

西安电子科技大学

硕士学位论文



面向老人的智能监控系统及行为识别的研究

作者姓名 李冬月

指导教师姓名、职称 千博 教授

申请学位类别 工学硕士

学校代码 10701
分 类 号 TP391

学 号 1604121928
密 级 公开

西安电子科技大学

硕士学位论文

面向老人的智能监控系统及行为识别的研究

作者姓名：李冬月

一级学科：控制科学与工程

二级学科：控制理论与控制工程

学位类别：工学硕士

指导教师姓名、职称：千博 教授

学 院：机电工程学院

提交日期：2019 年 6 月

Research on Intelligent Monitoring System and Behavior Recognition for the Elderly

A thesis submitted to
XIDIAN UNIVERSITY
in partial fulfillment of the requirements
for the degree of Master
in Control Theory and Control Engineering

By

Li Dongyue

Supervisor: Qian Bo Title: Professor

June 2019

西安电子科技大学
学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

本人签名： 李冬月 日期： 2019.06.13

西安电子科技大学
关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名单位为西安电子科技大学。

保密的学位论文在____年解密后适用本授权书。

本人签名： 李冬月 导师签名： 于博

日期： 2019.06.13 日期： 2019.06.13

摘要

智能监控系统的快速发展满足了人们在保障社会安全、家庭安全、生产安全等方面的需求，其不仅可以实现对危险场景的实时监控，还可以实现对意外事件的自动预警。在人口老龄化趋势下，适用于老年人的智能监控系统已有深入的研究，但多数的应用和研究仅考虑某一监控区域、某一监控方式，因而系统在智能化、便捷性、全面性上还存在一定局限性。

本文在老年人对监护有极大需求且现有系统存在不足的背景下，研究了以下三项内容。

1. 设计和实现全方位监护的监控系统。提出的智能监控系统由监控终端、监控节点、监控平台、用户交互四部分组成。监控终端是指视觉传感器、气体传感器、温湿度传感器等信息采集器件，用来提供采集居住环境信息、身体健康数据、视觉图像、实时视频等功能；监控节点分为树莓派监控节点、可穿戴设备监控节点两种，用来提供收集数据、通信传输、信息基本分析等功能；监控平台是由云服务器、云数据库、云对象存储组成，用来提供智能分析、图片存放、信息存储等功能；用户交互通过 Web 页面访问、短信预警的方式提供交互功能。

2. 提出的智能监控系统采用分布式架构，方便在不同应用场景中扩展。在系统中分布多个有数据分析功能的监控节点，利用监控节点 CPU 资源分担系统压力。由于系统主要运用监控终端收集的信息分析被监护的老年人是否存在潜在生命危险，因此监控终端会产生大量的数据。然而，分布式系统中监控节点承担的分析功能可以有效的缓解监控平台的压力，方便系统进一步扩展。同时，系统中的多个监控节点存在信息互补的优势，比如相近节点获取到的多视角视觉图像可以提高行为识别的准确度，进而可以保障系统在行为识别方面的准确性。

3. 为进一步发掘视觉信息的作用，论文在行为识别算法方面做了改进。运用卷积网络的行为识别算法是目前效果较好的方法，但由于此类方法的输入是静态图像，因此无法更多的获取视频的时序信息。而动态图像是多视频帧的融合，拥有较多时序信息，因此提出一种结合动态图像概念的改进方法，将其作为 Temporal segment networks (TSN) 的输入。同时，为充分发挥分布式系统中监控节点的信息互补优势、提高行为识别的准确度，本文提出一种多视角下结合 Borda 的融合决策方法，用以对多个视角下的识别结果做综合判断。

关 键 词： 监控系统， 行为识别， 老人， 动态图像， 分布式

ABSTRACT

The rapidly developing intelligent monitoring system has met people's needs in the aspect of guaranteeing social security, family safety and production safety. It can realize not only real-time monitoring of dangerous scenes, but also automatic early warning of accidents. In the trend of population aging, the intelligent monitoring system for the aged has been deeply studied, but most of the system applications and research aim at a specific monitoring area and monitoring mode. Hence, the system still has some limitations in terms of intellectualization, convenience and comprehensiveness.

In the context of the aged have great demand for guardianship and the existing system has defects, this paper studies the following three items.

Firstly, designing and implementing a fully functional monitoring system. The proposed intelligent monitoring system is composed of four parts, including monitoring terminals, monitoring nodes, monitoring platform and user interaction. The monitoring terminal refers to information acquisition devices such as visual sensors, gas sensors, temperature and humidity sensors, which are used to provide residential environment information, fitness data, visual images, real-time videos and so on; monitoring nodes include Raspberry Pi monitoring nodes and wearable devices, which are used to provide data collection, basic information analysis, communication transmission and other functions; monitoring platform consists of cloud server, cloud database and cloud object storage, which are used to provide intelligent analysis, image storage, information storage and other functions; user interaction mainly provides web page interaction and short message early warning.

Secondly, proposing an intelligent monitoring system using a distributed architecture to facilitate expansion in different application scenarios. There are multiple monitoring nodes with data analysis function in a monitoring system. The system mainly uses the information collected by wearable devices, gas sensors and visual sensors to analyze whether the aged under guardianship have potential life danger. The monitoring terminal will generate a mass of data, and the analysis function born by the monitoring nodes in the distributed system can effectively alleviate the pressure of the monitoring platform. Meanwhile, multiple monitoring nodes in the monitoring system have the advantage of information

complementation. For example the multi-view visual images acquired by nearby nodes can enhance the accuracy degree of behavior recognition, thus ensuring the accuracy of the system in behavior recognition.

