分类号：  **TP391.4**  密级: **公开**

****U D C：  **004.8**  编号:

**硕 士 学 位 论 文**

基于半监督协同训练和集成学习的人体动作识别研究

**RESEARCH ON HUMAN ACTION RECOGNITION**

**BASED ON SEMI-SUPERVISED CO-TRAINING AND ENSEMBLE LEARNING**

指导教师 \* \* \* 教授

作者姓名 \* \* \*

申请学位级别 工学硕士 专业名称 计算机科学与技术

论文提交日期 2017年4月 论文答辩日期 2017年6月

学位授予单位和日期 江苏大学2017年6月

答辩委员会主席 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

评阅人 \_\_\_\_\_\_\_\_\_\_\_\_\_\_

独 创 性 声 明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已注明引用的内容以外，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，也不包含为获得江苏大学或其他教育机构的学位或证书而使用过的材料。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。

学位论文作者签名：

年 月 日学位论文版权使用授权书

江苏大学、中国科学技术信息研究所、国家图书馆、中国学术期刊（光盘版）电子杂志社有权保留本人所送交学位论文的复印件和电子文档，可以采用影印、缩印或其他复制手段保存论文。本人电子文档的内容和纸质论文的内容相一致，允许论文被查阅和借阅，同时授权中国科学技术信息研究所将本论文编入《中国学位论文全文数据库》并向社会提供查询，授权中国学术期刊（光盘版）电子杂志社将本论文编入《中国优秀博硕士学位论文全文数据库》并向社会提供查询。论文的公布（包括刊登）授权江苏大学研究生院办理。

本学位论文属于不保密 □ 。

学位论文作者签名： 指导教师签名：

年 月 日 年 月 日

**摘 要**

随着科学技术的发展，人体动作识别逐渐成为人工智能和机器视觉领域一个重要的研究方向，具有广阔的发展前景和很强的实用价值。可应用于日常的视频监控、智能医疗、运动分析、人机智能交互等。同时，由于视频中场景的复杂性、动作类内变化，以及需要大量的有标注样本来训练泛化性能强的识别模型，这些都给人体动作识别的研究带来了挑战性。

本文对人体动作识别的若干问题，特别是基于半监督的人体动作识别进行了较深入的研究。首先阐述了人体动作识别的选题背景与研究的目的和意义；其次概述了人体动作识别的关键技术，如关键帧提取技术、特征提取技术以及人体动作识别技术等。在目前人体动作识别的理论研究基础上，本文提出了基于混合式协同训练的人体动作识别方法和基于半监督集成学习的人体动作识别方法，并在以上两种算法的基础上设计开发人体动作识别原型系统，主要研究内容如下：

1）提出了基于混合式协同训练的人体动作识别方法。针对目前人体动作视频中有标记数据不足的问题，提出了一种基于混合式协同训练的新型人体动作识别方法。该方法利用动作识别领域不同类型的识别方法来构建基分类器，并进行迭代的相互训练以提高泛化性能，可以降低标注成本并实现不同识别方法的优势互补，进而提高人体动作的识别精度。实验结果表明，本文所提出的算法可以有效地识别视频中的人体动作。

2）提出了基于半监督集成学习的人体动作识别方法。针对协同训练类算法随着迭代次数的增加，基分类器的差异性会越来越小，以及迭代训练中产生的基分类器没有被充分利用的问题。提出了基于协同训练和集成学习相结合的人体动作识别方法。该方法对每个基分类器设置一个集合。将基分类器迭代训练过程中产生的中间分类器加入到各自的集合中，然后利用这个集合来选择伪标号数据。并定义了一个基于置信度的最大证据边缘函数来选择伪标号数据，最终利用该算法对人体动作进行识别。该方法能有效克服协同训练迭代过程中基分类器差异退化的问题，进一步提高人体动作识别的准确率。

3）设计并实现了基于半监督协同训练和集成学习的人体动作识别的原型系统。采用面向对象语言C#和MATLAB进行编程，通过原型系统的运行测试，表明所提的方法可用于相应的人体动作识别，并且该原型系统界面友好、功能齐全、

可维护性好。

关键字：人体动作识别，协同训练，集成学习，半监督

**ABSTRACT**

With the development of science and technology, human action recognition has been becoming an important research topic in the field of artificial intelligence and machine vision. It has been already applied in a lot of areas such as video monitoring, intelligent medical, sports analysis, man-machine intelligence etc. Meanwhile, because of the complexity of video, motion diversity, and a large number of labeled data are needed to train the strong recognition model, it faces the huge challenge.

This thesis studies several topics in human action recognition, in particular, an unsupervised learning approach for analysis of human action recognition. Firstly, this thesis elaborates the goal and significance of background and research in human action recognition. Secondly, some key technologies in human action recognition such as key frame extraction, feature extraction, feature extraction and action recognition are briefly described. Finally, based on the current theoretical research on human action recognition, this thesis presents a human action recognition algorithm based on hybrid collaborative training and a human action recognition algorithm based on semi-supervised collaborative training and ensemble learning. A human action recognition prototype system is also developed. The main contents are as follow:

1)We propose a human action recognition algorithm based on hybrid collaborative training. Aiming at the problem of insufficient of labeled data on the current human action recognition methods, we proposed a human action recognition algorithm based on hybrid collaborative training. Different types of recognition methods for action recognition field are employed in this method to build the base classifiers, which are then iteratively retrained to increase their generalization abilities. In general, our method can decrease the labeling cost and achieve complementary advantages of different recognition algorithms, improve the recognition performance. The experimental results show that the proposed algorithms can identify human action in the video more effectively.

2)We propose a human action recognition algorithm based on semi-supervised collaborative training. With the increasing of iterations on hybrid collaborative training, The differences of base classifiers will get smaller and smaller. And base classifiers are underutilized. Aim at these problems, we proposed a human action recognition algorithm based on hybrid collaborative training and ensemble learning. This method set up a collection for each classifier, and add middle classifiers that are generated in the process of iterative training to their own collection. Then use this collection to select pseudo labeled data. In addition, we define a most evidence edge function which is based on confidence to select the pseudo labeled data. In the end, we use this method to identify human action. This method can overcome the problem that difference will get smaller and smaller in the process of iterative training, and further improve accuracy on human action recognition.

3)A prototype system of human action recognition based on semi-supervised method is designed and implemented. We use C# that it is based on object-oriented language and MATLAB to program. By operating prototype system，result indicates that it can be used in human action recognition. Moreover, the prototype system functions well with a friendly interface and good maintainability.

**Key words**: human action recognition, Co-training, ensemble learning, semi-supervised

目 录

[第一章 绪论 1](#_Toc480456167)

[1.1 课题研究背景及意义 1](#_Toc480456168)

[1.1.1 课题研究背景 1](#_Toc480456169)

[1.1.2 课题研究意义 3](#_Toc480456170)

[1.2 国内外研究现状 3](#_Toc480456171)

[1.3 论文的研究内容与主要工作 4](#_Toc480456172)

[1.4 论文的结构安排 5](#_Toc480456173)

[第二章 人体动作识别技术概述 7](#_Toc480456174)

[2.1 引言 7](#_Toc480456175)

[2.2 人体动作视频关键帧提取 7](#_Toc480456176)

[2.3 人体动作视频特征提取与融合 8](#_Toc480456177)

[2.3.1 颜色特征 9](#_Toc480456178)

[2.3.2 纹理特征 10](#_Toc480456179)

[2.3.3 形状特征 11](#_Toc480456180)

[2.3.4 多尺度LBP特征 12](#_Toc480456181)

[2.3.5 多特征融合 13](#_Toc480456182)

[2.4 人体动作识别方法 13](#_Toc480456183)

[2.4.1 基于形状模板的方法 13](#_Toc480456184)

[2.4.2 基于概率统计的方法 14](#_Toc480456185)

[2.4.2 基于语法的方法 15](#_Toc480456186)

[2.5 人体动作识别技术研究难点 16](#_Toc480456187)

[2.6 本章小结 17](#_Toc480456188)

[第三章 基于混合式协同训练的人体动作识别方法 18](#_Toc480456189)

[3.1 概述 18](#_Toc480456190)

[3.2 传统协同训练算法的分析 18](#_Toc480456191)

[3.3 混合式协同训练的人体动作识别算法 20](#_Toc480456192)

[3.2.1 基分类器的构造 20](#_Toc480456193)

[3.2.2 混合式协同训练算法Co-KNN-SVM 21](#_Toc480456194)

[3.2.3 结合KNN和SVM 23](#_Toc480456195)

[3.4 实验结果与分析 23](#_Toc480456196)

[3.4.1 人体动作数据集 23](#_Toc480456197)

[3.4.1 评价标准 24](#_Toc480456198)

[3.4.2 实验步骤及方法 25](#_Toc480456199)

[3.4.3 实验结果分析与讨论 26](#_Toc480456200)

[3.5 本章小结 29](#_Toc480456201)

[第四章 基于半监督集成学习的人体动作识别方法 30](#_Toc480456202)

[4.1 概述 30](#_Toc480456203)

[4.2 集成学习 30](#_Toc480456204)

[4.2.1基分类器的创建 31](#_Toc480456205)

[4.2.2基分类器的组合 31](#_Toc480456206)

[4.3 半监督集成学习算法 32](#_Toc480456207)

[4.4 集成学习与Co-KNN-SVM的结合 33](#_Toc480456208)

[4.4.1基于集成学习与Co-KNN-SVM的结合的ECKS算法 33](#_Toc480456209)

[4.4.2伪标号数据的选择与处理 36](#_Toc480456210)

[4.4.3中间分类器差异性和准确性分析 37](#_Toc480456211)

[4.4.4集成多个基分类器 37](#_Toc480456212)

[4.5 实验结果与分析 38](#_Toc480456213)

[4.5.1实验设置 38](#_Toc480456214)

[4.5.2 实验结果分析与讨论 38](#_Toc480456215)

[4.6 本章小结 41](#_Toc480456216)

[第五章 人体动作识别原型系统的设计与实现 42](#_Toc480456217)

[5.1 引言 42](#_Toc480456218)

[5.2 系统的总体设计 42](#_Toc480456219)

[5.3 人体动作识别原型系统的实现 43](#_Toc480456220)

[5.3.1 系统开发环境 43](#_Toc480456221)

[5.3.2 系统实现简介 44](#_Toc480456222)

[5.4 本章小结 48](#_Toc480456223)

[第六章 总结与展望 49](#_Toc480456224)

[6.1 工作总结 49](#_Toc480456225)

[6.2 工作展望 50](#_Toc480456226)

[参考文献 51](#_Toc480456227)

[致 谢 57](#_Toc480456228)

[攻读硕士学位期间已发表的论文 58](#_Toc480456229)

**第一章 绪论**

**1.1 课题研究背景及意义**

**1.1.1 课题研究背景**

近年来，随着手机等数码产品的不断普及，使得人们更容易地创建视频。另一方面，社交网络的不断发展以及互联网云存储的便利性和廉价性，使得我们能够轻松地将视频保存到网络中并与他人分享。相关统计表明，截止去年12月份，中国网络视频用户规模达5.45亿，成为互联网行业成长最快的市场。视频中大多数记录的都是作为社会主体人的活动，无论是从个人存储、安全监控，还是交通监测的角度，对视频中人体动作识别进行研究都具有重要的价值[1]。

人体动作识别是当今计算机视觉领域的研究热点，涉及了多个交叉学科的知识，主要包含机器学习、模式识别、人工智能、图像处理以及心理学等[2][3]。其主要研究通过计算机和视频采集设备对人体的动作模式进行分析，进而理解和识别不同的动作类别。当前，人体动作识别在许多领域得到应用，下面对它们进行介绍：

1）自动视频监测

人体动作识别的一个广泛的应用领域是自动视频监测。自动视频监测就是利用相关智能设备采集原始动作视频或者图像，经过计算机快速处理，智能视频分析模块能够提取并分析视频画面中的有效信息，识别出不同的人体动作。它不仅可以应用在交通监控等现实生产实践中，而且还可以应用在军事上，如军情勘测，目标追踪等。

2）体育动作分析

将人体动作应用到体育动作分析中，一方面，在比赛过程中，裁判员可以改变传统的人工的判断方式，使得判罚更加准确、公正；另一方面，在平常的训练过程中，教练员可以分析运动员的动作，通过对获取的数据进行量化处理，进而改进不理想运动的训练模式，可以提高运动员的比赛成绩。

3）基于视频内容的查找

针对目前结构化查询语言查找技术不能有效满足较大规模的视频数据库的检索，使得基于视频内容的多媒体查找技术得到快速地发展。人体动作为视频中重要的内容之一，通过对其进行识别与分析，正确辨别出不同的类型。可以帮助人们在视频库中快速找到目标视频片段，从而提高获取有效视频信息的速率。

4）人机交互

人机交互是指人们为了完成某些确定的任务，以交互的形式，通过使用某种语言与计算机进行“对话”。如人们在笔记本、平板电脑、手机上利用手指触摸的方式与设备进行交互。随着技术的逐渐发展，人们发现人机交互方式更加符合生活、学习以及工作习惯。因此，人机交互技术也是人体动作识别的应用领域之一。

5）虚拟现实

虚拟现实是指用计算机构建一个模拟真实世界的三维虚拟空间，并在该空间中融合了多源信息，是一种交互式的三维动态视景。用户在该视景中能感受真实的实体行为动作，并进行一些交互操作，各种体感游戏、3D影视等都是动作识别技术在虚拟现实中的应用。图1-1展示的是体感运动游戏。



图1-1体感运动游戏

目前，对人体动作识别建模的方法大致分为三种：非参数的方法（nonparametric）、体积分析的方法（volumetric）、参数时间序列分析的方法（parametric time-series approaches）[1]。非参数方法通常从视频片段的每个关键帧中提取一组特征，然后用一个事先存好的模板匹配这些特征；立方体的分析方法不是提取每帧图像的特征，而是考虑将视频片段作为一个三维立方体来进行分析；参数时间序列分析方法是构建一个关于运动的时序动态模型，然后根据训练数据得到模型参数来识别视频中的人体动作。上述方法大多是有监督的。若是要构造一个性能良好的动作模型，需要大量的有标记数据，但在人体动作的识别应用中，对数据进行标记，费时又费力。与此同时，随着存储技术的不断进步，人们更容易地获得无标记数据。因此，研究在利用少量的有标记数据的同时结合大量的无标记数据来进行半监督学习成为人体动作识别领域亟待解决的问题。

