，即

$$T \approx t(g_c)t(g_0 - g_c, g_1 - g_c, ..., g_{m-1} - g_c) \tag{3-4}$$

虽然这么做会造成图像纹理信息一定程度上的丢失，但利用公式(3-4)可以使得$t(g_c)$不再与邻域点产生关联，同时可以使得局部纹理的灰度值具有平移不变性。

$t(g_c)$是描述整幅图像在g_c点的亮度分布情况，说明g_c点的分布函数$t(g_c)$与g_c的邻域点没有相关性，即$t(g_c)$对于图像的纹理表述没有实际意义，可以将其忽略，即可将公式(3-4)可以表示为

$$T \approx t(g_0 - g_c, g_1 - g_c, ..., g_{m-1} - g_c) \tag{3-5}$$

在图像纹理信息频率较高或者在图像边沿时，公式(3-5)中某些维度的差分值会比较大，其它差分值会比较小。在图像纹理频率较低的情况下，公式(3-5)中的 m 维差分值都会较小或者趋于零。对于某个单一的点，则其在公式(3-5)任意一个方向上的差分值都会比较大。

通过公式(3-5)，我们可知$g_m - g_c$的差分值将不会受图像的光照改变和纹理信息变化而产生畸变，这便是我们所需的灰度平移不变性。如果图像像素值的大小同时扩大或者缩小，可知图像的纹理信息必然发生改变，但是差分值的符号却始终没有发生改变。所以，我们可以使用符号函数 sign 来取代

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$g_m - g_c$的差分符号。即

$$T \approx t(s(g_0 - g_c), s(g_1 - g_c), ..., s(g_{m-1} - g_c)) \tag{3-6}$$

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0. \end{cases}$$

我们使用与原始 LBP 类似的方法将 T 所表示的 m 个二进制数转换成十进制数，即对应位乘以2^m。可用公式(3-7)进行近似表达

$$LBP_{m,r} = \sum_{m=0}^{m-1} s(g_m - g_c) 2^m \tag{3-7}$$

所产生的十进制数即为 LBP 值，也可称之为 LBP 模式。当图像灰度保持不变或者同时发生变化时，图像上各个点的 LBP 值将维持不变。

3.3.4 旋转不变 LBP 算子

上一章节提出的灰度不变 LBP 算子所生成的 LBP 值具有唯一性，可以很好的描述图像的纹理信息。但是当图像发生旋转时，g_c的邻域点位置就会发生改变，g_c的 LBP 值也就会相应发生改变。为了增加 LBP 算子的旋转不变性，Maenpaa 提出了将所得 LBP 值的二进制序列按位进行位移，位移方向自定，并选取这些位移后的最小十进制数作为g_c的 LBP 值，用公式表达如下

$$LBP_{m,r}^{i} = \min\{ROR(LBP_{m,r}, i) \mid i = 0,1,...,m-1\} \tag{3-8}$$

式中$ROR(x,i)$表示为旋转函数，含义是将 x 循环位移 i(i<m)位。$LBP_{m,r}^{i}$表示 LBP 的旋转不变算子的最小值。本文将以中心点左上角的点作为起始点，按顺时针方向循环位移。如图 3-6 展示了旋转不变 LBP 算子的计算方法。

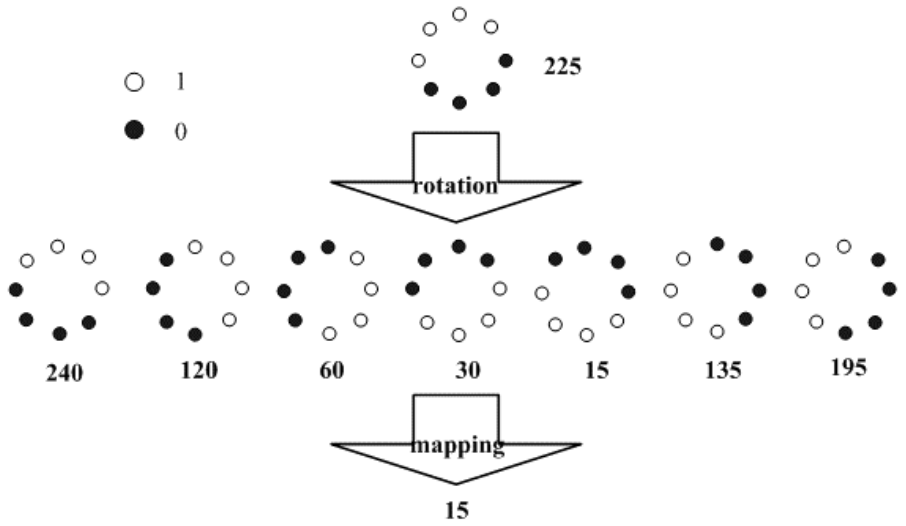

图 3-6 旋转不变 LBP 算子运算方法

旋转不变 LBP 算子使得在图像识别时，特征描述将不再受限于图像旋转，

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

可以使特征描述更加具有鲁棒性。虽然，这种算法会失去图像的方向信息，但是在很多图像分析与处理中，并不会有太多影响，所以这种方法是十分有效的。

3.3.5 深度运动图的 Uniform 模式 LBP 算子

从上述章节的介绍，LBP 算子的优势已经非常明显了，但随着邻域点数目的增加，产生的二进制数种类逐渐增多。如当半径 r=1，邻域点数 m=8 时，会有 $2^8=256$ 种二进制值，当半径 r=2，邻域点数 m=16 时，会有 $2^{16}=65536$ 种二进制值，可见随着半径 r 和邻域点数 m 的增加，二进制值的种数会按几何级数的形式不停增长，这会给计算机的计算速度和存储带来极大的灾难。鉴于此，Ojala 提出了 Uniform 模式，用于降低二进制数的种类个数。LBP 算子对于图像特征的描述并不是直接利用 LBP 算子所产生的图像进行特征描述，而是以 LBP 算子所计算的局部区域的各个 LBP 值的统计直方图作为纹理描述的特征向量，因此减少二进制数的种类数量可以使得直方图统计的各类二进制种类相对集中，避免纹理特征描述过于稀疏，不利于特征分类。

Uniform 模式的基本思想是鉴于实际图像应用中，有些 LBP 的二进制数值出现的频次数非常低，而有些数出现的频次却非常多，并且能够包含大多数的纹理信息。这些频次出现较多的二进制数，经分析后发现其二进制数从 0 变 1 或从 1 变 0 的次数一般不超过 2 次，根据此条规律我们将这些出现频次较高的 LBP 模式的二进制数称为 Uniform 模式。依据 Ojala 的描述定义，如 0000000,11111111 都属于 Uniform 模式。Ojala 提出了判断某一 LBP 模式是否是为 Uniform 模式的判别公式，如公式(3-9)所描述，

$$U(G_p)=|s(g_{p-1}-g_c)-s(g_0-g_c)|+\sum_{p=1}^{p-1}|s(g_p-g_c)-s(g_{p-1}-g_c)| \quad (3\text{-}9)$$

当所计算的 LBP 模式的 $U(G_p)$ 小于或等于 2 时，就可以称此 LBP 模式为 Uniform 模式。当所计算的 LBP 模式的 $U(G_p)$ 大于 2 时，我们统一将它们归为一类，称为混合模式类。如下图 3-7，白色表示 1，黑色表示 0。图中可以看出 Uniform 模式类占模式总数的大多数。

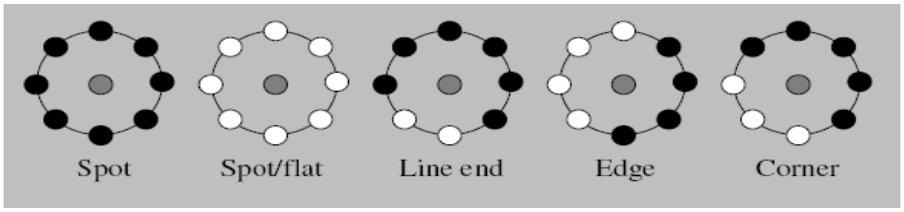

图 3-7 LBP 检测的基本纹理单元

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

使用 Uniform 模式可以极大的提高 LBP 算子的统计特性，以半径 r=1，邻域点数 m=8 的情况为例，LBP 的二进制数种类数量从原来的 256 种可以减少到 58 种，可见 LBP 模式种类数量骤减。这样可以使特征向量的维度种类大幅度下降，并减少噪音的高频成分，提高了计算速度并减少存储空间。

本文将使用 Uniform 模式的 LBP 算子对深度运动图的特征进行提取。如图 3-8 所示，展示了“双手挥动”的前视视角 DMM_f 深度运动图经过 Uniform 模式的 LBP 特征提取情况。

图 3-8 利用 Uniform 模式的 LBP 算子提取“双手挥动”的特征提取情况

从图 3-8 可以看出，经过 LBP 算子的 Uniform 模式所提取的特征识别度非常很高，已经将动作的基本轮廓表达清楚，这将对分类器的分类带来极高的识别效率。

一般通过得到的 LBP 值转换成 Uniform 模式即可，经 LBP 算子处理后的图像纹理特征一般使用直方图进行表达，即将 LBP 算子所计算的每个 block 块内的各个 Uniform 模式类别进行统计，并以直方图的形式表示该 block 块的纹理特征。为了将深度运动图的纹理特征更加清晰的表达出来，我们将 LBP 算子的 block 块相互交叠起来，每个 block 块交叠的大小为 block 块的一半，如图所示 3-9 所示，