Thirdly, in order to further explore the role of visual information, this paper makes improvement in the behavior recognition algorithm. At present, the behavior recognition algorithm based on convolution network has good effect, but the input of this method is static image, it is impossible to obtain more video temporal information. The dynamic image integrates multiple video frames and contains more temporal information. Therefore, the proposed improved algorithm combines the concept of dynamic image and uses it as the input of temporal segment networks (TSN). Meanwhile, in order to allow full play to the information complementary advantage of monitoring nodes in the distributed system and enhance the accuracy degree of behavior recognition, this paper proposes a fusion decision method based on Borda from multi-view, which makes a synthetic judgment of the recognition results from multiple perspectives.

Keywords: Monitoring System, Behavior Recognition, The Elderly, Dynamic Image, Distributed System

插图索引

图 1.1	人体行为识别方法	4
图 2.1	HOG 特征提取步骤 ^[38]	11
图 2.2	LeNet-5 示意图 ^[20]	13
图 2.3	双流法基本模型	14
图 3.1	智能监控系统功能图	20
图 3.2	智能监控系统组成图	22
图 3.3	智能监控系统框架图	22
图 3.4	树莓派监控节点部署示意图	23
图 3.5	云平台关系示意图	24
图 3.6	云服务器软件模块图	25
图 3.7	网页通信流程图	29
图 4.1	动态图像效果图 ^[67]	34
图 4.2	TSN 模型原理图 ^[68]	35
图 4.3	DP-TSN 模型框架图	36
图 4.4	结合 Borda 决策的多视角行为识别框架图	38
图 4.5	生成的动态图像示意图	38
图 4.6	DP-TSN 模型数据测试图	40
图 4.7	结合 Borda 决策的多视角行为识别数据测试图	40
图 5.1	环境传感器硬件连接图	43
图 5.2	监控节点数据上传流程图	45
图 5.3	可穿戴设备数据获取流程图	47
图 5.4	客户端与服务器交互示意图	51
图 5.5	可穿戴设备获取的部分数据图	53
图 5.6	系统注册界面	54
图 5.7	系统登录界面	54
图 5.8	监控界面示意图	55
图 5.9	异常发生时监控视频图	55
图 5.10	系统短信预警图	55
图 5.11	系统节点绑定图	56
图 5.12	系统节点升级图	56
图 5.13	周内温湿度变化图	57

图 5.14 周内睡眠时长变化图	57
图 5.15 周内心率变化图	58
图 5.16 用户账号管理图	58
图 5.17 监护对象管理图	59

表格索引

表 1.1	1960 年-2060 年部分国家人口中值变化 ^[1]	1
表 2.1	典型人工特征表示方法	10
表 2.2	视频行为识别常用 CNN 模型.....	13
表 2.3	单视角下常用行为识别数据集	16
表 2.4	多视角下常用行为识别数据集	16
表 3.1	监控终端所用传感器	27
表 3.2	云服务器中网站部署方式	28
表 4.1	网络训练参数	39
表 5.1	sys_user 表设计	49
表 5.2	gur_user 表设计	49
表 5.3	sys_node 表设计	50
表 5.4	data 表设计	50
表 5.5	温度数据	52
表 5.6	湿度数据	52
表 5.7	烟雾传感器测试数据	53

符号对照表

符号	符号含义
∇	梯度
argmin	目标函数取最小值时的 变量值
\exp	以 e 为底的指数函数
$\max\{x,y\}$	x,y 中的最大值

缩略语对照表

缩略语	英文全称	中文对照
CCD	Charge Coupled Device	电荷耦合器件
CMOS	Complementary Metal Oxide Semiconductor	互补金属氧化物半导体
CPU	Central Processing Unit	中央处理器
CSI	CMOS Sensor Interface	相机串行接口
DBN	Dynamic Bayesian Network	动态贝叶斯网络
DDoS	Distributed Denial of Service	分布式拒绝服务
GPIO	General-purpose input/output	图形处理器
HMM	Hidden Markov Model	隐马尔科夫模型
HOG	Histogram of Oriented Gradient	方向梯度直方图
IP	Internet Protocol	网际协议
MIPI	Mobile Industry Processor Interface	移动行业处理器接口
PIR	Passive Infrared	被动红外
SCCB	Serial Camera Control Bus	串行摄像机控制总线
SIFT	Scale-invariant feature transform	尺度不变特征变换
SSH	Secure Shell	安全外壳协议
TCP	Transmission Control Protocol	传输控制协议
UDP	User Datagram Protocol	用户数据报协议
URI	Uniform Resource Identifier	统一资源标识符
USB	Universal Serial Bus	通用串行总线
Web	World Wide Web	互联网总称
WSN	Wireless Sensor Network	无线传感器网络

目录

摘要	I
ABSTRACT	III
插图索引	V
表格索引	VII
符号对照表	IX
缩略语对照表	XI
第一章 绪论	1
1.1 研究背景及意义	1
1.2 国内外研究现状	2
1.2.1 国外研究现状	2
1.2.2 国内研究现状	5
1.3 论文研究内容与创新点	6
1.4 章节安排	7
第二章 行为识别技术基础	9
2.1 基于传统方法的行为识别	9
2.1.1 传统行为识别方法	9
2.1.2 图像特征的选择与分析	10
2.1.3 基于 HOG 与 SVM 的行人识别	11
2.1.4 改进的密集轨迹算法	11
2.2 基于深度学习的行为识别	12
2.2.1 基于 CNN 的行为识别方法	12
2.2.2 双流法及衍生方法	14
2.3 多视角行为识别	14
2.4 公共数据集	15
2.5 本章小结	17
第三章 分布式智能监控系统设计	19
3.1 系统设计原则	19
3.2 系统设计需求	19
3.3 系统功能分析	20
3.4 系统总体设计	21
3.4.1 监控平台设计	23