**1.1.2 课题研究意义**

通过对半监督的人体动作进行识别研究，本课题具有重要的理论和实践意义。

1）理论意义

通过对半监督的人体动作识别进行识别研究，可以进一步丰富人体动作识别和半监督学习等基础理论，为半监督的人体动作识别模型的构造方法研究和应用研究打下扎实的基础。

2）实践意义

通过将基于半监督的人体动作识别模型引入到实际应用中，为实际应用中的基于半监督学习的人体动作识别提供了新的方式和途径，推动了视频监控中如何利用未标记样本进行学习的问题的研究和应用，进一步强化人体动作识别和半监督学习的应用能力，具有重要的实践意义。

**1.2 国内外研究现状**

由于人体动作具有广阔的发展前景、巨大的经济潜力和社会价值，国内外越来越多的大学、科研学者、政府乃至商家都加入到这一行业进行研究、实践以及应用研发和推广中。

相较于国内而言，国外的研究起步较早，尤其在美国、欧洲和日本已经进行了大量的项目研究。在20世纪末，美国国防高级研究项目署（简称DARPA）成立了长达三年的视频监控项目VSAM[4]，该项目以美国Carnegie Mellon University为首，并联合了Massachusetts Institute of Technology等美国十几所高等学校和科研组织。该项目能够定位和检测出人的形状轮廓，从而实现行人的轨迹跟踪，以增强机场、民用等场合免受恐怖袭击的保护能力。美国的Maryland大学研发了一款实时的视觉检测系统w4[6]，通过采用单目灰度外摄像机传感器对人体进行目标检测和跟踪，并监视人的行为动作。英国University of Reading设立了ETISEO、ISCAPS等项目，研究包含人体和车辆等相关场景的识别理解[5]。

由于政府的支持，以及人体动作识别潜在的价值被挖掘，国内越来越多的高校、科研机构、相关企业都投入到该研究领域中，并在应用和理论上都取得了一定的成果。目前在人体动作识别研究领域的机构主要有清华大学、浙江大学、国防科技大学、中国科学院自动化研究所等。清华大学的徐光祐、任海兵对人体动作识别进行了研究[7][8]，前者对人体动作识别进行了综述，后者针对在人体动作识别中运动参数丢失和会产生非特定人之间差异的问题，提出了基于基元特征的动态贝叶斯网络的多信息融合策略的动作识别方法，该方法有效提高了系统的鲁棒性和动作识别的准确性。浙江大学与微软视觉感知实验室研发出了一款视频动画系统，该系统通过双摄像机形成立体视觉，通过用标签在人体关节点处获取人体的信息，来进行动作识别。其在3 D动画制作中具有很高的应用价值[9]。

在人体动作识别领域，针对目前标注瓶颈的问题，不少学者提出了半监督的学习算法。半监督学习是结合监督学习和无监督学习，利用少量的有标记数据来构造模型假设，并通过大量的无标记数据进一步改进模型泛化性能的学习算法[10][11]。其主要分为以下三类：基于生成式模型的半监督学习算法(SSL with generative models)[12][13][14]、基于图的半监督学习算法(SSL with graphs)[15][16]、协同训练算法（Co-training）[17]。而协同训练算法作为基于不一致性的半监督学习的典型的代表[17]，在Blum和Mitchell提出之后，许多学者对其进行了拓展，使其成为半监督学习最重要的算法之一。Gupta等人[18]提出了一种基于半监督的Co-training算法，分别利用视觉（visual）和语言信息（linguistic information）作为两个不同的视角，用于识别图像和视频序列中的场景和人体动作。并获得了较好的识别性能。随后，Yu等人[19]提出了基于贝叶斯无向图模型的协同训练识别方法，为了同时优化两个视角，该方法为高斯过程分类器引进了一种新型的co-training内核，并进一步提高了性能。为了解决Co-training需要视图冗余（view-sufficiency）和独立性（view-dependency）要求的问题，Liu和Yuen[20]提出了一种基于两种新的置信度评价方法的人体动作识别方法，同时也可以用来度量不同视图间的独立性。Tang等人[21]提出了一种多分类器协同训练的人体动作识别方法，通过委员会来评估无标记数据的置信度，并选择较高的加入有标记数据中进行重复训练。最终通过基分类器组进行决策输出。

上述算法都是利用不同视角来构建基分类器，而且大多是仅根据置信度来选择伪标记样本，而未考虑伪标记数据与实际样本空间的分布差异（分布噪音）。因此本文提出了一种基于动作识别中不同类型的算法来构建基分类器的新思路，可以实现动作识别领域中不同方法的优势互补。更进一步，引进了集成学习的思想，来改进随着迭代次数的增加，协同训练中基分类器的差异逐渐消失的缺点。

**1.3 论文的研究内容与主要工作**

综合大量的国内外有关半监督学习下的人体动作识别的研究资料，并对其中比较成功的方法进行分析和探讨。本文对人体动作识别相关技术进行了深入研究，提出了自己的算法应用于人体动作识别，设计并开发了基于半监督协同训练和集成学习的人体动作识别系统。本文的研究主要包含以下三个方面内容：

1）提出基于混合式协同训练的人体动作识别方法

针对目前动作视频中有标记数据不足的问题，提出一种混合式协同训练的新型人体动作识别算法Co-KNN-SVM。该方法利用动作识别领域不同类型的方法来构建基分类器，并进行迭代的相互训练以提高泛化性能，以降低标注成本并实现不同识别方法的优势互补。进而提高人体动作的识别精度。

2）提出基于半监督集成学习的人体动作识别方法

针对协同训练算法训练过程中产生的中间临时分类器没有被充分利用，以及随着训练次数的增多，分类器之间差异性逐渐减小的问题。本文提出协同训练和集成学习相结合的动作识别方法ECKS。该方法首先将迭代训练过程中产生的临时分类器加入到不同的基分类器集合中，然后分别利用各自的分类器集合标记和选择无标记数据。并且结合置信度，定义了一个最大证据边缘函数来选择伪标记数据。最终使用该方法对视频中的人体动作进行识别。该方法能有效克服混合式协同训练中分类器差异退化的问题，进一步增强系统的泛化性能。

3）设计实现人体动作识别的原型系统

根据以上所提的方法，本文采用统一建模语言UML对动作识别系统进行建模，并利用面向对象语言C#和MATLAB进行编程，实现基于半监督协同训练和集成学习的人体动作识别原型系统，来验证所提方法的可用性。

**1.4 论文的结构安排**

全文研究内容和章节安排如下：

第一章首先简要介绍了本课题的研究背景、意义以及国内外研究现状，总结了该研究方向的应用领域，并对半监督的人体动作识别方法进行了介绍。最后阐述了论文的主要研究内容和各个章节的组织架构。

第二章简要概述了人体动作识别技术：人体动作视频中关键帧的提取、特征提取与融合、人体动作识别方法，并分析了人体动作识别技术的研究难点，提出了本文所研究的方向。

第三章首先对协同训练进行了简单的描述，讨论并分析了其算法步骤和缺点，提出了一种新型的混合式人体动作识别算法Co-KNN-SVM。该方法利用动作识别领域不同类型的方法来构建基分类器，并进行迭代的相互训练以提高泛化性能。此外，改进了协同训练中对伪标记数据的选择方法和迭代训练策略，有效控制了伪标记数据的噪声影响，提高协同训练中人体动作的识别效果。并且在人体动作视频数据集上，验证所提方法的有效性。

第四章结合集成学习相关理论，提出基于半监督协同训练和集成学习相结合的人体动作识别方法。利用迭代训练过程中产生的中间分类器形成分类器组，对无标记数据进行标记、选择，有效克服了协同训练方法中分类器差异性逐步退化的弊端。并在人体动作数据集上进行实验，验证了所提算法的可用性和有效性。

第五章根据第三、四章所提的方法，设计并开发了基于半监督协同训练和集成学习的动作识别原型系统，并对各个模块进行了说明。

第六章为总结与展望。总结了本文的研究成果，并且对今后的工作进行展望。

**第二章 人体动作识别技术概述**

**2.1 引言**

从本质上来说，人体动作识别是一个模式识别问题。其重要的一步就是能够从含有人体视频片段中获取动作状态的特征，特征提取的合理性直接影响动作识别模型的有效性和准确性。鉴于真实场景下视频动作的复杂性，人体动作识别技术主要包含关键帧的提取、人体动作视频特征提取、人体动作识别方法等。视频的关键帧代表的是视频片段中能反映内容信息图像的集合。特征提取即动作表示，通过从视频片段中提取有效的信息来对人体动作进行表示。理想的动作特征应该能应对光照、视角变化、执行快慢等因素的影响，而且还应该减少冗余性。动作识别方法是指根据提取的特征构造相应的识别模型，理想的模型不仅能够在复杂场景下区分不同的动作类型，还应该具有鲁棒性。

**2.2 人体动作视频关键帧提取**

为了减小计算代价，通常从动作视频中选择若干个关键帧图像来代表整个视频段信息。关键帧获取的好坏直接影响人体动作识别的准确性，于是要求选取的图像集合能有效表达视频中人体动作的信息。正因如此，关键帧提取技术得到了越来越多国内外科研人员的关注，并取得了令人瞩目的成果。现有的视频关键帧提取方法大致可以分为5类[22]：

1）基于视频片段边界的方法：该方法通常选取视频片段的第一帧、最后一帧或者中间帧作为关键帧。该方法计算简单，适用于内容变化小以，基本保持不变的视频片段。缺点是提取复杂视频关键帧时，效果不够理想。

2）基于视觉内容的方法：该方法主要利用视频图像的形状、颜色、纹理等静态信息来选择关键帧图像[22][23]。比较典型的有直方图平均法和帧平均法，前者根据视频中所有图像统计直方图的平均值来选取视频关键帧，而后者直接计算所有帧图像在某个像素位置的平均值，以选取最接近该值的图像作为关键帧。由于这类方法中经验阈值的选择的难度较大，当视频片段不稳定时，容易选出冗余的关键帧。

3）基于运动分析的方法[22]：该方法通过人体的运动信息来获取视频关键帧，代表的有运动极小值算法，该方法通过光流来计算人体运动量，并根据计算最小值选取关键帧，计算较为复杂。

4）基于聚类分析的方法：该方法通常使用K均值和模糊C均值作为聚类算法，初始确定一个类心，然后根据视频图像帧与类心的距离来选取视频关键帧。前者选取的关键帧是清晰的，每幅图像只被分配到一个聚类中。后者针对每幅图像都有一个聚类函数，分类是模糊的。此类方法能较好地减少视频片段的相关性，但不能保持图像帧的顺序性和保存视频片段的运动信息。

5）基于压缩视频流的方法：不同于以上四种基于全图像序列的关键帧提取方法，基于压缩视频流的方法直接从压缩的视频流上提取关键帧，有效降低了计算复杂度，目前大致分为两种方法，一种方法是通过提取视频流中的I帧作为关键帧，另一种方法是通过利用压缩视频流中离散余弦变换的系数和运动矢量来提取关键帧。

上述的各种方法都具有优缺点，本文实验中，我们选取的数据集是UCF Sports视频数据库，视频连续帧的内容变化不是很大。因此，我们选择基于视频片段边界的方法来进行动作视频的关键帧提取，选取动作视频的前中后位置的图像作为关键帧，这样既能代表整个视频片段的主要内容，又能反映视频序列的时变信息。

**2.3 人体动作视频特征提取与融合**

由于不同的特征对人体动作的表述能力不同，选择特征时不仅要考虑提取的复杂程度，而且还要从易区分的角度考虑。因此，依据视频应用场景的不同，往往选择不同的特征。比如，在KTH[24]、Weizmann[25]等受限场景的动作识别数据库中，往往选择提取目标整体的运动轨迹进行轨迹分析，而在UCF YouTube Action[26]、PETS[27]等真实场景的动作识别数据库中，往往选择静态的颜色、形状、纹理等进行特征提取。人体动作特征主要有以下四种：基于形状的静态特征、基于运动的动态特征、基于时空特征点的时空特征以及描述性特征。静态特征一般包括颜色[28]、纹理、轮廓[29]、形状[30]等，较完整地体现人体运动以及场景的信息，例如：视频中人体轮廓（Contour）信息、人体形状（Shape）、包含人体头部，四肢，躯干等身体部位、以及动作视频图像的纹理、颜色等；动态特征一般包括光流[31][32]、速度[33]、轨迹[34]等，主要描述人体在场景中运动路径、速度等信息；时空特征是将视频图像序列看成时空域上的三维立方体，然后选取特征点计算其特征并进行动作识别，如：时空形状[35]、时空兴趣点[36]等。描述性特征即充分利用动作视频中丰富的语义信息来进行动作识别[37][38][39]。由于实验中使用的人体动作视频数据库是在真实环境中的，场景复杂、类内变化较大、人体动作外观不固定等原因。因此，本节主要介绍实验中所采用的基于全局的颜色特征、纹理特征、形状特征和多尺度LBP特征的提取方法。

**2.3.1 颜色特征**

颜色特征是视频的关键帧图像中最重要的特征之一，能简化目标提取和分类的重要表示。颜色特征有颜色直方图[40]、颜色矩（Color Moments，CM）和颜色聚合向量(Color Coherence Vector，CCV)等。颜色直方图反映的是图像中不同颜色所占的比例，具有对图像尺度缩放、平移、旋转等的不变性的优点。

通过颜色特征来进行视频中人体动作识别，首先考虑选择的颜色模型。常用的颜色模型有RGB颜色模型和HSV颜色模型。由于前者往往分布不均，并且与人类视觉感知的差异很大。因此，我们选择HSV颜色模型来进行特征提取。HSV颜色模型符合人眼的视觉特性，并且直接与人眼颜色视觉特性三要素相对应，分别为色调H(Hue)、饱和度S(Saturation)和明度V(Value)，模型如图2.1所示。该模型用一个倒立正圆锥来描述，色调H表示色彩信息，从0。到360。。饱和度S代表的是颜色的纯度，从圆锥的中心到圆锥的外框取值范围变化从0到1。明度V代表的是颜色的明亮程度，取值范围变化也为0到1。在获得关键帧的颜色特征后，需要进行颜色模型转换，将RGB转换为HSV。定义，，则转换公式如下所示：

 (2.1)

 (2.2)

 (2.3)

 (2.4)