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

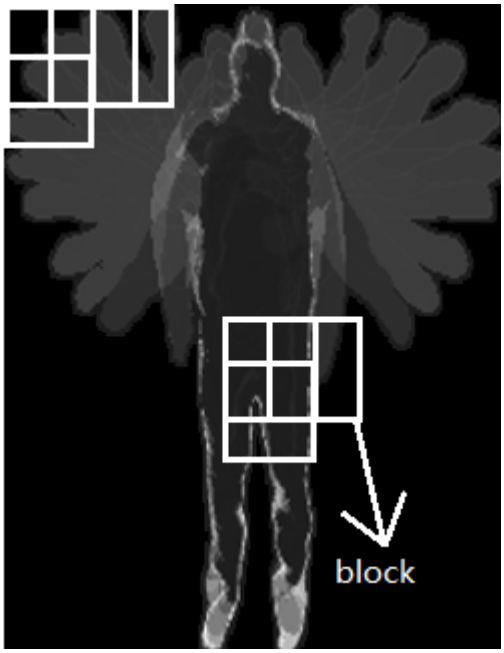

图 3-9 block 块的交叠形式

实验中，DMM_f，DMM_s 和 DMM_t 各自的 block 块大小分别为 25*27,25*25 以及 25*27。LBP 算子在每个 block 块中都会进行直方图统计，即统计该 block 块中各模式的数量，然后将各个 block 块的直方图顺次联结起来形成深度运动图的 LBP 特征。本文中将各个深度运动图的前视视角，侧视视角及俯视视角的 LBP 特征，h_{LBP}^{f}，h_{LBP}^{s}，h_{LBP}^{t} 顺序联结形成该动作的特征相量。如图 3-10 所示。

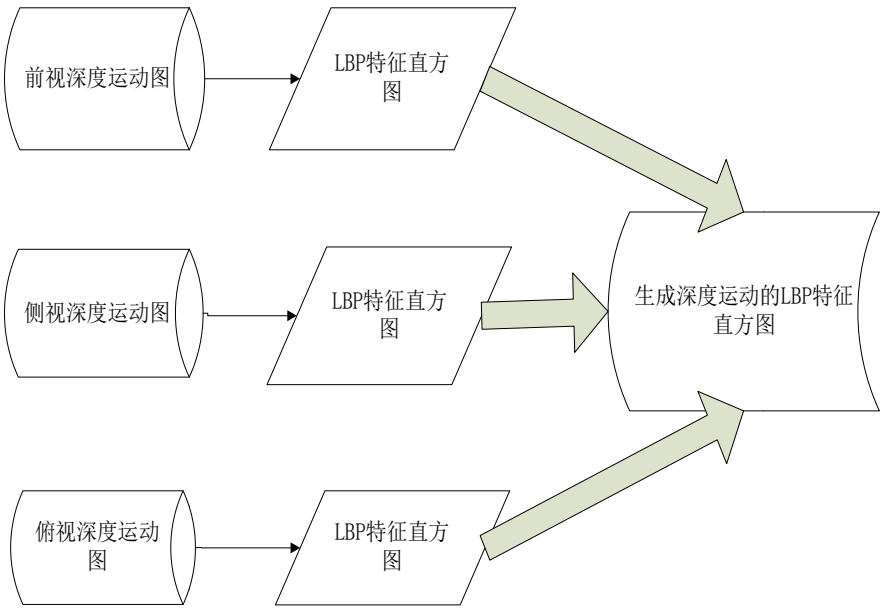

图 3-10 LBP 特征生成流程图

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

3.4 本章小结

本章首先介绍了 Kinect 的基本原理以及 Kinect 所获取的深度映射图。然后，介绍了本文将使用深度映射图的三维投影，形成识别所需的深度运动图。最后，本章介绍了 LBP 算子的基本理论，并使用 Uniform 模式对深度运动图进行特征提取，可以得到较为理想的动作特征。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

第 4 章 基于深度运动图的人体行为识别算法

4.1 稀疏表示方法

4.1.1 稀疏表示简介

稀疏表示最早是由 David Hubel 和 Toresten Wiesel 于上个世纪五十年代末，研究猫的视觉条纹皮层简单细胞感受野时所得，David Hubel 和 Toresten Wiesel 发现猫的视觉皮层 V1 区神经元的感受野能够对接收的视觉信息产生所谓的一种“稀疏表示”。这种“稀疏表示”本质上就是猫的视觉皮层对外界事物的一种降维处理，方便猫的大脑处理视觉信息。到了上个世纪八十年代末，学者 Michison 正式提出神经稀疏编码理论，并由英国牛津大学学者 E.T.Roll 正式引用到信息处理领域。在 1996 年，D.J.Field 和 B.A.Olshausen 在 Nature 上发表了一篇该方面的重要论文“Emergence of simple-cell receptive field properties by learning a sparse code for nature images”，这个实验结果，真正的从理论上证明了人类的视觉系统是可以使用最少的神经元来获取自然场景中的视觉图像信息。次年，D.J.Field 和 B.A.Olshausen 这两位学者正式提出了利用过完备基进行稀疏编码的算法理论。

进入新世纪，稀疏表示被称为信号处理领域近些年来十分流行的技术之一。同样，在图像去噪和图像复原等图像处理等相关领域中，稀疏表示亦取得了不错的成绩。Wright 等人[33]将稀疏表示应用到人脸识别中，即利用训练样本的特征向量组成字典，然后用字典线性稀疏的编码测试样本，最后计算得到最小的类别误差即可得知测试样本所属类别。其使用稀疏表示[34]的核心思想是使用由训练样本生成的过完备字典，并用过完备字典对测试样本进行稀疏编码，并计算出测试样本与稀疏编码后的残差，以其中最小残差的所属类别作确认测试样本的类别。

4.1.2 稀疏表示分类

首先，假设每一类人体行为的训练样本里含有 n_i 个人体行为特征样本，设矩阵 A_i 表示第 i 类人体行为训练样本里的人体行为特征，每个训练样本的特征维度为 m，即可由 $A_i=[v_{i,1},v_{i,2},...,v_{i,n_i}]\in R^{m\times n_i}$ 表示。设测试样本为 g，并设其所属类别为第 i 类，测试样本 $g\in R^m$。此时，g 可以利用第 i 类训练样本线性近似的表示，即可由式(4-1)，按如下表达，

$$g=\alpha_{i,1}v_{i,1}+\alpha_{i,2}v_{i,2}+\cdots+\alpha_{i,n_i}v_{i,n_i} \tag{4-1}$$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

式中 α 为标量，且 $\alpha \in R^m$。

若当测试样本 g 起初类别未知时，依上文理论推知，定义一个矩阵 A，使其包含数据库中所有人体行为类别的训练样本，类别种数设为 n，即可按公式(4-2)表示如下，

$$A = \left[A_1, A_2, ..., A_n \right] = [v_{1,1}, v_{1,2}, ..., v_{n,n_n}] \in R^{m \times n} \tag{4-2}$$

式中 A_i 表示一个第 i 类的训练样本，m 表示每个训练样本中人体行为的特征维度。在实验中，本文的 m 值则是指经 LBP 算子提取的特征 h_{LBP}^f， h_{LBP}^s，h_{LBP}^t 顺次组成的列向量，并通过 PCA 降维后的维度总和。测试样本 g 可以利用训练样本的稀疏线性组合来进行表示，按公式(4-3)表示，即

$$g = Ax_0 \tag{4-3}$$

x_0 是一个 $n \times 1$ 的训练样本的系数向量，若测试样本 g 属于第 i 类，则系数向量中只有第 i 类的位置是非零的，可按下式情形进行表达，即 $x_0 = [0, ..., 0, \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,v_i}, 0, ..., 0]^T \in R^n$。

当矩阵 A 里的训练样本足够多时，系数向量 x_0 里所包含的 0 元素个数也就越多，说明 x_0 的稀疏特性越好，依据稀疏表示理论，矩阵 A 可视为过完备字典，有利于判断测试样本 g 所属类别。依据基于稀疏表示分类方法[35] x_0 的求解问题可以利用 L0 范式进行优化，以求取 x_0 的稀疏解，即可按公式(4-4)按如下方式计算，

$$\hat{x}_0 = \arg\min \| x \|_0 \quad s.t. Ax = g \tag{4-4}$$

考虑到在真实环境中，图像采集会产生一定的噪声，仅仅通过我们自身所使用的训练样本无法精准表达测试样本，所以本文将公式进行一定修改，生成新的公式(4-5)，即

$$\hat{x}_0 = \arg\min \| x \|_0, \quad s.t. \| g - Ax \|_2 < \varepsilon \tag{4-5}$$

ε 为常量，表示可容忍的最大噪声幅值，一般可视为稠密噪声且幅值较小。

但是在稀疏表示中，上式很难通过一般手段直接求解 x_0 的值，这是一个 NP 难问题。随着近些年压缩感知[36]技术的发展，如果过完备字典 A 满足 RIP 条件，上式的 L0 范式最小化问题就可以转化为 L1 范式最小化，鉴于 L1 范式最小化拥有良好的凸优化属性，所以使用 L1 范式最小化，来计算测试样本 g 的系数向量 $\hat{x}$，公式(4-5)可转换为公式(4-6)，如下所示，

$$\hat{x}_1 = \arg\min \| x \|_1, \quad s.t. \| g - Ax \|_2 < \varepsilon \tag{4-6}$$

再将公式(4-6)进行整合，可得公式(4-7)，如下所示，

$$\hat{x} = \arg\min_x \left\{ \| g - Ax \|_2^2 + \theta \| x \|_1 \right\} \tag{4-7}$$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

θ是正则化尺度参数，它是用来平衡稀疏项的影响。

在理想状态下，稀疏解 $\hat{x}$ 里非零元素出现的位置正好对应过完备字典 A 中与测试样本 g 相同类别的位置上，以此就可以识别出测试样本 g 所属类别。但是在实际情况中往往事与愿违，由于人体行为之间的某些动作具有高度的相似性以及图像采集中所包含的一些噪声，这些都会给识别过程带来相当的影响，造成非 $\hat{x}$ 中非对应类别中也会出现非零元素出现误判的情况。因此，使用人体行为的测试样本与训练样本之间的重构残差，求取重构残差中的最小残差作为判断测试样本类别的依据。可定义 δ_i 为 $R^n \to R^n$ 的映射函数，j 为第 i 类人体行为中第 j 个实验者所做的动作。可按公式(4-8)表示如下，

$$\delta_i(x) = \begin{cases} x_{i,j}, j = 1,2,...,n_i \\ 0, \text{其他} \end{cases} \tag{4-8}$$

