

分类号	
UDC	
密 级	
学 号	

西安理工大学

硕士学位论文

基于 Kinect 的人体姿态估计与动作识别

卫永波

学 科 门 类: 工 学

学 科 名 称: 模式识别与智能系统

指 导 教 师: 朱虹 教授

申 请 日 期:

独创性声明

本人所提交的学位论文是在导师指导下进行的研究工作及取得的成果。尽我所知，除特别加以标注的地方外，论文中不包含其他人的研究成果。与我一同工作的同志对本文的研究工作和成果的任何贡献均已在论文中作了明确的说明并已致谢。

本论文及其相关资料若有不实之处，由本人承担一切相关责任。

论文作者签名：卫永波 2016年4月8日

学位论文使用授权

本人作为学位论文作者了解并愿意遵守学校有关保留、使用学位论文的规定，即：在导师指导下创作完成的学位论文的知识产权归西安理工大学所有，本人今后在使用或发表该论文涉及的研究内容时，会注明西安理工大学。西安理工大学拥有学位论文的如下使用权，包括：学校可以保存学位论文；可以采用影印、缩印或其他复制手段保存论文；可以查阅或借阅。本人授权西安理工大学对学位论文全部内容编入公开的数据库进行检索。本学位论文全部或部分内容的公布（包括刊登）授权西安理工大学研究生院办理。

经过学校保密办公室确定密级的涉密学位论文，按照相关保密规定执行；需要进行技术保密的学位论文，按照《西安理工大学学位论文技术保密申请表》内容进行保密（附《西安理工大学学位论文技术保密申请表》）。

保密的学位论文在解密后，适用本授权。

论文作者签名：卫永波 导师签名：朱中 2016年4月8日

论文题目：基于 Kinect 的人体姿态估计与动作识别

学科名称：模式识别与智能系统

硕士生：卫永波

(签名)：卫永波

导师：朱虹教授

(签名)：朱虹

答辩日期：2016 年 4 月 7 日

摘要

人机交互是当前人工智能学科与计算机学科一个非常重要的研究方法方向，赋予计算机以人类视觉功能，使机器能够自动识别和感知周围环境中人和事物的活动，是下一代智能计算机的重要组成部分^{[1][2]}。而为了探索一条更加有效人机交互方法，微软推出了深度视觉传感器 Kinect，可以有效的克服光线变化，解决复杂背景条件下的人机交互可靠性差的问题，因此基于 Kinect 的人体姿态估计和姿态识别也成了当下一个非常重要的研究方向。

首先本文分析了 Kinect 的特点，利用微软提供的 Kinect SDK 采集人体的三维骨骼信息，并且提出了一种双 Kinect 人体姿态数据融合的方法，解决了 Kinect 在获取人体三维数据出现自遮挡的问题，给出完整的人体姿态骨骼数据，并给出了一种基于角度描述的人体姿态特征描述方式。

对于人体多个姿态的估计问题，本文确定了由一个多分类问题转换为多个二分类问题，有效解决了识别姿态过多时，识别准确率不能保证的问题。

最后结合人体姿态特征向量以及分类策略的特点确定提出了基于逻辑回归的人体姿态识别以及基于随机森林的人体姿态识别两种方法，并且针对样本的特点，对原始的随机森林方法稍加改进，得到了很好地识别效果，最后通过实验数据对比两种识别方法，最终确定基于随机森林的人体姿态识别能够在一般环境中有效识别出人体姿态。

关键词：Kinect；姿态数据融合；人体姿态识别；逻辑回归；随机森林

Title: POSE ESTIMATE AND RECOGNITION BASED ON KINECT**Major: Pattern Recognition and Intelligent Systems****Name: YongBo Wei****Signature:** Yongbo. Wei**Supervisor: Prof. Hong ZHU****Signature:** Hong Zhu**Abstract**

HCI problem is the current computer science is a very important direction, giving the computer to the human visual function, so that it can automatically identify, analyze and perceive the surrounding environment and the activities of things, is the next generation of intelligent computer an important part. In order to explore an effective method of human-computer interaction, Microsoft launched Kinect depth vision sensor can effectively overcome the light changes, the problem of poor reliability of man-machine interaction in complex background conditions, based on Kinect human pose estimation and gesture recognition has become the moment a very important research direction.

Firstly, it analyzed the Kinect features, use Microsoft's Kinect SDK use Kinect collection of three-dimensional human skeleton information, and propose a method of double Kinect body posture data fusion to solve the problem of self-occlusion Kinect in obtaining the body three-dimensional data. After the use of the human body and acquires three-dimensional skeleton data fusion is proposed based on body posture angle feature vector depicted.

In this paper, and then determine the classification scheme of human gesture recognition, the classification can make the body gesture recognition by a multi-classification problem is converted to a binary classification problem, an effective solution in the interactive process, there needs to recognize the attitude of too much, the classification is the result when the classification problems caused by unsatisfactory.

Finally, body posture and the characteristics of the feature vector classification scheme is proposed to determine the human gesture recognition based on logistic regression and two human gesture recognition method based on random forest, and for the characteristics of the sample, the original random forest method slightly modified to give a good recognition results. Finally, the experimental data comparison of the two identification methods, and ultimately determine the body gesture recognition method based on random forests in the general environment can effectively identify the body posture.

Key: Kinect, Pose Data Fusion, Pose Recognition, Logistic Regress, Random Forests.

目录

摘要	1
Abstract	4
1 绪论	1
1.1 课题研究的背景和意义	1
1.2 国内外的研究现状	2
1.3 本文的研究内容	4
1.4 论文的组织结构	7
2 人体姿态估计	9
2.1 Kinect 简介	9
2.2 基于 Kinect 的人体骨骼关节点识别	10
2.3 双 Kinect 的人体姿态数据融合	12
2.3.1 搭建采集系统	12
2.3.2 人体姿态数据融合	13
2.4 实验与结果分析	16
2.5 小结	19
3 人体姿态特征提取与分类策略	21
3.1 人体姿态特征描述	21
3.2 人体姿态识别的分类策略及问题分析	23
3.3 数据集	25
3.4 小结	26
4 基于逻辑回归的人体姿态识别	27
4.1 逻辑回归模型	27
4.2 模型训练	28
4.3 分类识别	29
4.4 实验与结果分析	30
4.5 小结	35
5 基于随机森林的人体姿态识别	37

5.1	随机森林算法.....	37
5.1.1	算法的随机特性	38
5.1.2	泛化误差与袋外数据估计	39
5.2	随机森林模型的理论概述.....	40
5.2.1	随机森林的数学定义	40
5.2.2	随机森林的性质	41
5.3	人体姿态识别的随机森林模型.....	42
5.3.1	随机森林的训练子空间	42
5.3.2	随机决策树的构建	44
5.3.3	分类识别	47
5.3.4	随机决策树个数的选择	49
5.4	实验结果分析.....	50
5.4.1	随机森林中决策树个数的确定	50
5.4.2	逻辑回归与随机森林的实验结果对比	51
5.5	小结.....	54
6	总结与展望.....	55
6.1	课题总结.....	55
6.2	研究展望.....	56
	致 谢.....	57
	参考文献.....	58
	已发表论文、专利及科研情况	63

1 绪论

1.1 课题研究的背景和意义

使计算机拥有人类视觉功能,使其能够自动识别和感知周围环境中人和事物的活动,是下一代智能计算机与机器人领域的重要组成部分。而要计算机理解人的动作任务就是人体姿态估计,其主要任务是让计算机能够自动地识别出场景中的人在哪里以及人处于一个什么样的姿态^[9]。然后在得到人体姿态之后,然后对估计出的人体姿态进行特征提取,最后利用模式识别的方法对人体姿态进行识别,最终使得计算机可以理解场景中的人在做什么^{[3][4]}。

传统的人体姿态识别是通过运动目标检测,然后对运动目标的轮廓进行分析得到人体的姿态,但是这种方法对鲁棒性比较低,从而一直没有得到很好的实用效果。而近几年来,由于 RGBD 深度相机的普及,凭借其可以获取深度信息的强大功能已经在图像领域被广泛使用^[15],并且在三维建模领域也被广泛研究开发。因为深度相机可以获取深度信息,所以用它来获取人体姿态信息在近年来也备受关注^{[11][14]}。在所有 RGBD 相机中,微软的 Kinect 依靠其强大的 SDK 和廉价性,在有些领域已经有了非常广泛的应用前景,主要表现在人机交互领域^[18]、智能视频监控^[16]、虚拟现实^[14]等方向,甚至在体育训练和辅助医疗也有很好的使用价值^[5]。

众所周知,未来智能计算机的目的就是可以做到与人之间的交互应该更加友好和便捷,键盘、鼠标等输入设备也终将会退出历史舞台。人跟计算机的交流将会跟人与人之间交流一样简单方便,计算机可以理解人的表情、手势、动作以及各种身体语言等信息,并通过分析做出相应的操作。但是这也要求计算机需要具备独立感知外部场景的能力,特别是需要理解人的手势和动作^{[6][7]}。手势识别在近几年里发展特别迅速,已经有很多算法可以有效地识别人的手势动作。但是人体姿态识别由于其超高的非线性化,还一直处于研究状态,虽然没有比较全面的人体姿态识别,但是在有些方面,例如体育训练时,从图像序列中推测运动员的动作,然后进行量化分析,再结合人体生理学和物理学原理科学有效地提高运动员的水平和比赛成绩,使体育训练摆脱纯粹的单纯依靠经验的训练。还有在临床

医学的康复训练中,通过估计人体姿态从而得到人体关节角度数据,可以辅助医生判断受伤者的受伤情况和骨骼畸形程度,为随后医生的诊治和康复治疗提供行之有效的帮助^{[5][19][20]}。

因此,基于 RGBD 深度相机的人体姿态估计和姿态识别逐渐成为计算机视觉领域一个备受关注的研究方向,最后通过行为分析与语义理解使计算机具备场景感知能力,最后可以使得计算机更加智能,使人机之间的交互更加方便、快捷,更加和谐。

1.2 国内外的研究现状

近几年来人体姿态估计和姿态识别方法发展迅速,虽然利用传统相机的人体姿态估计和识别方法已经取得了一些研究进展,但是随着计算机硬件技术的飞速发展,深度相机凭借其强大的能力为人体姿态估计和识别提供了一条新的研究道路,微软为 Kinect 提供的 SDK 已经可以很准确的获取到人体姿态,依靠运动捕获设备价格不断降低,三维运动捕获技术已进入实用化阶段,并且逐渐发展成为一种重要的三维运动信息获取手段。Kinect 利用其 SDK 也已经成功地应用于三维游戏领域,并且在运动分析等领域显现出很大的应用前景。



图 1-1 人体姿态采集装置
Fig 1-1 Pose collection device

对于人体姿态信息的常用骨架模型和三维运动信息的采集,目前广泛使用的方法,按照其感知方式主要有如下四类:一、机械式运动捕捉系统;二、无线传

传感器式运动捕捉系统；三、基于光学的运动捕捉设备；四、无标记非接触式运动捕获系统。接触式的人体姿态采集设备主要使用穿戴传感器或是颜色标签来实现，通过传感器获取肢体运动中各部位的加速度和位置信息，如图 1-1(a)所示；非接触式的人体姿态识别主要是基于视觉的人体姿态识别，通常分为普通相机和深度相机两种，如图 1-1(b)所示，由于基于非接触式的人体姿态识别价格低廉，而且适合场景多，成为当前计算机视觉的热门研究方向，是机器人平台下人机交互的关键技术^{[21][22]}。

本文的主要研究方向是无标记非接触式运动捕获系统，而非接触式又主要分为普通相机和深度相机两种研究方向。

■ 基于普通相机的人体姿态识别

利用普通相机，一般过程对采集到的彩色图像进行图像分割技术去除背景区域，然后对彩色图像进行特征提取，最后进行特征融合，然后利用分类器进行识别，但是现有的研究都是在实验室的理想光照，简单背景条件下完成的，环境的变化都会直接影响特征提取及姿态分类的准确率，这还无法在实际人机交互应用中取得较好的效果。

例如，Ankur Gupta 等在文献^[2]中提出了一种利用一个普通摄像头在固定视角拍摄出的大量视频，然后对视频中的人体进行骨骼点标记训练，再使用非线性循环编码算法，对任意视角下的人体姿态进行估计以及识别，但是由于需要大量的训练数据，而且当人物出现自遮挡时，种方式对于自遮挡没有很好的解决方案。Vasileios Belagiannisti 等在文献^[8]中利用 4 个摄像头对同一场景中的人进行姿态分析，进而达到姿态估计与动作识别的目的，但是其识别的人体骨骼点只有 11 个，而且对于场景的布置有很高的要求，因此对于多场景下人体姿态估计还有比较多的问题^{[23][24]}。

■ 基于深度相机的人体姿态识别

采用上面提到的深度相机，通过获取人体骨架的深度位置信息^[15]，便可以很方便地获得人体三维骨架信息，因此基于 Kinect 的人体动作识别在人机交互领域已经展现出出了巨大的商业价值和应用价值，因此本文的重点也是利用

Kinect 获取人体姿态数据以进行人体姿态识别的研究。下面介绍国内外基于 Kinect 的姿态估计和识别的研究成果。

Min-Chun 等在文献^[4]中提出一种利用 Kinect 采集到的骨骼数据,对各个骨骼点的角度进行分析以达到辅助病人进行肢体康复治疗,提示病人做的不标准的动作。Packer 等在文献^[16]中提出一种先利用 Kinect 采集到的人体的姿态信息,然后对骨骼点进行轨迹分析,以及对操作物品的识别,然后结合两方面信息对人体行为的识别分析。Pisharady PK^[17]将 Kinect 采集的人体骨骼数据转换成为关节角度描述,再利用多项式核函数的支持向量机模型对人体姿态进行分类识别。Xiao Z 等人^[20]利用 Kinect 传感器获得人体深度图像,结合场景信息进行多特征数据融合方法来识别 3D 人体姿态。Liu H 等^[21]使用 Kinect 获得的深度图像信息,将其与人体骨骼点的特征相融合,并通过动作模型的训练学习,提高了人体行为识别的准确性。

综上所述,早期利用动作采集捕获技术获取人体姿态数据大多是通过普通摄像机获得彩色图像得到的,但是由于彩色图像很容易受到复杂背景和光照强度的影响,这就导致了目标的跟踪和分割的困难^{[26][27]}。由于微软深度相机 Kinect 的推出,并且微软提供了获取人体骨骼三维数据的 SDK,可以获取比较稳定的人体骨骼点三维坐标数据信息。Kinect 在获取关节的过程中,是通过骨架拟合的方式来定位人体的各个部位的,然后利用机器学习的方法,根据每一个可能的像素点来确定各个关节的位置,但是 Kinect 在获取人体姿态识别也会出现自遮挡以及人体由于个体差异导致姿态特征描述困难的问题。

目前国内对于基于深度相机的人体姿态估计和姿态识别的研究方向,还都处于起步阶段或者说是实验阶段,要实时准确的获取人体姿态估计以达到人机交互的地步还有很多问题要解决,本文的重点也在提高 Kinect 的人体姿态估计准确率,以及给出一种人体姿态识别的方法。

1.3 本文的研究内容

在实际应用中,人机交互的场景情况多种多样,很难找一个通用算法解决在任何情况下都能满足要求人体姿态识别方法,因此本文研究的重点,是通过两个

标定了相互关系的 Kinect 获取人体三维骨骼信息后，确定一种姿态分类识别策略和识别算法，以解决使用 Kinect 时的自遮挡和人体姿态识别的问题^[24]。

本文的研究内容主要分为以下几个方面。

■ 双 Kinect 的人体姿态数据融合

Kinect 提供的人体三维骨骼信息是以 Kinect 自身为坐标原点，得到人体各个骨骼关节点在空间的三维坐标，Kinect 可以得到 20 个人体的关节节点构成的骨骼模型，可利用骨骼模型进行人体姿态识别^[25]。但是当人的站姿出现相对于 Kinect 的位置，发生自遮挡时，骨骼模型的 20 个骨骼关节点信息就会出现缺损，对缺损骨骼关节点的估计若出现偏差，则会直接导致后续人体姿态识别时的准确性^{[26][27]}。因此本文针对该问题，首先提出了一种双 Kinect 人体姿态融合方法，对缺损的自遮挡的骨骼关节点信息，由另外一个 Kinect 的信息投影获取，解决了当骨骼模型出现信息缺损的问题。

■ 人体姿态特征提取

通过 Kinect 获取的人体姿态数据是三维坐标信息，但是在实际的应用中，不同人体的结构特性的差异，关节点的原始坐标数据过于粗糙，不同的人在同一姿态下，其各个关节点在全局空间中的三维坐标值也会有很大的不同，并且人体在距离 Kinect 不同的方向和距离都会对人体姿态估计产生很大的影响，会使不同人人体姿态数据都会有很大的差异，导致很难真实地反馈出人体的姿态^[28]。

因此，本文利用人体三维坐标信息，建立以人体姿态质心为坐标系原点的方法，求出人体各个骨骼点与原点之间的向量与坐标轴的夹角关系，然后利用这些角度信息构造出人体姿态特征，这样得到的不同人的姿态特征向量能够对姿态的三维坐标的差异具有鲁棒性。

■ 人体姿态识别

人体姿态识别，在这里定义为对预先定义好的若干种人体姿态进行正确分类。利用人体骨架信息作为姿态分类，现有的常规方法把人体骨架的信息作为特征向量来进行姿态分类，而实际场景中，人体的姿态的随意性很大，同一个人做几次同一个动作，姿态都会有差异，不同的人，做同一个动作，姿态也会有差异。当