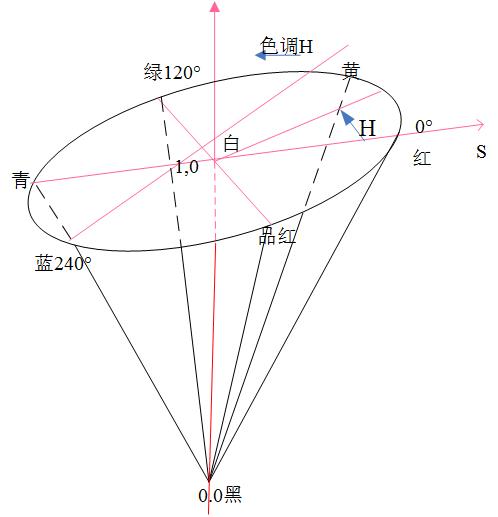


图2.1 HSV颜色模型

由于计算整幅图像，运算量较大。因此需要对提取的关键帧图像进行处理。根据注意力模型理论[41]，图像的中心部分往往更能引起人们的重视。因此，我们将关键帧图像分成4×4个区域，并赋予不同的权重，如表2.1所示。

表2.1 颜色权重分配表



此外，根据人的视觉分辨能力，在把RGB模型转为HSV模型后，通常把色调H空间分为9份，饱和度S和亮度空间分为3份，并根据主观感知和不同颜色色彩范围对其进行量化。从而分成81（9H×9S×3V）个子空间。

最终结合各个分块对应的颜色权重，统计每个子块中各颜色出现的频数并进行归一化处理，得到每帧图像的颜色特征向量。

**2.3.2 纹理特征**

纹理特征[42]是一种反映图像中同质现象的视觉特征，通过像素及其周围空间邻域的灰度分布来反映物体的质地。在20世纪70年代，Haralick等人[43]提出了灰度共生矩阵(GLCM)，它通过灰度图像来得到共生矩阵，进而得到矩阵的特征值。该值能反映图像灰度关于方向、相邻间隔以及幅度变化的综合信息。经过理论证明，GLCM是目前最常见、效果最好的一种统计分析方法，被广泛应用于将灰度值转化为纹理信息。本文选取常用的六个特征来提取图像的纹理特征：

1. 一致性：

 (2.5)

1. 平滑度：

 (2.6)

1. 标准差：

（u为均值） (2.7)

1. 熵:

 (2.8)

1. 平均值：

 (2.9)

1. 三阶矩：

 (2.10)

每帧图像的纹理特征向量为： 。

**2.3.3 形状特征**

形状特征比较符合人眼的感知特性，通常情况下有两种表示形式，一种是基于轮廓的特征，另一种是基于区域的特征。前者主要针对图像中事物的边界来获得轮廓信息，或者则关系到事物的整个形状区域。几种典型的形状特征有边界特征法、傅里叶形状描述符法、几何参数法、形状不变矩法等。径向Tchebichef矩特征[44]作为基于区域的形状特征，具有平移、缩放不变性，处理简单等特点。因此我们选择其作为人体动作视频关键帧图像的形状特征。

在计算径向Tchebichef矩之前，我们先对Tchebichef多项式进行简化，公式(2.11)展示的是离散Tchebichef多项式的递推算法：

 (2.11)

 (2.12)

当n≥2时，

 (2.13)

其中表示常量，与变量无关，令。

另外，Tchebichef多项式的对称性如公式(2.14)所示：

 (2.14)

结合上述Tchebichef多项式的两种特性可以化简Tchebichef矩算法，然后通过Clenshaw化简多项式二维求和中涉及到的正余弦运算。本文实验结合上述方法提取到5维径向Tchebichef矩特征向量。

**2.3.4 多尺度LBP特征**

LBP特征是由T.Ojala[45]等人在1994年提出，是一种局部二值模式，它主要反映像素与周围像素之间的关系。首先在3×3的领域像素内定义一个LBP算子，把中心的像素值与周围的进行比较。若某一像素值大于中心像素值，则该位置被填上数字1，否则为0，即就可以产生8位二进制数。并将这8位二进制数按逆时针排成一个二进制数字，这个二进制就代表了中心像素的LBP值，共有256（28）种可能。此中心像素值即反映了该区域的纹理信息。该过程如图2.4所示。



图2.4 LBP算子示意图

随后，Ojala[46]等总结了LBP特性，并对其进行了改进，将3×3正方形邻域替代为圆形邻域，改进后的LBP允许在以半径为R的圆形内有多个像素点，设gc为图像中的任何中心像素点，则周围像素点的纹理特征T定义如下：

 (2.15)

以中心点为阈值对周围像素做二值化处理，则可写为：

 (2.16)

其中。根据公式（2.16）可以得到一组二进制数，然后对不同位置的像素点加权求和，得到以为中心，为半径的区域对应的LBP值，计算公式如下所示：

 (2.17)

在本文实验中，我们使用多尺度LBP来进行人体动作视频关键帧的特征提取。通过选择（8,2），（8,3），（8,4）三种不同尺度来提取特征，然后将得到的LBP特征组合在一起得到图像帧的特征向量：LBP=（LBP8,2，LBP8,3，LBP8,4）。

**2.3.5 多特征融合**

由于单一的特征不能充分表示视频中的人体动作信息，本文采用特征级融合方法来对提取的特征进行融合。特征级融合[46]分为串行融合和并行融合，串联特征融合指的是将提取的视频关键帧的个特征联合在一起，形成一个特征向量，然后选择相应的模型进行分类处理。后者把不同特征组合成复向量，以进一步进行识别。实验中我们采用串联方式对人体动作视频特征进行融合。首先提取动作视频关键帧图像的各种底层视觉特征，这里我们选择颜色特征、纹理特征、形状特征、多尺度LBP特征，然后融合所有的特征形成特征矢量。融合模型如公式（2.18）所示：

 (2.18)

其中表示四种特征融合后生成的122维特征的集合，表示81维颜色特征的集合，表示6维纹理特征的集合，表示5维形状特征的集合，表示30维多尺度LBP特征的集合。

**2.4 人体动作识别方法**

人体动作识别过程是一个动态化的识别过程，本质上是对视频中随时间变化的数据的分类问题。图2.5展示的是动作识别方法的分类：



图2.5 人体动作识别方法分类

**2.4.1 基于形状模板的方法**

基于形状模板的动作识别方法一般会预先定义人体动作的模板（2D或者3D），然后将视频序列中待识别样本的人体动作表示信息与已知人体动作进行比较，通过计算相似度来度量最终的分类效果。依据识别的对象不同，该方法又可为：模板匹配 [47[48][49]和动态时间规整 [2]。前者匹配的是一个对象，后者匹配的是一组静态模式。

1）模板匹配

模板匹配是一种简单、直观的模式识别方法，它将事先已有的模板与输入的样本数据直接进行比较。该方法首先给视频中的动作建立一个或者多个模板组成模板库。接着将输入的待识别样本特征与模板库中的模板进行匹配，通过计算相似度来进行动作识别。

在模板匹配中常用的特征有颜色、形状轮廓、时域信息等。一般通过距离测量的方式来计算相似度，常用的距离测量方法有马氏距离(Mahalanobis distance)和欧氏距离(Euclidean distance )，前者计算两者之间的协方差，而后计算两点间的真实长度。David等人[50]在动作识别的训练阶段将视频中帧序列转换成运动能量图像(MEI)和运动历史图像(MHI)，提取了基于矩的特征，并作为动作类型模板库。在识别时，通过计算最小马氏距离来判定待识别视频所属的类别。[51][52] 等人提出了采用K-近邻/最近邻的人体动作识别方法，通过计算测试视频关键帧的特征数据与训练视频关键帧数据的距离来进行动作识别，该类方法本质上也是模板匹配方法。

虽然模板匹配的方法计算简单，比较容易实现，但是对动作时间变化较为敏感，不能解决执行时间长度不一致的问题，缺乏鲁棒性。此外，构造识别模板存在主观性，难免会影响最终的分类结果。

2）动态时间规整

动态时间规整算法（Dynamic Time Warping）[47][53]最初由日本学者在20世纪60年代提出，旨在把两个代表同一类型的事物的不同长度序列进行时间上的规整对齐。针对模板匹配方法对时间的敏感性，将识别模板在时间上进行规整是很有必要的。该算法鲁棒性较好，有效解决了视频中人体动作在时间上的不确定性。但是缺点是易受噪声的影响，计算量会随着训练数据的增多而增多，且在样本过多时，无法获取有代表性的识别模板。

**2.4.2 基于概率统计的方法**

概率统计方法是把视频中的人体动作表达成一组连续的状态序列，然后通过转移函数来构造状态变化规律。常用的概率统计方法包含两类：产生式模型（Generative Model）和判别式模型（Discrimitive Model）。

1）产生式模型

产生式模型有较强的灵活性，其方法主要是观测识别样本与状态之间的联合概率分布，常用的有：隐马尔可夫模型（HMM）、动态贝叶斯网络（DBN）。鉴于在语音识别方面的成功，Yamato最早将HMM应用到人体动作识别中[54]，之后许多学者提出了各种改进模型。Park等人提出了基于树状结构DBN动作识别方法，通过把人体姿势划分为不同的层次等级，来识别两个人之间的交互动作[55]。但因为网络的参数过多，构造比较复杂，并且也需要标注大量的有标号数据来得到函数的联合概率分布，难以在动态系统的识别中得到广泛应用。

2）判别式模型

产生式模型对马尔可夫性和输出独立性有较强的依赖性，并且为了得到联合概率分布，通常需要大量的训练样本。而判别式模型能够根据提取的特征直接计算条件概率分布，并且不需要单独学习每个动作类别，很好地解决了产生式模型难于处理相似动作的分类。常见的判别式模型有支持向量机SVM（Support Vector Machine）、条件随机场CRF（Conditional Random Field）等。支持向量机是一种机器学习算法，在解决非线性和高维模式识别中有许多优势。Schuldt[56]最早将其应用到人体动作识别中，之后许多学者进行了拓展[57][58][59][60]。Schindler等人分别提取视频中人体动作的局部特征和运动特征，作为SVM的输入特征，并在动作视频库上取得了较好的识别结果。CRF是一种无向图的概率学习模型，通过计算整个观测序列的联合概率来拟合现实中的数据。Natarajan[61]用条件随机场来表示从多个角度提取的人体姿态信息，并计算转移概率。

虽然判别式模型有很强的分类能力，但它的结构固定，缺乏泛化性能，并且难以通过模型来体现变量之间的关系。

**2.4.2 基于语法的方法**

基于语法的方法高度抽象了动作的属性特征，为视频中人体动作建立一种高层次的描述。最具有代表性的是有限状态机（Finite State Machine，FSM）[62]模型。但是最初的有限状态机体现不了样本的分布规律，因为是通过有限的数据来手动建立模型，而不是对大量的样本进行训练以获得能表示各状态转移的拓扑结构。并且系统是按照时间顺序将不同时刻的观测值作为状态值添加到模型中，因此鲁棒性很差，如果某一状态出现错误，就会影响整个模型的识别率。

综上所述，不同的动作识别方法都有优缺点，基于形状模板的方法直观、简单，但是缺乏鲁棒性，一般用于静态姿势或是简单动作识别中；基于概率统计的方法应用广泛，但需要大量的训练数据来学习模型参数；基于语法的方法能够对先验知识进行有效利用，但计算复杂，难以得到推广。

**2.5 人体动作识别技术研究难点**

目前，在各个层次上，人体动作识别技术都取得了巨大的研究进展，涌现了许多优秀的学习算法，但如何设计一套高性能、鲁棒性的识别方案仍然存在许多问题。主要难点和挑战如下：

1）运动的分类和定义

针对不同的动作类别，学术界没有一个统一的划分标准，各种姿态、行为界限模糊，并且视频中完成一个动作，需要哪几个阶段，由哪几部分组成，如何界定动作的开始和结束，这些都需要研究人员根据经验自行决定和划分。

2）存在动态背景

在真实场景中，往往存在不同的目标同时运动，在这种情况下，如果背景处于不断变化之中，会使得背景和前景分离变得相当困难，并且由于动作目标存在部分遮挡，给提取的视频特征增加了大量噪声。

3）动作类型内部以及类型之间的变化

同一个动作类别，由于尺度的变化，即使由同一个人完成，在视频图像中的表现大小可能也是不一样的。例如步行，可以在不同的场景中完成，走路速度的快慢直接影响到步长的长短，并且从不同的角度观察，它们也表现出不同的时空特征。因此，这也是人体动作识别领域急需研究的问题。

4）模型的泛化能力

目前在动作识别领域大多采用单一的方法对动作进行识别，无论是基于模板的方法还是基于概率统计的方法都有自己的局限性，这一定程度上制约了模型的泛化能力。如基于模板的方法在标签不足时性能下降较为明显，而基于概率统计的方法在类重叠时性能下降较为明显。

5）标注瓶颈

人体动作识别的研究经历了从受限场景中简单动作的识别到电影中以及现实生活中等复杂场景的人体动作识别。而目前现有的动作数据库中样本数据有限，需要对大量的无标记视频信息进行标注。利用人工标注的方式费时费力，半监督学习方法是一个好的解决方案。

针对人体动作识别技术存在的问题，本文从减少大量的有标记训练样本和提高模型的泛化能力入手，引入了半监督下的协同训练和集成学习相关技术，并取得了较好的识别效果，接下来第三章、第四章将对提出的算法进行介绍。

**2.6 本章小结**

人体动作识别是目前研究的热点，本章对该领域的相关研究技术进行了概述。首先介绍了视频中关键帧的提取方法以及它们各自的特点，然后重点介绍了基于静态的人体动作特征提取方法和融合方法，接下来对人体动作识别方法作了具体介绍，并分析不同方法的优缺点。最后阐述了人体动作识别技术研究难点。

**第三章** **基于混合式协同训练的人体动作识别方法**

**3.1 概述**

现有的人体动作识别方法通常利用数量庞大的已标注样本来获得性能优异的分类器模型，而在实际应用中，对样本进行标记耗时且耗力，这也是限制监督学习方法发展的阻碍之一。近年来，不少学者对未标记样本的使用进行了研究，基于半监督的协同训练算法作为其重要的方法之一得到了快速的发展，并成功应用在人体动作识别领域。鉴于此，本章提出一种基于混合式协同训练的新型人体动作识别算法Co-KNN-SVM，该方法是一种协同训练方法，能够实现不同识别方法的优势互补，首先选择一种基于模板的分类器KNN和一种基于概率统计的分类器SVM作为基分类器，然后，每次迭代过程结束后利用产生的中间基分类器标记无标号数据，并按一定规则选取置信度靠前的加入到对方的有标记样本集中来更新基分类器，最后通过集成的基分类器组对人体动作进行识别。