3.4.2 监控节点设计	25
3.4.3 监控终端设计	27
3.4.4 Web 交互设计	28
3.5 本章小结	29
第四章 系统中行为识别算法的改进	31
4.1 基于动态图像的 TSN 模型	31
4.1.1 动态图像	31
4.1.2 TSN 模型	34
4.1.3 基于动态图像的 TSN 模型	36
4.2 多视角融合决策	37
4.2.1 Borda 投票	37
4.2.2 结合 Broda 的融合决策	37
4.3 实验和分析	38
4.4 本章小结	40
第五章 分布式智能监控系统实现	43
5.1 监控终端实现	43
5.2 监控节点实现	44
5.2.1 树莓派监控节点实现	44
5.2.2 可穿戴设备监控节点实现	46
5.3 监控平台实现	47
5.3.1 云服务器实现	47
5.3.2 云数据库实现	49
5.3.3 云对象存储实现	50
5.4 Web 页面实现	51
5.5 系统测试和分析	52
5.5.1 系统数据测试	52
5.5.2 系统功能测试	54
5.6 本章小结	59
第六章 总结与展望	61
6.1 研究内容总结	61
6.2 研究工作展望	62
参考文献	63
致谢	69
作者简介	71

第一章 绪论

本章首先介绍了适用于老人监护系统的研究背景及研究意义，其次介绍了老人监护技术、人体行为识别技术的国内外研究现状，最后指出了本文的研究内容、创新点以及章节安排。

1.1 研究背景及意义

联合国 2017 年的人口调查^[1]数据显示，全球大多数国家都存在人口老龄化不断加重的趋势。预计到 2060 年，欧洲国家的人口年龄中值将在 40 岁至 50 岁之间。虽然北美和南美人口老龄化的增长速率低于欧洲国家，但是到 2060 年加拿大人口中值不会低于一些欧洲国家。由于同为移民国家的性质，因此澳大利亚的人口变化趋势与美国相近。在亚洲国家中，韩国和日本的人口老龄化发展速度高于我国，韩国人口中值将在 2060 年上升到 54 岁。表 1.1 为 1960 年至 2060 年之间部分国家的人口中值变化。

表1.1 1960 年-2060 年部分国家人口中值变化^[1]

国家		1960 年	1980 年	2000 年	2020 年	2040 年	2060 年
亚 洲	中国	21.4 岁	21.9 岁	30.1 岁	38.7 岁	47 岁	48.5 岁
	日本	25.4 岁	32.6 岁	41.2 岁	48.2 岁	53.1 岁	53.0 岁
	韩国	18.6 岁	22.1 岁	31.9 岁	43.4 岁	51.1 岁	54.3 岁
	印度	20.3 岁	20.2 岁	22.7 岁	28.2 岁	34.5 岁	39.9 岁
欧 洲	丹麦	33.0 岁	34.3 岁	38.4 岁	42.3 岁	43.4 岁	44.2 岁
	瑞典	36.0 岁	36.3 岁	39.4 岁	41.0 岁	43.5 岁	43.7 岁
	英国	35.6 岁	34.4 岁	37.6 岁	42.0 岁	43.8 岁	44.7 岁
	德国	34.7 岁	36.5 岁	40.1 岁	46.6 岁	49.5 岁	49.4 岁
	荷兰	28.6 岁	31.3 岁	37.5 岁	43.2 岁	45.9 岁	46.9 岁
美 洲	美国	29.8 岁	30.0 岁	35.2 岁	38.3 岁	41.2 岁	42.5 岁
	加拿大	26.5 岁	29.2 岁	36.8 岁	41.4 岁	45.1 岁	45.5 岁
澳 洲	澳大利亚	29.6 岁	29.3 岁	35.4 岁	37.9 岁	41.0 岁	42.5 岁
	新西兰	27.4 岁	27.9 岁	34.3 岁	37.9 岁	42.0 岁	44.3 岁

我国自 2000 年迈入人口老龄化社会之后，人口老龄化的国情日益严重。2010 年第六次全国人口普查数据显示，全国大于等于 60 岁和大于等于 65 岁的人口比值分别

为 13.26%、8.87%，超过了老龄化社会判断 60 岁及以上人口比值 10% 或 65 岁及以上人口比值 7% 的标准。尽管我国进入人口老龄化社会比较晚，但是老龄化发展速度比较快。同时，少子化以更快的速度下降，老龄化和少子化反向发展加快了人口老龄化程度，致使人口老龄化的问题和矛盾更加突出^[2]。

目前，全球人类平均预期寿命为 71 岁，而健康的寿命预期只有 62 岁，即人们在生命的最后十年可能需要稳定各种慢性疾病，意味着越来越多老年人的护理需求超越了子女可以提供的能力范围。各个国家中老年人护理方式存在巨大的文化和政治差异，但目标都是通过一种更加适合老年人的方法来确保老年人享有健康的照顾和尊严。虽然成人日托、长期护理、疗养院等服务可以为老年人提供医疗保健、饮食营养、社交和其他日常生活支持，但是老年人会存在失去独立的感觉。老年人更喜欢呆在舒适的家中，在家里他们会比在任何昂贵的护理机构、医疗机构更有舒适感。