需要识别的姿态种类比较多时，姿态识别才具有实际意义，可是，相似姿态之间会存在很大的干扰^{[29][30][31]}。所以，确定一个行之有效的人体姿态的分类策略来进行姿态识别是非常重要的。

因此，本文提出的姿态识别的分类策略是，将任意两个同类别姿态的特征向量的向量差的绝对值作为正样本，任意两个不同类别姿态向量的向量差的绝对值为负样本，采用学习算法构造出一个分类器，最后利用这个分类器去判别这两个输入的姿态数据是否属于同一类。这样便可以把一个多分类问题，变成了多个二分类问题，有效地解决了相似姿态间的相互干扰问题。

当确定了人体姿态识别的分类策略之后，然后就要选择一种二分类的分类器来识别两个姿态是同一姿态。因此本文先后使用两种分类器，分别是逻辑回归模型和随机森林模型，选择逻辑回归是因为它是比较经典而且使用最广泛的二分类分类器之一，而且该模型可以提高特征向量中对分类有用的特征的权值，但是在使用过程中，由于人体姿态的高度非线性特征，而导致模型对于相似姿态识别率过低的情况。在此基础上又选择随机森林模型来做该分类器，实验过程中发现随机森林方法有以下几点。

- 1) 随机森林模型在一些公共数据集上的分类准确率不输给其他学习算法，并且在高维数据上的表现能力优于 SVM、Boosting 和神经网络模型^{[32][33][34]};
- 2) 本文从模型的理论分析得到当随机森林中随着随机决策树数量逐渐增多时，模型仍不容易出现过拟合现象;
- 3) 随机森林模型可以很好继承树形模型学习、训练和测试都比较高效的特性，并且算法的随机性可以有效降低算法复杂度，适用于要求对算法实时性要求比较高的应用场合;
- 4) 随机森林模型对噪声样本具有很强的容忍性，它能够有效的检测出训练数据中的野点 (outliers);
- 5) 随机森林中的每棵随机树在训练过程中都是独立生长的，所以无论是在训练还是识别过程中都很容易对它实现并行化处理;

近年来，由于随机森林算法拥有的优良特性，使它在科学研究领域和工程应用方面都受到研究者和工程开发人员的重视，被广泛应用于计算机视觉的各种领

域，例如特征点的识别、目标检测和识别、人体姿态估计、和人体部件识别等。因此本文根据人体姿态特征向量的特征和样本分布的特点，对原始的随机森林算法进行改进，最终确定了本文提出的基于 Kinect 人体姿态估计与动作识别的算法。

1.4 论文的组织结构

本论文的章节内容安排如下：

第一章 绪论

本章介绍了人体姿态估计和姿态识别的基本含义，然后介绍了其研究背景与意义、国内外研究现状、研究现状等，并最后阐述本文对于该方向的创新点。

第二章 人体姿态估计

本章简单介绍了 Kinect 的基本原理以及通过 Kinect 如何获取人体的骨骼坐标点，然后介绍了 Kinect 的获取人体骨骼坐标点的一些弊端，进而提出一种双 Kinect 人体姿态数据融合的方法。

第三章 人体姿态特征提取与分类策略

本章给出一种提取 Kinect 三维坐标信息的人体姿态特征向量描述，进而分析传统人体姿态识别策略的问题，并给出了一种人体姿态识别的分类策略，进而使人体姿态识别由一个多分类问题转换为一个二分类问题。最后简单介绍了以下本文所使用的所有测试数据集。

第四章 基于逻辑回归的人体姿态识别

本章首先简单介绍了逻辑回归模型，然后利用逻辑回归模型进行人体姿态识别，最后对实验结果进行分析。

第四章 随机森林模型

本章首先介绍了随机森林模型，以及针对人体姿态特征向量的特点，给出一种有效的随机森林模型训练过程，最终实现人体姿态识别，并在最后对实验结果进行分析。

2 人体姿态估计

由于微软为 Kinect 提供了强大功能，因此本文的人体姿态数据信息也是通过微软 Kinect 提供的 SDK 来获取人体的姿态动作信息来进行人体姿态估计和姿态识别的实验研究的^[38]。因此，在展开人体姿态估计相关内容阐述之前，使了解 Kinect 设备的基本结构和获取人体三维关节点数据的原理是研究工作中的一个重要环节。

2.1 Kinect 简介

Kinect 是由微软推出的一个融合多种传感器技术的 RGBD 摄像机，是为 Xbox 360 开发的一套体感设备。Kinect 有两种摄像头，一个是 RGB 摄像头，另外一种红外摄像机，如图 2-1 所示。Kinect 不仅可以利用 RGB 摄像头来得到普通彩色图像，还可以利用红外摄像机来获取当前场景的深度图像，这样利用 Kinect 不仅可以获取该场景下的 RGB 彩色图像，还可以感知该场景下的物体与 Kinect 之间的距离。而且在摄像头的下面是一排阵列式麦克风，利用麦克风便捕得到该场景下的声音并可以定位到声源位置。Kinect 与 Kinect 的连接方式是通过 USB 接口与计算机连接的^{[40][41]}。



图 2-1 Kinect 设备结构图

Fig 2-1 Kinect Equipment Chart

我们先来介绍下 Kinect 的空间坐标系,如图 2-2 所示。假设人是面对 Kinect 站立,然后 Kinect 获取的坐标信息就是以红外摄像机为原点, X 轴正方向指向人的右侧, Y 轴的正方向指向人体正上方, Z 轴正方向与 Kinect 的检测方向一致 [39]。

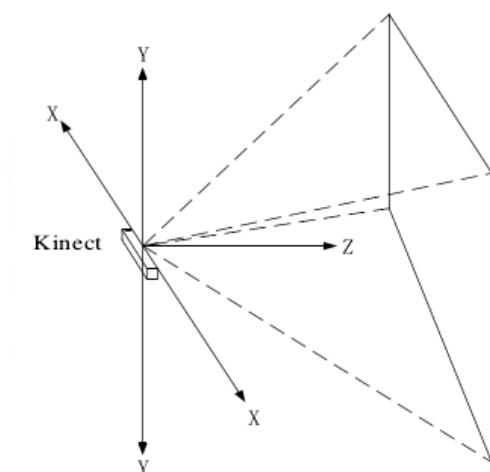


图 2-2 Kinect 三维空间坐标系

Fig 2-2 Three-dimensional coordinate system of Kinect

Kinect 采用了基于光编码理论的技术,可以直接通过红外摄像头获取人体与摄像头之间的三维距离信息。其基本思想是通过连续光对测量空间进行编码,然后通过感应器获取编码光线,然后将数据进行解码后,再从得到的深度数据流中获取该场景下深度图像。在产生的深度图像中每一个像素点都包含了该三维场景的三维深度信息。最后排除深度信息中的噪声,再通过识别算法找到人体的各个关节点的三维空间位置信息^{[37][40][41]}。

2.2 基于 Kinect 的人体骨骼关节点识别

Kinect 设备获取三维骨骼关节点数据的工作过程主要分为人体轮廓分割、人体部位识别和关节点定位三个步骤。

步骤 1 人体轮廓分割

Kinect 先将完整的人体区域从深度图像中分离出来。在此过程中,由于 Kinect 获取深度图像受到光照的变化、复杂背景的影响很小,因此, Kinect 在

通过深度图像提取的人体轮廓会比在彩色图像中提取的人体轮廓更加完整、效果更佳。而且 Kinect 再利用深度数据提取人体轮廓时，还融合了 RGB 彩色图像信息与深度信息，并且以人体结构的特征和 Kinect 内置的运动检测、人体关节识别、人脸识别等技术为依据，首先找到离 Kinect 较近区域中任何形状类似人体姿态的目标，然后便利目标内的深度像素点，再利用图像识别的技术来判断这些点是否属于人体的一部分，最后再将属于人体的像素点所构成的人体轮廓从深度图像中分离出来^{[40][43][45]}。

步骤 2 人体部位识别

在将人体轮廓从深度图像中分割出来后，还需要在人体轮廓中识别出人体的各个关节部位，本论文使用的 Kinect 识别出 20 个关节部位。而且在 Kinect 的识别关节的过程中，是通过人体骨架拟合的方式来找到人体的各个部位的。

步骤 3 骨骼关节点定位

在得到人体部位后，Kinect 还需要从人体不同的部位中进一步找到人体的各个部位骨骼关节点。Kinect 是利用机器学习的方法，判断每一个可能的像素点来确定关节点位置。

采用基于 Windows 的 Kinect SDK 提供的骨骼跟踪功能，通过上述三个步，便可以得到场景中人体的 20 个骨骼关节点在三维空间中的三维位置信息，Kinect 获取的人体 20 个关节点的示意图如图 2-3 所示。

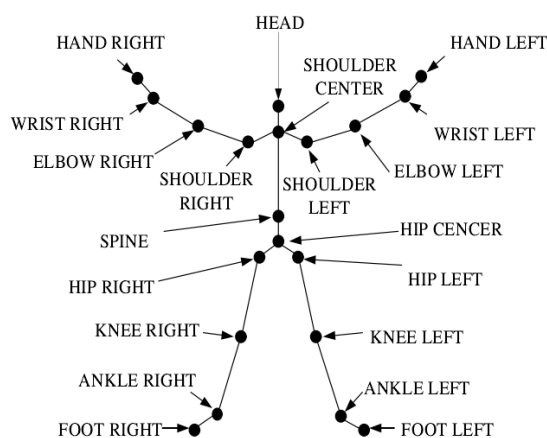


图 2-3 Kinect 捕获的人体关节点的示意图

Fig 2-3 The twenty points of Kinect pose

2.3 双 Kinect 的人体姿态数据融合

通过 Windows 的 Kinect SDK 提供的骨骼跟踪，便可以得到 20 个人体的骨骼模型，但是当人的站姿出现与 Kinect 发生自遮挡的问题时，骨骼模型的就会出现缺损，如图 2-4 所示，Kinect 只能识别出人体左侧部分骨骼点，这样在做姿态识别时会严重干扰的算法对人体姿态的识别。

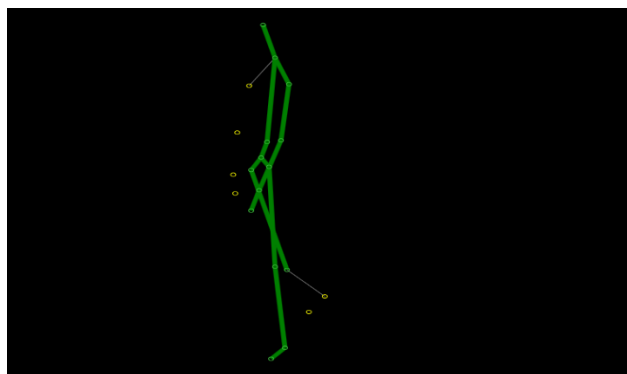


图 2-4 Kinect 发生自遮挡时的示意图

Fig 2-4 The Kinect self-occlusion Image

因此针对上述问题，本文提出一种融合两台 Kinect 骨骼数据的方法来解决该问题，以此来提升后面对人体姿态识别时的准确度。

2.3.1 搭建采集系统

首先，如图 2-5 所示搭建人体姿态数据采集系统，在相互垂直正交的位置上摆放两台 Kinect，使其面向拍摄区域 3，拍摄区域 3 大小由两台 Kinect 的摄像区域决定，Kinect 1 和 Kinect 2 分别与电脑 4 和电脑 5 连接，电脑 4 与电脑 5 连接之间建立局域网，以便进行数据传输。默认电脑 4 为主处理器，系统运行时，电脑 5 会实时将 Kinect 2 检测到的人物姿态骨骼数据通过局域网发送给电脑 4，同时电脑 4 也会实时得到 Kinect 1 的姿态骨骼数据。然后电脑 4 会同时调用融合算法进行多 Kinect 的数据融合。

系统运行时，系统按照以下步骤运行：

电脑 5 会实时将 Kinect 2 获取的 20 个骨骼数据通过局域网发送给电脑 4，同时电脑 4 实时获取 Kinect 1 的骨骼数据。其中每个骨骼数据包括 $G_{(i,j)}$ 和 $F_{(i,j)}$ 两部分： $G_{(i,j)}$ 是以第 i 个 Kinect 为坐标原点的人体第 j 个骨骼位置坐标； $F_{(i,j)}$ 是第

i 个 Kinect 是否跟踪到人体第 j 个骨骼点的标记信息。 $F_{(i,j)}$ 值为 0 时, 表示第 i 个 Kinect 没有跟踪到第 j 个骨骼位置; $F_{(i,j)}$ 值为 1 时, 表示第 i 个 Kinect 跟踪到第 j 个骨骼位置。其中 i 表示 Kinect 的标号, j 表示骨骼点的标号。

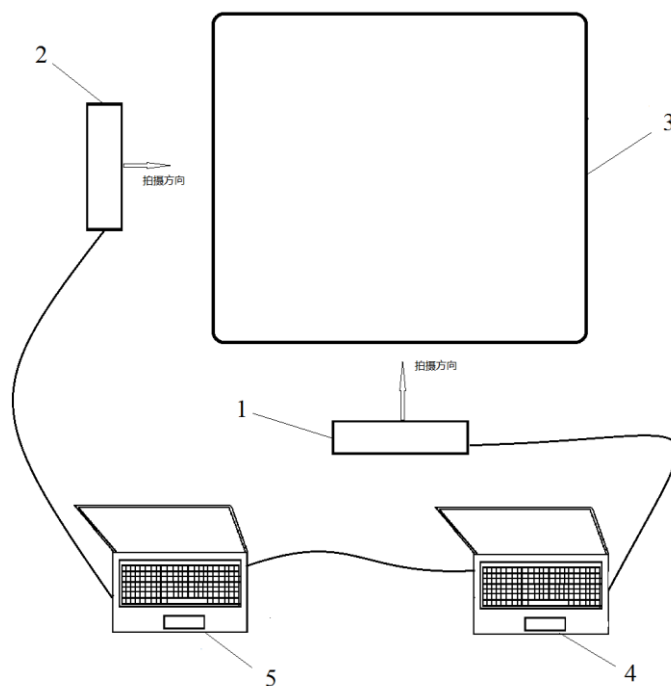


图 2-5 双 Kinect 人体姿态采集系统

Fig 2-5 The Pose Collection System of double Kinects

通过该数据采集环境, 可以基本满足人以何种方向面对 Kinect 时, 都可以采集到人各个方向的数据, 对 K 主 Kinect 中出现自遮挡的点, 通过辅 Kinect 的信息可以给出正确的估计。

接下来要做的就是融合两台 Kinect 的数据, 常规的算法融合两台摄像头的的数据, 需要对两台摄像头进行标定然后进行数据融合, 但是由于本文设置的两台 kinect 的位置可事先确定, 并且针对的是 20 个人体骨骼关节点的坐标关系, 因此, 本文给出了一种基于骨骼模型的人体姿态数据融合方法。

2.3.2 人体姿态数据融合

本文给出的人体姿态数据融合方法, 首先要对两个 Kinect 进行自标定, 标定方法如下:

假设两个 Kinect 获取人体骨骼点有如下转换关系，如公式（2.1）所示：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} \quad (2.1)$$

其矩阵表达式为：

$$m = Km' \quad (2.2)$$

其中 $(x, y, z)^T$ 是主 Kinect 检测到的人体骨骼点的人体坐标点，记为 m ； (x', y', z') 是辅 Kinect 检测到的人体骨骼点的三维坐标，记为 m' 。 $(k_{11}, k_{12}, \dots, k_{33}, k_{34})$ 是坐标转换矩阵，记为 K 。

公式（2.1）展开后可得方程（2.3）：

$$\begin{cases} k_{11}x' + k_{12}y' + k_{13}z' + k_{14} = x \\ k_{21}x' + k_{22}y' + k_{23}z' + k_{24} = y \\ k_{31}x' + k_{32}y' + k_{33}z' + k_{34} = z \end{cases} \quad (2.3)$$

将公式（2.3）变形，即可得方程（2.4）：

$$\begin{bmatrix} x' & y' & z' & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x' & y' & z' & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x' & y' & z' & 1 \end{bmatrix} \begin{bmatrix} k_{11} \\ k_{12} \\ k_{13} \\ k_{14} \\ k_{21} \\ k_{22} \\ k_{23} \\ k_{24} \\ k_{31} \\ k_{32} \\ k_{33} \\ k_{34} \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2.4)$$

则向量 $(k_{11}, k_{12}, k_{13}, k_{14}, k_{21}, k_{22}, k_{23}, k_{24}, k_{31}, k_{32}, k_{33}, k_{34})^T$ 为需要求出的参数，共计有 12 个参数需要确定。但是，公式（2.4）给出的只有 3 个方程，是无法求解的。

为此，本文采用了最小二乘的方法，假设获得了 n 组观察数据对 $\{(x_1, x_1', y_1, y_1', z_1, z_1'), (x_2, x_2', y_2, y_2', z_2, z_2'), \dots, (x_n, x_n', y_n, y_n', z_n, z_n')\}$ ，可以

得到如下的方程组：

$$\begin{bmatrix} x'_1 & y'_1 & z'_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x'_1 & y'_1 & z'_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x'_1 & y'_1 & z'_1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x'_n & y'_n & z'_n & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x'_n & y'_n & z'_n & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x'_n & y'_n & z'_n & 1 \end{bmatrix} \begin{bmatrix} k_{11} \\ k_{12} \\ k_{13} \\ k_{14} \\ k_{21} \\ k_{22} \\ k_{23} \\ k_{24} \\ k_{31} \\ k_{32} \\ k_{33} \\ k_{34} \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ \dots \\ x_n \\ y_n \\ z_n \end{bmatrix} \quad (2.5)$$