相比于目前的动作识别方法，本章的主要贡献如下：

1）本节提出的基于混合式协同训练的人体动作识别方法，对传统的协同训练算法进行了改进，可以有效利用丰富的未标记示例并辅以少量的有标记示例来提高系统的性能，并且该方法不要求数据集有两个冗余的视角，满足了一些现实的应用。

2）算法引进了伪验证集机制，并提出一种新的置信度度量方法来选择伪标记数据，有效控制了分类噪音和分布噪音。通过实验和理论分析，表明基于混合式协同训练的人体动作识别方法的有效性。

**3.2 传统协同训练算法的分析**

协同训练方法是一种典型的半监督学习算法，这里以Co-Training[64]，Tri-Training[63]为例进行简单的介绍。

Co-Training作为一种规范的协同训练类算法，由Blum等人在20世纪末提出，又被称为基于分歧的半监督学习。其基本流程为：选取两个不同的基分类器，然后对两个不同视角的有标记样本集进行学习，通过新生成的分类模型对无标记样本添加标号，形成伪标号数据。在伪标号数据中按照一定规则选取置信度排名靠前的作为信任样本加入进有标记数据集中，更新训练样本。这一过程不断迭代进行，直至到达某个条件。最终，组合两个新生成的分类器对测试集进行学习。如下图3.1所示展现的是标准协同训练的结构图。其中L，U分别代表标记的训练集和未标记的训练集，h1，h2代表不同种类的分类器，U’代表被标记的伪标号数据。



图3.1 协同训练结构图

Co-Training算法要求样本集有两个冗余的不同视角，需要同时符合以下两个条件：第一，每个视角的数据都能有效表述所提的问题，当样本量足够时，对其进行学习都能产生强分类器。第二，满足独立性要求，给定两个视图样本集，和类别标号，满足如下条件：

 (3.1)

 (3.2)

但是在实际应用问题中，条件独立性这一要求很难满足。所以，部分学者对不需要冗余的视角的协同训练类算法进行了探索总结，研究出一些改进的方法。比较典型的有Zhou和Li提出的Tri-Training算法，该方法是一种改进的多分类器半监督学习算法，不需要数据集有冗余的独立视角，并且对分类器的类型没有限制。首先通过对数据集重复采样，训练产生三个不同的分类器，接着在迭代训练过程中利用“少数服从多数”策略标记无标号数据，即对未标记样本来讲，当有两个基分类器预测结果是一致的，则认为该样本其在结果上具有较高的可信度，且对另一个分类器有所帮助。接着，我们将添加标记的无标记样本加进第三个分类器的训练样本集中。最后通过集成的分类器组对未见数据进行学习。

Tri-Training算法利用多个分类器，虽然可以解决未标号样本置信度的计算问题和测试样本的学习问题，并且能利用集成学习来提高模型的泛化性能，但是当初始分类器较弱时，无标号的示例被判别错误，进而为第三个分类器引入噪声。针对以上研究和问题，本文中提出的混合式协同训练算法，分别选择差异的分类器KNN和SVM作为基分类器，不需要数据有冗余的独立视角，此外通过定义伪标号样本的获取规则和控制训练类别比例降低了样本的分类噪音和分布噪音。下一节，我们将对该方法进行详细介绍。

**3.3 混合式协同训练的人体动作识别算法**

**3.2.1 基分类器的构造**

在本节，我们采用一个基于模板的和一个基于概率统计的分类器来构建协同训练的算法模型。

1）基于模板的分类器

由于KNN（K 最近邻法）方法主要依赖周围有限的邻近样本，对于待分类数据中有较多重叠来说，KNN算法较其他方法更为合适，并且可以有效地克服样本类别比例不均衡和类域重叠问题。因此，我们选择KNN作为基于模板的分类器。

KNN算法思想：在特征集中，假使其中以样本数据点与其K个最邻近的数据点的绝大多数同属于一个类别，那么就把这个样本数据划分为该类别。即有一个样本训练集L，对于输入的新的测试样本，在L中查找与该样本最邻近的K个有标记训练样本，如果这些样本属于某一类别，则把该样本也判断为这个类别。

2）基于概率统计的分类器

对于基于概率统计的分类器，我们选取支持向量机SVM。SVM具有解决高维问题的特点。SVM是一种线性学习方法，通常寻求建立一个由表示的决策面。当时，否则为0。其中参数可通过对下面式子的极大化进行估计：



  (3.3)



其中参数称为松弛变量，C是离群点的权重，越大表明离群点对目标函数的影响越大。

**3.2.2 混合式协同训练算法Co-KNN-SVM**

伪标记选择的好坏对协同训练的性能有着至关重要的影响，为了合理控制扩充后的有标记样本集的分类噪音和分布噪音。我们从以下两个方面对数据进行处理。

1）提高伪标记样本的准确性

为了控制产生的伪标记样本的分类噪音，我们从提高数据的准确性入手，首先计算无标记数据的预测置信度，然后对其进行排序，选取较高的添加到有标记样本中来扩充训练集。对于无标记数据，KNN分类器对每个动作识别类别都能计算出一个预测概率：

 (3.4)

其中，为设置的近邻样本的数量，为不同类别包含在测试样本近邻中的个数。为了计算无标记样本的置信度，常用的做法是，根据不同类别的计算概率作为判断依据，选取最大类的预测概率值作为置信度，即。但仅考虑类的最大预测概率作为置信度有时会不够合理。因此我们采用一种新的置信度计算标准：通过计算的最大的预测概率值减去第二大预测概率值来表示置信度的大小。如公式(3.5)所示：

  (3.5)

其次，对于基分类器SVM，由于它通过超平面来对样本进行分类，并不计算类别的预测概率。我们参考文献[17]中的做法来转化预测概率，然后同样把置信度表示为：

 (3.6)

这种方法计算置信度，能从伪标记数据集中剔除类别交界处的样本，并且将置信度较高的伪标记数据加入到有标记样本集中有助于解决KNN在小样本情况下的性能不足的问题，以及降低训练样本中类别不平衡的不利影响。

2）控制样本的类别比例

在训练过程中，由于不是独立地随机从伪标记数据集中选取样本，因此其分布与真实分布之间必然存在误差。我们采用控制样本类别比例的方法来降低数据的分布噪音，即在迭代训练过程中，按照原始的有标记样本的比例动态地向训练集中添加伪标记数据。

进一步地，我们改进了伪标记的添加策略，传统的协同训练往往根据置信度，单一地向有标记样本中添加伪标记数据。为了提高性能，我们选取初始的有标记训练集，以及伪标记数据集和置信度的前20%形成伪验证集。然后通过在验证集上的准确率估计基分类器的性能；另一方面，我们用估计其错误率：

 (3.7)

其中表示的是对无标记数据分类结果中类别的比例，而表示原始有标记样本中类别的比例。因此，通过以及，从伪标记数据集中剔除影响分类器性能的样本。下面我们给出具体的算法步骤：

**算法3.1 Co-KNN-SVM算法**

输入：有标记数据集，无标记数据集，以及伪标记数据集。设置; ; ，最大迭代次数为。

**Step1：**分别在在上训练KNN和SVM，然后分类中数据，并分别估计其、和、。

**Step2：**重复步骤3-5，直到。

**Step3：**SVM从中按标签比例选择与KNN分类相同并且置信度前的样本，加入中，并令;同理，KNN从中按标签比例选择与SVM分类相同并且置信度前的样本，加入中，并令。

**Step4：**分别在，上重新训练KNN和SVM，然后用它们估计其和，其中。

If  or 

{;;}

**Step5：**设置=，，。

输出：经过多次迭代训练后获得性能更优的分类器KNN和SVM经过迭代训练后得到更优的基分类器KNN和SVM。

**3.2.3 结合KNN和SVM**

上一节算法运行结束后，我们获得性能更优的分类器KNN和SVM。作为基分类器，我们通过一定策略组合两者来对测试样本进行最终的预测。由于3.2.2给出的分类器KNN和SVM对不同类别标签的预测概率值没有利用同一个判断标准上，缺乏参照性，因此我们归一化了两者的预测概率值，使其在同一评价尺度上，公式（3.6）给出了最终预测结果的计算方式，如下所示：

 (3.8)

**3.4 实验结果与分析**

本章中我们选择UCF YouTube Action Data Set 作为人体动作识别实验数据集来评估所提算法的性能，它是YouTube上收集的各种体育运动。不同于受限制场景的简单动作视频库，它包含了大量的光照变化、障碍物遮挡等复杂场景，具有挑战性。进一步地，我们与其他算法作了对比，并分析了本文所提算法的有效性。

**3.4.1 人体动作数据集**

UCF YouTube Action Data Set[25]是关于YouTube的动作数据集库。该数据库共包含投篮、骑车、跳水、挥杆、骑马、花式足球、荡秋千、网球挥拍、蹦床、排球扣球、遛狗等11种不同的动作。其中每一类由25个人做动作，每个人做4-7组，共有1600个视频序列。这些视频来自于YouTube的上传者，大多由手持设备拍摄，往往包括了大量视角变化，背景混杂以及动作的时空差异，其挑战性也是不言而喻的。该数据集的视频分辨率为320\*240 像素，图3.2给出了部分样图。



图3.2 UCF YouTube Action Data Set

在实验中，我们使用基于视频片段边界的方法来提取YouTube数据集的关键帧，分别选取每个视频的第一帧、最后一帧和中间的一帧作为关键帧，然后采用第二章所提的方法对图像的四类底层特征进行提取，分别为颜色、纹理、形状、多尺度LBP特征，最后采用串联融合的方法来获得视频关键帧的特征矢量，将该特征数据用于本章后续实验中。融合的特征向量为：

 (3.7)

其中代表融合后的矢量，表示颜色特征，表示纹理特征、表示形状特征，表示多尺度LBP特征。

**3.4.1 评价标准**

为了对所提方法的可靠性进行判断，本文采用交叉验证（Cross Validation）的思想来估计算法模型的准确率。交叉验证也称为循环估计，是一种将样本集划分成若干部分的统计学方法。其方法首先按一定规则把样本集分为训练部分（train set）和验证部分（validation set），前者对分类模型进行训练，使其具有一定的分类性能。最后，通过验证部分对模型进行测试，判断其准确性。常见的交叉验证有以下两种：

1）折交叉验证（-fold Cross Validation）

折交叉验证是指将样本数据均分成个部分，并轮换把每个子集部分作为一次验证集，其余部分作为训练集，然后把计算的平均正确率作为最终的结果，这里的取值一般为10.

2）留一交叉验证（Leave-one-out Cross Validation）

假设样本集共有个样本，留一交叉验证就是把每个样本集单独作为一次验证集，剩下的个样本作为训练集。由于每一次选取的几乎是全部的样本来作为训练集，该方法具有较高的可靠性。但是当样本数量庞大时，计算代价大，对硬件环境有较高的要求。

因此，针对于UCF YouTube Action Data Set数据集，我们使用十折交叉验证来计算算法模型预测的准确率。即把视频数据样本随机地分成十份，并轮换地选取一份作为训练集，其余部分作为验证集，并对模型进行训练和验证。

**3.4.2 实验步骤及方法**

本实验所用的计算机操作系统为64位的Windows 10系统，硬件参数如下：硬盘为三星 MZVLV256HCHP-000L2 256GB固态硬盘，内存为8.0GB DDR4 2133MHz，CPU 为主频2.30GHz的Intel Core i5-6200U。所有的实验过程均在MATLAB 中进行设计和实现，实验所用MATLAB版本为MATLAB R2014a。在MATLAB的实验中进行Co-KNN-SVM协同训练分类时，SVM采用的是Libsvm工具箱，该工具箱是台湾大学林智仁等开发的一套支持向量机库。KNN采用的是在MATLAB中手动编写的代码。

我们进行三组实验来验证本文方法的有效性，它们分别是：

1）在有标记样本为20%，未标记样本为80%下，对我们的算法与监督学习算法进行了对比，并分别给出了混淆矩阵。

2）与其它协同训练类方法进行比较，在协同训练类算法中，我们选择Co-training和MCM算法。Co-training是一种标准的协同训练算法，采用2种不同的分类器，分别在视觉和纹理两种不同的视角上训练数据。MCM采用文献10中多分类器协同训练的的标准设置，首先采用基于Q统计量的分类器选择算法选择一组基分类器，然后对基分类器进行协同训练。

3）此外，我们讨论了在同一训练数据集下，不同的有标记样本个数和无标记样本个数对分类器性能的影响。

**3.4.3 实验结果分析与讨论**

（1）与监督学习进行对比

监督学习是对训练集中有标记样本进行训练而得到分类模型，半监督学习是在利用有标记样本的同时并辅以大量的无标记样本来训练生成模型。本节采用十折交叉验证的方法来对所提的混合式协同训练算法Co-KNN-SVM与传统有监督算法的比较。为了统一评价标准，在对比试验中，我们把有标记样本比例设置为20%。表3.1，表3.2分别为数据集在Co-KNN-SVM算法上和在监督学习上的混淆矩阵。第个数据代表的是第类动作被判断为第类别的比例。

通过对表3.1和表3.2的分析，我们可以得到，在UCF动作视频库上，相比于传统的有监督的学习方法，本文所提方法取得了更高的识别准确率。在各个动作类别上，平均识别率达到了78%。在骑车、骑马、足球、排球等动作类别上提高尤为明显，究其原因是这些动作类别存在场景和动作之间的相似性，当有标记的训练样本数量不足时，混淆较为明显。而我们的方法是基于半监督的，可以通过大量的无标记样本来辅以少量的有标记样本来进行学习，提高模型的识别效果。

（2）与协同训练算法进行对比

表3.3给出在不同有标记样本数下对不同的动作识别算法的准确率比较，与Co-training方法进行比较，本文方法识别率提高较为明显。究其原因是Co-training方法通常需要两个相互独立且冗余的视角，而我们的提取的视频特征不能很好地满足这个要求。并且在进行协同训练时，没有有效控制样本的分类噪音，所以识别准确率低于我们的所提的方法；与MCM[20]方法进行比较，本文方法识别率提高了2.6%。究其原因是MCM选取一组Q统计量来对多个分类器进行协同训练，并根据多个基分类器来预测和选择伪标记样本，通过引进集成学习的思想，提高标注样本的准确率的同时较好地控制了训练样本的分类噪音。但是该方法没有效地解决分布噪音问题，所以分类结果略低于我们所提的方法。