δ_i 的含义是将人体行为特征中第 i 类行为的所有训练样本的特征元素保留，其它特征元素都清零。则测试样本 g 与第 i 类训练样本的重构公式(4-9)表达如下，

$$\hat{g}_i = A \times \delta_i(\hat{x}_1) \tag{4-9}$$

那么测试样本的重构误差为 $e_i = \| g - \hat{g}_i \|_2$ ，即可按公式(4-10)表达如下，

$$e_i = \| g - A \times \delta_i(\hat{x}_1) \|_2 \tag{4-10}$$

当求得测试样本与训练样本中各类人体行为的重构误差后，重构误差大的，说明测试样本与该类训练样本差别很大，不应属于此类，若重构误差小，则说明测试样本与这类训练样本差别较小，应属于该类动作。

利用公式(4-11)即可计算测试样本 g 所属类别，即

$$class(g) = \arg\min_i \{e_i\} \tag{4-11}$$

得出测试样本 g 所属类别。

4.2 Tikhonov 正则化方法

4.2.1 Tikhonov 正则化简介

Tikhonov 正则化[37][38][39]，是上个世纪前苏联院士 Andrey Tikhonov 研究所得成果，并以其姓氏进行命名。ill-posed 问题(不适定问题)是图像处理、量子力学等各种学科中都经常遇到的问题之一，为解决 ill-posed 问题，解决方法多以正则化为主。其核心思想是利用现有的先验知识作为依据，对 ill-posed 问题添加适当的约束条件项，使 ill-posed 问题转变成 well-posed 问题(适定问题)，利用 well-posed 问题的解最大程度的去接近 ill-posed 问题的真

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

实解。

4.2.2 Tikhonov 正则化

假设已知矩阵 A 和向量 b，计算 $Ax=b$ 中向量 x。一般的方法是使用最小二乘法进行计算，即按公式(4-12)表示，

$$\min \| Ax-b \|_2^2 \tag{4-12}$$

但是，如果 x 不存在或者解不唯一时，就会造成我们上文所讲的 ill-posed 问题。在线性代数中这种问题的一般求解方法为正则化，即通过先验知识来增加限制性条件来缩小 x 的范围，最后得到趋近于正解的近似解。

Tikhonov 正则化方法作为常规的正则化方法，一般是将 L2 范式的正则项添加入公式(4-12)中，形成目标函数(4-13)，即

$$T(x)=\| Ax-b \|_2^2+\lambda \| Lx \|_2^2 \tag{4-13}$$

求得公式 $\arg\min T(x)$ 的值即为公式(4-12)的优化解，参数 λ 为正则化参数，其值为正的常量，依据文献[40]，λ 一般在 0.1~0.3 效果较好，在实验中，本文使用 λ=0.25。公式中的 L 就是 Tikhonov 矩阵，标准的 Tikhonov 正则化中 L 一般有两种算子矩阵，分别是 $L=\nabla$ 或 I。依据文献[40]的方法，本文的 Tikhonov 矩阵使用对角矩阵，所表达的含义是，将测试样本与训练样本里相似的动作权重适当增加，不相似的动作权重适当减少。Tikhonov 矩阵便可形成一个权重对角矩阵 L，公式(4-14)表达如下，

$$L=\begin{pmatrix} \| b-h_1 \|_2 & & 0 \\ & \ddots & \\ 0 & & \| b-h_n \|_2 \end{pmatrix} \tag{4-14}$$

依据[41]方法可知，因为本文使用 Tikhonov 的标准方法解决问题，所以 x 值的求取可以转化为利用公式(4-15)进行计算，即

$$(A^TA+\lambda L^TL)x=A^Tb \tag{4-15}$$

来进行求解。

4.3 L2 范式正则化协同表示和行为分类

4.3.1 协同表示简介

协同表示[42][43]是以稀疏表示为基础研究得来，稀疏表示的核心是利用过完备字典中与测试样本同类的训练样本来稀疏的表示测试样本，然后通过求取训练样本重构误差的最小值来识别测试样本所属类别。但是在现实环境中，

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

我们所获取的人体行为数据库里各类动作的种类和数量毕竟有限，并且为了增加识别的准确性和鲁棒性，人体行为特征向量的维度一般不会太低，所以构成的字典往往是欠定的，这样往往就无法满足稀疏表示里过完备的要求。那么，该如何满足生成过完备字典的要求呢？通过对人体行为数据库的研究可知，人体在做各类动作时是具有一定的相似度的，如果可以利用这种相似度协同表示测试样本，便可以极大的丰富训练样本的过完备性。比如，人体做蹲下的动作和做坐下的动作之间往往会有一定的相似程度。

正因如此，协同表示逐步受到学者的重视。本文协同表示的核心思想是利用字典中各类动作间具有一定相似度这一特性，利用字典中所有训练样本来共同表示测试样本。最后，利用残差值来判断测试样本所属类别。

4.3.2 L2 范式协同表示和行为分类

首先，设训练样本为 $A=\left[A_1, A_2, ...A_i, ..., A_n\right]$，再设 A_i 表示字典中第 i 类动作的训练样本，因为人体行为的特征维度一般会多于该类动作的训练样本数量，即造成 A_i 是欠定的。这时利用 A_i 的稀疏表示测试样本 g，必然会造成残差过大，使得分类不准确。但是，由于人体行为的某些动作都做具有一定的相似度，比如 Pickup throw 和 High throw 在深度运动图中具有一定的相似度。这样可以在很大程度上解决训练样本数量比特征维度少的情况。所以，本文将利用 L2 范式协同表示的方法来应对此问题。

依据[43]中的讨论，我们进一步说明了 L2 范式协同表示[44]相对于稀疏表示的优势。为此，我们首先使用稀疏表示的公式(4-7)，并移除 L1 范式的稀疏项，使公式(4-7)转换成最小二乘法问题，生成公式(4-16)即

$$\hat{x}=\arg\min_{x}\| g-Ax\|_2^2 \tag{4-16}$$

从图 4-1 可以看出，g 在 X 的张成空间的垂直投影为 $\hat{g}$，$\hat{g}$ 可以利用如下公式(4-17)表示,

$$\hat{g}=\sum_i A_i\hat{x}_i \tag{4-17}$$

在稀疏表示中，使用类别计算公式计算 $e_i=\| g-\hat{g}_i\|_2$，依据此公式，可推导出公式(4-18)如下，即

$$e_i=\| g-A_i\hat{x}_i\|_2^2 \tag{4-18}$$

计算测试样本与各类训练样本间的重构残差[45]。进一步推导公式(4-18)可得公式(4-19)，如下

$$e_i=\| g-A_i\hat{x}_i\|_2^2=\| g-\hat{g}\|_2^2+\| \hat{g}-A_i\hat{x}_i\|_2^2 \tag{4-19}$$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

由于$\| g-\hat{g} \|_2^2$是一个常量，所以e_i的值由$\| \hat{g}-A_i\hat{x}_i \|_2^2$所决定。由此，便很容易就可以得到公式(4-20)，如下所示

$$e_i^* = \| \hat{g}-A_i\hat{x}_i \|_2 \tag{4-20}$$

利用公式(4-20)计算判断测试样本 g 所属类别。

再设$\chi_i = A_i\hat{x}_i$，表示χ_i为X_i张成空间里的一个向量。$\overline{\chi}_i = \sum_{j\neq i} A_j x_j$，表示$A_j$中所有类别(除去第 i 类)张成空间里的一个向量。从图中 4-1 可以看出，

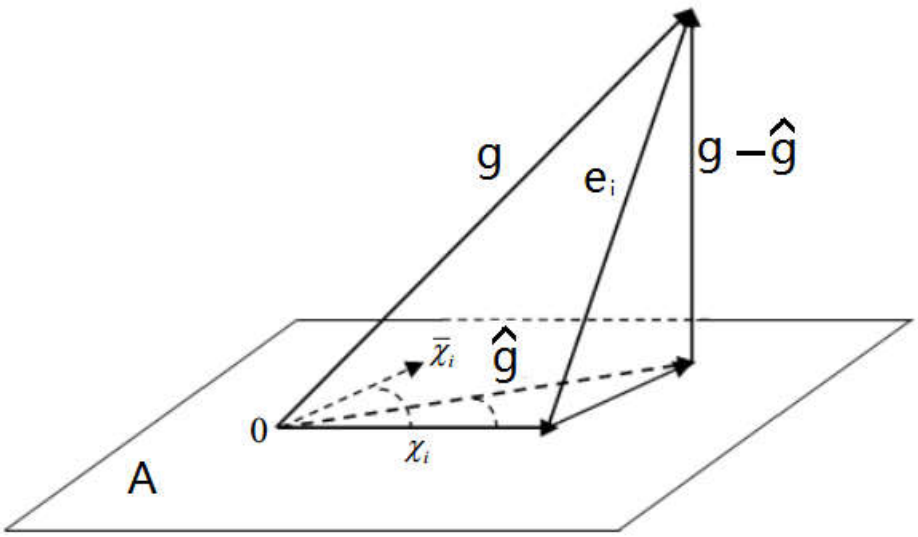

图 4-1 g 在 A 上表示的几何图解

$\overline{\chi}_i$与$\hat{g}-A_i\hat{x}_i$是相互平行的，所以可得下式(4-21)

$$\| \hat{g} \|_2 \times \sin(\hat{g},\chi_i) = \| \hat{g}-A_i\hat{x}_i \|_2 \times \sin(\chi_i,\overline{\chi}_i) \tag{4-21}$$