因为未知参数有 12 个，所以只需 $n>3$ ，便可利用最小二乘法来求解 K 的最优解。所以假设有 n 个点对，则得到方程(2.5)，记为 $M'K = M$ 。

则公式(2.5)对应的最小二乘解是公式(2.6)：

$$\hat{K} = (M'^T M')^{-1} M'^T M \quad (2.6)$$

根据上面的步骤，便可得到两个 Kinect 获取的人体姿态数据的相互转换关系。

因此本文提出的双 Kinect 人体姿态数据融合方法步骤如下：

步骤 1，获取双 Kinect 观察点对

采集两个 Kinect 在 10 帧内检测到的人体姿态数据，并将对应帧、对应人体姿态对应位置的骨骼点组成坐标点对。

步骤 2，获取双 Kinect 投影矩阵参数

根据步骤 1 得到的两个 Kinect 的坐标点对，采用最小二乘法，按照公式(2.6)求解方程，得到双 Kinect 的标定投影矩阵参数 $(k_{11}, k_{12}, \dots, k_{33}, k_{34})$ 。

步骤 3，融合双 Kinect 的姿态数据

姿态数据融合策略：

- a. 对于主 Kinect 检测到的骨骼关节点，则认为已得到该点的骨骼点信息，

将忽略辅 Kinect 检测到的骨骼关节点；

b. 对于主 Kinect 没有检测到，而辅 Kinect 检测到的骨骼关节点，根据两个 Kinect 的标定投影矩阵和该骨骼关节点在辅 Kinect 中的位置信息，利用公式(2.7)，求得在主 Kinect 在该骨骼点的位置：

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (2.7)$$

c. 对于在主 Kinect 以及辅 Kinect 中，均未检测到的骨骼关节点，则需要对这些骨骼关节点坐标信息进行预测。

因考虑到时效性以及人物动作的连续性，本方法采用对坐标数据的运动变化量求简易平均值进行预测。

根据连续 5 帧该骨骼关节点位置的平均位移为 $\delta = (\Delta x, \Delta y, \Delta z)$ ，在该帧中，该骨骼关节点的位置为：

$$G_i = G_{i-1} + \delta \quad (2.8)$$

通过以上步骤，就可以保证所有的人体骨骼点坐标信息都得到，并且误差范围可以满足后面人体姿态的识别。

2.4 实验与结果分析

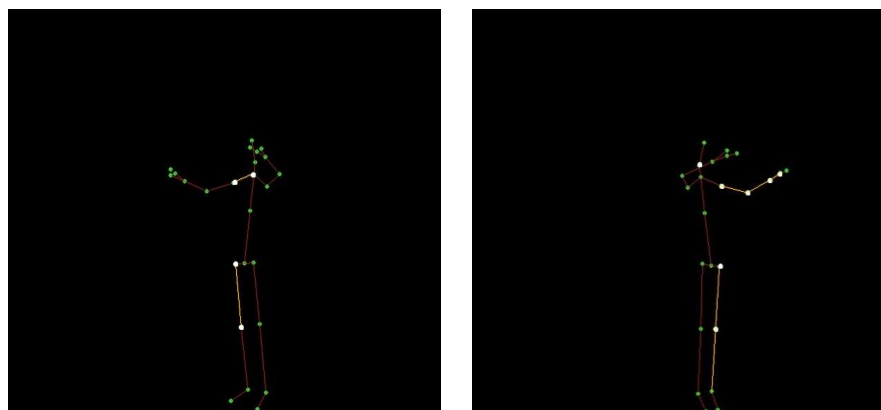
本实验首先搭建如图 2-5 所示的数据采集环境，再根据两个 Kinect 在前 10 帧共同采集的同一个人的不同姿态，然后利用公式(2.5)求得两个 Kinect 的标定投影矩阵如下公式：

$$K = \begin{bmatrix} -1.09768 & -0.32258 & 1.07236 & -2.10996 \\ -1.3338 & 0.712778 & 1.12177 & -2.23886 \\ -1.47173 & 0.166766 & 0.02723 & 2.47457 \end{bmatrix} \quad (2.9)$$

由此，便可以利用该标定投影矩阵对主 Kinect 未获取到的人体骨骼关节点，利用辅 Kinect 检测到的人体骨骼关节点的信息进行估计。

如图 2-6 所示，图 2-6(a)是主 Kinect 获取的人体姿态，图 2-6(b)是辅 Kinect

获取的人体姿态，图中白色骨骼关节点为 Kinect 未检测到的人体骨骼关节点。



(a) 主 Kinect 获取的人体姿态

(b) 辅 Kinect 获取的人体姿态

图 2-6 人体姿态数据融合

Fig 2-6 Pose Data Fusion

本实验首先选取一部分主 Kinect 和辅 Kinect 都检测到的点，用来验证计算求得的两个 Kinect 的标定投影矩阵的准确性，实验结果见表 2-1。

表 2-1 预测坐标与实际检测坐标对比表

Tab 2-1 Comparison between the predicted coordinates and detected coordinates

骨骼关节点索引	主 Kinect	辅 Kinect	预测的主 Kinect
SPINE_BASE	X: 0.1014 Y: -0.3623 Z: 2.1396	X: 0.2382 Y: -0.3926 Z: 2.1859	X: 0.0993 Y: -0.3843 Z: 2.1180
ELBOW_LEFT	X: -0.0916 Y: 0.1003 Z: 2.5696	X: -0.0041 Y: 0.2162 Z: 1.9429	X: -0.0916 Y: 0.1003 Z: 2.5696
HEAD	X: 0.0916 Y: 0.4076 Z: 2.2482	X: 0.2646 Y: 0.4006 Z: 2.3921	X: 0.0355 Y: 0.3771 Z: 2.2170

从表 2-1 可以看出，通过辅 Kinect 和标定投影矩阵预测的主 Kinect 坐标点的值，与主 Kinect 实际检测到的坐标点的位置偏差很小，几乎可以忽略。由此，验证了计算得到的标定投影矩阵的准确性。

表 2-2 列出了在对主 Kinect 未检测的骨骼关节点，利用辅 Kinect 和标定投

影矩阵预测得到的主 Kinect 的骨骼关节点坐标的估计值。

表 2-2 预测到的主 Kinect 未检测到的骨骼点坐标

Tab 2-2 The predicted skeleton coordinates that the main Kinect not detected

骨骼点索引	辅 Kinect	预测主 Kinect
NECK	X:0.2117 Y: 0.2792 Z: 2.3475	X: 0.0849 Y: 0.3111 Z: 2.2734
SHOULDER_RIGHT	X: 0.3549 Y: 0.0499 Z: 2.2913	X: -0.0585 Y: -0.1063 Z: 2.0229
HIP_RIGHT	X: 0.3008 Y: -0.3989 Z: 2.1623	X: 0.0072 Y: -0.4989 Z: 2.0241
KNEE_RIGHT	X: 0.1650 Y: -0.7206 Z: 2.0611	X: 0.1516 Y: -0.6605 Z: 2.1676

经过上面过程，便可得到一付完整的人体姿态，如图 2-7 所示。对比图 2-6(a) 和图 2-7，可以看出来，经过本文采用的双 Kinect 人体姿态融合方法可以很准确的进行人体姿态估计。



图 2-7 融合后的人体姿态数据

Fig 2-7 The pose of fusion

通过以上步骤，所有骨骼的位置坐标都被跟踪或预测到，得到了完整的人体姿态骨骼信息，解决了当人物出现自遮挡以及 Kinect 受到干扰时没有检测到人物骨骼信息的问题，为后期处理中姿态识别、人机交互、虚拟现实等提供了便利。本方法可以有效解决现有 Kinect 在提取人体骨骼关节点信息时，受到干扰骨骼

关节点数据出现跳变或者出现自遮挡时姿态缺失的问题。

2.5 小结

本章首先介绍了 Kinect 获取人体姿态的基本原理，并介绍了 Kinect 在获取人体姿态时的一些问题，然后提出一种双 Kinect 人体姿态数据融合的方法，以得到完整的人体姿态骨骼信息，解决了当人物出现自遮挡以及 Kinect 受到干扰时没有检测到人物骨骼关节点信息的问题，为后期处理中姿态识别、人机交互、虚拟现实等提供了便利。本方法可以有效解决现有 Kinect 在提取人体骨骼关节点信息时，受到干扰骨骼骨骼关节点位置数据出现跳变或者出现自遮挡时姿态缺失的问题。

3 人体姿态特征提取与分类策略

通过人体姿态数据融合之后,我们便可以得到人体所有骨骼点相对于 Kinect 的三维信息,但是在实际的应用中,若直接采用 Kinect 设备捕获的 20 个关节点的三维数据来表征人体姿态,还是会有很多问题,由于不同人体的结构特性的差异,关节点的原始坐标数据过于粗糙,不同的人在同一姿态下,其各个骨骼点在全局空间中的三维坐标值的差异,距离 Kinect 不同的方向和距离,这样得到三维坐标数据都会有很大的不同,会对姿态识别产生很大的影响,即直接采用人体三维空间数据描述人体姿态特征太过粗糙,导致同类之间特征不具有有很好的一致性^{[47][48]}。

因此选择一个有效的姿态特征描述方式对后面的姿态识别至关重要,如果选择的特征过少,特征向量维数太低,不能准确地表示人的姿态,会增加人体姿态理解的难度;如果选择的特征太多,会导致特征向量维数太高,增加分类识别的复杂度,造成维数灾难;因此,在提取 Kinect 获取的人体关节点的原始数据的特征时,必须尽可能在降低特征维数的情况下保留人体姿态的原始特征,因此,必须对人体关节点数据进一步分析和转换以提取更加丰富和准确地的人体姿态特征信息^{[49][50][51]}。

考虑到上述问题,本文选取 Kinect 捕获的骨骼关节点之间的角度信息进行描述。因为人体运动过程中,人体的骨骼结构不会发生变化,同一个姿态,骨骼关节点之间的角度变化受不同人体身体结构差异的影响较小。

3.1 人体姿态特征描述

接下来本文将描述对人体骨骼点的三维坐标数据进行处理以提取出人体姿态特征的过程。

利用姿态融合方法获取的人体姿态的 20 个人体骨骼点相对于 Kinect 的三维位置坐标,每个位置坐标记为 $G_i = (x_i, y_i, z_i)$,其中 G_i 表示是以 Kinect 为坐标原点的人体第 i 个骨骼位置坐标, x_i 表示第 i 个骨骼位置的 x 坐标, y_i 表示第 i 个骨骼位置的 y 坐标, z_i 表示第 i 个骨骼位置的 z 坐标。

首先,求出人体姿态的质心位置 $G_o = (x_o, y_o, z_o)$,然后建立以人体姿态质心

为原点的坐标系，如图 3-1 所示。

人体姿态质心 G_O 的计算公式如下：

$$x_o = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.1)$$

$$y_o = \frac{1}{n} \sum_{i=1}^n y_i \quad (3.2)$$

$$z_o = \frac{1}{n} \sum_{i=1}^n z_i \quad (3.3)$$

其中， x_o 表示人体姿态质心位置的 x 坐标， y_o 表示人体姿态质心位置的 y 坐标， z_o 表示人体姿态质心位置的 z 坐标。

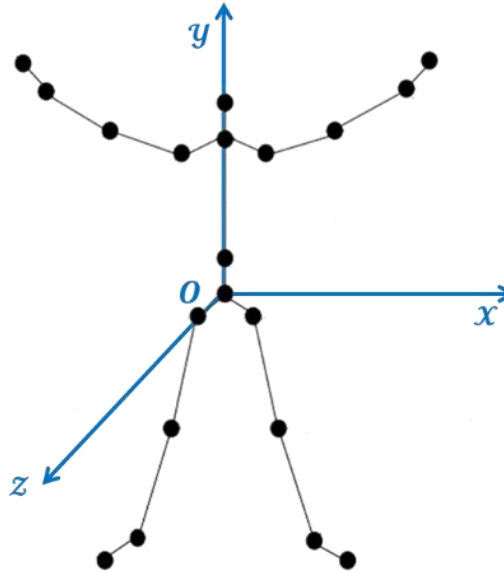


图 3-1 以人体姿态质心为原点的三维空间

Fig 3-1 Pose center for three-dimensional coordinate center

然后，在新的坐标系下，得到人体每个骨骼点与姿态质心的向量与坐标轴 x, y, z 在 $[0, 2\pi]$ 的夹角。

首先利用公式(3.4)、(3.5)、(3.6)求出人体姿态的每个骨骼关节点相对于新坐标原点 G_O 的方向余弦，记为 $R_i = (\alpha_i, \beta_i, \gamma_i)$ ， $i = 1, 2, \dots, 20$ ，其中 R_i 表示是人体第 i 个骨骼关节点方向余弦， α_i 表示第 i 个骨骼的 x 方向的方向余弦， β_i 表示第 i 个骨骼的 y 方向的方向余弦， γ_i 表示第 i 个骨骼的 z 方向的方向余弦。

计算公式如下：

$$\alpha_i = \arccos \frac{x_i - x_o}{\sqrt{(x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2}}, \quad i = 1, 2, \dots, 20 \quad (3.4)$$

$$\beta_i = \arccos \frac{y_i - y_o}{\sqrt{(x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2}}, \quad i = 1, 2, \dots, 20 \quad (3.5)$$

$$\gamma_i = \arccos \frac{z_i - z_o}{\sqrt{(x_i - x_o)^2 + (y_i - y_o)^2 + (z_i - z_o)^2}}, \quad i = 1, 2, \dots, 20 \quad (3.6)$$

由于反正弦求出的角度只能在 $[0, \pi]$ 之间，所以还需要根据 G_o 与 G_i 的相对关系，即知道每个骨骼点在新的坐标系的象限，这样就可以求出的人体每个骨骼点与人体姿态质心的向量与 x ， y ， z 轴在 $[0, 2\pi]$ 的夹角。

这样，可得到 $(\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \dots, \alpha_{20}, \beta_{20}, \gamma_{20})$ 这样一个 60 维向量，然后把这个向量作为人体姿态的特征向量，记作 $\phi = (x_1, x_2, \dots, x_n)$ ，其中 $1 \leq n \leq 60$ ， $x_i \in [0, 2\pi]$ 。

3.2 人体姿态识别的分类策略及问题分析

按照上面给出的计算方法，获得了人体姿态的 20 个骨骼关节点的三维坐标构成的 60 维特征向量，但是直接通过该特征进行姿态识别还是会遇到很多问题，主要体现在以下几点：

■ 训练集规模过小问题

现有的常规方法把人体骨架的信息作为特征向量来进行姿态分类，一般的训练集无法包含人体所有可能姿态，并且各种姿态的样本都有限。这样由于一般数据集无法包含所有的姿态，并且同种姿态也没有足够多的样本，会给人体姿态的识别带来大量的误判。

■ 姿态种类太多

实际场景中人体的姿态的随意性很大，当需要识别的姿态种类太多，相似姿态的之间也会存在很大的干扰，而且人体关节点数据除了含有能表征人体姿态的信息之外还含有无关的人体物理结构信息，这些与姿态无关的数据也会对人体姿态识别构成干扰，如果使用一种分类器同时输出多个节点，这样对分类器在训练过程成将会很难收敛，就算可以收敛，也常常会出现过拟合的现象。

因此,确定一个行之有效的人体姿态的分类策略来进行姿态识别是非常重要的。鉴于此,本文提出一种姿态识别的分类策略,即将任意两个同类别姿态的特征向量的向量差的绝对值作为正样本,任意两个不同类别姿态向量的向量差的绝对值为负样本,然后利用学习算法构造出一个分类器,最后利用这个分类器去判别对任意两个姿态是否为同一姿态,然后最后比对该姿态与哪一种姿态的置信度更高,便认为它是那一种姿态。这样便可以把一个多分类问题变成一个二分类问题,上面提出的两种问题便会迎刃而解。

该分类策略的具体流程图如图 3-2 所示。

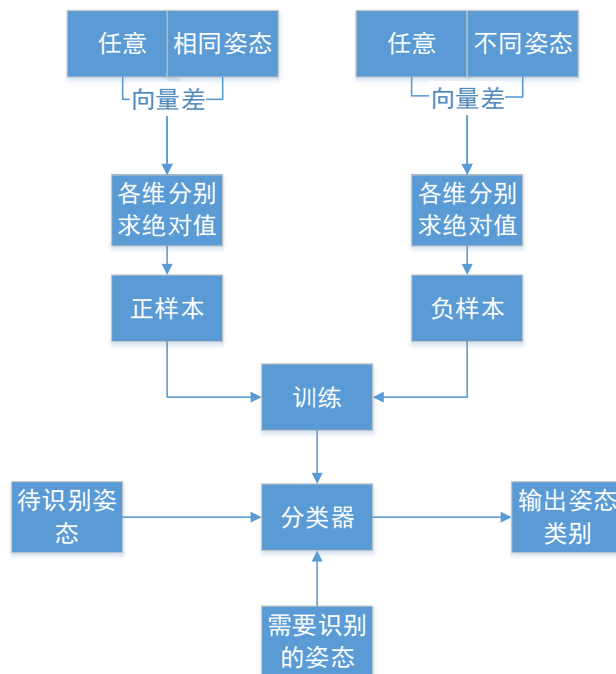


图 3-2 人体姿态识别流程图
Fig 3-2 Human gesture recognition flowchart

具体流程如下:

首先,对样本库中的姿态进行标记分类,同一姿态分为同一类,并且提取其人体姿态的 20 个三维骨骼关节坐标构成的 60 维特征。

然后,构造正负样本,利用公式 (3.7) 求出任意两个姿态 1 和姿态 2 的姿态特征向量的向量差的绝对值。当姿态 1 和姿态 2 为相同姿态时, ϕ_s 标记为正样本; 当姿态 1 和姿态 2 为不同姿态时, ϕ_s 标记为负样本。

$$\phi_s = \phi_1 - \phi_2 = (|x_{11} - x_{21}|, |x_{12} - x_{22}|, \dots, |x_{1n} - x_{2n}|) \quad (3.7)$$

其中, $n = 60$ 是特征向量的维数, ϕ_s 是姿态识别的样本, ϕ_1 、 ϕ_2 为样本库中任意两个人体姿态的姿态特征向量。

3.3 数据集

下面简单介绍本论文采用的数据集。

本文人体姿态的实验数据是利用双 Kinect 系统采集了 20 个人的日常行为姿态来构建数据库, 为了保证数据库的完备性, 该数据库中包含各种自遮挡的问题, 所有姿态均是通过本文提出的双 Kinect 人体姿态估计方法获取的人体姿态。

最后实验中使用到的人体姿态共有 7 万帧, 其中部分姿态如图 3-3 所示。其中也包含了相似姿态, 以评测其姿态识别方法的性能。

为了验证人体姿态识别的算法, 本文从这 7 万帧姿态数据中挑选出 15 种需要识别的人体姿态, 如图 3-4 所示。

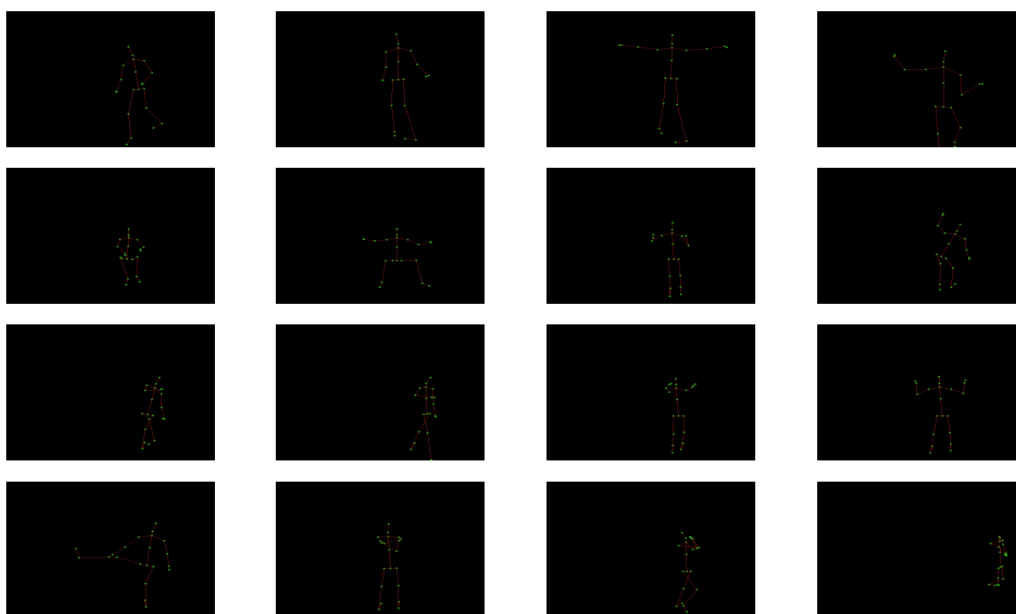


图 3-3 人体姿态数据集部分姿态
Fig 3-3 Some pose samples of the pose data base

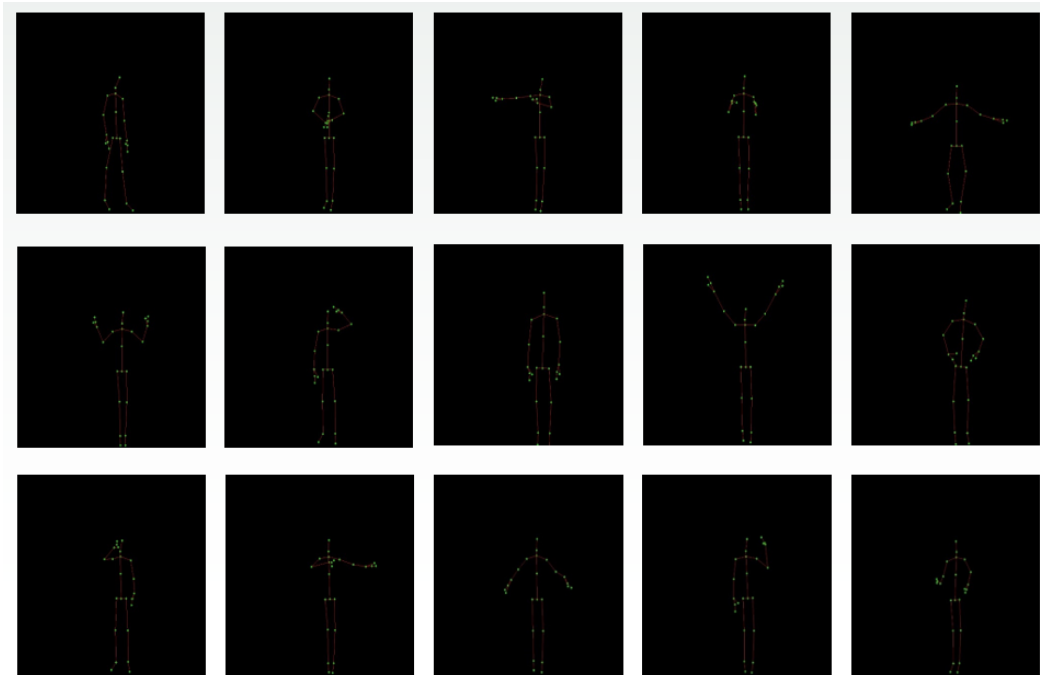


图 3-4 需要识别的 15 种人体标准姿态

Fig 3-4 The 15 standard poses that need to recognize

3.4 小结

本章首先介绍了 Kinect 获取的三维人体骨骼关节点坐标转换成为角度数据，然后将人体所有骨骼关节点的角度信息作为人体姿态的特征描述向量。

然后根据人体姿态识别和人体姿态特征向量的特点，提出一种人体姿态识别的分类策略。

最后介绍了本文所有实验用到的实验数据集。

4 基于逻辑回归的人体姿态识别

通过上文的描述，本文提出的人体姿态识别已经将一个多分类问题转换为一个二分类问题，而逻辑回归模型是最为经典的二分类模型之一，所以本方法首先采用逻辑回归来进行人体姿态识别。

4.1 逻辑回归模型

逻辑回归模型是最常用的一种经典二分类算法，该模型可以根据自变量计算出该变量数据每种类别的概率，一般常用的是二项逻辑回归，即待识别的类别只有 0 和 1 两类。

其条件概率分布为：

$$P(Y = 1|x, w) = \frac{e^{w^T x + b}}{1 + e^{w^T x + b}} = \frac{1}{1 + e^{-(w^T x + b)}} \quad (4.1)$$

$$P(Y = 0|x, w) = \frac{1}{1 + e^{(w^T x + b)}} \quad (4.2)$$

其中， $x \in R^n$ 是输入，称为实例的特征； $Y \in \{0, 1\}$ 是输出，本文中两个分类对应的是同一姿态和不是同一姿态； $w \in R^n$ ， $b \in R$ 是参数， w 称为权值向量，对应于每个输入特征的权值， b 称为偏置。

将权值向量和输入向量加以扩充，即 $w = (w^{(1)}, w^{(2)}, \dots, w^{(n)}, b)^T$ ， $x = (x^{(1)}, x^{(2)}, \dots, x^{(n)}, 1)^T$ ，其中 $w^{(i)}, x^{(i)}$ 分别表示 w, x 向量的第 i 维。此时逻辑回归模型变为：

$$P(Y = 1|x, w) = \frac{1}{1 + e^{-w^T x}} \quad (4.3)$$

$$P(Y = 0|x, w) = \frac{1}{1 + e^{w^T x}} \quad (4.4)$$

权值向量 w 表示的是输入中每一维向量对分类结果的影响程度，当输入项对应的权重为正数时，表示此维特征增大了结果的概率，为负数时，表示此维特征减小了结果的概率；权重绝对值越大时，表示对结果的影响越强，反之权重绝对

值越小时，表示该项对结果影响越小。

为画出 $P(Y = 1|x, w) = \frac{1}{1+e^{-wTx}}$ 的图像，我们抽取出其原型 $f(x) = \frac{1}{1+e^{-x}}$ ，该函数被称为逻辑函数，曲线如图 4-1 所示，可见其值是一个介于 0 和 1 之间的小数。

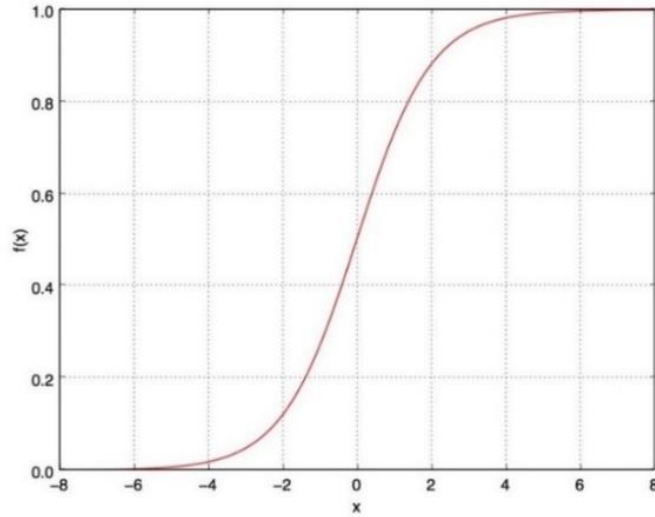


图 4-1 逻辑函数曲线图

Fig 4-1 The image of logical function

对于给定的训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，可以应用训练的方式求出模型参数 w ，从而得到逻辑回归模型，然后对于一个实例输入 x ，按照该模型可以计算出 $P(Y = 1|x, w)$ ，这是一个介于 0 和 1 之间的数，此时设定一个阈值 σ ，如果 $P(Y = 1|x, w)$ 大于阈值 σ 则判定该实例的类别为 1，如果小于阈值 σ 则判定该实例的分类为 0。

4.2 模型训练

首先构造训练数据 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，其中 $x_i \in R^n, y_i \in \{0, 1\}$ ，然后应用最大似然估计法来估计模型参数。

首先，利用最大似然估计可以得到样本预测的损失函数为：

$$\text{Cost}(h_w(x), y) = \begin{cases} -\log(h_w(x)) & y = 1 \\ -\log(1 - h_w(x)) & y = 0 \end{cases} \quad (4.5)$$

其中 $h_w(x) = \frac{1}{1+e^{-w^T x}}$ 。

则所有训练样本的损失函数和为公式(4.6)、(4.7)所示：

$$J(w) = \frac{1}{n} \sum_{i=1}^n \text{Cost}(h_w(x_i), y_i) \quad (4.6)$$

$$J(w) = -\frac{1}{n} \left[\sum_{i=1}^n y_i \log(h_w(x_i)) + (1 - y_i) \log(1 - h_w(x_i)) \right] \quad (4.7)$$

使 $J(w)$ 最小值时 w 的值，为要求的最佳参数，即：

$$w^* = \text{argmin}\{J(w)\} \quad (4.8)$$

利用梯度下降法得到则 w 的更新公式：

$$w_j = w_j - \alpha \frac{d}{d_{w_j}} J(w) \quad (4.9)$$

其中，

$$\frac{d}{d_{w_j}} J(w) = \frac{1}{n} \sum_{i=1}^n (h_w(x_i) - y_i) x_i^j \quad (4.10)$$

即逻辑回归的权值 w 的迭代方程为：

$$w_j(t+1) = w_j(t) - \alpha \frac{1}{n} \sum_{i=1}^n (h_w(x_i) - y_i) x_i^j \quad (4.11)$$

其中， t 为迭代次数。

4.3 分类识别

通过训练数据得到人体姿态识别的逻辑回归模型，则基于逻辑回归的人体姿态识别的过程如图 4-2 所示。

首先，按照下式求出待识别姿态与需要识别的姿态的特征向量的差的绝对值：

$$\phi_{input_i} = (|x_{input}^{(1)} - x_i^{(1)}|, |x_{input}^{(2)} - x_i^{(2)}|, \dots, |x_{input}^{(n)} - x_i^{(n)}|) \quad (4.12)$$

其中， ϕ_{input_i} 为逻辑回归模型的输入参数， $x_{input}^{(j)}$ 表示待识别姿态的姿态特

征向量的第 j 个特征, $x_i^{(j)}$ 表示第 i 个需要识别人体姿态特征向量的第 j 个特征。

之后, 将求得所有 ϕ_{input_i} 输入构建好的逻辑回归模型, 并输出预测值, 即待识别姿态与所有需要识别的姿态是否为同一姿态的置信度。

比较所有的逻辑回归的输出值, 并找到置信度最高的姿态, 并且如果置信度满足一定条件, 则认定待识别姿态是该姿态。

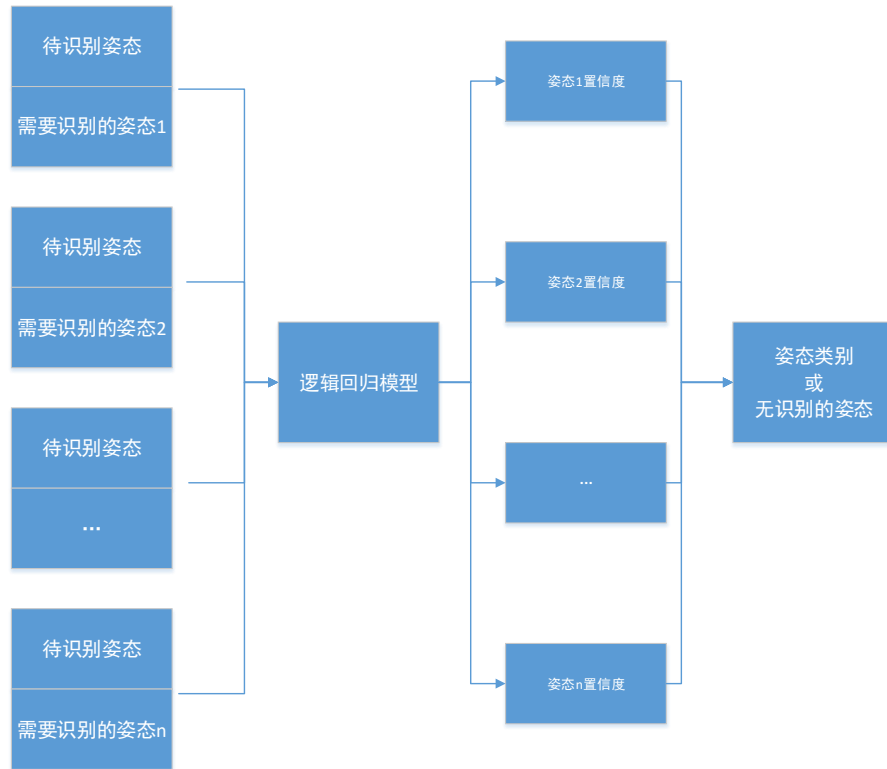


图 4-2 基于逻辑回归的人体姿态识别流程图

Fig 4-2 Pose Recognition based on Logical Regression

4.4 实验与结果分析

本实验采用前文提到的采集到的 7 万个人体姿态数据, 其中 35000 个数据用于逻辑回归模型的训练, 其余 3.5 万个数据用于测试。

首先通过实验得到逻辑回归的 ROC 曲线, 如图 4-3 所示。图中 x , y 轴的值计算如下:

$$X \text{ 轴: } FPR = \frac{\text{分类器错认为正类的负实例}}{\text{所有负实例}} \quad (4.13)$$

$$Y \text{ 轴: } TPR = \frac{\text{分类器所识别出的 正实例}}{\text{所有正实例}} \quad (4.14)$$

根据 ROC 的特性, 我们知道当 ROC 曲线距离左上角(100,100)越近越好, 因为距离左上角点最近的说明: 此时模型在拥有一个比较高的识别率的情况下的误事别率最低。从图 4-3 曲线, 可以看出该模型, 当 TPR 也就是 Y 轴到达 84% 的时候, 其 FPR 也达到了 13% 左右。因此可以看出利用逻辑回归模型的来判断两个人体姿态是否是同一姿态的效果并不理想。