无论是在单一国家之间还是在不同国家之间，人们都比以往任何时候更具流动性，而产生此现象的原因是多方面的、复杂的^[3]。概括来说，人们离父母更远。随着年龄的增长，老年人的身体机能会持续恶化，日常生活自理能力会下降，突发事件也会增多。保护老年人居家养老时的生命安全、满足老年人的情感需求需要子女和社会的关心。同时，也需要全面的监护系统、便捷的智能产品为老年人提供更便捷、更安全的保障。

老年人监护系统的核心任务是分析老年人在日常生活中是否存在意外事件和异常行为并提供相关帮助。正确分析出老年人的突发意外和行为异常，不仅可以有效地降低老年人因意外事故造成的生命风险，还可以消除老年人对未来生活的恐惧、提高晚年幸福指数。面向老人的监护系统虽然已被科研人员、产品规划者提出不同的解决方案，方案解决的核心途径包括老年人生理参数采集、生活环境数据采集、日常活动室内定位、数据分析算法优化等。然而，大部分系统仍存在如对老年人行为检测单一、功能不完全的问题^[4]。因此，开发一套应用便捷、功能齐全、能够为老年人提供全方位保护的监护系统很有必要。

1.2 国内外研究现状

1.2.1 国外研究现状

1. 老人监护系统研究现状

目前，通信技术、传感器技术、计算能力、数据分析算法的发展为长期医疗服务奠定了基础。同时，技术的发展也帮助老年人可以在家中独立生活更长的时间，家庭正成为未来可能与医院竞争的医疗保健创新的中心^[5]。基于智慧家庭的护理系统、通用基础技术研究是国外在老年人监护方面中常研究的内容，其中基于智慧家庭的监护

系统包括无线传感器网络（Wireless Sensor Networks, WSN）、机器人的综合运用；其中通用基础技术的研究也是用于智慧家庭监护系统的支撑通用技术研究。

在 2002 年，Gregory D 等提出了 AHRI (Aware Home Research Initiative) 项目^[6]，项目是利用感知和传输技术了解居住者的位置和活动；Costa 等^[7]在 2012 年提出的智能家庭系统结合了智能议程管理器和智能监控框架，为老年人提供了基于传感器的监控、活动检测、调度和决策的智能功能，比如跌倒检测；Dawadi 等^[8]在 2013 年提出的智慧家庭可以利用机器学习的方法评估活动者的健康状况，其中数据由部署的传感器收集，所用的传感器包括门窗传感器、温湿度传感器、电力消耗功率计等；Suryadevara 等^[9]在 2014 年提出了一种运用被动红外（Passive Infrared, PIR）传感器、具有普适计算功能的智能系统，以确定在家中独立生活的老年人的健康状况；Do 等^[10]在 2018 年提出的 RiSH (A robot-integrated smart home) 系统集成了家庭服务机器人、家庭传感器网络、身体传感器网络、移动设备、云服务器和远程护理人员。

适用于智慧家庭监护系统的基础研究内容包括通用通信框架研究、隐私策略研究。如在通信框架研究方面，Ghorbel 等^[11]在 2013 年提出了一种多模式通知框架，Armentia 等^[12]在 2015 年提出了一种基于多代理技术的通用中间件体系结构。在隐私保护方面，Cavoukian 等提出设计隐私（Privacy by Design，PbD）的概念^[5]，通用电气公司的 Quiet Care 项目和因特尔的 Health Guide 项目运用了 PbD 的概念。

总而言之，基于智慧家庭实现的老人监护系统侧重于使用各式各样的传感器，获取老年人行为数据。除了物品传感器之外，视觉传感器组成的智能视频监控系统也是监护老年人的有效途径。

国外现有的智能视频监控系统包括用于公共交通的 CROMATICA、PRISMATICA 项目、用于公共场所的 HESPERIA、VSAM 项目等。在适用于老人的智能视频监控系统设计方面，Tran 等^[13]2014 年提出了一种利用图像和音频信息来检测异常事件的系统，检测在地板上不动、长时间外出、语言异常和声音异常的情况；Stone 等^[14]在 2015 年提出了一种使用微软 Kinect 的两级跌倒检测系统，用以检测老人在家跌倒的情况。

在智能视频监控分析方面，Wiliem 等^[15]在 2012 年提出了一个使用上下文信息来检测视频监控中可疑行为的系统，该系统包含三个主要组件，即上下文空间模型，数据流聚类算法和推理算法；Bian 等^[16]在 2015 年提出了一种人体部位跟踪方法来检测老年人的跌倒，其中图像的采集使用了深度摄像头以确保在黑暗中也可以使用。

2. 人体行为识别研究现状

人体行为识别最早的学术报道是瑞典心理学家 Johansso^[17]在 1973 年提出的用于人体运动建模的 MLD (Moving Light Displays) 模型。从广义上，人体行为识别方法可以分为单模式方法和多模式方法，如图 1.1 所示。其中单模式方法表示处理如图像

单一模态数据的人类活动，其适用于基于运动特征识别人类行为的情况，可以进一步分类为：时空方法、随机方法、基于规则的方法、基于形状的方法；多模式方法用来处理结合不同来源收集的特征，其可以分情感、动作、社交网络的方法。