式中$(\chi_i,\overline{\chi}_i)$表示的是$\chi_i$与$\overline{\chi}_i$之间的夹角。$(\hat{g},\chi_i)$表示的是$\hat{g}$与$\chi_i$之间的夹角。从空间向量的角度，我们可得到公式(4-22)，如下

$$e_i^* = \frac{\sin^2(\hat{g},\chi_i)\| \hat{g} \|_2^2}{\sin^2(\chi_i,\overline{\chi}_i)} \tag{4-22}$$

从公式(4-22)中，我们可以得知，若要判断测试样本 g 是否属于第 i 类，不仅需要考虑$(\hat{g},\chi_i)$之间夹角范围，而且也需要考虑$(\chi_i,\overline{\chi}_i)$之间的夹角范围。说明误差的大小要从这两个方面共同决定，也进一步说明，协同表示确实可以充分利用训练样本的价值，并提高对人体行为特征的识别效果，同时利用协同表示也可以增加分类效果的鲁棒性。

在稀疏表示中，我们一般使用 L1 范式作为正则化项来求取我们所需的解。但是随着字典 A 的数量逐渐增多时，相较于使用 L2 范式，L1 范式的计算时长和复杂度都会提高很多，所以，协同表示中一般使用 L2 范式进行正则化求解，并且拥有与 L1 范式几乎相同的准确率。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

设 Y_p 为一个测试样本，设已训练完成的字典 $A=[h_1,h_2,...,h_n]$，n 表示所用的训练样本数量。鉴于字典 A 可能是欠定的，依据上文所提的 L2 范式协同表示求解 $Y_p = Ax$。利用最小二乘法，同时使用 L2 范式正则化 x，即可得到公式(4-23) 如下

$$\hat{x} = \arg\min \| Y_p - Ax \|_2^2 + \lambda \|x\|_2^2 \tag{4-23}$$

式中 λ 为正则化参数，本文将其设为 0.25。为了进一步增加公式(4-23)的正则化效果，得到更优解，本文使用 Tikhonov 正则化[46][47]方法来优化系数向量 x 中各类系数的权重，则可得公式(4-24)如下，

$$\hat{x} = \arg\min_x \left\{ \| Y_p - Ax \|_2^2 + \lambda \| Lx \|_2^2 \right\} \tag{4-24}$$

λ 表示正则化参数[48]，L 表示 Tikhonov 矩阵，L 的表达式按式(4-25)表达，如下

$$L = \begin{pmatrix} \| Y_p - h_1 \|_2 & & 0 \\ & \ddots & \\ 0 & & \| Y_p - h_n \|_2 \end{pmatrix} \tag{4-25}$$

对角矩阵 L 表达的含义为测试样本 Y_p 与字典 A 中各类特征向量的距离，在后续的系数向量计算中，对角矩阵 L 有利于将与测试样本 Y_p 与字典 A 中相似的动作给予较高的权重，对于不相似的动作给予较低的权重。

通过对公式(4-23)的推导，可得向量系数 $\hat{x}$ 的计算公式(4-26)，即

$$\hat{x} = \left(A^T A + \lambda L^T L \right)^{-1} A^T Y_p \tag{4-26}$$

当利用测试样本 Y_p 计算出系数向量 $\hat{x}$ 后，即可利用公式(4-10)，按公式(4-27)计算测试样本的重构残差[49]，

$$e_j = \| Y_p - A_j \widehat{x_j} \|_2, \quad j \in (1,2,...,n) \tag{4-27}$$

求出测试样本 Y_p 与字典 A 中各类动作之间的重构残差值后，利用公式(4-11)，计算测试样本所属类别。即可按公式(4-28)计算 Y_p 所属类别，

$$class(Y_p) = \arg\min_j \{ e_j \} \tag{4-28}$$

即可求出 g 的所属类别。

那么，本文实验中所使用的准确率的运算公式为识别准确的动作除以测试样本中所有动作数量，即公式(4-29)，

$$\text{准确率} = \frac{\text{识别准确的动作数量}}{\text{测试样本中所有动作数量}} \tag{4-29}$$

下面，我们将列出本文所提的算法流程图，如图 4-2

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

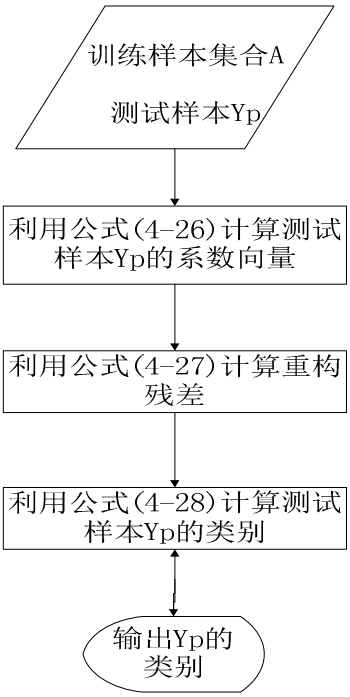

图 4-2 本文算法流程图

4.4 本章小结

本章主要利用稀疏表示和 Tikhonov 正则化矩阵作为基础，引出本文提出的 L2 范式正则化协同表示方法。将稀疏表示的稀疏项使用 L2 范式代替，同时利用 Tikhonov 正则化矩阵计算测试样本与字典中的特征向量间的距离，形成对角矩阵 L，并利用对角矩阵 L 计算测试样本的系数向量。然后利用稀疏表示中的类别计算方式，计算测试样本所属类别。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

第 5 章 实验过程、结果与分析

5.1 实验环境与数据库简介

为了对本文所提算法进行仿真实验，我们使用 Windows7 操作系统，并使用 Matlab2015b 作为仿真环境进行仿真实验。本文使用 MSRAction3D 人体行为数据库作为标准数据库进行对比实验。

MSRAction3D 数据库是由 Kinect 传感器所获取的人体行为深度图像。文章所使用的深度图像分辨率为 320*240。MSRAction3D 数据库中包含有 20 种动作类别，分别由十位实验者做出这 20 种动作类别，每位实验者分别将这 20 个动作做 2 到 3 次，而且每个实验者在做相同动作时，动作的幅度和频率都会略有不同，其目的在于丰富 MSRAction3D 数据库中类内样本的丰富性。MSRAction3D 数据库中包含的动作的总数量为 557 个。下图 5-1 为 MSRAction3D 数据库中几类动作的深度映射图。

图 5-1 MSRAction3D 数据库的深度图像示例

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

本文参照相关论文的对比方法，提出了两种实验对比方式。

实验一：将 MSRAction3D 数据库分成三个子集数据库，分别是 ActionSet1(AS1)，ActionSet2(AS2)和 ActionSet3(AS3)。分类情况如表 1 所示。从表 1 可知，ActionSet1(AS1)和 ActionSet2(AS2)里的动作相对较为简单，但其动作间的相似程度较高。而 ActionSet3(AS3)里的动作则复杂度相对较高。这种分类可以较为准确的评价测试算法的鲁棒性和泛化能力。同时，对于每个动作子集我们使用三种测试方法，Test One 是将动作子集中的三分之一作为训练集，剩余的作为测试集；Test Two 是将动作子集中的三分之二作为训练集，其余作为测试集；Test Three 又称为 Cross Test，它是将子集中一半实验者所做的动作作为训练集，另一半作为测试集，一般我们会将实验者 1，3，5，7，9 作为训练集，实验者 2，4，6，8，10 作为测试集。

表 1 MSRAction3D 数据库的子数据库分类情况

Action set1(AS1)	Action set2(AS2)	Action set3(AS3)
Horizontal wave(2)	High wave(1)	High throw(6)
Hammer(3)	Hand catch(4)	Forward kick(14)
Forward punch(5)	Draw x(7)	Side kick(15)
High throw(6)	Draw tick(8)	Jogging(16)
Hand clap(10)	Draw circle(9)	Tennis swing(17)
Bend(13)	Two hand wave(11)	Tennis serve(18)
Tennis serve(18)	Forward kick(14)	Golf swing(19)
Pickup throw(20)	Side boxing(12)	Pickup throw(20)

实验二：本文将不划分 MSRAction3D 数据库，而是把数据库中 1，3，5，7，9 号实验者所做的动作作为训练集。2，4，6，8，10 号实验者所做的动作作为测试集。这样实验二的难度会大大超过实验一，也就更能说明被测算法的鲁棒性和泛化能力。

5.2 实验结果展示与分析

5.2.1 LBP 算子的确定

首先，本文将利用 MSRAction3D 数据库作为基准，以实验二的方式确定 LBP 算子中参数半径 r 和环形邻域 m 的个数。获得了如表 2 所示的识别准确率的结果。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 2 LBP 算子各个参数的识别率

	r=1	r=2	r=3	r=4	r=5	r=6
m=4	0.901	0.894	0.890	0.879	0.886	0.875
m=6	0.901	0.897	0.894	0.890	0.886	0.894
m=8	0.894	0.894	0.890	0.890	0.890	0.879
m=10	0.901	0.894	0.890	0.894	0.879	0.875

从表 2 可知，当半径 r=1，m=4，6 或者 10 时，其准确率都是相同的。所以，在此基础上本文将从特征提取速度方面考虑可选取的 LBP 参数。下面的表 3 为，当 r=1 时，不同环形邻域个数下特征提取速度。

表 3 特征提取速度

	r=1
m=4	1.105ms
m=6	1.974ms
m=8	2.190ms
m=10	3.171ms

依据表 3 的结果，本文将使用 r=1，m=4 作为实验的特征提取参数，所得的 LBP 特征 h_{LBP}^{f}，h_{LBP}^{s}，h_{LBP}^{t} 的维度分别为 315，525 和 225。但是，过多的维度可能会造成信息冗余和计算效率过差等缺点。因此，本文会将所获取的 LBP 特征利用主成分分析(PCA)[50]进行降维处理，并保留 95%的特征向量。