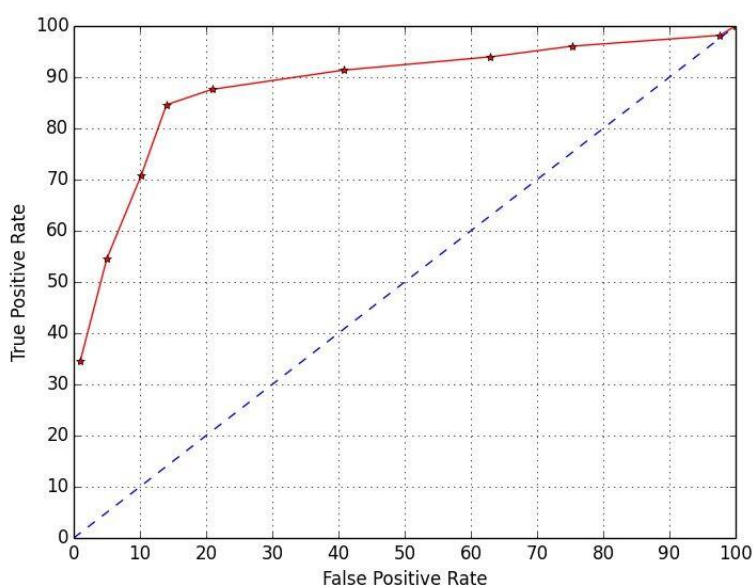


图 4-3 基于逻辑回归的人体姿态识别结果

Fig 4-3 The recognize result based on logical regression

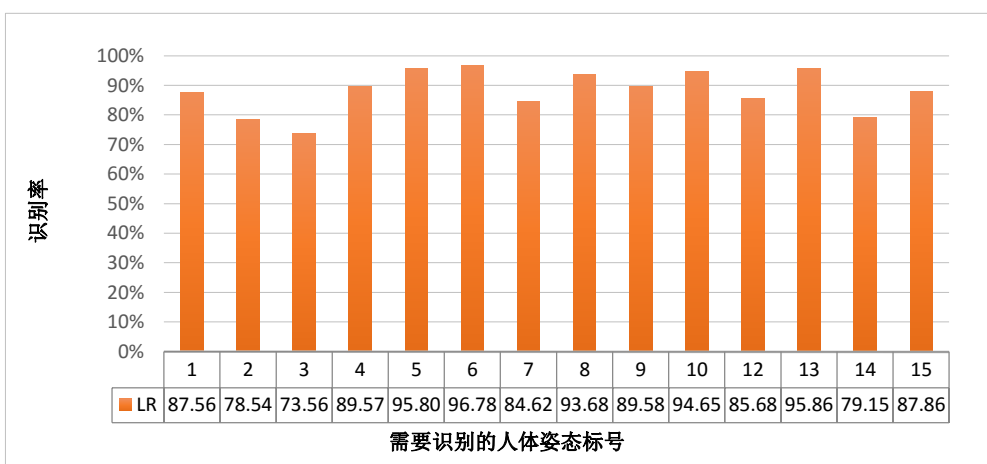


图 4-4 15 个需要识别的人体姿态的识别率

Fig4-4 The recognition rate of 15 poses

然后实验按图 4-2 所示流程对需要识别的 15 个姿态进行姿态识别，最后统计的识别率如下图 4-4 所示。

从图 4-4 中可以看出，基于逻辑回归的人体姿态识别的识别率不稳定，即有些人体姿态的识别率比较高，如姿态 5 的识别率可以达到 95%以上，但是有几种人体姿态的识别率非常低，如姿态 2，姿态 3，姿态 14 的识别率还不到 80%，这么低的识别率是无法实用化的。

因此在上面实验的基础上，又取出几种识别率比较低并且比较相似的姿态，这几个姿态的标号为 2、3、7、12、14。如图 4-5 所示，观察可知，这几个姿态是具有较大的相似性。

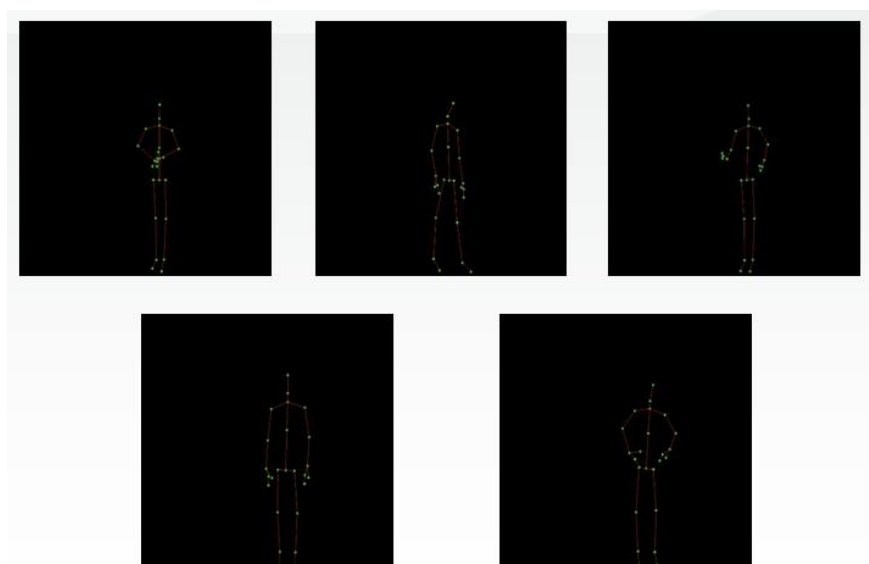


图 4-5 姿态 2、3、7、12、14

Fig4-5 The pose 2、3、7、12、14

为了对其进行客观评价，表 4-1 列出了姿态的标号为 2、3、7、12、14 的混淆矩阵。从表 4-1 可以看出，基于逻辑回归的人体姿态识别在相似姿态之间的识别会产生特别大的干扰，有些姿态，例如姿态 2 和姿态 7，姿态 3 和姿态 12，识别成其相似姿态的比率会超过 10%。可以看出当我们需要识别的姿态增多时，而且相似姿态也增加时，其识别率的影响会越来越大。

表 4-1 姿态 2、3、7、12、14 的混淆矩阵

Tab 4-1 The Confusion matrix between pose 2、3、7、12 and 14

预测值(%) 实际值(%)	姿态 2	姿态 3	姿态 7	姿态 12	姿态 14
姿态 2	78.5	2.4	10.7	2.1	4.9
姿态 3	4.5	73.5	2.2	13.2	4.3
姿态 7	7.2	1.5	84.6	1.3	4.4
姿态 12	1.8	3.5	0.4	85.6	1.4
姿态 14	7.7	5.9	2.3	1.2	79.1

为了更加清晰的看出相似姿态之间的互相影响，我们又做如下实验，利用逻辑回归模型，计算事先标注好为姿态 2 的所有测试样本，与姿态 2，以及姿态 7 是否是同一姿态的逻辑回归值。姿态 2 和姿态 7 如图 4-6 所示。姿态 2 是一个拍手的姿态，而姿态 7 是一个跑步的起手动作。

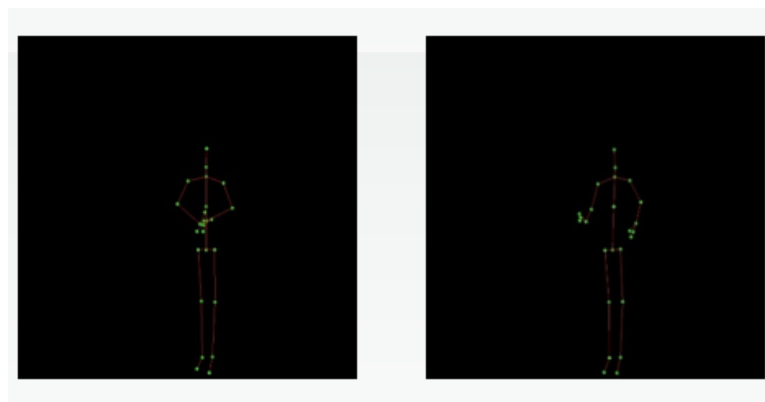


图 4-6 标准姿态 2、7

Fig4-6 The standard pose 2、7

然后，按照这些测试样本与姿态 2 的逻辑回归值对所有样本从低到高进行排

序，得到如图 4-7 所示结果。

图 4-7 中，红色点为测试样本姿态与姿态 2 的逻辑回归值，蓝色点为测试样本姿态与姿态 7 的逻辑回归值，绿色线条为本文采用的认为是同一姿态的逻辑回归模型的阈值。

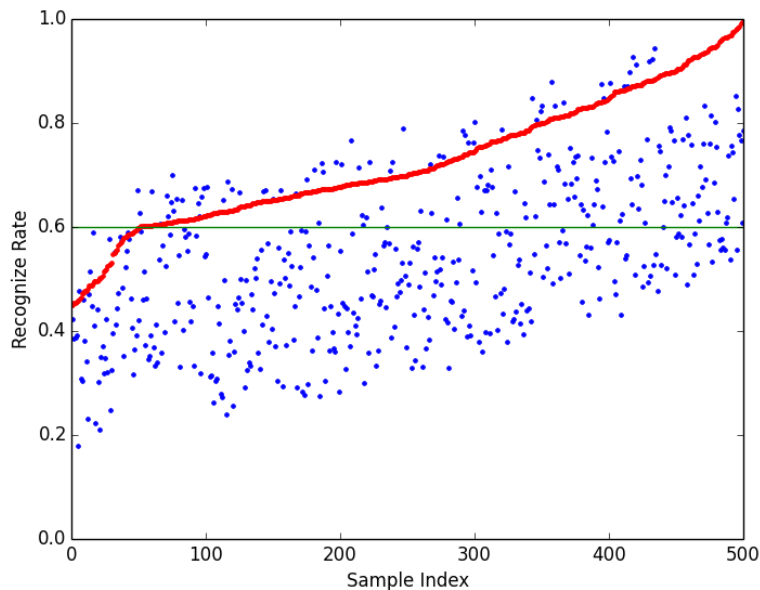


图 4-7 姿态 2 样本集与标准姿态 2 和标准姿态 7 的逻辑回归值

Fig 4-7 The logistic value between pose 2、7 and samples of pose 2

从图 4-7 中可以看出，有 40%左右的蓝色点在绿色阈值线之上，即该模型会把接近 40%姿态 2 的测试样本与姿态 7 认为是同一姿态；还有 10%左右的红色点在绿线以下，即该模型会把接近 10%的事先标注为姿态 2 的测试样本认为与姿态 2 认为不是同一姿态，导致了约 10%的漏检率。虽然通过曲线可以看出，大部分红色的点都在蓝色点之上，也就是说大部分样本对姿态 2 的置信度会高于姿态 7，但是仍有 10%左右的样本会出现姿态 7 的置信度高于姿态 2，当出现这种情况的时候，就会出现识别错的情况。

综上所述，逻辑回归模型很难满足本文提出的人体姿态识别的分类策略的需求，因为该模型对于逻辑回归值得选取比较敏感，但是从 ROC 曲线可以看出，如果选择的阈值太低，就会导致错误率提高，而如果阈值太高，准确度又不够。而且该模型对于相似姿态没有很好地区分度。

4.5 小结

本章首先介绍了逻辑回归模型，然后简单介绍了逻辑回归模型的参数求解过程。

后面介绍了如何利用逻辑回归判断两个任意姿态是否是同一姿态，然后根据本文采用的实验数据进行人体姿态识别。

最后对实验数据进行分析，发现逻辑回归模型虽然比较简单快捷，但是当识别姿态增加或需要识别的姿态比较相似时，该模型的识别率是下降很快。

5 基于随机森林的人体姿态识别

通过对逻辑回归的人体姿态识别的实验数据的分析,确认其存在识别率偏低的问题,为此,本文提出了利用随机森林算法来训练分类器来判断两个任意姿态是否是同一姿态。

5.1 随机森林算法

随机森林算法是一种决策树的集成学习算法,它是由统计学习届的大师级人物 Leo Breiman 提出的^[51]。随机森林模型是由多个分类决策树组集成的分类器,每一个决策树都是独立同分布的随机向量来决定树的生长。最终由所有决策树的共同表决结果来决定模型的最终输出结果。

随机森林模型是多个随机决策树(Randomized Tree)的组合,这些随机决策树同分布且互相独立,而且模型在训练时,对样本的选择和树的生长过程中引入随机性,以此来降低各个树结构分类器的较高方差。

如图 5-1 所示,随机森林是由多个决策树构成的分类器,模型在构建每一个

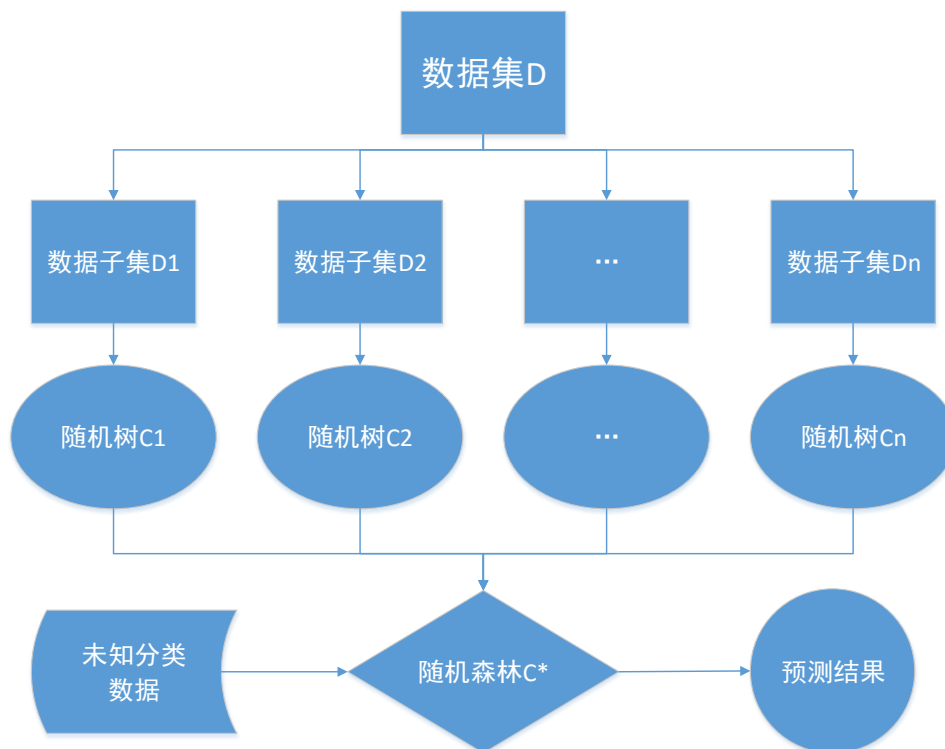


图 5-1 随机森林模型

Fig 5-1 Rondon forest model

分类器时，都从原始的训练集中有放回地随机取出一部分样本子集作为该决策树的训练子空间，然后再在这个随机子空间上建立决策树，而且在树的生长过程中，每一次分裂使用的特征都是在随机选取的特征子空间中选取的特征，最后在每一个样本子集上都建立一个决策树模型，然后最终的模型可以认为是这些树的合集，分类结果这些树模型的分类结果来共同决定的。

5.1.1 算法的随机特性

针对不同的问题，随机森林算法包括分类和回归两类，由于本文给出的识别模型只涉及分类问题，因此下文提到的“随机森林”不加说明即指随机森林分类算法。对随机森林算法的完整描述参考文献^{[51][52]}。

随机森林是一组随机决策树的组合而成，并且它们彼此独立且有较大差异。其中的随机树在训练时，会按传统分类回归树的训练方式生长到最大深度，但是不进行剪枝。其中其随机性主要体现在以下三个方面。

第一，随机选择训练样本子集

每个决策树使用的训练样本集是采用多次有放回的抽样方式构成的，这样森林中每棵树使用的训练样本都是有差异的，其本质上是 **Bagging** 集成学习思想。**Bagging** 可以有效提高不稳定学习算法的分类准确性，即降低学习算法的方差^[51]。

第二，随机抽取特征子集

森林中每棵树使用的特征都是从所有样本中随机抽取的特征集合，这样可以使每棵树之间都存在差异性。

第三，不剪枝

所有的树模型都是让其自由生长，不进行剪枝。但是每棵树的深度都是一定的，这样就可以使得每棵树都是一个弱分类器，使得模型更加不会出现过拟合的情况。

由于随机森林这几方面的随机性，使得随机森林模型相对于传统决策树模型而言，在效果上有很程度的改进，传统的决策树模型在数据量稍微大点的样本集合上，很容易出现枝繁叶茂的情况，使得生成的最终模型过于复杂，也就很容易出现过拟合的情况。而随机森林的思想就是训练出很多个在某一方面有很强决

策能力的决策树集合，这样每个决策树几乎不会出现多度复杂和过拟合的问题，相对于总体样本数据而言，每个决策树都是一个弱决策树，但是多个方面的弱分类器集合就可以形成一个异常强大的分类器。

5.1.2 泛化误差与袋外数据估计

首先，对相关的概念进行介绍。

■ 泛化能力

机器学习的目的是提取出隐藏在数据后的固有规律，对具有同一规律的学习样本集之外的数据，学习出的分类器也可以给出准确的预测结果，该能力称为泛化能力。泛化能力是指模型对未在训练集中出现的样本作出正确预测结果的能力，概括地说，所谓泛化能力（**generalization ability**）是指学习的模型对新样本的适应能力^{[50][51][52]}。

■ 泛化误差

泛化误差（**Generalization Error**）是泛化能力的一个指标，当泛化误差越小，则该模型性能越好，反之则性能越差。

在实际环境中，样本数据的期望输出和样本的分布情况一般都是未知的，因此很难直接通过计算泛化误差来得到随机森林模型的泛化能力。目前用来评估机器学习模型泛化误差的方法主要有两种，一种是分析数学模型（**Analytical Model**），第二种就是交叉验证（**Cross-Validation, CV**）。