同时我们还把所选取的基分类器KNN和SVM的预测结果加入进行比较，它们本身是监督类学习方法。从表中可以看到，在不同有标记样本数下，我们方法的准确率均得到了较明显提高，这也进一步验证了上一节的观点。

表3.1 Co-KNN-SVM算法

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类别 | 投篮 | 骑车 | 跳水 | 挥杆 | 骑马 | 足球 | 秋千 | 网球 | 蹦床 | 排球 | 遛狗 |
| 投篮 | 0.81 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.04 | 0.00 | 0.10 | 0.00 |
| 骑车 | 0.00 | 0.79 | 0.00 | 0.00 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 跳水 | 0.05 | 0.00 | 0.80 | 0.00 | 0.03 | 0.02 | 0.03 | 0.00 | 0.04 | 0.00 | 0.03 |
| 挥杆 | 0.00 | 0.00 | 0.00 | 0.70 | 0.00 | 0.10 | 0.00 | 0.12 | 0.00 | 0.08 | 0.00 |
| 骑马 | 0.00 | 0.20 | 0.00 | 0.00 | 0.75 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 |
| 足球 | 0.10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.84 | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 |
| 秋千 | 0.00 | 0.00 | 0.00 | 0.12 | 0.00 | 0.00 | 0.83 | 0.00 | 0.05 | 0.00 | 0.00 |
| 网球 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.79 | 0.00 | 0.15 | 0.00 |
| 蹦床 | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.77 | 0.00 | 0.03 |
| 排球 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.05 | 0.00 | 0.82 | 0.00 |
| 遛狗 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.12 | 0.10 | 0.00 | 0.00 | 0.74 |

表3.2 监督学习算法

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类别 | 投篮 | 骑车 | 跳水 | 挥杆 | 骑马 | 足球 | 秋千 | 网球 | 蹦床 | 排球 | 遛狗 |
| 投篮 | 0.75 | 0.10 | 0.00 | 0.05 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 |
| 骑车 | 0.10 | 0.15 | 0.00 | 0.15 | 0.15 | 0.00 | 0.10 | 0.10 | 0.00 | 0.00 | 0.25 |
| 跳水 | 0.00 | 0.00 | 0.80 | 0.00 | 0.05 | 0.00 | 0.10 | 0.05 | 0.00 | 0.00 | 0.00 |
| 挥杆 | 0.05 | 0.00 | 0.00 | 0.80 | 0.00 | 0.05 | 0.05 | 0.05 | 0.00 | 0.00 | 0.00 |
| 骑马 | 0.00 | 0.10 | 0.00 | 0.05 | 0.55 | 0.00 | 0.15 | 0.00 | 0.05 | 0.00 | 0.10 |
| 足球 | 0.15 | 0.00 | 0.00 | 0.25 | 0.10 | 0.45 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 |
| 秋千 | 0.10 | 0.00 | 0.00 | 0.10 | 0.15 | 0.00 | 0.60 | 0.00 | 0.05 | 0.00 | 0.00 |
| 网球 | 0.00 | 0.08 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.87 | 0.00 | 0.00 | 0.00 |
| 蹦床 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.05 | 0.70 | 0.00 | 0.05 |
| 排球 | 0.20 | 0.05 | 0.05 | 0.05 | 0.10 | 0.00 | 0.17 | 0.00 | 0.00 | 0.40 | 0.05 |
| 遛狗 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.35 | 0.00 | 0.00 | 0.00 | 0.00 | 0.40 |

表3.3 本文方法与其他方法的比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 有标记样本数 | SVM | KNN | Co-raining | MCM | 本文方法 |
| 220 | 0.47 | 0.52 | 0.61 | 0.76 | 0.78 |
| 440 | 0.65 | 0.67 | 0.70 | 0.78 | 0.81 |
| 660 | 0.74 | 0.75 | 0.78 | 0.78 | 0.80 |
| 平均准确率 | 0.62 | 0.67 | 0.70 | 0.77 | 0.80 |

（3）不同数量的有标记数据和无标记数据对分类性能的影响

半监督学习方法的一个重要的参数就是样本的数量，接下来我们就不同的有标记样本数量和无标记样本数量对分类器性能的影响多做一分析。我们对有标记样本个数分别取220，440和660，无标记样本个数分别取0，440和880。

图3.3给出的是不同的有标记样本数量对分类器性能的影响。从图中可以看出：一方面，随着无标记样本数量的加多，分类器的精度均得到了提升，最高达到了81%；当有标记样本较少时，由于选择置信度较高的伪标记数据，通过协同训练充分挖掘这些数据隐藏的样本分布信息，来帮助提高模型的性能。另一方面，虽然半监督学习方法能够利用无标记样本数据来提高分类器的准确率，但是样本的数量对于分类器还是很重要的。当有标记样本数量为220时，分类器的性能提升程度最快。而当样本数量为660时，利用已有的有标记样本数据本身就可以训练性能较好的分类器，通过继续添加无标记数据，不能继续有效提高模型的性能。



图3.3 不同的无标记样本数对Co-KNN-SVM的影响

图3.4给出的是不同的无标记样本数量对分类器性能的影响。当无标记样本个数为880时，在不同的有标记样本数量下，我们所提的方法都能取得较好的识别效果。这也进一步证明了，如何控制训练集中有标记数据和无标记数据的比例，是半监督协同训练算法重要的一环。



图3.4 不同的有标记样本数对Co-KNN-SVM的影响

**3.5 本章小结**

本章提出了一种基于混合式协同训练的人体动作识别方法，利用人体动作识别领域不同类型的算法来构造基分类器，并进行迭代的相互训练以提高泛化性能，可以降低标注成本并实现不同识别方法的优势互补，提高了人体动作识别的准确性。

但是在分类器迭代训练的时候，随着次数的增加，基分类器的差异性也会逐渐消失。下一步工作将引入集成学习的思想来遏制差异性的退化，并根据D-S证据理论进行分类器融合，进一步提高人体动作识别系统的泛化性能。

**第四章 基****于半监督集成学习的人体动作识别方法**

**4.1 概述**

上一章中，通过对人体动作特征的带标记的样本和不带标记的样本进行半监督学习，得到两个性能更优的分类器，然后通过融合后的学习模型对视频中动作进行分类。但是该方法迭代训练时，当循环次数不断增加时，基分类器之间的差异逐渐变小。当添加到有标记样本集中的无标记数据基本相同时，基分类器便不能有效利用差异性来实现互补。而集成学习的一个重要方向就是研究如何创造有差异的分类器。所以，集成学习与半监督学习能形成良好的互补。鉴于此，本章提出的基于半监督集成学习的人体动作识别算法ECKS，在分类器迭代过程中，利用联合的分类器组为无标记数据添加标号。该方法有效遏制了迭代训练过程中基分类器差异性的退化问题，并且通过基于置信度的最大证据边缘函数和样本类别概率组合的方式来选择伪标号数据，以增强系统的泛化性能。

本章的主要贡献如下：

1）针对协同训练算法随着迭代次数的增加基分类器差异性退化的问题，提出了Co-KNN-SVM与集成学习相结合的新型人体动作识别算法ECKS（Ensemble Co-KNN-SVM），该方法把协同训练过程中生成的中间分类器加入到两个不同的分类器组中，接着用分类器组对无标记数据进行预测标注，并选择置信度高的加入到对方的训练集中，该方法提高了分类器的差异性。

2）定义了基于置信度的最大证据边缘函数，并将最大证据边缘函数与预测概率相结合，来选择伪标号数据。最后对测试样本集的预测结果也是通过其计算得来的。

3）提出一种基于伪验证集的分类器选择方法，在迭代过程中，根据新生成的中间分类器在伪验证集上的准确性和预测的接近程度来决定是否加入分类器组中。

**4.2 集成学习**

建立一个有效的集成学习系统一般包含两个步骤，基分类器的创建和基分类器的组合[64][65]。并且构造的基分类器之间存在较大的差异性，以及对样本的预测有较高的准确率，其根据是不同的分类器对同一样本产生的预测错误一般不同时发生。因此，可以通过这种性质，对分类器进行集成，提高模型的学习性能。

**4.2.1基分类器的创建**

基分类器差异性是衡量集成学习方法的重要标准之一，创建不同的基分类器对集成学习的性能起着至关重要的作用。以下对常用的四种创建方法进行介绍。

（1）对训练数据集进行处理

集成学习最直接的方式就是利用重采样技术对样本进行多次抽样来获得不同的训练子集[66]，然后使用不同的分类算法对这些子集进行学习以建立基分类器，一个示例能否被选择为训练子集中的数据与具体的抽样分布有关。流行的方法有装袋（bagging）和提升（boosting）。前者通过对样本赋予权值的方式来改变数据分布；而后者一次选择若干个不同的子集来产生多个分类器。这两种方式在有标记样本较多时能更好地获得有差异的分类器。

（2）对输入特征进行处理

这种方式通过对样本的不同特征子集进行学习，从而产生不同的基分类器。特征子集的获得又被称为集成特征选取，常用的方法有随机子空间方法（Random Subspace Method）、基于遗传算法的特征选择方法。随机子空间是利用样本的随机部分特征来训练分类器，从而减少模型的相似性；而基于遗传算法的特征选择方法是通过对可能的特征子集进行评价，来选择最优的子集组合。

（3）构造不同类型的分类器

该方法通常选择类型不同的学习算法进行集成，也成为异态集成方式。其主要有叠加法（Stack Generalization）和元学习法（Meta Learning）。元学习法是得到一个元分类器，然后对所有分类器的预测作分析，最终获得判别结果。而叠加的方法是利用分布在不同层次的分类器来完成判别任务。

（4）调整分类器的结构

通常在每个基分类器的训练过程中引入随机性，来改变分类的构成。只要分类模型对随机性敏感，则能产生多个不同的学习算法。该类方法多用于人工神经网络的集成学习。例如，在使用后向传播算法对神经网络进行学习时，随机初始化不同的参数，都能产生不同分类结构的基分类器。该类方法较容易地与其它集成学习进行结合，以取得更好的学习性能。

**4.2.2基分类器的组合**

在得到不同的基分类器之后，下一步就是对这些分类器进行集成。通常有两种集成方式，一种是融合，另一种是选择。前者在现实中使用较为普遍，可以划分为数据融合、特征融合、决策融合三个层次。常用的融合方法有投票法、D-S证据理论[67]、模糊积分以及回归方法等。D-S证据理论是一种可信度较高的推理方法，在不需要知道先验概率的条件下，能够很好地结合各分类器提供的“证据”表示“不知道”和“不确定”，并且推理形式简单。将D-S证据理论应用于多分类器决策层融合，有效克服了传统融合机制对各分类器具体特性考虑不足的缺点。

证据理论是一种解决不确定问题的推理方法，它是建立在一个非空识别框架（或假设空间）上的理论，由一些互斥且穷举的元素组成，代表的是某一问题所有可能取值的集合。对于问题域中的任一假设，其中为Θ的幂集，给出概率赋值函数，满足

 (4.1)

 (4.2)

其中表示证据支持假设发生的程度，体现了证据对假设的可信度，式（4.1）表示在空集情况下可信度为0，而式（4.2）表示不论给不同假设赋多少可信度值，它们所有值的都为1，即总信任度为1。证据理论给出一个有效的计算公式，来合成不同证据源产生的证据。假设n个独立的证据其所相应的赋值函数为，则可以根据式子来计算合成后的概率赋值函数。Dempster证据理论合成公式为

 (4.3)

 (4.4)

其中表示的是冲突因子，衡量证据之间的冲突程度， 表示归一化因子，防止在空集中引入非零元素。

**4.3 半监督集成学习算法**

半监督学习通过挖掘和利用未标记样本来使得分类器具有更好的分类性能，而集成学习通过组合多个不同的分类器来提高模型性能。两种不同的方法并行发展，且存在互补性，所以把半监督学习和集成学习相结合是一个重要的方向。最先对其研究的是Miller等人[68]提出的专家组合分类器模型，作者使用概率密度函数来拟合样本数据，并通过期望最大值算法训练有标记样本和未标记样本。该方法是对集成学习和半监督学习的浅层结合。之后，Curie等人在文献[69]中提出一种半监督集成算法Margin Boost，并将其应用到图像识别中。算法是基于AdaBoost的最小化确定成本的一个版本，文中对边缘（Margin）进行了扩展，使其可以用于未标记样本。实现结果表明，该方法在少量训练样本可用的情况下，比传统的AdaBoost取得更好的性能。Bennet对Margin Boost进行了改进，提出了一种新的半监督学习和集成学习的结合算法ASSEMBLE[70]，它通过对有标记样本和未标记样本训练迭代的方式来构建基分类器组，克服了Margin Boost算法的一些局限性。Zhou[63]等人在2005年提出的Tri-Training也是一种半监督集成学习算法，通过使用三个分类器来改善学习性能。该方法不需要训练数据集有充分冗余的视图，并且对学习算法没有约束条件。

上面的方法大多使用当前的分类器组对无标号数据进行预测和标记，接着添加进有标记样本集中进行迭代训练。其本质都是分类器集合本身的自训练。因此该方法存在一些局限性，就是当初始基分类器集中分类器准确率不高时，随着迭代的进行，往往会不断累积错误，导致泛化误差增加，对模型的性能提升有限。

对于集成学习和协同训练相结合的方法，目前研究还较少。作为一种最初尝试，Liu等人将boosting引入到协同训练框架下，通过在协同训练过程中改变样本的权值来控制样本的分布。该方法使得新产生的学习器聚集在难于分类的样本上，可扩展性差，并且在迭代训练过程中没有通过当前集成的分类器对无标记数据进行判断和选择，影响了性能的提高。针对以上研究，我们结合Co-KNN-SVM算法与集成学习，提出了一种新型的半监督集成学习的人体动作识别方法ECKS。下一节，对其进行详细介绍。

**4.4 集成学习与Co-KNN-SVM的结合**

**4.4.1基于集成学习与Co-KNN-SVM的结合的ECKS算法**

Co-KNN-SVM将分类器的多差异性引入到传统的协同训练模型中，并改进了对伪标记样本的获取方法，降低了分类噪音和分布噪音，通过最终的基分类器组进行人体动作识别。方法本身就融入了一些集成学习的思想。但Co-KNN-SVM在协同训练过程中舍弃了产生的临时中间分类器，并且仅仅根据当前基分类器预测的置信度来选择伪标记数据，没有充分发挥集成学习的优势。