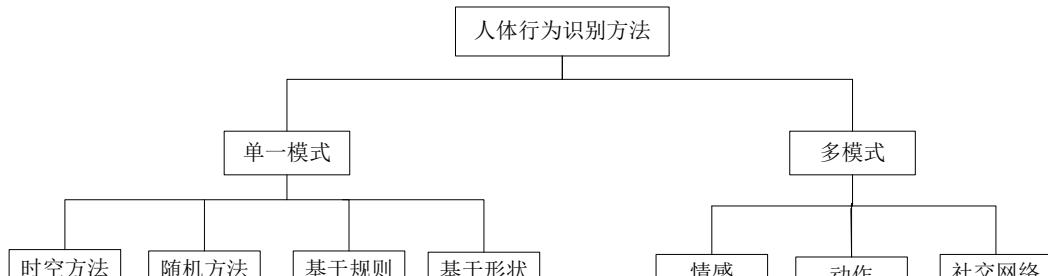


图1.1 人体行为识别方法

在图像、视频数据对应的计算机视觉领域中，人体行为识别与分析是一个活跃的研究热点，其在智能视频监控、人机交互、基于内容的视频压缩与检索、增强现实和游戏等领域都有着广泛的应用，通俗来讲就是通过视频画面来判断人或者感兴趣的物体在进行什么行为。通常，在文献中术语“活动（activity）”和“行为（behavior）”可互换使用^[18]。

人体行为识别技术在短暂的几十年中已取得突破性的进展。Roach 等^[19]在 1980 年运用最小二乘法检测图像中的运动对象，标志着计算机视觉技术运用在人体行为识别领域；LeCun 等^[20]在 1998 年提出的 LeNet 模型，标志着卷积神经网络（Convolutional Neural Networks, CNN）的正式形成；Alex 等^[21]在 2012 年提出 AlexNet 模型在图像分类中取得了历史性的突破，分类的准确率超过了人类。至此，运用深度学习方法对包括行为识别在内的计算机视觉技术的研究成为趋势。

目前，利用深度学习进行行为识别的研究主要从网络结构、数据输入、网络层次连接三个方面改进。在网络结构方面，主流的结构都是基于双流卷积网络和 3D 卷积网络发展而来，如 TSN 模型、C3D（3D Convolutional Neural Networks, 3D 卷积神经网络）模型，除此之外，也有基于 inception 等模型、网络层次的改进；在数据输入方面，包括对数据类型、数据格式、数据增强等进行改进，典型的研究包括关键帧、光流特征；在网络层次连接方面，主要是指双流法中时空信息的交互，可以是单网络内部各层之间交互、也可以是双流网络之间的交互，如 Feichtenhofer、Girdhar^[22-23]等人的研究。

国外在计算机视觉方面的科研机构有卡内基梅隆大学双目实验室、斯坦福大学人工智能实验室、伯克利人工智能实验室、蒙特利尔大学机器学习研究所、加州大学伯克利分校视觉实验室等。

1.2.2 国内研究现状

1. 老人监护系统研究现状

国内研究领域和产业领域都关注着人口老龄化的社会现状。从产业角度分析，用于老年人监护的产品类型主要包括智能手环、智能手表、远程摄像头等；从研究角度分析，在老年人监护方面的课题主要包括智能视频监控系统的设计、智能产品的设计等。

许多以老年人监护为背景的智能产品因为设计目标重点不同，选择的实现路径也不同。贾立平^[24]在 2013 年针对老年人的监护需求设计了一款智能腕表，其提供了行为识别、数据监护两大功能，系统拥有小型化、携带便捷的特点；吕莎莎^[25]在 2016 年提出了一种手腕式血压仪的设计，目的是通过人性化、情感化、互动化的设计帮助老年人监督自己身体健康；贾国忠^[26]在 2018 年提出了适用于老年人的智能音箱的设计，为老年人提供便利的交互方式和情感陪伴。

除了适用于老年人智能产品的设计，国内也有许多监护系统的设计。Huang 等^[27]在 2014 年设计了适用于老年人的智能空间系统并提出了一种能够识别老年人异常行为的算法，其中设计的智能空间包括智能节点、视频采集设备以及服务机器人；马宝庆^[28]在 2014 年设计的监护方案是在视频监控范围内利用视频画面判断异常行为，在视频监控范围外利用日常生活规律分析异常行为，并且系统采用虚拟人形方法保护老人隐私；武传胜等^[29]在 2016 年提出了一种运用声音监护老人的系统，具体是利用声音信号是否具有基频、泛频结构来判断声音信号为语音信号或非语音信号，若语音信号的字句内容为求救信息，或非者语音信号为异常声响，则主机对外送出警示信息给照顾护理人或单位；余文明^[30]在 2016 年设计的系统定位是实现老年人的实时监控功能，其提供了紧急拨号、视频语音通话、室内定位、环境监控等功能，满足了情感陪伴、安全保障的基本需求。

对比国内外研究现状发现，国外团队更侧重于系统设计和基础研究，在系统设计中充分运用传感器及传感器的数据，也更倾向于提供通用的平台或框架；国内团队更侧重于实现产品自身智能化，通过图像数据判断分析监护系统是最常用的方式。

2. 人体行为识别研究现状

中文数据库中关于人体行为识别方面的研究资料开始于 2000 年左右，近些年来相关的期刊、论文等呈逐步上升的趋势。在 2005 年左右，国内研究机构发表在英文数据库中关于人体行为识别的研究成果逐步增多。目前，国内在计算机视觉方面的研究机构包括北京大学智能科学系、浙江大学图像技术研究与应用团队、中国科学院自动化所、上海交通大学图像处理与模式识别研究所、电子科技大学的图像与视频处理实验室、西安交通大学人工智能与机器人研究所等。