分析数学模型一般只能对随机森林中一些简单的单一性线性分类问题比较有用，而对于随机森林的有效参数个数很难做出很好的估计，因此一般很少采用这种方式。

而交叉验证是通过将所有的样本集划分为训练数据集和测试数据集，并且所有的数据集都是标注过准确输出的，因此可以通过测试集来估计泛化误差。而在交叉验证的所有算法中，一般采用 **k-fold** 交叉验证的方法。

■ 袋外数据估计

而在估计随机森林模型泛化误差时，有一种比较方式是使用袋外数据估计。

随机森林是使用 Bagging 的方法对整体样本集进行有放回的随机抽取生成的，因此抽取的这些数据集中，有一些样本是不会从原始数据集中被抽取到的，这些样本的个数是初始数据集的 $(1 - 1/N)^N$ （其中 N 为初始训练集中样本的个数）。而且当 N 足够大时， $(1 - 1/N)^N$ 将收敛于 $1/e \approx 0.368$ ，因此把这些不能从初始数据集中抽取出来的约 37% 样本组成的集合，称之为袋外数据，简记为 OOB。而使用 OOB 数据来估计随机森林模型的泛化能力，称之为 OOB 估计。

Breiman 通过实验已经证明，OOB 估计是随机森林模型泛化误差的一个无偏估计^[49]。而且有大量的实验数据表明^{[51][52]}，OOB 估计与交叉验证相比，OOB 估计的效率比交叉验证的效率要高很多，而在估计效果上两者是差不多的，因此，在对随机森林进行泛化误差的估计时，使用 OOB 估计就可以了。

5.2 随机森林模型的理论概述

5.2.1 随机森林的数学定义

设一组随机决策树 $h_1(x), h_2(x), \dots, h_k(x)$ 形成的随机森林和两个随机向量 X 和 Y 。

定义 1 边缘函数

边缘函数定义如下：

$$mg(x, y) = av_k(I(h_k(X) = Y)) - \max_{j \neq Y} av_k(I(h_k(X) = j)) \quad (5.1)$$

其中， $I(\cdot)$ 为示性函数， Y 为正确分类的向量， J 为不正确分类的向量， $av_k(\cdot)$ 表示平均。

边缘函数表示的是正确分类得票数大于不正确分类得票数的程度。因此，从共识可以看出当边缘函数的值越大，分类器的分类置信度就越高。

定义 2 泛化误差

泛化误差定义如下：

$$PE^* = P_{X,Y}(mg(X, Y) < 0) \quad (5.2)$$

其中，下标 X, Y 表示概率定义空间。

定义 3 随机森林模型的边缘函数

随机森林的边缘函数定义如下：

$$mr(x, y) = P(h_k(X) = Y) - \max_{\substack{j \neq Y \\ j=1}} P(h_k(X) = j) \quad (5.3)$$

其中， $P(h_k(X) = Y)$ 表示正确分类的概率， $\max_{\substack{j \neq Y \\ j=1}} P(h_k(X) = j)$ 为错误分类概率的最大值。

5.2.2 随机森林的性质

■ 随机森林的收敛性

定理 1

根据边缘函数和泛化误差的定义，以及大数定律和随机决策树结构可知，随机森林模型的泛化误差将收敛于公式(5-4)：

$$\lim_{n \rightarrow \infty} PE^* = P_{xy}(P_\theta(k(X, \theta) = Y) - \max_{j \neq Y} P_\theta(k(X, \theta) = j) < 0) \quad (5.4)$$

其中， n 为随机森林中决策树的个数。

定理 1 表明随着随机决策树个数的增加，随机森林模型的泛化误差 PE 会收敛于一个值，这表示随机森林对未知新样本有很好的泛化能力，也就是说它可以有效抑制模型过拟合。这也说明了为什么在随机森林模型在随着决策树个数的增加的同时没有发生过拟合的情况。

■ 决策树的相关度与强度影响算法的泛化误差

定理 2

泛化误差存在上界：

$$PE^* \leq \frac{\bar{p}(1-s^2)}{s^2} \quad (5.5)$$

其中， \bar{p} 为决策树之间相关度平均值， s 是决策树平均强度。

从定理 2 可以看出要使随机森林模型的泛化性越好,就应该尽量减少每个决策树与别的决策树的相关性,并且尽量增加每棵决策树分类性能。

5.3 人体姿态识别的随机森林模型

由于本文提出的人体姿态识别的识别方法是对任意两个人体姿态,去识别它们是否是同一姿态,所以我们需要构造的随机森林模型其实是一个二分类问题,输入节点 $\phi = (x_1, x_2, \dots, x_n)$, 其中 x_i 是任意两个姿态特征向量的差的绝对值,输出节点是 0 或 1, 其中 0 代表两个姿态不是一个姿态, 1 代表两个姿态是同一姿态。

随机森林的分类效果是多棵决策树的融合结果,而在构建每一棵决策树时都要从训练集中随机抽取一部分样本,然后进行特征提取进行训练。而由于本文提出的人体姿态识别的分类策略,导致我们构造出来的训练样本中正负样本极不平衡,正样本(相同姿态的样本)明显少于负样本(不同姿态构成的样本)。

因此,在构建人体姿态识别的随机森林模型之前,我们还需要解决以下几个问题:

1) 当训练的正负样本数不平衡的时,这样决策树特征选取的结果会选择对负样本有利的特征而掩盖对区分正样本有帮助的特征,所以我们需要选择一个合适的抽取训练子空间的方法来解决正负样本不平衡的问题。

2) 随机决策树在构建过程中特征子空间、分裂准则、最大深度以及停止分裂的问题。

3) 由于随机森林是由很多个决策树构成的,所以选择一个合适的森林规模(随机决策树的个数)对于最后的识别率来说影响很大。

4) 多个决策树的投票方式的问题。

对于以上问题,我们首先解决由于样本不平衡导致的问题。

5.3.1 随机森林的训练子空间

为了选出不平衡数据集中对正样本影响较大的特征,我们可以借助集成学习的思想来做,通过调整正负样本采样的比例来解决该问题,让采样出来的负样本数量与正样本数量相近,这样的话,采样出来的负样本和原有数据的正样本就能

够组成一个相对来讲比较平衡的数据子集。

因此本文采用了对负样本采样，而正样本会参加所有决策树的构建，以降低负样本的规模所带来的对特征选取算法的影响，这样同时极大的提高了正样本参与有效特征选取的概率，这样每次 Bagging 的多次采样过程中正样本是全部参与 Bag 子集中去。而且使得在训练的时候，每一棵树的输入样本都不是全部的样本，使得相对不容易出现过训练。

由于采样方式是有放回的，负类样本被选择的概率是一样的，一些负类样本可能在同一个数据集中出现多次，而还有一些负样本可能就不会出现在所有的数据子集中，但是即便是出现这种情况，也不会对模型的分类效果行程太大的影响。而且对于有噪声的数据，Bagging 的方式能够很好的抑制过拟合现象。流程图如图 5-2 所示。

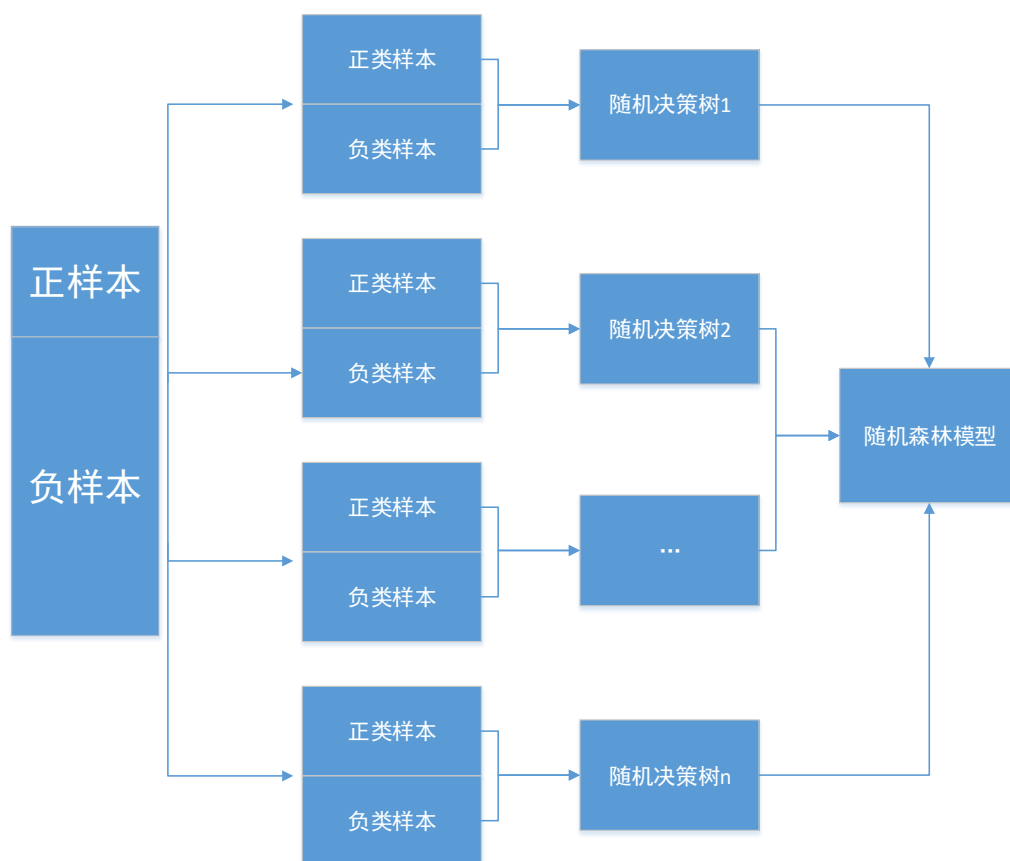


图 5-2 随机森林模型训练子空间选择示意图

Fig 5-2 Random Forest model training subspace selection schematic

通过上面的 Bagging 方法，可以达到稀释了每个 Bag 中负样本的数据的目的，

使负样本和正样本在参与特征选取时可以作为同等性的考量，产生的多个 Bag 也必然导致正类样本很大程度上能够被特征选择算法考量，同时综合考量采样的负样本，最终也不会减弱算法选取对负样本有影响力的特征集合的能力。

5.3.2 随机决策树的构建

通过上 Bagging 方法，这样在构建每个随机决策树的训练样本就是利用上面的方法产生一个训练样本子集作为该决策树的训练样本。

在构建随机决策树之前，我们先对决策树做简单的介绍，决策树算法是一种分类预测算法，它实际上是将特征空间用超平面进行划分的一种方法，每次分割的时候，都将当前的空间一分为二，以图 5-3 所示的决策树为例，根节点只判输入的整数的值是大于等于 3，还是小于 3；之后，再判大于等于 2，还是小于 2；大于等于 4，还是小于 4；最后，这个决策树，能够判别输入的数字究竟是 1，还是 2，3，4。

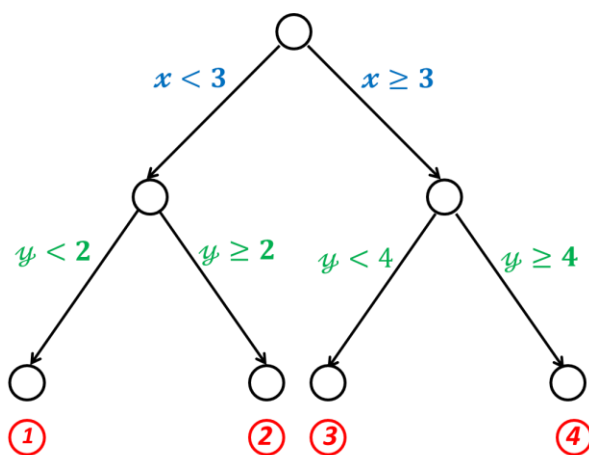


图 5-3 决策树示意图

Fig 5-3 The decision tree diagram

按照图 5-3 的思路，可以拓展认为，算法通过样本的训练将特征空间划分成图 5-4 所示。这样使得每一个叶子节点都是在空间中的一个不相交的区域，在进行决策的时候，会根据输入样本每一维 feature 的值，一步一步往下，最后使得样本落入 N 个区域中的一个（假设有 N 个叶子节点）。

在构造决策树时必须找到一种合适的属性选择度量方法来进行树的分裂。本

文算法采用经济学中的基尼指标（Gini）作为属性选择度量，选择最大化不纯度降低（具有最小 Gini 指标）的属性选作分裂属性。

Gini 指标度量训练元组集合 D 的不纯度，定义为：

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (5.6)$$

其中， p_i 是数据集 D 中属于 C_i 类的概率。设是 n_i 是 D 中属于类 C_i 的样本数，则 $p_i = n_i/|D|$ 。

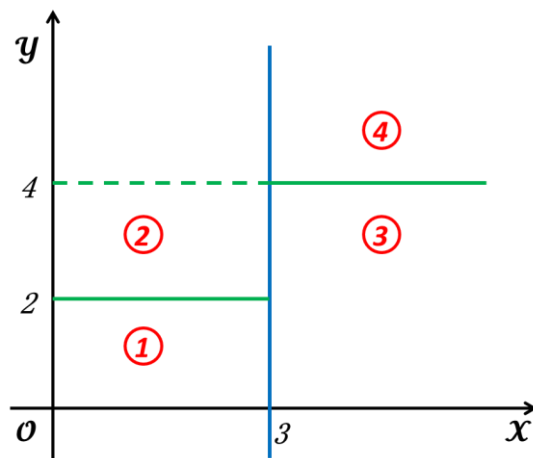


图 5-4 决策树特征空间划分
Fig 5-4 The feature space partition of decision tree

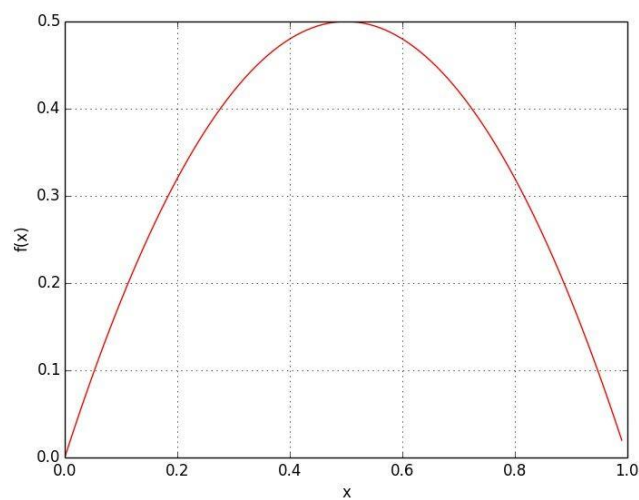


图 5-5 基尼不纯度曲线
Fig 5-5 The gini curve

在决策树生长的过程中,假设属性 A 的某种二元分裂将 D 划分成 D_1 和 D_2 ,则此次分裂的不纯度下降落差为:

$$\Delta Gini(D) = Gini(D) - \frac{|D_1|}{|D|}Gini(D_1) - \frac{|D_2|}{|D|}Gini(D_2) \quad (5-7)$$

其中, $\Delta Gini(D)$ 为不纯度下降落差。

本文随机森林的决策树训练采用的分裂方式是采用那个使不纯度下降最快的那个属性,即在决策树分裂过程中遍历所有特征向量的属性,找出那个使不纯度下降最快的属性作为此次分裂属性。

但是在建立随机决策树的过程中,还有三点需要注意:

■ 特征采样

特征采样就是随机从特征向量中随机采样一部分特征作为该决策树的分类特征,本文的特征向量是 60 维的人体姿态特征,我们采用的采样方式是无放回的在 60 维特征中随机选取 40 个特征。

■ 决策树的最大深度

根据 Leo Breiman 的建议,假设特征总数量为 m ,则树的最大深度比例可以是 \sqrt{m} , $2\sqrt{m}$, $\sqrt{m}/2$,本方法最终采用最常用的 \sqrt{m} 的方式来确定每棵决策树在训练时的最大深度。

所以本方法的在每个 Bag 的特征选择的方法是采用在所有 m 个特征中随机的无放回的抽取 \sqrt{m} 个特征作为该 Bag 的特征集合。

■ 完全分裂

就是对采样之后的数据使用完全分裂的方式建立决策树,使决策树在某一个叶子节点上要么是无法分裂,要么里面的所有样本都指向的同一类别。

综上所述,本文每棵决策树的构建过程如图 5-5 所示。

首先,采用有放回采样的方式,从负类样本中随机抽取与正样本个数相同的样本 N,则每棵随机树样本总数为 2N。

然后,在 60 个人体姿态特征中随机抽取 40 个特征,作为此次随机树的分裂特征。

最后,进行决策树训练,决策树的分裂准则以及停止条件采用上文提到的方

法。

按这种算法得到的随机森林中的每一棵都是很弱的，但是每一棵决策树都是一个精通于一个不同领域的专家，因此在随机森林模型中就有了很多个精通于不同领域的专家，这样，在对一个新的问题，模型就可以用不同的方向去理解它，最终由所有专家投票得到最终结果。