本章提出的基于集成学习与Co-KNN-SVM的结合的人体动作识别方法结构如图4.1所示。相对于协同训练Co-KNN-SVM，该方法进行了如下改进：



图4.1：ECKS系统结构图

1）为基分类器建立两个不同的集合，这里选用的基分类器分别为协同训练中的KNN和SVM。并把迭代过程中生成的临时中间分类器分别添加到指定的集合中，接着通过包含多个分类器的集合来预测和筛选伪标号样本。

2）引入证据理论的思想并结合当前的置信度计算方法，定义了一个基于置信度的最大证据边缘函数来获取伪标号样本。然后通过最大证据函数对数据集进行最终的结果预测。

3）提出了一种有效的基分类器筛选策略；在迭代训练中，根据新生成的基分类器在伪验证集上的性能和预测接近程度来决定是否加入该基分类器组中。

接下来我们给出详细的算法步骤。，代表的是人体动作特征的有标记样本集和无标记样本集。是人体动作的类别集合。，是给定的基分类器。P1，P2分别为，的伪标号数据集。，是KNN和SVM迭代过程中产生的临时分类器的集合，是一致集，当两个基分类器对里面每个无标记数据预测都一致时则加入到该集合中。为了评估学习器的性能，我们建立了一个伪标号验证集，该验证集由初始的以及（，代表伪标记数据集中）中置信度排名靠前的30%组成。然后计算分类器在上的准确率来估计其性能；另一方面，我们用估计当前分类器与上一个之间的接近程度。

**算法4.1 ECKS算法**

输入：有标记数据集，无标记数据集，两个基分类器，（这里代表KNN分类器，代表SVM分类器），以及伪标记数据集。设置；；，，，最大迭代次数为。

**Step1：**分别在上训练KNN和SVM，然后分类中数据，并分别估计其、和、。

**Step2：**重复步骤3-4，直到

**Step3：**SVM从中按类别比例选择个置信度最大的并且与KNN预测一致的数据，加入中，并令;KNN从中按类别比例选择个置信度最大的并且与KNN预测一致的数据，加入中，并令。

**Step4：**分别在，上重新训练KNN和SVM，然后用它们估计其和，其中=1,2。

If  or >0.9

{;}

Else

{

把加入到，中（当=1时加进集合中，=2时加进集合中）；Set 

}

**Step5：**更新伪验证集，并设置;;

输出：对集合和集合中的基分类器进行集成，输出分类器集合。

**4.4.2伪标号数据的选择与处理**

为了获取合适的伪标号样本集，同样，我们从控制伪标号样本的分类噪音、分布噪音、和构造有差异的分类器三种方法进行处理。

1. 降低分类噪音

为了控制产生的伪标记样本的分类噪音，我们从提高数据的准确性入手。不同于协同训练中仅通过KNN或SVM来计算置信度，我们采用当前生成的新分类器与集合中历史的临时中间分类器来估算伪标号样本的置信度。我们定义了一个新的基于证据的边缘函数（见4.4.3节），并与当前的分类器预测概率进行结合，来估计K和S对多标签中学习任务中无标号样本集中数据的分类置信度。具体公式如下所示：

 (4.5)

 (4.6)

这里，分别表示基分类器KNN和SVM。K是基分类器KNN的集合，S是基分类器SVM的集合。此外，为了进一步提高伪标记的准确率，我们从一致集中选择伪标号数据，并根据和把可能会导致分类器退化的伪标号数据及时移除。

1. 降低分布噪音

为了降低分布噪音，这里同样使用Co-KNN-SVM中的方法，通过按不同标签在训练集中的比例添加伪标号样本集，从而尽量减少分布噪音。并且我们不是把所有的伪标号样本都添加到训练集中，这也有效控制了样本数量对分类性能的潜在影响

1. 构造有差异的分类器

现有的大部分半监督集成学习算法都是自训练的，随着迭代训练的进行，分类器之间的差异性越来越小。因此，在我们ECKS算法中，我们保证每个基分类器组都有自己的训练样本集，并且通过各自的分类器组集合来对无标号样本集进行预测和标注形成伪标号数据。该方法在扩大生成分类器之间差异性的同时也使得算法保持协同的特点。

**4.4.3中间分类器差异性和准确性分析**

由于在迭代训练过程中，通过集成学习生成的中间分类器难免会受各种因素的影响而使分类准确率下降，并且如果与其它分类器之间的差异过小，那么对集成后的分类模型的性能提升也有限。为了剔除这些性能不好的分类器，我们通过评价它们之间的差异性和准确性来进行操作。

对于分类器的准确性，我们通过计算中间分类器在伪验证集上预测值来得到，用表示。当下一次迭代训练时，同样会得到一个准确性值，我们比较它们的大小，即如果的值小于的值，那么认为该新生成中间分类器性能有所下降，从而将其剔除掉；另一方面，为了尽量选择有差异的分类器，我们定义一个接近程度函数来衡量当前分类器与上一分类器之间的差异：

， (4.7)

其中表示当前基分类器对伪标记集的预测结果与表示上一分类器对伪验证集的预测结果。这样通过差异值就能将差异性较大的分类器从迭代训练过程中剔除。从而扩大了集成分类器之间的差异性，进一步提高模型的性能。

**4.4.4集成多个基分类器**

在算法迭代训练后，会得到基分类器集合H。假设里面有n个基分类器，我们把每一个分类器对样本的预测结果看成一个证据，利用证据推理方法对它们进行融合。设为待识别类别，即人体动作视频库的11个类别，令辨别框架为，对于同一个样本，每一个基分类器对其都有一组预测输出值，我们把预测输出值归一化为。因此，n个证据的基本概率赋值函数分配如下

 (4.8)

式中，；。

最后，给定一个集合H，对于每个类别j，利用证据融合公式（2），对证据进行合成，得到一个证据函数，如下公式所示：

 (4.9)

其中

进一步，我们定义了多标签下的边缘函数为：

 (4.10)

即标签p的证据值与其余标签中最大的证据值的差。函数值越大，则表示集成分类器把测试集中数据判断为标签的可信度越高。

**4.5 实验结果与分析**

**4.5.1实验设置**

为了验证所提模型的有效性，本小节进行了一些综合实验，实验环境和实验数据集与第三章实验部分相同。使用UCF YouTube Action数据集选取训练视频和测试视频进行实验，其中训练样本中有标记样本统一比例为20%，保证实验的一致性。然后计算十次识别准确率的平均值当为最终结果，每次训练的初始样本集都是随机抽取的，使测试结果更具有准确性和可靠性。同时，将本小章所提方法与传统的半监督集成学习CoBC[71]和MCSSB[72]方法作了对比。和上一章相同，我们采用KNN和SVM作为模型的基分类器。

**4.5.2 实验结果分析与讨论**

（1）在UCF YouTube Action数据集上的准确率

在UCF YouTube Action人体动作数据集上，结果如表4.1所示。可以看出，所有类别的识别准确率都稳定在80%左右，也验证了可以通过半监督的集成训练来实现不同基分类器的优势互补。相对于其它类别，骑马和骑车、挥杆和足球、秋千和蹦床等混淆较为明显，其原因是这些类别的场景和动作存在相似性，进而导致识别上的混淆。

表4.1 UCF YouTube Action数据集每一类动作的准确率

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类别 | 投篮 | 骑车 | 跳水 | 挥杆 | 骑马 | 足球 | 秋千 | 网球 | 蹦床 | 排球 | 遛狗 |
| 投篮 | 0.81 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.04 | 0.00 | 0.10 | 0.00 |
| 骑车 | 0.00 | 0.82 | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 跳水 | 0.05 | 0.00 | 0.80 | 0.00 | 0.03 | 0.02 | 0.03 | 0.00 | 0.04 | 0.00 | 0.03 |
| 挥杆 | 0.00 | 0.00 | 0.00 | 0.78 | 0.00 | 0.12 | 0.00 | 0.02 | 0.00 | 0.08 | 0.00 |
| 骑马 | 0.00 | 0.15 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 |
| 足球 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.82 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 |
| 秋千 | 0.00 | 0.00 | 0.00 | 0.12 | 0.00 | 0.00 | 0.81 | 0.00 | 0.07 | 0.00 | 0.00 |
| 网球 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.79 | 0.00 | 0.15 | 0.00 |
| 蹦床 | 0.00 | 0.00 | 0.17 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.78 | 0.00 | 0.03 |
| 排球 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.05 | 0.00 | 0.82 | 0.00 |
| 遛狗 | 0.00 | 0.00 | 0.05 | 0.04 | 0.00 | 0.00 | 0.10 | 0.02 | 0.00 | 0.00 | 0.79 |

（2）与混合式协同训练的识别方法进行比较

与使用基于混合式协同训练的识别方法Co-KNN-SVM比较的结果如表4.2所示。从表中可以看出，采用基于半监督集成学习的方法平均识别率要比采用混合式协同训练的方法的识别效果要好，平均识别准确率提高了2.6%。其中挥杆和遛狗类别提高最为明显，分别为11.4%和6.8%。

表4.2 与混合式协同训练的识别方法进行比较

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类别  方法 | 投篮 | 骑车 | 跳水 | 挥杆 | 骑马 | 足球 | 秋千 | 网球 | 蹦床 | 排球 | 遛狗 | 平均 |
| Co-KNN-SVM | 0.81 | 0.79 | 0.80 | 0.70 | 0.75 | 0.84 | 0.83 | 0.79 | 0.77 | 0.82 | 0.74 | 0.78 |
| ECKS | 0.81 | 0.82 | 0.80 | 0.78 | 0.80 | 0.82 | 0.81 | 0.79 | 0.78 | 0.82 | 0.79 | 0.80 |

其原因是Co-KNN-SVM算法在协同训练过程中，随着迭代次数的增加，基分类器的差异性逐渐消失，因此，更多的无标记数据不能继续提高分类器的预测准确率。而我们所提的方法采用集成学习的思想，在训练过程中通过基分类器组中不同的分类器来对无标记样本进行标记，并加入到对方有标记训练样本中，以生成新的分类器，进一步扩大了分类器的差异性。相比于第三章所采用的Co-KNN-SVM方法，本章的方法能够有效提高识别系统的可靠性和稳定性。

（3）与传统的半监督集成方法进行比较

与其他的一些半监督集成学习的方法进行了对比，结果如表4.3所示。从表4.3中可以看出，本章所提算法的识别率要高于传统的CoBC和MCSSB算法，识别准确率分别提高了1.3%和2.6%。其原因是前者首先通过AdaBoost方法生成多个分类器，然后对无标记数进行预测，并选择置信度高的加入训练集中以对基分类器作进一步训练，期间不生成新的分类器。因此，随着循环次数的增加，分类器之间的差异性逐渐消失。并且如果初始的样本数据不充足时，也会影响分类器的准确率。而后者是基于Boosting方法，虽然在迭代训练过程生成了新的分类器，并且通过结合置信度和样本之间的相似性来选择伪标号数据，但是它是利用集合中所有的分类器来标注样本，由于早先的分类器性能往往较差，这种方法预测的效果不如选择性识别的好。从本质上说，这两种方法都是自训练的方式。我们提出的半监督集成方法ECKS，在迭代训练过程中保留新生成的中间分类器，然后通过不同的筛选后的基分类器组来预测和选择无标记样本，扩大了分类器之间的差异性。最后基于证据理论对基分类器进行集成，提高系统的识别性能。

表4.3 与其余一些半监督集成学习方法比较

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 识别方法 | SVM | KNN | CoBC | MCSSB | 本文方法 |
| 识别率 | 0.47 | 0.52 | 0.79 | 0.78 | 0.80 |

（4）不同迭代次数下，本文所提方法与其余方法的比较

图4.2和图4.3给出的是在有标记样本数的比例分别为20%和40%时，随着迭代次数的增加，准确率变化的折线图。通过图4.1和图4.2可以看出，CoBC和 MCSSB算法在迭代训练初期准确率提高较快，但是后期相比于我们所提出的的算法较为平坦，主要原因就是分类器差异性减少。我们提出的方法本质上是两个不同的基分类器组的协同训练，通过基于伪验证集的方式来选择性能优良的分类器，不仅可以减少自训练过程中的误差累计效应，还能有效减缓迭代训练过程中基分类器差异性减小的问题。从图中可以看到我们所提的算法取得了更高的准确率。



图4.2 有标记样本比例为20%



图4.3 有标记样本比例为40%

**4.6 本章小结**

本章在Co-KNN-SVM算法基础上，提出了基于半监督集成学习的人体动作识别方法。该方法首先把迭代过程中生成的临时中间分类器分别添加到指定的集合中，通过最大证据边缘函数来选取伪标号数据以提高分类器的性能并提出一种基于伪验证集的分类器选择方法。在得到基分类器组后采用证据理论原理对基进行融合。实验结果表明，相比于传统的自训练的半监督集成学习方法，本章提出的方法提高了准确率，增强了协同训练识算法的鲁棒性。

**第五章 人体动作识别原型系统的设计与实现**

**5.1 引言**

人体动作视频整体结构较为复杂，有着自身的特点，关键帧提取、特征提取也各有差异。对用户而言，人体动作识别系统与其他管理系统相同，都需要有一个友好的操作界面，具有可维护，易扩展的特点。

本系统采用面向对象的方法进行开发实现，面向对象编程思想是计算科学技术的主流思想，其方法是借鉴现实世界中的概念，通过对事物属性和行为封装来思考解决问题。根据前两章对人体动作识别方法的深入研究，结合视频关键帧提取、特征提取与融合、协同训练、半监督集成学习、人体动作识别等技术，设计并实现了基于半监督协同训练和集成学习的人体动作识别系统。通过该系统验证本文方法的可行性。

**5.2 系统的总体设计**

本节详细阐述了人体动作识别原型系统的整体框架和功能模块，系统包括人体动作视频预处理、模型训练和人体动作识别三个子系统，系统框架图如图5.1所示。从功能角度划分，可以把人体动作系统划分为以下几个模块部分：关键帧提取、特征提取、特征融合、协同训练、集成学习、动作识别。