国内在利用深度学习进行行为识别方面同样做了许多研究，对研究工作从网络结

构、数据输入、网络层次连接三个方面概述。

在结构改进方面，国防科技大学的 Qiu 等^[31]在 2017 年提出了基于 ResNet 改进的 P3D ResNet 新结构，实现以经济有效的方式模仿 3D 卷积网络。

在数据输入方面，Zhao 等^[32]在 2018 年提出了一种用于视频进行学习的新卷积网络架构，它可以纯粹地从原始视频帧中获得动态的信息，而不需要光流特征；王等^[33]在 2018 年提出了一种新颖的混合视频存储器机器，使用较少的静止图像来增强动作识别。

在网络连接方面，王云波等^[34]在 2016 年提出了时空金字塔网络（Spatiotemporal Pyramid Network）模型，以融合金字塔结构中的空间和时间特征，使特征能够相互加强；Wei 等^[36]在 2016 年提出了 Hypotheses-CNN-Pooling（HCP）框架，可以使用大规模单标签图像数据集进行良好的预训练；Zhao 等^[35]在 2017 年提出了一种简单的网络架构，通过深度合并和运行神经网络进一步减少训练难度。

1.3 论文研究内容与创新点

面向老年人的监护系统不仅有越来越多的产品投入使用，并且在学术研究中受到越来越多的关注。基于智慧家庭的监护系统监控全面，但系统部署复杂；基于智能视频监控系统的方案结构简单，但系统设计未考虑老年人真实活动范围；基于智能产品的老人监护轻巧便携，但功能单一。

针对以上问题，论文设计并实现了一套应用于老人监护的智能监控系统，系统基于分布式架构设计，包含监控平台、监控节点、监控终端、用户交互四部分，并且为了提高系统准确性，对行为识别算法做了改进。具体研究内容如下：

1. 设计和实现高效率的监护系统。由于传感器产生大量的数据、算法运行消耗大量的资源，而监护系统需要具备良好的实时性来保障老年人安全。因此设计的系统采用分布式架构，在监控节点处理一部分数据，充分利用了监控节点剩余的 CPU 资源。同时，也有效缓解了大量数据背景下监控平台的数据分析压力。

2. 设计和实现监护全面性的系统。由于生活环境的复杂性和意外的多样性，全方位保障老年人的养老安全需要在老年人居住范围内充分监测老年人活动信息。所设计的监护系统充分考虑了老年人的生活状态、活动场景，并运用传感器、摄像头、可穿戴设备全方位监护老年人健康状况、居住环境、日常行为。

3. 提高视频中行为识别算法的准确性。所设计的智能视频监控系统利用传感器、智能可穿戴设备检测老年人的异常情况，但是老年人的行为异常主要是通过视觉传感器采集的视频图像数据判断。因此本文提出了一种结合动态图像、TSN 模型的改进方法，提高视频中异常行为识别的准确度。

4. 提高多视角下行为识别的准确性。在设计的分布式监控系统中存在多个监控节点，由多个监控节点摄像头构成的多视角场景可以检测到遮挡物、光线不适等复杂场景。本文提出一种结合 Borda 投票的融合决策方法，将单视角下的行为识别判断做融合决策，进一步行为识别提高准确性。

论文的创新点体现如下：

1. 智能监控系统设计。设计的分布式监护系统拥有良好的伸缩性，不仅可以适用于独立的家庭环境，而且可以适用于大型护理中心。同时，监护节点承担分析的作用，平衡了系统实现的复杂度与算法的高效性，进而保证了监控系统的实用性。
2. 提出了单视角下的行为识别改进方法。视频比图像拥有更多的信息，而常用的 TSN 模型的输入是运用静态图像获取光流、空间流信息。动态图像是多帧图像的融合，拥有更多的帧间信息，可以更充分表达视频携带的数据量，因此利用动态图像改进 TSN 模型。
3. 提出一种多视角下的行为识别融合决策。多视角行为识别改进方法的提出进一步发掘了系统中监控节点的信息互补的优势。在视频监控过程中，存在某一视角与监控目标角度更合适的情况，提出的融合决策方法充分保证了投票的公平性。

1.4 章节安排

本文共分为七章，具体内容安排如下：

第一章是绪论，详细阐述了论文的研究背景、监控系统和行为识别算法的研究现状，并在确定本文研究价值的基础上指出了课题的研究内容。

第二章主要介绍了行为识别的理论技术，包括基于传统方法的行为识别和基于深度学习的行为识别，并介绍了多视角下行为识别的研究基础和常用的公共数据集。

第三章介绍了面向老人的智能化监控系统总体设计，首先介绍了系统设计原则、设计需求、功能需求，然后介绍了设计的分布式智能监控系统，并详细描述了系统中各个组成部分的设计。

第四章介绍了研究的行为识别算法。首先，提出了基于动态图像的 TSN 模型；其次，针对多视角下行为识别提出了结合 Borda 的融合决策方法；最后，在公共数据集下测试了提出方法的准确性。

第五章是智能化监控系统的实现部分，主要介绍了监控终端、监控节点、监控平台、用户交互四个组成部分的实现，并测试了系统的运行结果。

最后总结了本课题的研究成果，指出了其中的不足之处，并展望了未来的研究方向。

第二章 行为识别技术基础

行为识别之所以是较高层次的视觉研究，是因为其以运动特征、环境信息、行为模型为基础。基于视频的行为识别是通过已建立的行为模型和已知的环境信息对实时输入的视频流判断有什么行为发生。目前，对于视频行为识别主流的处理方法可以分为两大类：基于传统方法的行为识别和基于深度学习的行为识别。