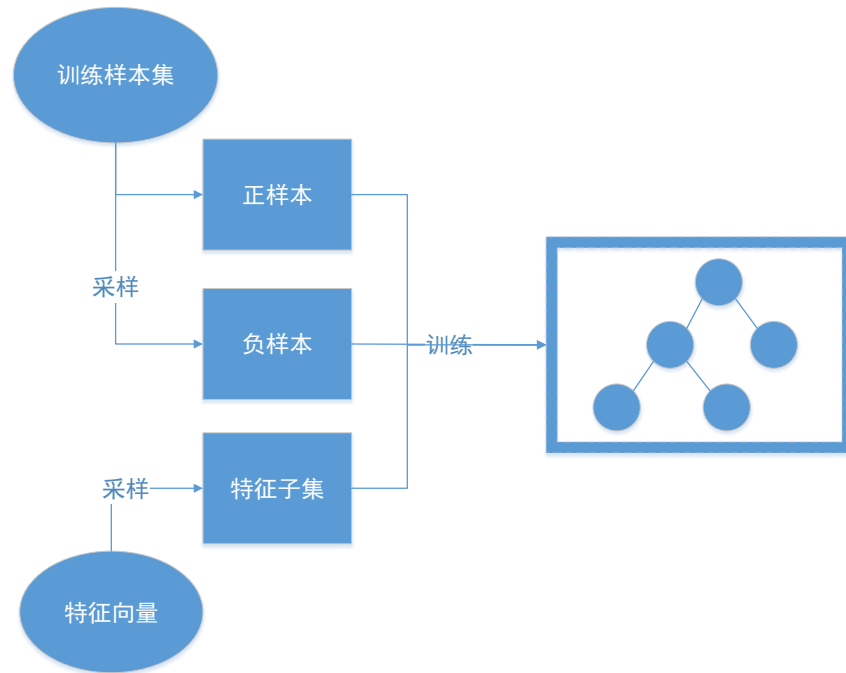


图 5-5 随机决策树构建过程
Fig 5-5 The build process of decision tree

5.3.3 分类识别

通过上述方式，便可以构造出很多个随机决策树用于人体姿态的分类识别，假设随机决策树的个数为 N ，则可以利用这 N 个决策树构造出来的随机森林来判断任意两个人体姿态是否为同一姿态，如图 5-6 所示。

1) 求出两个姿态特征向量的差的绝对值：

$$\phi_{input} = \phi_1 - \phi_2 = (|x_{11} - x_{21}|, |x_{12} - x_{22}|, \dots, |x_{1n} - x_{2n}|), \quad n = 60$$

2) 从每棵决策树的根节点开始，按照训练好的每棵树所随机抽取的 40 个特征，以及每棵树当前节点的阈值，判断 ϕ_{input} 是进入左节点还是进入右节点，直到到达某个叶子节点，并输出预测值，预测值为姿态 1 和姿态 2 属于同一姿态的置信概率。

3) 重复执行步骤 2 直到 n 棵树都输出了预测的姿态 1 和姿态 2 属于同一自他的置信概率。

4) 对所有树输出的置信概率求平均值，则该平均值就是姿态 1 和姿态 2 是否为同一姿态的置信度。

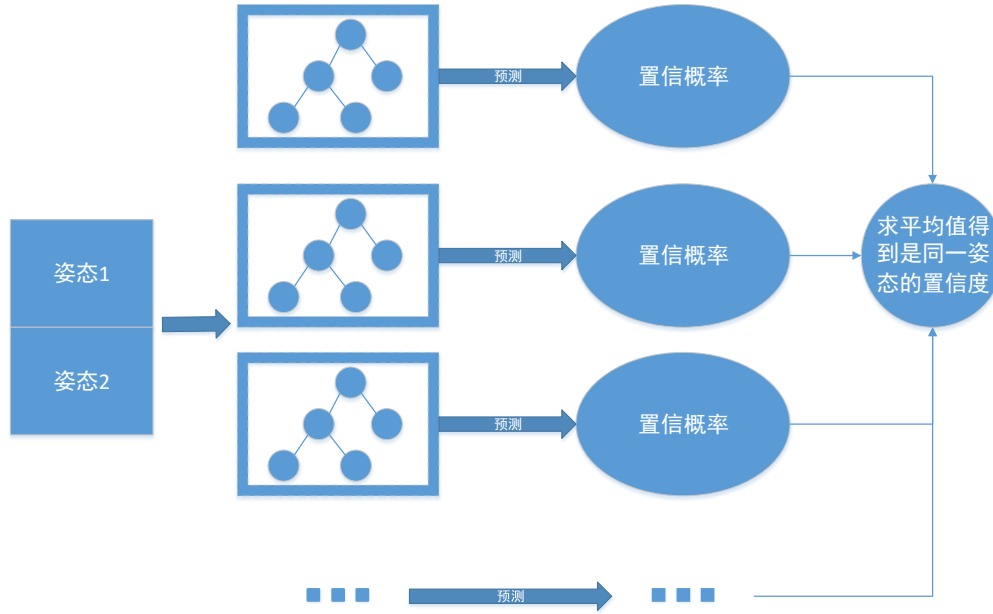


图 5-6 随机森林模型的识别过程

Fig 5-6 The recognition process of the random forest

通过上面流程可以通过构建好的随机森林模型判断任意两个姿态是否是同一姿态，所以用作人体姿态识别时，本文构建了与待识别姿态树的个数相同数量的随机森林模型，每个模型只负责给出待识别数据与某个姿态属于同一姿态的置信概率。识别流程如图 5-7 所示。

1) 首先求出待识别姿态与需要识别的姿态的特征向量的差的绝对值：

$$\phi_{input_i} = (|x_{input}^{(1)} - x_i^{(1)}|, |x_{input}^{(2)} - x_i^{(2)}|, \dots, |x_{input}^{(n)} - x_i^{(n)}|)$$

其中， ϕ_{input_i} 为随机森林模型的输入参数， $x_{input}^{(j)}$ 表示待识别姿态的姿态特征向量的第 j 个特征， $x_i^{(j)}$ 表示第 i 个需要识别人体姿态特征向量的第 j 个特征。

2) 将求得的所有 ϕ_{input_i} 输入构建好的与待识别姿态个数相同的随机森林模型中，按照图 5-6 给出的流程，分别给出待识别数据与所有需要识别的姿态是否

为同一姿态的置信概率。

3) 比较所有的随机森林模型的输出的属于同一姿态的置信概率，并找到置信概率最高的姿态，为可能的归属姿态。

4) 若这个最高的姿态自信概率足够大，满足预设条件，则认定该测试数据为该姿态，否则，认为该测试数据的姿态不在给定的姿态中，是一个新的姿态。

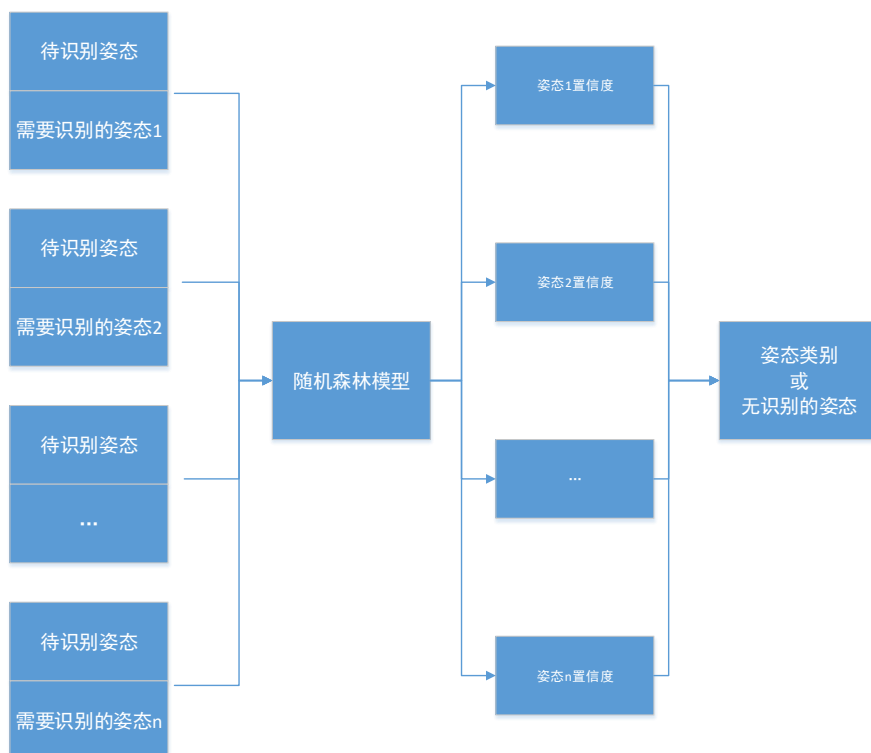


图 5-7 基于随机森林的人体姿态识别流程图

Fig.5-7 Pose Recognition based on Random Forests

5.3.4 随机决策树个数的选择

在确定了随机森林中每棵决策树的构建方式以及分类识别方法之后，最后就是要确定随机森林的模型大小，也就是随机决策树的个数。因为对于模型来说，模型对未知参加训练的样本的识别效果以及泛化能力最好的模型才是最优的模型，本文采用的方法是区域搜索加交叉验证的方式。

具体方法如下：

1) 设决策树的数量范围大概在 $[n_{min}, n_{max}]$ ，在这个范围内，从小到大，每

次增加 Δ ，则可以构建得到 $(n_{max} - n_{min})/\Delta$ 个候选的随机森林模型。

2) 利用训练样本分别构建步骤 1 得到的不同个数的随机森林模型，然后再用测试样本去测试模型，最后在这些候选的随机森林模型中，综合考虑时间和识别率挑选一个最合适的决策树个数。

5.4 实验结果分析

本节主要对上文的算法运行效果进行验证，以证明本文识别算法的有效性。实验数据为第三章说明的采集到的姿态数据，然后根据数据对本文提出的方法进行验证，并进行最后实验结果分析。

在进行算法印证之前，首先需要得到本文提出的随机森林模型，表 5-1 是本文最后使用的模型训练及识别参数的汇总表。

表[5-1] 随机森林参数汇总表
Tab[5-1] The parameters of the random forest

随机森林训练及识别参数	参数值
训练样本子集获取方式	对负样本抽样，正样本使用所有样本
样本特征总数	$K=60$ (两个姿态的向量的差的绝对值)
随机决策树特征子集	$k=0.6K=36$ (无放回随机抽取)
随机决策树最大深度	$D=\sqrt{K}=7$
决策树分裂方式	使 Gini 系数下降最快的那个特征
随机森林规模（决策树个数）	区域搜索的方法
随机森林投票方式	所有决策树的输出结果的比率

从表 5-1 可以看出，需要得到随机森林模型首先需要得到随机森林中决策树的个数，因此本文先通过区域搜索的方式得到决策树的个数。

5.4.1 随机森林中决策树个数的确定

首先初始化决策树的个数范围在 60~200 个，每次递增 10 个，这样便可以得 15 个随机森林模型，然后取出样本库中一半的数据进行训练，另外一半的数据进行验证，最后得到的识别率曲线如图 5-8 所示。

从图 5-8 可以看出随机森林模型中，当一个随机森林模型的决策树的个数只有 60 个时，由于每个决策树都是很弱的分类器，模型对人体姿态的特征空间描述能力太差，识别率很低，只有不到 80%。而随着随机树个数的增加，识别率大幅度提升，当随机森林模型的规模到达 150 个随机树之后，模型开始稳定，识别率已经很高，达到了 98% 以上，而且当随机树个量再增多时，该模型仍不容易发生过拟合。所以最终综合考虑时间和识别的效率，最后随机森林的模型使用 150 个决策树。

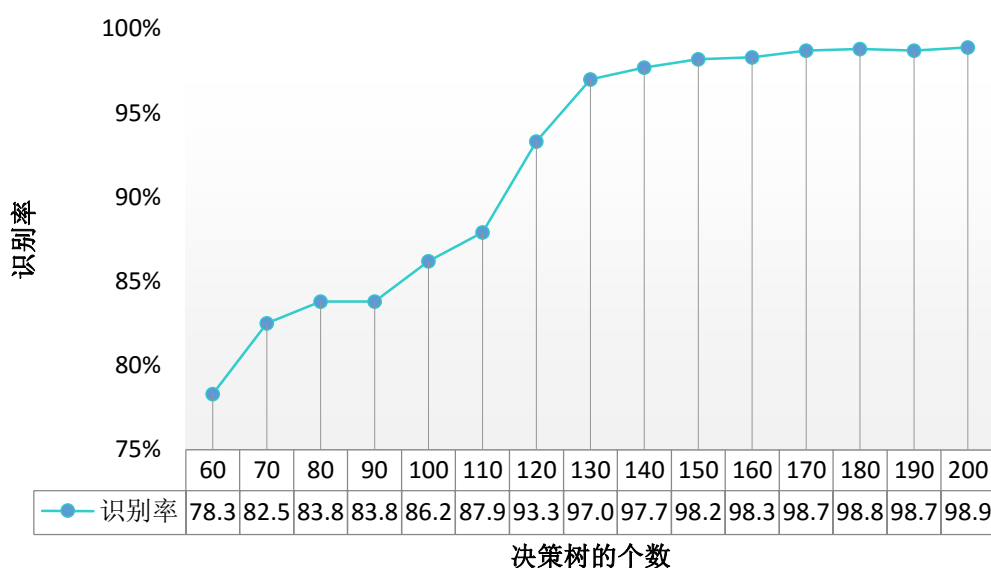


图 5-8 随机决策树个数与识别率的关系

Fig 5-8 The relationship between the tress number and recognition rate

5.4.2 逻辑回归与随机森林的实验结果对比

本实验首先对得到的随机森林模型，与第四章提出的逻辑回归模型的 ROC 曲线进行比对试验。对比试验采用的人体姿态数据与第四章使用的数据相同，然后分别对两种算法分别进行测试。

如图 5-9 所示，为两种模型的 ROC 曲线，其中绿色为基于随机森林的人体姿态识别模型的 ROC 曲线，红色为基于逻辑回归的人体姿态识别模型的 ROC 曲线，从曲线可以看出随机森林的模型明显优于逻辑回归的模型，因为基于随机森林的模型，当正样本的识别率达到 98% 以上时，其识别错的比例也不超过 6%，

说明随机森林模型的特征选择可以使正负两类之间的类间距更大。

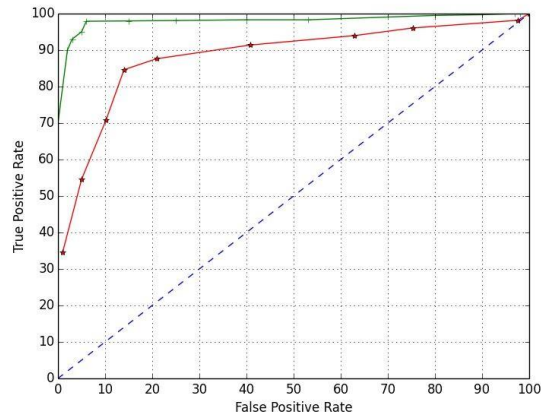


图 5-9 逻辑回归与随机森林的 ROC 曲线对比

Fig5-9 The ROC curve between LR and RF

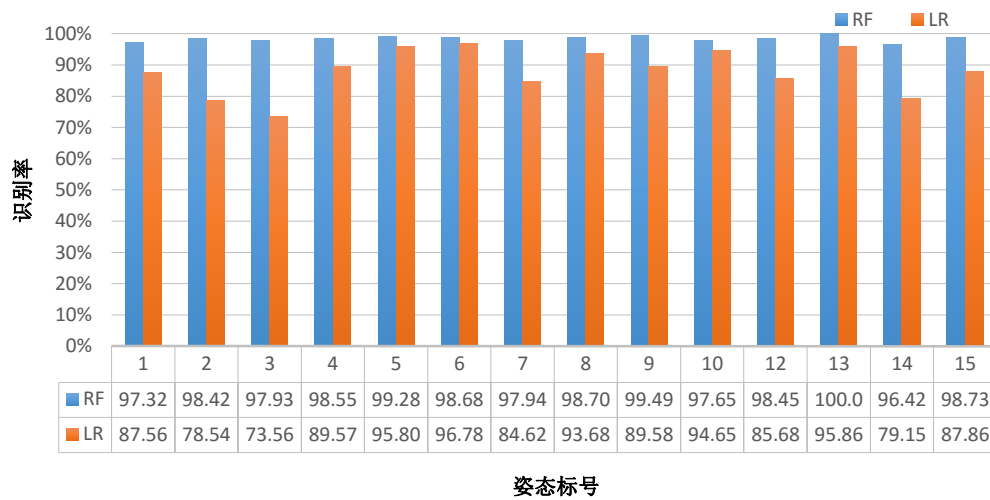


图 5-10 逻辑回归与随机森林各个姿态的识别率

Fig5-10 The recognition rate between LR and RF

然后实验对两种模型对 15 种需要识别的人体姿态的识别率进行对比试验。如图 5-10 为两种模型对 15 种需要识别的人体姿态的识别率统计。其中蓝色为随机森林模型的各个姿态的识别率，橙色为逻辑回归的各个姿态识别率，可以看出随机森林在每个姿态上的识别率都优于逻辑回归模型。并且随机森林模型对每一种姿态的识别率都高于 97%，已经可以基本满足一般的人体姿态识别需求。

为了看出随机森林模型在相似姿态之间的数据分布情况，我们同样做了利用

随机森林模型，将事先标注为姿态 2 的所有测试样本，与姿态 2 和姿态 7 是否是同一姿态的随机森林输出值，如图 5-11 所示，姿态 2 是一个拍手的姿态，姿态 7 是一个跑步的起手动作。