图5.1 人体动作识别系统框架图



图5.2 人体动作识别系统用例图

图5.2所示的用例图表示的是人体动作识别原型系统中功能模块之间的操作关系。关键帧提取采用第二章介绍的基于视频片段边界的方法，从视频片段中提取具有代表意义图像帧。特征提取模块主要含有四类特征，分别是颜色特征、纹理特征、形状特征、多尺度LBP特征。系统从视频关键帧中提取这四种底层特征，来表示动作的运动属性。并将特征进行串联，形成具有全面概括性质的特征矢量。协同训练和集成学习的模型分别是第三章、第四章提出的Co-KNN-SVM以及ECKS方法，系统管理员通过协同训练来产生性能更优的基分类器以及通过集成学习来减少分类器的之间的差异并集成多个分类器以增强系统的泛化能力，最终根据组合的基分类器组判定视频的动作类型。

**5.3 人体动作识别原型系统的实现**

**5.3.1 系统开发环境**

人体动作识别原型系统使用的操作系统为Window10，并采用Visual Studio 2013进行开发。Visual Studio是微软公司推出的开发环境，是目前最流行的Windows平台应用程序的集成开发环境，包括程序的设计、开发、调式等功能。我们在系统开发过程中将各个功能模块封装成类，并设计了清晰的人机交互界面，方便用户操作人体动作识别系统。

Matlab是美国MathWorks公司于1982年出品的一款商业数学软件。主要用于矩阵计算、算法开发、数值分析和信号处理等等。Matlab作为高效科学的计算软件，一经推出就受到许多高等院校的热爱。不仅是开发研究人员的开发工具也是高级课程教学的教学工具。其主要由Matlab语言、工作环境、图像处理、函数库以及应用程序界面等五个部分组成。MATLAB为不同的应用设计了专业的工具箱，用户只需导入工具箱即可调用相关的函数，具有简单易学，操作简便的特点。

**5.3.2 系统实现简介**

进入系统后，用户首先指定路径下的视频库。本系统选用的动作视频库为UCF YouTube Action Data Set。模型训练完成后，选择待测试的视频，利用Co-KNN-SVM和ECSK算法进行人体动作识别，识别完成后用户可以将识别结果保存在本地计算机以待验证。如果用户需要验证视频，可以对其进行抽取并播放，以判断内容是否与识别的动作相吻合。详细的操作流程如图5.3所示。



图5.3 人体动作识别操作流程图

下面展现系统的运行界面。

（1）系统主界面

基于半监督协同训练和集成学习的人体动作识别系统由模型训练和动作识别两大模块构成，两个模块相互独立，便于其他功能的拓展，系统的主界面如图5.4所示。界面左上部分视频播放窗口用于播放用户选中的视频。点击“点击选择动作视频库”按钮会让用户在磁盘中选择需要训练的视频库，视频选定后点击“模型训练”按钮进行模型训练。右边的列表框供用户选择需要测试的视频。右下部分是一个日志栏，显示系统运行时间、运行状态、以及分类结果等。

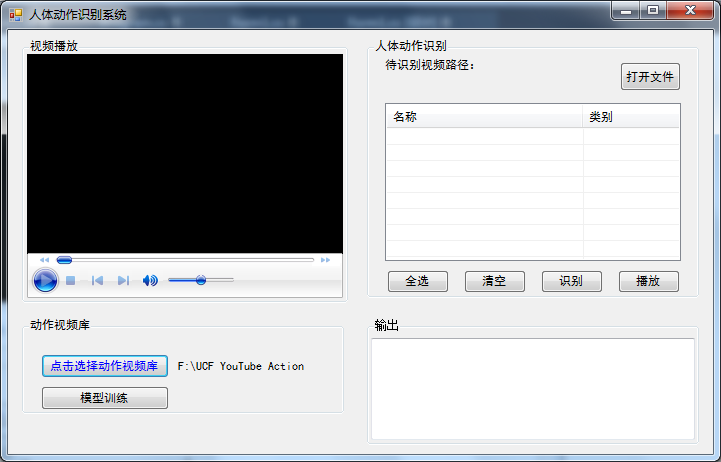


图5.4 系统主界面

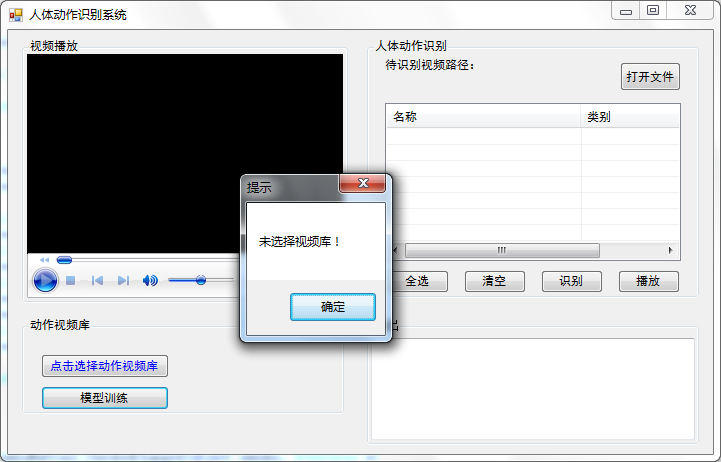


图5.5未选择视频提示图

（2）模型训练识别简介

如图5.5，用户首先点击选择动作视频库按钮选择相应的视频库，这里我们选择UCF YouTube Action视频数据库，然后点击“模型训练”按钮进行模型训练，如果用户未选择任何动作视频库，则界面会弹出“未选择视频库”的提示。

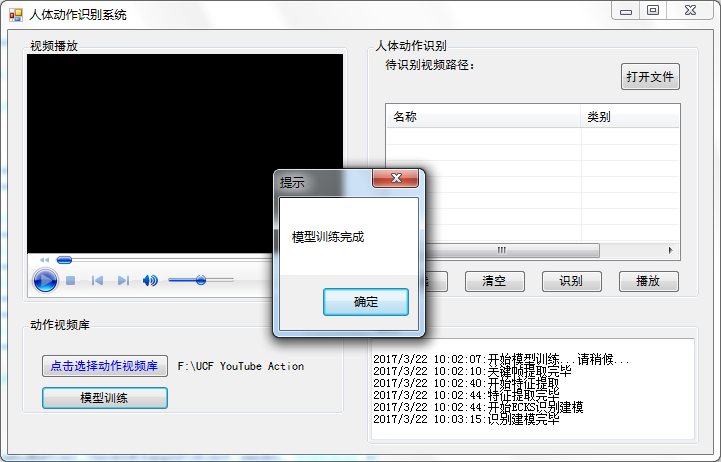


图5.6 模型训练图

如图5.6所示，模型训练开始时，系统将会对视频库中的视频进行关键帧提取、特征提取、使用基于半监督协同训练和集成学习方法，最后得到人体识别模型。识别过程中的每一个具体步骤都会在日志框中显示，模型训练完成后自动弹出“模型训练完成”提示。

1. 人体动作识别实现简介

1）在人体动作识别模块。用户点击“打开文件”按钮，选择文件存储路径打开存储在本地磁盘上的待训练视频文件，如图5.7所示。

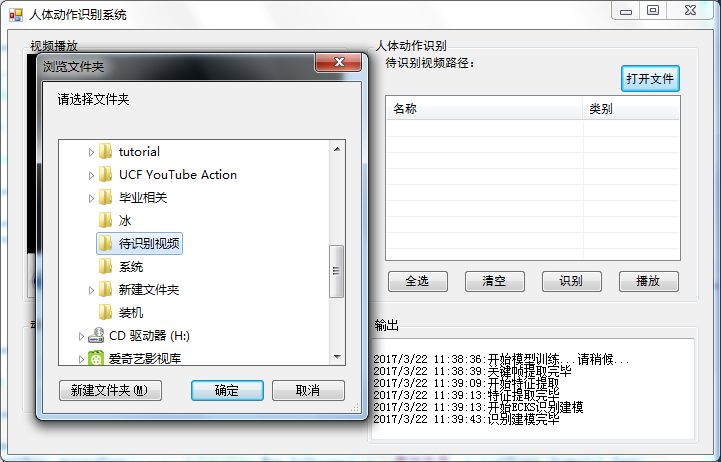


图5.7 选择待识别视频界面

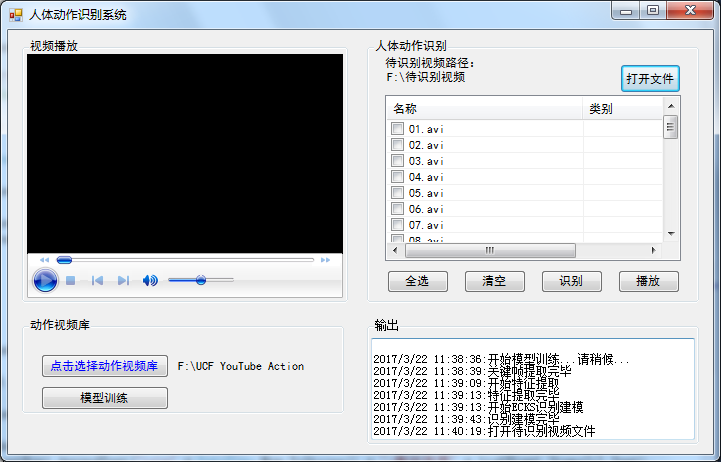


图5.8 载入待识别视频文件界面

2）如图5.8所示，文件打开后会显示出文件所在路径，路径下面的列表中会被导入已经打开的文件，显示文件的名称及所属类别。下面的四个按钮：“全选”、“清空”、“识别”、“播放”用于对视频文件进行各项操作。“全选”是选中列表中所有文件；“清空”是取消全部选中的视频文件；它们方便于对视频进行批量操作。用户选中视频文件后点击“播放”“识别”按钮分别进行对视频内容播放和人体动作识别。

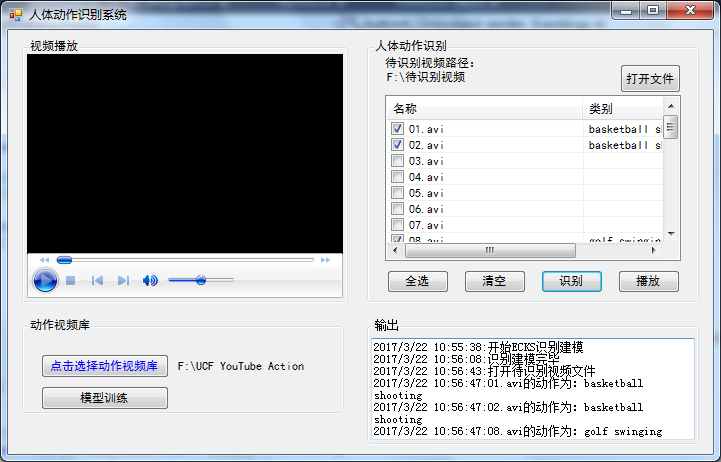


图5.9 人体动作识别界面

3）在图5.9中的人体动作识别模块部分，用户在右上部分的视频列表中勾选需要识别的视频片段，点击“识别”按钮进行识别，此时，系统会对视频进行处理以及基于半监督协同训练和集成学习的动作识别分析，并将所得结果显示在视频对应的类别列表中，输出框中同步输出视频文件对应的日期、时间及动作信息。

4）如图5.9所示，完成对人体动作识别后，如果用户想知道系统识别的正确与否，可以打开播放原来的视频文件进行比较，判断准确性。界面关闭后，系统会自动将所有日志文件保存起来，以备日后需要时查看。



图5.10 视频播放界面

**5.4 本章小结**

本章采用Visual Studio和Matlab等开发工具，将前两章所提出的协同训练和集成学习的思想应用于人体动作识别的研究中，并对基于半监督协同训练和集成学习人体动作识别系统的模块结构和界面做了详细的介绍，且针对每一个模块都做了界面展示和功能说明。大量的试验验证了本文所提方法在人体动作识别应用上的有效性和可靠性。

**第六章 总结与展望**

**6.1 工作总结**

对于庞大而复杂的人体动作视频，现如今的大多数识别方法是基于监督学习框架。监督学习的前提是通过对采集的足够多的样本进行训练以得到识别模型，然后根据模型对测试样本进行识别分类。然而在现实生活中小样本问题是一个重大的挑战，并且采集样本费事费时。此外，在动作识别领域使用的模型比较单一，无论是基于模板的方法还是基于概率统计的方法都有自己的局限性。因此，基于半监督集成学习的人体动作识别成为模式识别、动作识别等领域的研究热点。

本文第一章中对选题意义及背景、国内外研究状况做了简要的介绍，阐明此研究工作的重要性。第二章介绍了人体识别技术的步骤方法。第三、四章对此论文提出的方法做了详细的介绍，同时分析人体动作识别中关键技术的相关理论知识。第五章介绍模型系统的总体设计及必要的开发环境。第六章对整篇论文做了总结与拓展。基于对现有经典方法的分析研究，本文提出了基于混合式协同训练的人体动作识别方法和基于半监督集成学习的人体动作识别方法，取得了一些进展。本文的主要工作和创新之处如下：

1）提出了基于混合式协同训练的人体动作识别方法，有效解决有监督学习下人体动作识别需要大量有标记样本和单一模型下系统泛化能力不足的问题。该方法利用动作识别领域不同类型的方法来构建基分类器，并进行迭代的相互训练以提高泛化性能，可以降低标注成本并实现不同识别方法的优势互补。实验结果表明，基于混合式协同训练的人体动作识别方法能够在有标记样本数不足的情况下仍然能提高动作识别的准确率。

2）提出了基于半监督集成学习的人体动作识别方法。鉴于协同训练在随着迭代次数的增加，基分类器的差异性逐渐退化，本文受集成学习的启发，利用集成学习来创造基分类器的差异性。该方法将协同训练迭代过程中产生的中间临时分类器加入到每个分类器集合组中，并定义了基于置信度的最大证据边缘函数，然后分类器集合根据置信度来选择伪标号数据，最后通过选择集成对个分类器对人体动作进行识别。有效地解决了分类模型退化的问题，从而提高了人体动过识别的识别率。

3）设计实现了基于半监督协同训练和集成学习的人体动作识别原型系统，使用面向对象的设计思想，结合MATLAB较强的数据处理能力以及C#中友好的人机交互界面进行混合编程，进一步验证了本文所提算法的有效性。

**6.2 工作展望**

人体动作识别是目前炙手可热且具有挑战性的研究课题，虽然本文在半监督的识别方法上取得了一定的研究成果，但是在人体动作识别领域还存在着许多问题亟需解决，主要体现在以下几个方面。