2.1 基于传统方法的行为识别

在深度学习方法广泛应用在行为识别领域之前，传统的行为识别方法也取得了良好的效果。传统行为识别方法主要是根据专家知识选取有区分能力的特征描述子，再根据提取的特征进行识别和分类。

2.1.1 传统行为识别方法

广义上，行为识别包括目标检测、目标跟踪和行为识别分类三个过程，传统方法研究点通常在于特征的选取、融合与分析。

1. 目标检测

目标检测是检测出视频序列中一组连接的动态目标像素，其在时间上具有相干性并且在空间上具有语义相似性，即意味着通常应该考虑像素的时空关系来很好地检测运动物体。常用的目标检测方法包括帧间差分法、背景减除法、光流法，从近些年相关文献可知，还有一些其他方法用于目标检测如基于主动轮廓线模型方法、基于小波的方法。

帧间差分法的基本思想是在一系列图像帧上通过图像序列的前一帧或下一帧减去当前图像帧来识别移动目标。给定的阈值被应用于与差异图像比较，以消除由于相机噪声、小的照明变化等引起的像素变化，该方法非常适应于动态场景。

背景减除法的是将图像与图像的估计进行比较，在观察到的和估计的图像之间存在的差异区域中提取前景对象。背景减除法在进行原始减法和阈值处理之后需要应用形态学操作。

光流法是一种基于矢量的方法，通过匹配多个帧上的对象上的点来估计视频中的运动，光流法的精确度依赖于需要适度高的帧速率。由于算法的复杂性，因此光流法的实时实现通常需要专用硬件。使用光流的一个好处是它对多个同时的相机和物体运动的情况判断结果很稳健，是人群分析和包含密集运动的条件的理想选择。

2. 目标跟踪

目标跟踪是在给定的一系列图像中推断目标的运动，然后标记目标的位置。常用的几种跟踪方法分别以特征、3D、主动轮廓、运动估计为出发点。

基于特征的跟踪方法是由于采样时间间隔小，使得可以认为相邻图像中单个特征的运动形式是平滑的，进而在连续帧图像中匹配特征点跟踪目标；基于3D的跟踪过程是利用摄像机与场景之间的几何关系，把三维模型投影成二维图像，并跟踪二维图像中位置的变化；基于主动轮廓的跟踪方法中常用的是基于边缘信息跟踪，因为运动目标边缘特征能提供不依赖于运动形式、物体的目标信息；基于运动估计的跟踪算法是在光流约束条件下检测目标运动参数，通常有基于特点光流场和全局光流场两种方法。

3. 行为识别算法

人体行为识别算法主要分为模板匹配、状态空间两类。基于模板匹配的方法是将视频画面序列转换为一组静态模板，然后对定义的模板进行匹配，并与模板进行相似度比较，通过行为分析选择与视频画面序列匹配度最高的模板，大概可以分为灰度相关匹配、特征相关匹配两类；基于状态空间的方法首先是用状态表示运动中的静态姿态，其次对状态进行概率关联，然后把视频中的运动作为状态的遍历，最后以联合概率的最大值作为行为分析的基础，其中最具代表性的方法是动态贝叶斯网络（Dynamic Bayesian Network, DBN）和隐马尔科夫模型（Hidden Markov Model, HMM）。

2.1.2 图像特征的选择与分析

在传统行为识别算法实现过程中提取合适的特征可以决定识别的准确度，合适的特征可以降低数据维度、提取有效信息进而提高图像匹配精准度。

人工选择特征可以分成梯度特征，图案特征，形状特征和颜色特征，表2.1中列

表2.1 典型人工特征表示方法

	典型的特征表示方法
梯度特征	SIFT, PCA-SIFT, 图像局部显著性一点特征（GLOH）, SURF, v-HOG, CoHOG, spHOG, 积分直方图
图案特征	Gabor, LBP, CS-LBP, NR-LBP, Semantic-LBP, Fourier-LBP, multi-LBP, Co-LBP, Haar-like 特征
形状特征	形状上下文, 角度矩阵, 定向线性轮廓, 边界, 轮廓, 局部曲线, 射线特征, kAS
颜色特征	交叉皮层模型, 协方差特征, C-SIFT, HSV-SIFT, Hue-SIFT, Opponent-SIFT, 颜色名称

中列出了属于不同类的典型特征^[37]。梯度特征从梯度滤波输出构建直方图，常用的有尺度不变特征变换（Scale-invariant feature transform, SIFT）特征和方向梯度直方图（Histogram of Oriented Gradient, HOG）特征；图案特征是指分析相邻像素或子区域与局部图像表示关系的模式特征，包括 Gabor 特征、类 Haar 特征和局部二值模式（Local Binary Pattern, LBP）特征；形状特征主要集中在形状描述上，包括形状上下文等；颜色特征主要关注强度、颜色空间中构造的概率表示，常用的颜色特征包括颜色 SIFT、颜色名称等。

2.1.3 基于 HOG 与 SVM 的行人识别

HOG 特征由法国学者 Dalal 和 Triggs 在 CVPR (Computer Vision and Pattern Recognition) 会议上提出^[38]，它是利用强度梯度或边缘方向的分布表达图像内的局部对象外观、形状。这些描述符的实现可以通过将图像划分为单元来实现，首先，对于每个单元计算单元内像素的梯度方向或边缘方向的直方图，然后，将直方图的进行组合来表示描述符。HOG 开始被用于直立人体检测，后来扩展到通用物体检测，HOG 特征提取步骤如图 2.1 所示。