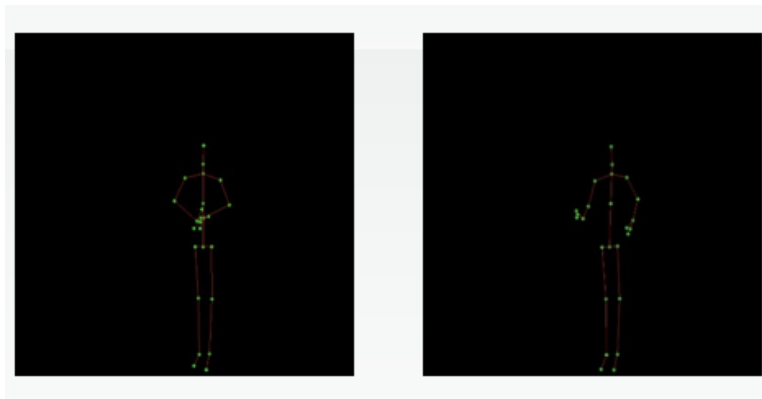


图 5-11 标准姿态 2、7
Fig5-11 The standard pose 2、7

然后按照与标准姿态 2 的随机森林输出值对所有样本进行排序，得到如下图 5-12 所示结果。图 5-12 中，红色点为事先标注为姿态 2 的测试样本与姿态 2 的随机森林模型输出的置信概率，蓝色点为事先标注为姿态 2 的测试样本与姿态 7 的随机森林模型输出的置信概率，绿色线条为本文采用的认为是同一姿态的随机森林模型的阈值。

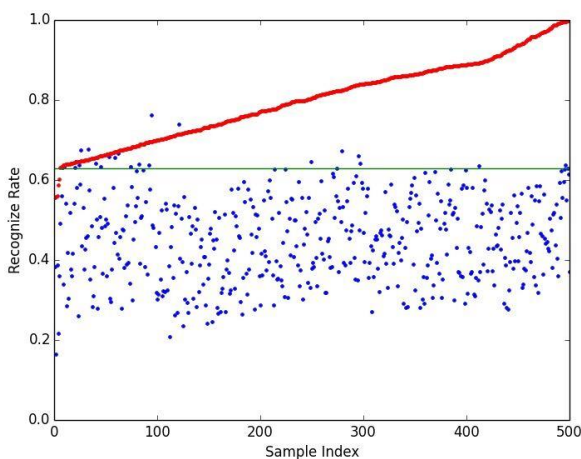


图 5-12 姿态 2 样本集与标准姿态 2 和标准姿态 7 的逻辑回归值
Fig 5-12 The logistic value between pose 2、7 and samples of pose 2

从图 5-12 中可以看出，只有极个别的红色点在绿色阈值线之上，即该模型

认为样本库中姿态 2 的 99%以上的样本与标准姿态 2 是同一姿态;虽然有一些蓝色点在阈值上面,但是其置信度大部分也低于红色点,也就是说 98%以上样本对姿态 2 的置信度会远高于姿态 7。

综上所述,该随机森林模型可以有效的识别出相同姿态,并且对于相似姿态有很好的区分度,可以基本满足要求。

5.5 小结

本章首先介绍了随机森林算法,并简单介绍了随机森林模型的构建过程。后面针对人体姿态识别的特点,在构建随机森林模型时,做了以下工作。

1) 由于正负样本的不平衡性,本文采用在构建随机决策树时,对负样本进行采用,而正样本不采样的方式,解决了数据不平衡的问题;

2) 随机决策树的个数采用区域搜索的方式,根据统计信息找到适合人体姿态识别特点的最优的决策树个数。

最后根据构建出的随机森林模型,进行人体姿态识别,并与之前的逻辑回归模型进行结果比对,结果表明,随机森林模型更加适合本文提出的人体姿态识别的分类策略,该模型可以有效抑制不同人的体态差异,并且可以有效地区分开相似姿态。

6 总结与展望

6.1 课题总结

本文首先分析讨论了如何使用 Kinect 获取人体姿态数据，以及这些姿态数据如果使用在人体姿态识别上的一些限制。

第一，Kinect 获取的是人体骨骼关节点的三维坐标，但是当人的站姿出现自遮挡时，会对姿态识别有很大的影响，基于此问题本文提出了一种双 Kinect 人体骨骼关节点数据融合的方法，解决了使用 Kinect 时出现的自遮挡问题。

第二，在实际应用中，由于不同人的体型差异很大以及人与 Kinect 之间角度和距离的不同，如果直接用三维坐标去做人体姿态识别，会有很多的问题，因此本文利用三角转换公式将人体骨骼关节点三位坐标信息转换为角度数据描述，然后用人体的 20 个骨骼关节点的 60 个角度信息作为人体姿态的特征向量，这样的话就可以有效抑制上面所提出的问题。

然后分析人体日常行为姿态的多样性，会导致当需要识别的人体姿态太多时，一般的分类算法会很难满足要求，所以本文提出一种人体姿态识别的分类策略，就是训练分类器去判断任意两种姿态是否是同一姿态，这样在姿态识别时，无论有多少种需要识别的姿态，都只需要判断待识别姿态和所有需要识别的姿态是否为同一姿态，这样就会解决由于需要识别的人体姿态太多而导致分类模型过度复杂的问题。

根据上述分类策略，本文提出两种分类器，一种是经典的二分类分类器—逻辑回归，另一种是本文根据人体姿态特征的特征而构建的随机森林模型。逻辑回归其实就是根据训练数据得到一组特征向量的权值，然后经过 sigmoid 函数映射而判断两个姿态是否是同一姿态，但是通过实验发现当人体姿态特征增多时，并且有相似需要识别的姿态时，就会出现待识别姿态和几种相似的需要识别的姿态都判定为同一姿态，并且置信度都很高，严重干扰识别结果。

所以最后本文提出基于随机森林模型的人体姿态识别算法，并针对人体姿态特征和分类策略的特点，重新设计的随机森林在构建时怎样解决数据不平整的问题，以及参数优化的方法，并通过实验数据结果验证算法的有效性，并对逻辑回

归和随机森林两个分类器做了实验对比,结果表明随机森林模型可以满足人体姿态识别的要求。

6.2 研究展望

人际交互一个很重要的问题就是时效性的问题,由于本文提出的分类策略,待识别姿态会和所有的待识别姿态都要进入构建好的随机森林模型进行判断是否是同一姿态,这样会导致当识别姿态特别多时,随机森林分类器会被调用很多次,所以很难达到实时性。而且本文目前识别的人体姿态都是静态姿态,而人机交互还需要识别出对人体动作。所以只有解决实时性、以及识别出人体姿态和动作的问题,才能达到真正的人机交互。

因此,本文未来进一步研究的方向,可以分为几个方面:

- 1) 找到更有效的人体姿态特征向量,这样就可以利用更少的决策树来构建随机森林。
- 2) 优化算法流程,可以使用多线程或者 GPU 来加速算法速度,已解决实时性问题。
- 3) 寻找好的利用人体姿态对人体动作进行编码描述的方式,以及动作识别分类器,以实现动作识别,达到真正的人机交互。

致 谢

在这篇硕士论文即将完成之际，回首自己攻读硕士这三年的学习经历，我由衷的感激导师、同学和家人给予我的帮助和关心。

首先我要感谢我的导师朱虹教授，硕士三年，导师在我学术水平和综合能力的提高等方面给予了我全面而细致的指导与关怀。在学术上，是朱老师带我走进了机器视觉的大门，从而让我有幸进入当前最火的学术领域。老师从我论文的开题、研究方案的确定，到如今完成研究生课题和论文，这些都离不开朱老师的耐心的指导和不厌其烦的帮助。在生活中，朱老师更是待我们如自己的孩子一样，是老师的平易近人和对我们无时无刻的关心，让我们时刻感觉教研室更像是一个家，也是老师用自身的人格魅力教会我们怎样做一个有能力的好人。在此，我要衷心的给朱老师说一句：“老师，谢谢您！”祝老师和老师的家人永远健康幸福！

感谢师兄王栋、康凯、苟荣涛在项目上给我的指导和帮助，在学习和生活上给我的关心，感谢你们给教研营造了这么适合大家生活和学习的地方，祝你们生活幸福，工作顺利！

感谢黎路师兄，虽说我们交集不多，但是您教会了我男孩子应该以何种态度和勇气去面对生活和工作，从你身上我收获很多，谢谢。

感谢我的同窗张静波、赵永飞、权甲、谢凡凡、张晨、白帅，是你们的陪伴，让研究生的生活更加丰富多彩。是你们在我刚来学校时，在各个方面给予我最直接的帮助，我永远也忘不了这一段时光有你们陪我走过，祝你们以后生活幸福、工作顺利。

感谢教研室所有的师弟师妹们，是你们带给了教研室生气和活力，让我们能够快乐的学习，开心的生活。

感谢我的家人，谢谢你们在生活上和学业上对我的关心和爱，是你们让我拥有一个最温暖的家，可以让我在遇到困难的时候，都可以挺起胸坦然面对，无所畏惧。

最后，感谢母校西安理工大学三年以来对我的培养，愿母校的明天更加辉煌！

参考文献

- 【1】 Ibrahim Radwan, Abhinav Dhall, Roland Goecke. Monocular Image 3D Human Pose Estimation under Self-Occlusion[A]. ICCV[C]. IEEE, 2013:1888-1895.
- 【2】 Ankur Gupta, Julieta Martinez, James J. Little, Robert J. Woodham. 3D Pose from Motion for Cross-view Action Recognition via Non-linear[A]. CVPR[C]. IEEE, 2014:2601-2608.
- 【3】 Jiang Wang, Xiaohan Nie, Yin Xia, Ying Wu, Song-Chun Zhu. Cross-view Action Modeling, Learning and Recognition[A]. CVPR[C]. IEEE, 2014: 2649-2656.
- 【4】 Min-Chun Hu, Chi-Wen Chen, Wen-Huang Cheng, Che-Han Chang, Jui-Hsin Lai, Ja-Ling Wu. Real-Time Human Movement Retrieval and Assessment With Kinect Sensor[J]. IEEE, 2014(99).
- 【5】 Shahriar Nirjon, Chris Greenwood, Carlos Torres. Kintense: A Robust, Accurate, Real-Time and Evolving System for Detecting Aggressive Actions from Streaming 3D Skeleton Data[A]. PerCom[C]. IEEE, 2014:2-10.
- 【6】 Bojan Dikovski, Gjorgji Madjarov, Dejan Gjorgjevikj. Evaluation of different feature sets for gait[A]. MIPRO[C]. IEEE, 2014:1304-1308.
- 【7】 Genzhi Ye, Yebin Liu, Yue Deng, Nils Hasler, Xiangyang Ji. Free-Viewpoint Video of Human Actors[J]. IEEE, 2013(5):1370-1382.
- 【8】 Vasileios Belagiannis, Sikandar Amin, Mykhaylo Andriluka. 3D Pictorial Structures for Multiple Human Pose Estimation[A]. CVPR[C]. IEEE, 2014: 1669-1676.
- 【9】 吴心筱. 图像序列中人的姿态估计与动作识别[D]. 北京: 北京理工大学, 2010.
- 【10】 Mykhaylo Andriluka, Stefan Roth, Bernt Schiele. Monocular 3D Pose Estimation and Tracking by Detection[A]. CVPR[C]. IEEE, 2010:623-630.

-
- 【11】 战荫伟, 张昊. 基于 Kinect 传感器的人体行为分析算法[J]. 传感器与微系统, 2015,01:142-144.
- 【12】 殷海艳. 基于深度图像的人体姿态识别[D].北京工业大学,2013.
- 【13】 费诗龙. 基于运动捕捉的人型机器人动作生成框架[D].清华大学,2011.
- 【14】 Ari Shapiro,Andrew Feng, Ruizhe Wang, Gerard Medioni, Mark Bolas, Evan A. Suma. Automatic Acquisition and Animation of Virtual Avatars[A]. VR[C]. IEEE, 2014:185-186.
- 【15】 Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J]. The International Journal of Robotics Research, 2012, 31(5): 647-663.
- 【16】 Packer B, Saenko K, Koller D. A combined pose, object, and feature model for action understanding[C]. CVPR. 2012: 1378-1385.
- 【17】 Pisharady P K, Saerbeck M. Kinect based body posture detection and recognition system[C]. 2012 International Conference on Graphic and Image Processing. International Society for Optics and Photonics, 2013.
- 【18】 Shotton J, Sharp T, Kipman A, et al. Real-time human pose recognition in parts from single depth images[J]. Communications of the ACM, 2013, 56(1): 116-124.
- 【19】 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: 20-27.
- 【20】 Xiao Z, Mengyin F, Yi Y, et al. 3D human postures recognition using Kinect[C]. Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2012 4th International Conference on. IEEE, 2012, 1: 344-347.
- 【21】 A.F.Bobick, J.W.Davis. The recognition of human movement using temporal templates[J]. IEEE, 2001(3):257-267.

-
- 【22】 C.Chen, K.Grauman. Watching Unlabeled Video Helps Learn[A]. CVPR[C]. IEEE, 2013:572-579.
 - 【23】 I. N. Junejo, E. Dexter, I. Laptev, and P. Perez. View-independent Action Recognition from Temporal Self-similarities[C]. TPAMI, 33(1), 2011.
 - 【24】 Saiyi Li¹, Pubudu N. Pathirana². Multi-Kinect Skeleton Fusion for Physical Rehabilitation Monitoring[C]. IEEE, 2014(3):5060-5063.
 - 【25】 C.-Y. Chen and K. Grauman. Watching unlabeled video helps learn new human actions from very few labeled snapshots[C]. CVPR,2013.
 - 【26】 N. Dalal and B. Triggs. Histograms of oriented gradients for human detection[C]. In CVPR, 2005.
 - 【27】 R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon. Efficient regression of general-activity human poses from depth images[C]. ICCV, 2011.
 - 【28】 N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas. The i3DPost multi-view and 3d human action/interaction database. In Proc. Conf. Visual Media Production, pages 159–168, 2009.
 - 【29】 A. Iosifidis, A. Tefas, and I. Pitas. View-invariant action recognition based on artificial neural networks[C]. TNNLS, 2012.
 - 【30】 I.-H. Jhuo, D. Liu, D. Lee, and S.-F. Chang. Robust visual domain adaptation with low-rank reconstruction[C]. In CVPR, 2012.
 - 【31】 I. Laptev and T. Linderberg. Space-time interest points[C]. In ICCV, 2003.
 - 【32】 Z. Lin, M. Chen, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices[R]. Technical report, UIUC, 2009.
 - 【33】 J. Liu, S. Ji, and J. Ye. Multi-task feature learning via efficient $l_{2,1}$ norm minimization[R]. UAI, 2009.
 - 【34】 S. Maji, L. Bourdev, and J. Malik. Action recognition from a distributed representation of pose and appearance[C]. In CVPR, 2011.

-
- 【35】 P.Matikainen, M.Hebert,and R.Sukthankar. Representing pair wise spatial and temporal relations for action recognition[C]. In ECCV,2010.
- 【36】 O. Oreifej and Z. Liu. HON4D: Histogram of oriented 4d normals for activity recognition from depth sequences[C]. In CVPR, 2013.
- 【37】 S. Pan and Q. Yang. A survey on transfer learning[C]. TKDE, 2010.
- 【38】 R. Poppe. A survey on vision-based human action recognition. IVC, 2010.
- 【39】 J.K.Aggarwal and M.S.Ryoo. Human Activity Analysis: A Review[J]. ACM Computing Surveys, 43(3):16:1–16:43, 2011. 1, 2
- 【40】 R. Chaudhry, F. Ofli, G. Kurillo, R. Bajcsy, and R. Vidal. Bio-inspired Dynamic 3D Discriminative Skeletal Features for Human Action Recognition[C]. In CVPRW, 2013. 3
- 【41】 M.Hussein, M.Torki, M.Gowayyed, and M.El-Saban. Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations. In IJCAI, 2013. 2, 7.
- 【42】 D. Weinland, M. "Ozuysal, and P. Fua. Making action recognition robust to occlusions and viewpoint changes[C]. In ECCV, 2010.
- 【43】 Z.Wen and W.Yin. A feasible method for optimization with or thogonality constraints[J]. Technical report, Rice University, 2010.
- 【44】 X.Wu,D.Xu,L.Duan, and J.Luo. Action recognition using context and appearance distribution features[C]. In CVPR, 2011.
- 【45】 L. Xia and J. Aggarwal. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera[C]. In CVPR, 2013.
- 【46】 Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition[C]. In CVPR, 2011.
- 【47】 G.Z hao and M.Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions[R]. TPAMI, 2007.
- 【48】 X. Wu, Z. Wu and K. Li, Classification and Identification of Differential Gene Expression for Microarray Data: Improvement of the Random Forest

- Method[J]. Bioinformatics and Biomedical Engineering, 2008.
- 【49】 Breiman L. Random Forests[J]. Machine Learning, 2001, 45: 5-32.
- 【50】 P Bermejo, J M Pureta, J A Gamez. Improving the performance of naive bayes multinomial in e-mail folding by introducing distribution-based balance of datasets[J]. Expert Syst Appl, 2011, 38(3): 2072-2080.
- 【51】 Yang P, Liu W, Zhou B B, et al. Ensemble-based wrapper methods for feature selection and class imbalance learning[M]. Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2013: 544-555.
- 【52】 Xu B, Huang J Z, Williams G, et al. Hybrid random forests: advantages of mixed trees in classifying text data[M]. Advances in Knowledge Discovery and Data Mining. Springer Berlin Heidelberg, 2012: 147-158.

已发表论文、专利及科研情况

发明专利两项：

1. 一种基于多 Kinect 的人体姿态数据融合方法（201510363867.7）

朱虹 卫永波 谢凡凡 权甲

2. 一种复杂背景下的车牌定位方法（201510478623.4）

朱虹 卫永波 谢凡凡 权甲 白帅