5.2.2 实验一

接下来，我们使用如表 1 数据库作为基准，将本文方法分别与主流的方法如 Bag of 3D Points[51]， DMM-HOG，HOJ3D[52]，Space-Time Occupancy Patterns[53]进行准确率评价。如表 4 所示。

表 4 实验一中 Test One 各种算法识别率对比情况

	Test One			
	AS1	AS2	AS3	AVG
Bag of 3D Points	0.895	0.890	0.963	0.916

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

DMM-HOG	0.973	0.922	**0.980**	0.958
HOJ3D	**0.985**	0.967	0.935	0.962
Space-Time Occupancy Patterns	0.982	0.948	0.974	0.968
本文算法	0.978	**0.974**	0.973	**0.975**

表 5 实验一中 Test Two 各种算法识别率对比情况

	Test Two			
	AS1	AS2	AS3	AVG
Bag of 3D Points	0.934	0.929	0.963	0.942
DMM-HOG	0.987	0.947	**0.987**	0.974
HOJ3D	0.986	0.972	0.949	0.972
Space-Time Occupancy Patterns	**0.991**	0.970	**0.987**	0.983
本文算法	0.981	**0.987**	0.986	**0.984**

表 6 实验一中 Cross Test 各种算法识别率对比情况

	Cross Test			
	AS1	AS2	AS3	AVG
Bag of 3D Points	0.729	0.719	0.792	0.747
DMM-HOG	**0.962**	0.841	**0.946**	0.916
HOJ3D	0.880	0.855	0.636	0.790
Space-Time Occupancy Patterns	0.847	0.813	0.884	0.848
本文算法	0.952	**0.910**	0.901	**0.921**

从表 4 可以看出，在 Test One 中，本文方法的平均准确率最高，且比 Space-Time Occupancy Patterns 提高了接近 1%。本文方法在这三个动作子集中的识别稳定度也是最占优势的。在 Test Two 中，在平均识别率上，本文方法虽然仅比 Space-Time Occupancy Patterns 高 0.1%左右，但从三个动作子集的整体情况看，DMM-HOG，HOJ3D，Space-Time Occupancy Patterns 和本文方法整体的识别率不仅都高达 95%以上，且整体稳定度都很均衡，可见 Test Two 的识别难度并不大。在 Cross Test 中，本文方法依然有着不错的表现，平均识别率比 DMM-HOG 的准确率平均提高了大约 0.5%。从三个动作子集的识别稳定度方面考虑，本文方法依然表现优异，AS1，AS2，AS3 都

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

将平均识别准确率超过了 90%，并且各个实验中的识别准确差别不大，识别稳定度表现良好。

接下来，我们将列举出本文方法在实验一各种情况下的混淆矩阵图，进一步分析本文算法的识别情况。

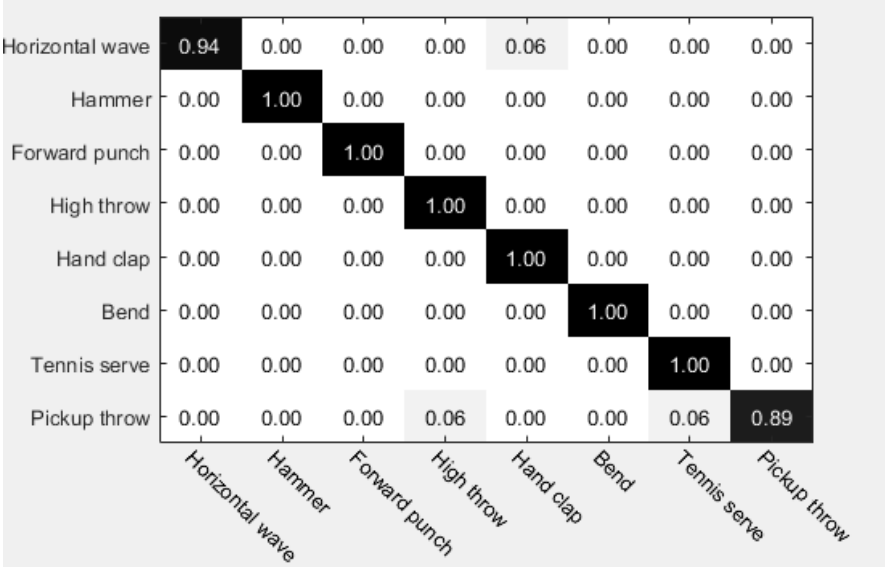

图 5-2 Test One 条件下 AS1 的混淆矩阵

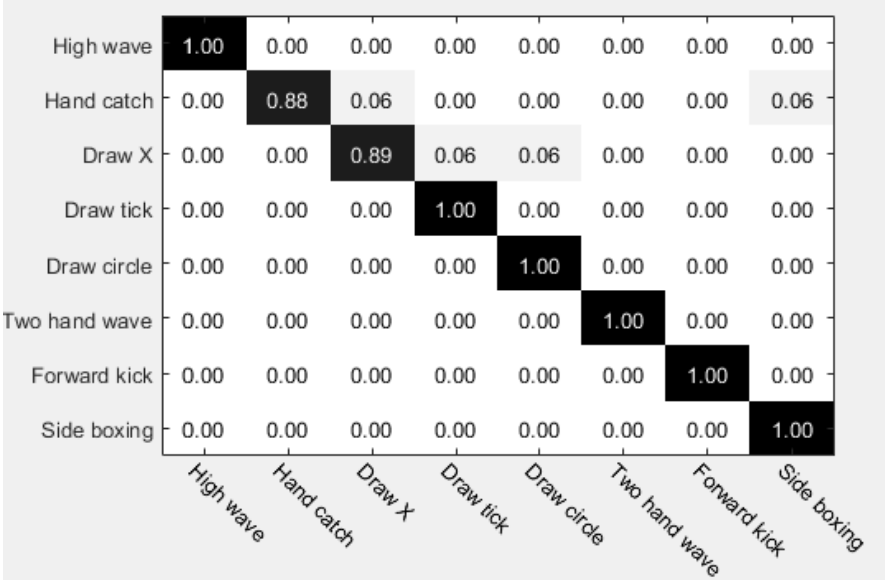

图 5-3 Test One 条件下 AS2 的混淆矩阵

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

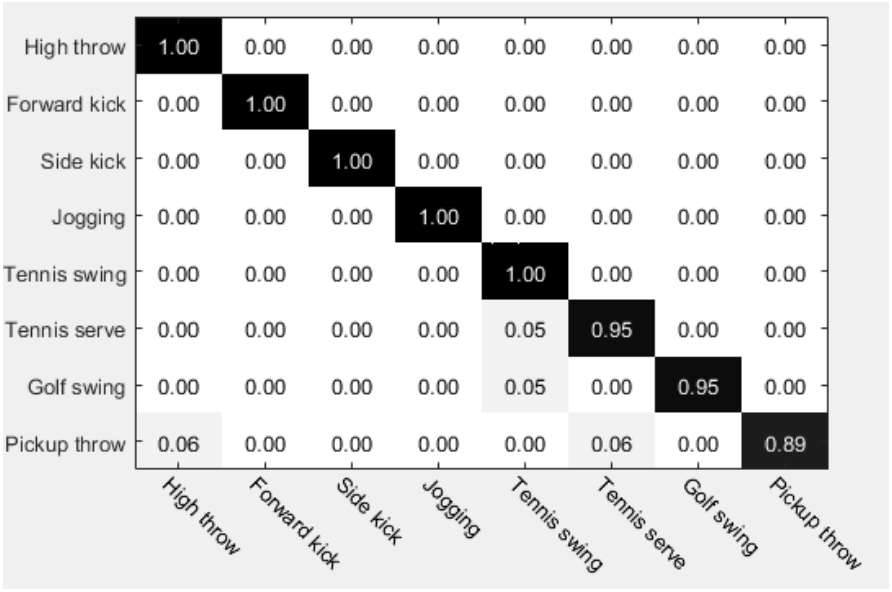

图 5-4 Test One 条件下 AS3 的混淆矩阵

从 Test One 的 AS1 中可以看出，除 Horizontal wave 与 Pickup throw 未做到完全正确外，其余动作识别率都完全正确。在 AS2 中，除 Hand catch 和 Draw X 外，其余动作都识别成功。在 AS3 中，只有 Tennis serve，Golf swing 和 Pickup throw 没有完全识别成功。