（1）如何实现静态特征与动态特征、时空特征完美的结合，形成多模态融合的人体动作识别系统，促进视频中人体动作更精确地识别仍是目前研究的重点和难点。因此，这在未来的长时间内仍是人体动作识别的重要研究方向。

（2）现有的人体动作识别系统大都是设计和应用在受约束的场景下。而在真实场景下，受噪声、遮挡、阴影等复杂因素的影响情况下，识别效果往往不尽如人意。并且，由于实际存储系统存储能力的限制，往往记录的都是低分辨率的人体动作视频。虽然不少学者已经针对这些问题进行了研究，但是仍然缺乏系统和通用的方法。因此，还需要更多的研究来解决这些实际问题。

（3）大部分识别方法检测和识别的都是已经发生的行为和动作。而在某些应用场景中往往需要预判即将发生的人体动作，这也向人类的意图推理和推断提出了一个重要的智力挑战，尤其是在安全监测领域。这也是人体动作识别领域未来的一个重要研究方向。

**参考文献**

1. Turaga P, Chellappa R, Subrahmanian V S, et al. Machine Recognition of Human Activities: A Survey[J]. Circuits & Systems for Video Technology IEEE Transactions on, 2008, 18(11):1473-1488.
2. 胡琼, 秦磊, 黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报, 2013, 36(12):2512-2524.1.
3. Aggarwal J K, Ryoo M S. Human activity analysis: A review[J]. Acm Computing Surveys, 2011, 43(3):16.
4. [4] Retal C. A System for Video Surveillance and Monitoring: VSAM final report[J]. Carnegie Mellon University: Technical Report CMU, Pittsburgh USA, 2000.
5. Haritaoglu I, Harwood D, Davis L S. W4: Real-Time Surveillance of People and Their Activities[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(8):809-830.
6. Remagnino P, Tan T, Baker K. Multi-agent visual surveillance of dynamic scenes[J]. Image & Vision Computing, 1998, 16(8):529-532.
7. 黄飞跃, 徐光祐. 自然的人体动作识别[C]// 全国普适计算)学术会议. 2007.
8. 任海兵. 非特定人自然的人体动作识别[D]. 清华大学, 2003.
9. 潘云鹤, 庄越挺, 陈纯. 基于视频的人体动画系统: CN, CN 1356671 A[P]. 2002.
10. 鲁珂, 赵继东, 叶娅兰,等. 一种用于图像检索的新型半监督学习算法[J]. 电子科技大学学报, 2005, 34(5):669-671.
11. 孙广玲, 唐降龙. 基于分层高斯混合模型的半监督学习算法[J]. 计算机研究与发展, 2004, 41(1):156-161.
12. Shahshahani B M, Landgrebe D /. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon[J]. IEEE Transactions on Geoscience & Remote Sensing, 1994, 32(5):1087-1095.
13. Miller D J, Uyar H S. A Mixture of Experts Classifier with Learning Based on Both Labelled and Unlabelled Data[J]. Medical Imaging IEEE Transactions on, 1996, 9(5):571-577.
14. Nigam K, Mccallum A K, Thrun S, et al. Text Classification from Labeled and Unlabeled Documents using EM[J]. Machine Learning, 2000, 39(2):103-134.
15. Blum A, Chawla S. Learning from Labeled and Unlabeled Data using Graph Mincuts[C]// Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 2001:19-26.
16. Zhu X, Ghahramani Z, Lafferty J D. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions.[C]// Machine Learning, Proceedings of the Twentieth International Conference. DBLP, 2003:912-919.
17. Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]// Eleventh Conference on Computational Learning Theory. ACM, 2000:92-100.
18. Gupta, Sonal, et al. Watch, Listen & Learn: Co-training on Captioned Images and Videos. [C]. European Conference on Machine Learning & Knowledge Discovery in Databases Springer-Verlag,2008:457-472.
19. Yu S, Krishnapuram B, Rosales R, et al. Bayesian Co-Training［J］. Journal of Machine Learning Research, 2007, 12(3):2649-2680.
20. [19] Liu C, Yuen PC. A Boosted Co-Training Algorithm for Human Action Recognition[J]. Circuits & Systems for Video Technology IEEE Transactions on, 2011, 21(9):1203-1213.
21. Tang C, Wang WJ, Li W, Li G, Cao F. Multi-Learner co-training model for human action recognition[J]. Journal of Software, 2015,26(11):2939−2950.
22. 章毓晋. (2003). 基于内容的视觉信息检索[M]. 第1版. 科学出版社. 221-223.
23. [22] Flickner M, Sawhney H, Niblack W, et al. Query by Image and Video Content : The QBIC System[J]. Computer, 1995, 28(9):23-32.
24. Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach[C]// International Conference on Pattern Recognition. IEEE, 2004:32-36 Vol.3.
25. Blank M, Gorelick L, Shechtman E, et al. Actions as space-time shapes[C]// Tenth IEEE International Conference on Computer Vision. IEEE, 2005:1395-1402 Vol. 2.
26. Rodriguez M D, Ahmed J, Shah M. Action MACH a spatio-temporal Maximum Average Correlation Height filter for action recognition[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. DBLP, 2008:1-8.
27. <http://pets2010.net/>.
28. Carlsson S, Sullivan J. Action Recognition by Shape Matching to Key Frames[C]// IEEE Computer Society Workshop on Models versus Exemplars in Computer Vision. 2001.
29. Cheung K M G, Baker S, Kanade T. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture[C]// Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. IEEE, 2003:I-77-I-84 vol.1.
30. Shotton J, Fitzgibbon A, Cook M, et al. Real-time human pose recognition in parts from single depth images[C]// The, IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011, Colorado Springs, Co, Usa, 20-25 June. DBLP, 2011:1297-1304.
31. Johansson G. Visual perception of biological motion and a model for its analysis[J]. Attention, Perception, & Psychophysics, 1973, 14(2):201-211.
32. Stauffer C, Grimson W E L. Learning patterns of activity using real-time tracking[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(8):747-757.
33. Wang X, Tieu K, Grimson E. Learning Semantic Scene Models by Trajectory Analysis[M]// Computer Vision – ECCV 2006. Springer Berlin Heidelberg, 2006:110-123.
34. Raptis M, Soatto S. Tracklet Descriptors for Action Modeling and Video Analysis[C]// Computer Vision - ECCV 2010, European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings. DBLP, 2010:577-590.
35. Davis J W, Bobick A F. The Representation and Recognition of Action Using Temporal Templates[J]. Proc of Cvpr, 2000, 23(3):928--934.
36. Yuan J, Liu Z, Wu Y. Discriminative subvolume search for efficient action detection[C]// Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:2442-2449.
37. Liu J, Kuipers B, Savarese S. Recognizing human actions by attributes[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2011:3337-3344.
38. Yao B, Jiang X, Khosla A, et al. Human action recognition by learning bases of action attributes and parts[C]// IEEE International Conference on Computer Vision, ICCV 2011, Barcelona, Spain, November. DBLP, 2011:1331-1338.
39. Maji S, Bourdev L, Malik J. Action recognition from a distributed representation of pose and appearance[C]// Computer Vision and Pattern Recognition. IEEE, 2011:3177-3184.
40. Swain Michael J，Ballard Dana H．Color indexing[J]．International Journal of Compmer Vision, 1991, 7(1): 11-32.
41. Ma, YuFei, Lu, et al. A user attention model for video summarization[J]. 2002.
42. 杨静宜. 基于颜色与纹理的图像检索技术研究与实现[D]. 武汉理工大学, 2009.
43. Haralick R M, Shanmugam K, Dinstein I. Textural Features for Image Classification[J]. Systems Man & Cybernetics IEEE Transactions on, 1973, smc-3(6):610-621.
44. Mukundan R. Radial Tchebichef Invariants for Pattern Recognition[C]// Tencon 2005 2005 IEEE Region. IEEE, 2005:1-6.
45. Ojala T, Harwood I. A Comparative Study of Texture Measures with Classification Based on Feature Distributions[J]. Pattern Recognition, 1996, 29(1):51-59.
46. 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, 2002, 24(7):971-987.
47. Lin Z, Jiang Z, Davis L S. Recognizing Actions by Shape-Motion Prototype Trees[C].IEEE International Conference on Computer Vision. 2009:444-451.
48. Jiang Z, Lin Z, Davis L S. A Tree-Based Approach to Integrated Action Localization, Recognition and Segmentation[M].Trends and Topics in Computer Vision. Springer Berlin Heidelberg, 2010:114-127.
49. Liu J, Ali S, Shah M. Recognizing human actions using multiple features[J]. 2008:1-8.
50. Davis J W, Bobick A F. The Representation and Recognition of Action Using Temporal Templates[J]. Proc of Cvpr, 2000, 23(3):928--934.
51. Efros A A, Berg A C, Mori G, et al. Recognizing action at a distance[C]// IEEE International Conference on Computer Vision, 2003. Proceedings. IEEE, 2003:726-733 vol.2.
52. Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features[C]// IEEE International Workshop on Visual Surveillance and PERFORMANCE Evaluation of Tracking and Surveillance. IEEE, 2005:65-72.
53. Ji R, Yao H, Sun X. Actor-independent action search using spatiotemporal vocabulary with appearance hashing[J]. Pattern Recognition, 2011, 44(3):624-638.
54. Yamato J, Ohya J, Ishii K. Recognizing human action in time-sequential images using hidden Markov model[C]// Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92. 1992 IEEE Computer Society Conference on. IEEE Xplore, 1992:379-385.
55. Park S, Aggarwal J K. Recognition of two-person interactions using a hierarchical Bayesian network[C]// 2003:65-76.
56. Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach[C]// International Conference on Pattern Recognition. IEEE, 2004:32-36 Vol.3.
57. Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition[C]// International Conference on Multimedia 2007, Augsburg, Germany, September. DBLP, 2007:357-360.
58. Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. DBLP, 2008:1-8.
59. Natarajan P, Nevatia R. Online, Real-time Tracking and Recognition of Human Actions[C]// IEEE Workshop on Motion and Video Computing. IEEE Computer Society, 2008:1-8.
60. Matikainen P, Hebert M, Sukthankar R. Trajectons: Action recognition through the motion analysis of tracked features[C]// IEEE, International Conference on Computer Vision Workshops. IEEE, 2009:514-521.
61. Natarajan P, Nevatia R. View and scale invariant action recognition using multiview shape-flow models[C]// Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008:1-8.
62. Golson S. State Machine Design Techniques for Verilog and VHDL[J]. Synopsis Journal ofHigh-Level Design, Sep, 1994, 26(7):97-102.
63. Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]// Eleventh Conference on Computational Learning Theory. ACM, 2000:92-100.
64. Zhou Z H, Li M. Tri-Training: Exploiting Unlabeled Data Using Three Classifiers[J]. Knowledge & Data Engineering IEEE Transactions on, 2005, 17(11):1529-1541.
65. Dietterich T G, Bakiri G. Solving multiclass learning problems via error-correcting output codes[J]. Journal of Artificial Intelligence Research, 1994, 2(1):263-286.
66. 于玲, 吴铁军. 集成学习:Boosting算法综述[J]. 模式识别与人工智能, 2004, 17(1):52-59.
67. Breiman, L.. Bagging predictors [J]. Machine Learning, 1996, 24 (2), 123–140.
68. 蓝金辉, 马宝华, 蓝天,等. D-S证据理论数据融合方法在目标识别中的应用[J]. 清华大学学报自然科学版, 2001, 41(2):53-55.
69. Cox I J, Miller M L, Minka T P, et al. The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments.[J]. IEEE Transactions on Image Processing, 2000, 9(1):20.
70. Curie U P E M, Grandvalet Y, Ambroise C. Semi-Supervised MarginBoost[J]. Advances in Neural Information Processing Systems, 2002:553-560.
71. Bennett K P, Demiriz A, Maclin R. Exploiting unlabeled data in ensemble methods[C]// ACM Int Conf Knowledge Discovery & Data Mining. 2002:289-296.
72. Liu R, Cheng J, Lu H. A robust boosting tracker with minimum error bound in a co-training framework[C]// IEEE, International Conference on Computer Vision. IEEE, 2009:1459-1466.
73. Valizadegan H, Jin R, Jain A K. Semi-Supervised Boosting for Multi-Class Classification[C]// European Conference on Machine Learning and Knowledge Discovery in Databases. Springer-Verlag, 2008:522-537.

**致 谢**

七年既漫长又短暂的校园生活已接近尾声。回首过去，在学习上，本科四年多数时间是接受老师传授的知识，而研究生阶段则更注重于发现知识。生活上少了几份稚气多了几份成熟。即将步入社会，却怀念美好的校园时光。感情很深的师弟师妹们，细心教导我孜孜不倦的老师们等，这些都将是我以后人生道路上的最珍贵的回忆。

首先，我要感谢的是我的导师\*\*\*教授。詹老师和蔼可亲，学识渊博，经常来实验室与我们讨论学术问题，督促我们奋发图强。毕业设计是对研究生三年来所研究知识的一次综合性考察。在毕业设计初期，詹老师就给我安排了各个阶段应该完成的具体任务，督促我毕业设计的完成进度，时不时的查看我的完成情况，给我提出了许多宝贵的建设性意见，使我设计中的难题迎刃而解。从论文的选题、算法的设计、毕业论文的撰写等各个阶段，詹老师都带着我踏踏实实、一步一个脚印，我才能顺利的完成毕业设计。因此我向我的指导老师表示由衷的感谢！此外，感谢\*\*\*副教授对我实验和论文的关心和帮助。

其次，我要感谢课题组\*\*\*、\*\*\*、\*\*\*、\*\*\*等众多老师们。在我研究生学习期间，无论是在学术讨论方面，还是在论文撰写方面，他们都热心帮助我，给了我很多有价值的意见和指导。另外，我要感谢419实验室的同学们，感谢有你们的陪伴，使我的毕业设计过程中不再那么枯燥，使我能够高效率的完成每项任务。

最后感谢我的家人们，爸爸妈妈含辛茹地苦教育和培养我，从小让我接受高质量的教育，在我碰到难题的时候，辅导我和开导我，让我的人生一路明亮，在我求职时，给我支持和鼓励，使我顺利找到满意的工作。

在毕业即将来临之际，我深深地感谢那些在我研究生三年里给予我帮助的人，有了你们我的生活更加精彩。未来，我会带着你们的祝福，继续前进。

攻读硕士学位期间已发表的论文

1. 基于混合式协同训练的人体动作识别算法研究.计算机科学 已录用，本人第一作者。