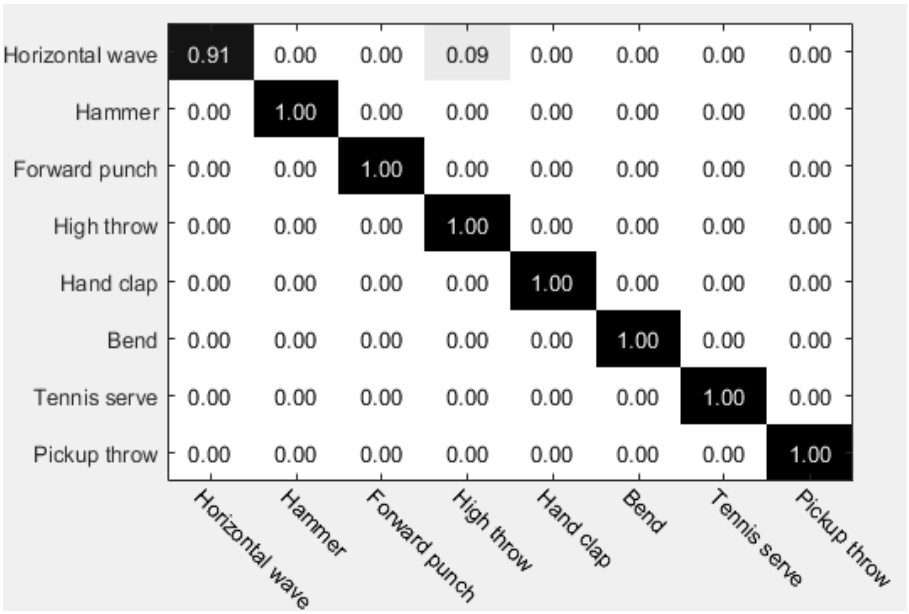

图 5-5 Test Two 条件下 AS1 的混淆矩阵

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

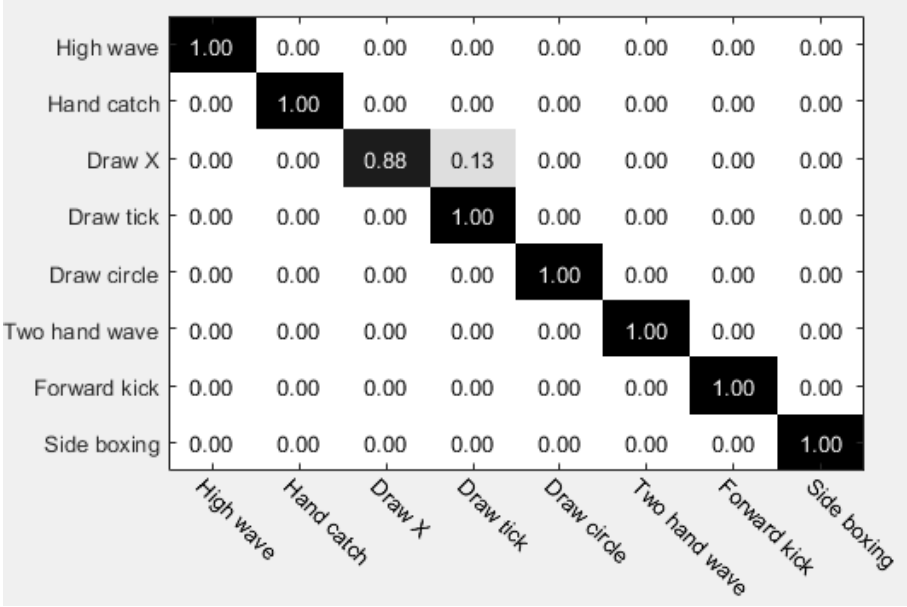

图 5-6 Test Two 条件下 AS2 的混淆矩阵

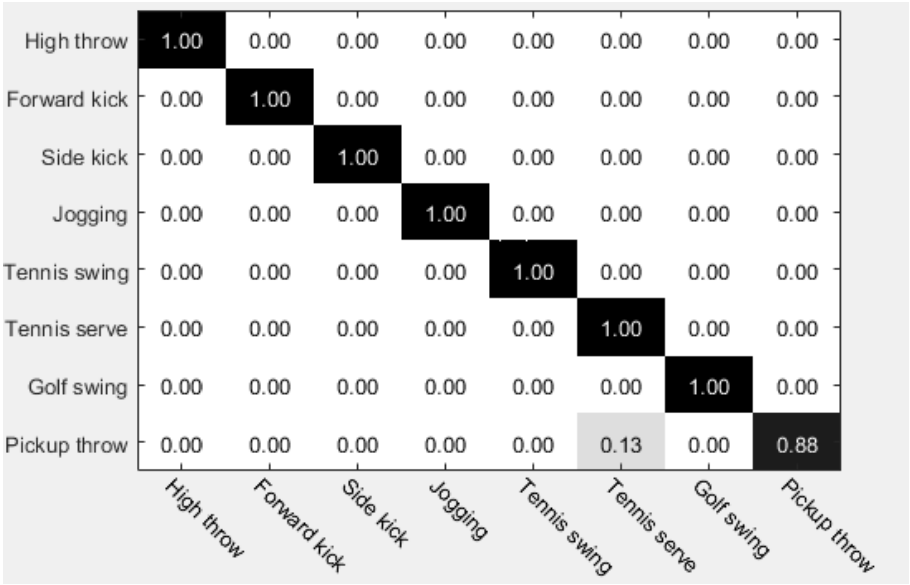

图 5-7 Test Two 条件下 AS3 的混淆矩阵

在 Test Two 条件下，AS1 数据库中只有 Horizontal wave 没有完全识别成功，其余动作都识别成功。在 AS2 中，只有 Draw X 没有完全识别成功，其余都成功识别出来了。在 AS3 中，也是只有 Pickup throw 没有完全识别成功，其余都成功识别出来了。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

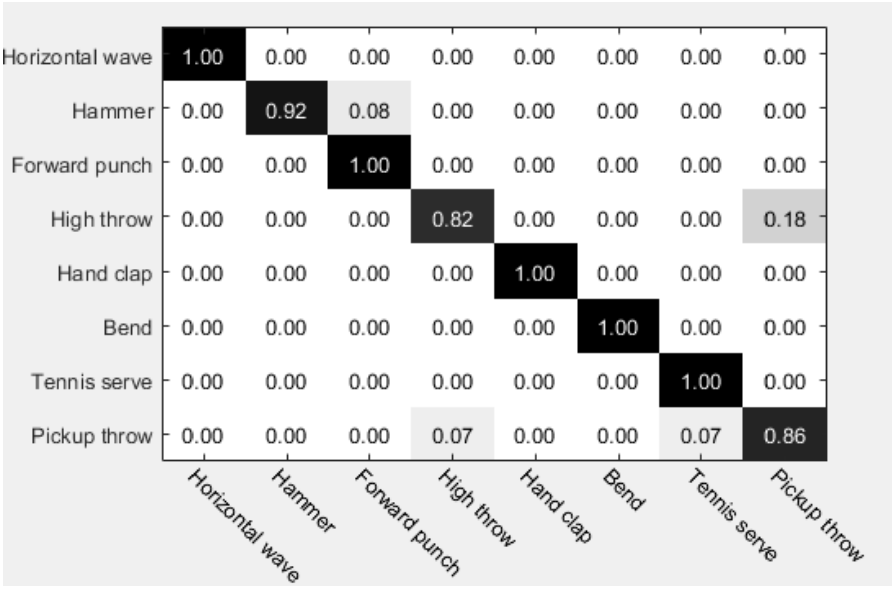

图 5-8 Cross Test 条件下 AS1 的混淆矩阵

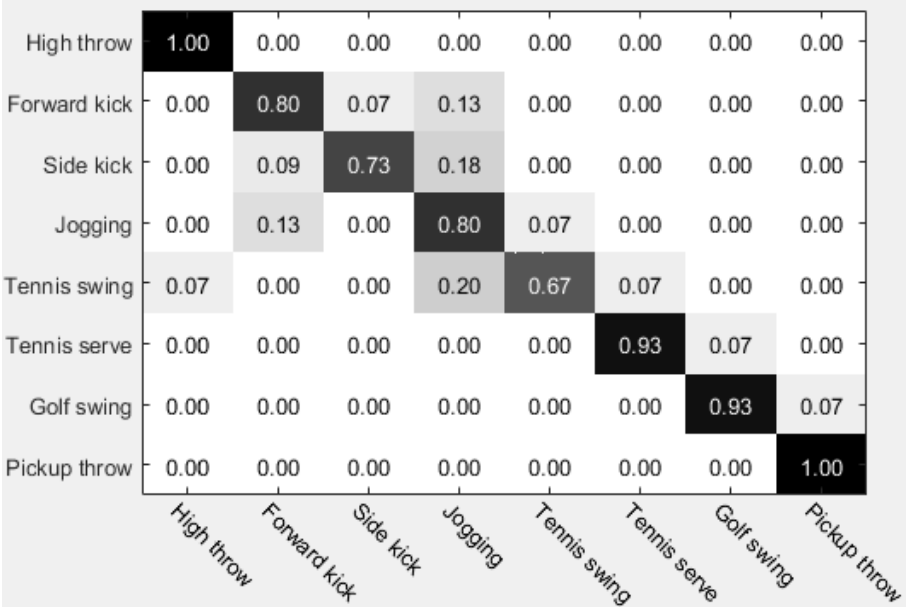

图 5-9 Cross Test 条件下 AS2 的混淆矩阵

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

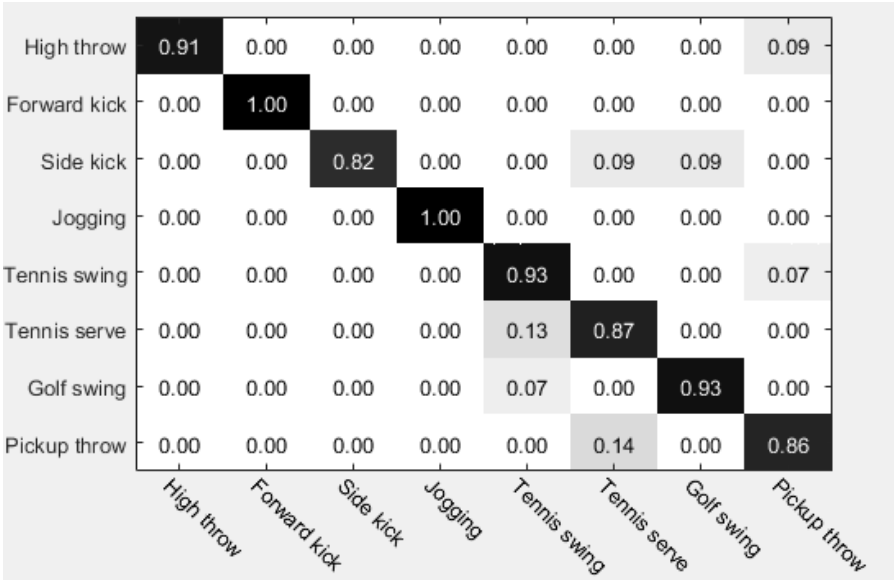

图 5-10 Cross Test 条件下 AS3 的混淆矩阵

在 Cross Test 条件下，整体效果差于 Test One 和 Test Two。相比较之下，在 AS1 中，Hammer，High throw，Pickup throw 没有完全识别成功。在 AS2 中，除了 High throw 和 Pickup throw 全部识别成功，其他均没有完全识别出来，其中 Tennis swing 稍差，准确率只有 67%，但是其他动作识别完成度相对来说不错。在 AS3 中，也只有 Forward kick 和 Jogging 完全识别成功，其他动作也没有完全识别出来，但识别的准确度还是不错的，Side kick 识别度相对较低，但也达到了 82%，说明识别效果不错。

5.2.3 实验二

在实验二的环境中，本文算法将与 DMM-HOG，Random Occupancy Patterns[54]和 Actionlet Ensemble[55]相比较，其算法准确率如表 5 所示。

表 7 实验二中各个算法识别率对比情况

所使用的算法	识别率
DMM-HOG	85.5%
Random Occupancy Patterns	86.5%
Actionlet Ensemble	88.2%
本文方法	90.1%

可见，本文算法比 DMM-HOG 和 Random Occupancy Patterns 准确率分别高出接近 5%和 4%，比 Actionlet Ensemble 相比准确率也高出了接近 2%。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

可见本文算法在实验二这种严苛环境下依然有着不错的鲁棒性和泛化能力。

接下来，我们将以 Random Occupancy Patterns，Actionlet Ensemble 和本文算法的混淆矩阵图，进一步分析各自算法的识别情况。

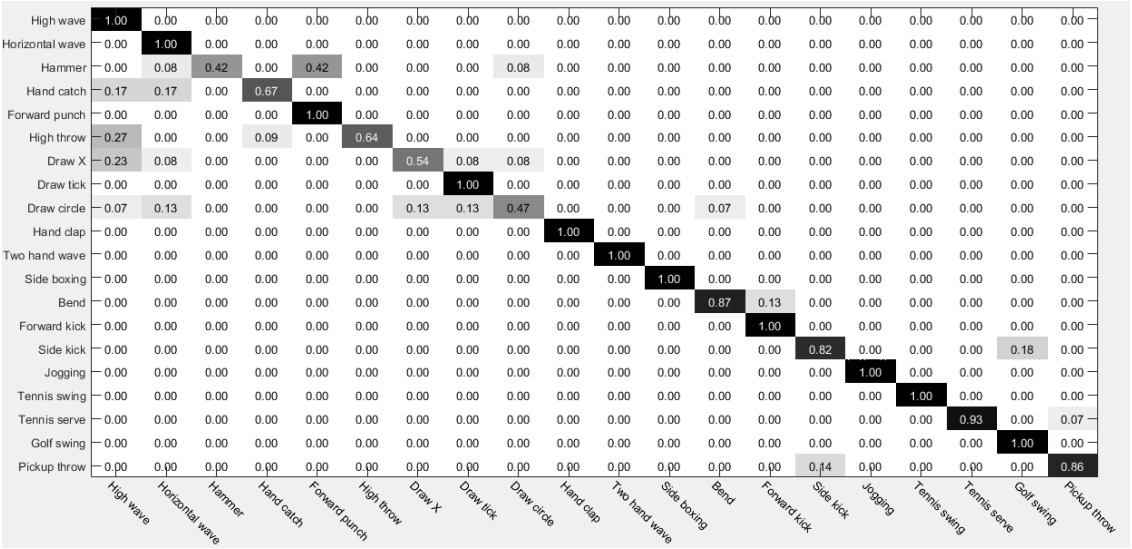

图 5-11 实验二条件下 Random Occupancy Patterns 的混淆矩阵

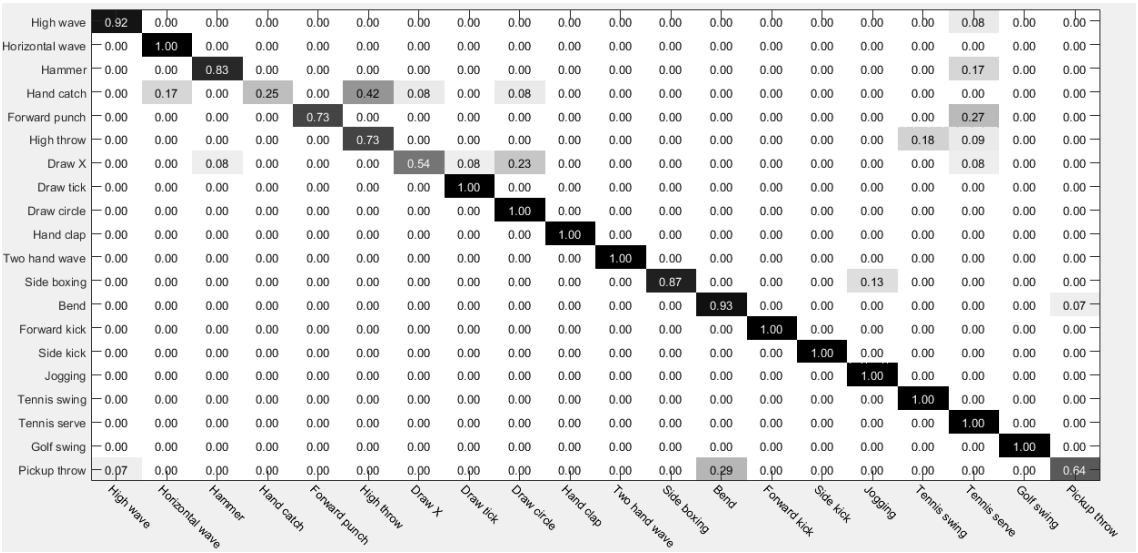

图 5-12 实验二条件下 Actionlet Ensemble 的混淆矩阵

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

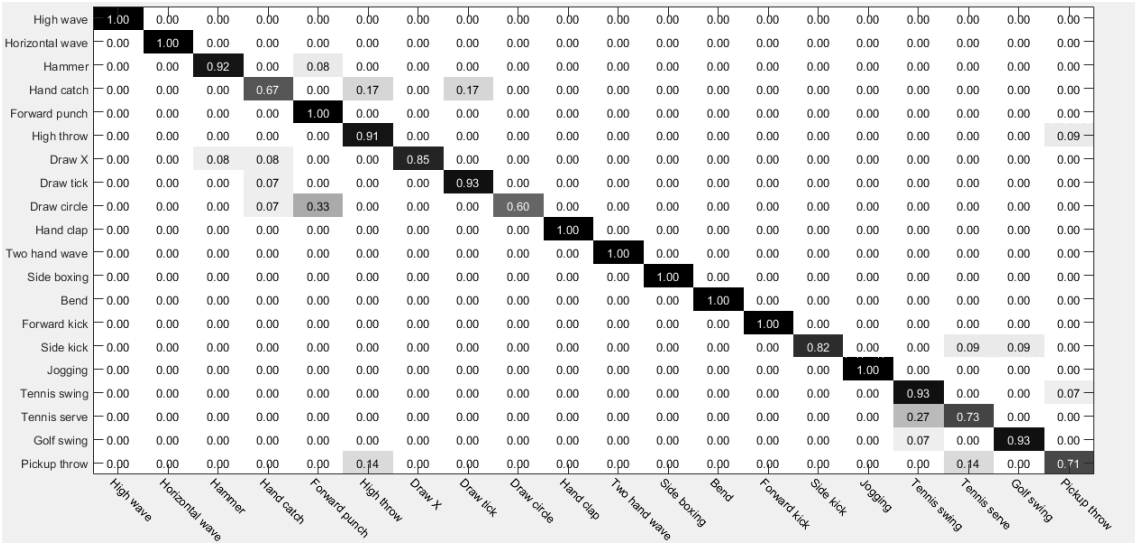

图 5-13 实验二条件下本文算法的混淆矩阵

在实验二的实验条件下，Random Occupancy Patterns 算法在 Hammer，Draw X，High throw 以及 Draw circle 表现不佳，基本维持在 50%左右的识别率。Actionlet Ensemble 算法在 Hand catch 和 Draw X 表现稍差，在 Forward punch 和 High throw 的动作识中与本文算法相比也表现略差，其它动作识别率保持不错。本文算法与其他两种算法相比在各个动作的识别准确率上都有不错的提升，尤其是在 Hammer，High throw，Draw X 有明显的提升。

5.3 本章小结

本章首先对实验环境作了简要介绍，然后主要介绍了 MSRAction3D 数据库并将此数据库按实验一和实验二的条件分别进行实验数据采集。在实验中，我们对 LBP 算子所需要的参数 r 和 m 进行了实验分析和确定。最后，利用实验一和实验二分别对不同算法进行识别率对比。可以得出本文所提算法在识别鲁棒性以及对相似动作的泛化能力上都有着不错的提升。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

第6章 结 论

本文从现实社会的需求入手，着重强调了当前社会的安全形势不容乐观，公共安全经常受到各方面的威胁。正在此种条件下，人体行为识别逐步被各国政府，高校以及科技公司等重视。现阶段，人体行为识别方法多样，处理数据具有相当的局限性，如何突破这种局限性，一直是各国学者积极研究的方向。

本文在积极研究各国学者的内容基础上，选取其中具体的一个方向，即深度图像。由于深度图像有着较好的抗光照和遮挡的能力。天然的具有比一般图像更为优异的信息处理维度。基于此，本文在前人的研究内容中，提出了更为良好的特征提取方法和特征识别方法。

本文在实验中提出了在深度运动图的基础上，融合 LBP 算子，其目的在于增加深度运动图中细节纹理信息，使得深度运动图的运动轮廓更为清晰，让人体行为中相似动作可以得到进一步的区分。由于人体行为特征多，本文使用 Uniform 模式的 LBP 算子，以期使提取的特征向量具有一定的集中性，不至于出现特征向量分布过于稀疏的情况。便于下一步的特征识别算法的准确率和识别效率。

本文在特征识别的内容中，提出了使用 L2 范式的协同表示对生成的特征向量进行识别。由于使用的 MSRAction3D 数据库数据容量并不是很大，而人体行为的特征向量提取维度却十分庞大，即使利用 PCA 进行降维，依然会的到比较庞杂的特征向量。因此会出现 ill-posed 问题，所以本文使用 L2 范式协同表示，以期通过相似动作间的协同表达，来提高识别效果。在 L2 范式协同表示中，我们应用 Tikhonov 正则化来优化 ill-posed 问题的解。本文中，Tikhonov 矩阵里，我们会把测试样本与生成的训练样本中，相似动作给予较高的权重系数，不相似的动作给予较低的权重系数。这样，可以使得分类器对于动作的识别准确度得以进一步的提升。

从本文的实验结果看，本文的实验结果确实优于其他方法的实验结果。在实验一中，由于实验样本的较小，各个方法间的差别并不明显，本文方法一般只比其它方法高 0.5%到 1%。在实验二中，由于实验样本进一步扩大，各种动作的相似又较为接近，使得本文方法对于相似动作的识别优势得到了进一步的显现。

由于 MSRAction3D 数据库的动作都是单人的单一动作，整体难度与实际场景中的动作有些许差别，同时此数据库中的图像是已经经过去噪的图像。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

所以，下一步重点是将现实中的动作加入到数据库中，并考虑深度图像的去噪方法，以期进一步加强人体行为识别的识别效果。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

致 谢

短短三年的研究生时光匆匆而过，在这里我要感谢我的导师，我的同学，我的亲友们对我真挚的关怀与无私的帮助。在此，我向他们致以我最衷心、最诚挚的感谢。

充实而又快乐的三年研究生生活让我学到了很多，我要感谢我的导师史东承教授，他在学术上的严谨认真，给我的人生观、价值观都带来了前所未有的指导与教育。在论文研究过程中，史老师对我的论文题目、算法优化、论文内容等各个方面都进行了细致、有效的指导。让我在此期间学习了很多，受益匪浅。。

三年的时间，我也得到了梁老师的细心的指导，他认真的学风，高超的能力让我对我的研究领域有了更全面的认识。同时，在我论文撰写过程中，我的同学衡瑶瑶、高珊、王磊以及我的学弟乔新宇都给我带来了许多新颖的思路和写作灵感。他们都让我的学术能力与学术思维有了进一步的提高与升华。

最后，我要感谢我的亲友对我学业的大力支持，在工作两年后，依然支持我重回校园，再次充电，让我的人生可以再次重新起航。在这其中，我亲友的支持是必不可少的，是他们的支持让我在学习的道路上一往无前。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

参考文献

[1] Cheng G, Wan Y, Saudagar A N, et al. Advances in Human Action Recognition: A Survey[J]. Computer Science, 2015.

[2] Wang H, Klaser A, Schmid C, et al. Action recognition by dense trajectories[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2011:3169-3176.

[3] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]// International Joint Conference on Artificial Intelligence. Morgan Kaufmann Publishers Inc. 1981:674-679.

[4] Ahad M A R. Motion History Image[M]// Motion History Images for Action Recognition and Understanding. Springer London, 2013:31-76.

[5] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786):504.

[6] Wang L, Suter D. Informative Shape Representations for Human Action Recognition[C]// International Conference on Pattern Recognition. IEEE, 2006:1266-1269.

[7] Mahbub U, Imtiaz H, Ahad M A R. Action recognition based on statistical analysis from clustered flow vectors[J]. Signal, Image and Video Processing, 2014, 8(2):243-253.

[8] Seo H J, Milanfar P. Action Recognition from One Example[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 33(5):867-882.

[9] 章毓晋. 图像工程:图像分析[M]. 清华大学出版社, 2012.

[10] Laptev I, Lindeberg T. On Space-Time Interest Points[J]. International Journal of Computer Vision, 2005, 64(2):107-123.

[11] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE, 2005:886-893 vol. 1.

[12] Harris C. A combined corner and edge detector[J]. Proc Alvey Vision Conf, 1988, 1988(3):147-151.

[13] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[M]. Kluwer Academic Publishers, 2004.

[14] Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features[J]. Computer Vision & Image Understanding, 2008, 110(3):404-417.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[15] Tola E, Lepetit V, Fua P. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2010, 32(5):815-30.

[16] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition[J]. 2007:357-360.

[17] Thi T H, Cheng L, Zhang J, et al. Integrating local action elements for action analysis[J]. Computer Vision & Image Understanding, 2012, 116(3):378-395.

[18] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2):121-167.

[19] Martínez A M, Kak A C. PCA versus LDA[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2001, 23(2):228-233.

[20] Chang C C, Lin C J. LIBSVM: A library for support vector machines[M]. ACM, 2011.

[21] Ghahramani Z. An Introduction to Hidden Markov Models and Bayesian Networks.[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2001, 15(1):9-42.

[22] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4):193-202.

[23] 李飞腾. 卷积神经网络及其应用[D]. 大连理工大学, 2014.

[24] 李红波, 丁林建, 冉光勇. 基于 Kinect 深度图像的人体识别分析[J]. 数字通信, 2012, 39(4):21-26.

[25] 董珂. 基于 Kinect 的人体行为识别研究[D]. 武汉科技大学, 2015.

[26] 陈万军, 张二虎. 基于深度信息的人体动作识别研究综述[J]. 西安理工大学学报, 2015(3):253-264.

[27] Wang J, Liu Z, Chorowski J, et al. Robust 3d action recognition with random occupancy patterns[M]// Computer Vision – ECCV 2012. 2012:872-885.

[28] 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, 2012:1057-1060.

[29] Zhao Y, Liu Z, Yang L, et al. Combing RGB and Depth Map Features for human activity recognition[C]// Signal & Information Processing Association Summit and Conference. IEEE, 2012:1-4.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[30] Xia L, Aggarwal J K. Spatio-temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera[C]// Computer Vision and Pattern Recognition. IEEE, 2013:2834-2841.

[31] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 24(7):971-987.

[32] 黄非非. 基于 LBP 的人脸识别研究[D]. 重庆大学, 2009.

[33] Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(2):210-27.

[34] 蔡家柱. 基于稀疏表达的人脸识别算法研究与实现[D]. 电子科技大学, 2015.

[35] Wright J, Ma Y, Mairal J, et al. Sparse Representation for Computer Vision and Pattern Recognition[J]. Proceedings of the IEEE, 2010, 98(6):1031-1044.

[36] Chen C, Tramel E W, Fowler J E. Compressed-sensing recovery of images and video using multihypothesis predictions[C]// Signals, Systems and Computers. 2011:1193-1198.

[37] 商俊国, 焦斌亮. 多帧图像的 Tikhonov 正则化重建算法研究[J]. 计算机应用研究, 2011, 28(2):785-787.

[38] 张迪. 基于图像复原的一种正则化方法[D]. 电子科技大学, 2013.

[39] Fuhry M, Reichel L. A new Tikhonov regularization method[J]. Numerical Algorithms, 2012, 59(3):433-445.

[40] Tramel E W, Fowler J E. Video Compressed Sensing with Multihypothesis[C]// Data Compression Conference. IEEE, 2011:193-202.

[41] Golub G H, Hansen P C, O'Leary D P. Tikhonov Regularization and Total Least Squares[J]. Siam Journal on Matrix Analysis & Applications, 1997, 21(1):185-194.

[42] 林国军. 基于流形学习和协同表示的人脸识别算法研究[D]. 电子科技大学, 2014.

[43] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition?[C]// IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November. DBLP, 2011:471-478.

[44] 周昊, 火元莲. 一种改进协同表示字典的人脸识别方法[J]. 济南大学学报(自然科学版), 2016(1):29-35.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[45] Lei Z, Meng Y, Feng X, et al. Collaborative Representation based Classification for Face Recognition[J]. Computer Science, 2014.

[46] Tikhonov A N, Arsenin V Y. Solution of Ill-Posed Problems[J]. Mathematics of Computation, 1978, 32(144):491-491.

[47] Jiang J, Chen C, Huang K, et al. Noise robust position-patch based face super-resolution via Tikhonov regularized neighbor representation[J]. Information Sciences, 2016, 367(C):354-372.

[48] Girod B. Motion-compensating prediction with fractional-pel accuracy[J]. IEEE Transactions on Communications, 1993, 27(3):604-612.

[49] Mun S, Fowler J E. Residual Reconstruction for Block-Based Compressed Sensing of Video[C]// Data Compression Conference. IEEE Computer Society, 2011:183-192.

[50] Shlens J. A Tutorial on Principal Component Analysis[J]. 2014, 51(3):219-226.

[51] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3D points[C]// Computer Vision and Pattern Recognition Workshops. IEEE Xplore, 2010:9-14.

[52] 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. IEEE, 2012:20-27.

[53] Vieira A W, Nascimento E R, Oliveira G L, et al. STOP: Space-Time Occupancy Patterns for 3D Action Recognition from Depth Map Sequences[M]// Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. 2012:252-259.

[54] Wang J, Liu Z, Chorowski J, et al. Robust 3d action recognition with random occupancy patterns[M]// Computer Vision – ECCV 2012. 2012:872-885.

[55] Wang J, Liu Z, Wu Y, et al. Mining actionlet ensemble for action recognition with depth cameras[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012:1290-1297.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

作者简介

李延林，男，1990 年 2 月出生，朝鲜族
工作单位：长春工业大学

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

攻读硕士学位期间研究成果

[1] 《基于深度运动图的人体行为识别》 长春工业大学学报 (自然科学版) 第二作者

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net