在 Dalal 等利用 HOG 特征结合支持向量机（Support Vector Machine, SVM）在行人检测任务上取得成功之后，基于此方法的改进不断被提出。Fatih Poriki^[39]提出了积分直方图，大大提高了 HOG 特征提取的速度；文献[40]中将 HOG 和 LBP 组合用于处理具有部分遮挡的行人检测；文献[41]验证了 RBF-SVM(径向基函数支持向量机)的分类效果明显优于线性 SVM 的效果，缺点是更耗时；文献[42-43]中以不同的方式结合 adaboost 和 SVM，有效地提高了算法的速度和准确性；文献 [44-45]中以多种方式执行多个 SVM 分类器的组合，例如运用二叉树结构或 SVM 网络的方式。

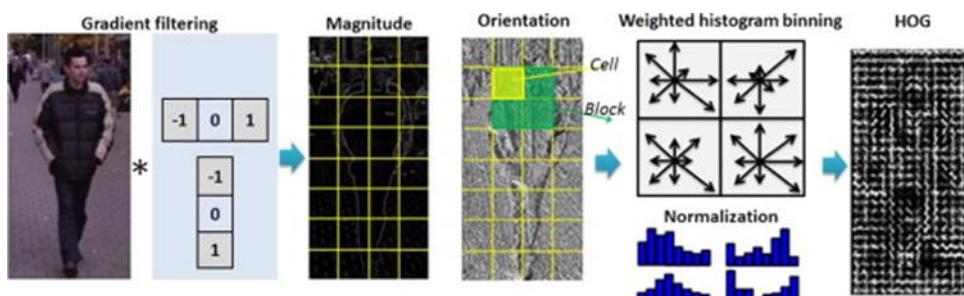


图2.1 HOG 特征提取步骤^[38]

2.1.4 改进的密集轨迹算法

改进的密集轨迹（improved Dense Trajectories, iDT）方法是行为识别领域运用传统方法中的经典、高效算法。在深度学习方法作为研究重点之前，iDT 算法的效果是最好的。IDT 算法是 H. Wang 基于密集轨迹（Dense Trajectories, DT）算法上改进得

到的^[46], 算法思想是利用相机运动减弱时背景运动生成的光流与轨迹, 在除去背景影响的情况下获取更有价值的前景物体的运动特征。

运用深度学习的行为识别方法比传统方法得到的准确率要高, 但是传统算法与深度学习的结合可以进一步提升算法准确度。就 iDT 算法而言, 与深度学习方法结合可以得到更好的效果。如 Afshar 等^[47]融合人脸局部小波二值模式-三维正交平面 (Local Gabor Patterns on Three Orthogonal Planes, LGBP-TOP)、几何特征, 将 iDT 算法用于人脸表情识别; Li^[48]等提出了一种利用改进的凸形密集轨迹进行视频表示的方法, 过程是首先检测出运动凸形区域, 其次在每个空间尺度上分别跟踪感兴趣点提取密集轨迹, 然后通过运动显著性分析对密集轨迹进行抑制; Shu^[49]等提出的方法是融合 iDT 算法与多视觉特征符、特征袋, 并将此用于一种交互系统。

2.2 基于深度学习的行为识别

深度学习最早正式出现在学术刊物上是 1986 年加州大学尔湾分校计算机系教授 Rina Dechter 在文献中提到 Deep Learning 一词^[50]。目前, 人们对深度学习方法的兴趣激增, 原因是它们在一些任务中的表现优于传统的算法。深度学习允许多个处理层的计算模型在多个抽象级别学习和表示数据, 进而模拟大脑如何感知和理解多模态信息, 从而隐式地捕获大规模数据的复杂结构。促成深度学习快速发展的因素是出现了大量高质量、公开可用的带标签数据集, 同时, 用于并行计算的 GPU 的出现, 使得基于 CPU 的训练能够过渡到基于 GPU 的训练, 加快深度模型的训练。另外一些因素也起到了推动作用, 例如新正则化技术的出现、TensorFlow 等强大框架的出现。

2.2.1 基于 CNN 的行为识别方法

基于深度学习的行为识别可以分为基于无监督或半监督学习的行为识别、基于卷积神经网络的行为识别、基于循环神经网络的行为识别。

CNN 的灵感来自于生物视觉皮层组织的结构, 尤其是来自文献[51]中提出的卷积神经网络模型。Yann LeCun 及其合作者^[20]设计了卷积神经网络模型 LeNet-5, 并在各种模式识别任务中获得了非常好的结果。如图 2.2 所示, LeNet-5 包含三种主要类型的神经层, 即卷积层、池化层、全连接层。LeNet-5 有两个卷积层、两个池化层、三个全连接层, 每一层都扮演着不同的角色, LeNet-5 的每一层都将输入量转换为神经元激活的输出量, 最终形成完全连接的层, 从而将输入数据映射到一维特征向量。

除了 LeNet 之外, 常用于视频行为识别的神经网络架构还包括 AlexNet、VGGNet、GoogleNet、ResNet、双流模型 (Two Stream Networks)、C3D 模型、长短期记忆网络架构 (Long Short-Term Memory, LSTM) 等。将常用的卷积网络模型的过程和特点,

总结如表 2.2 所示。

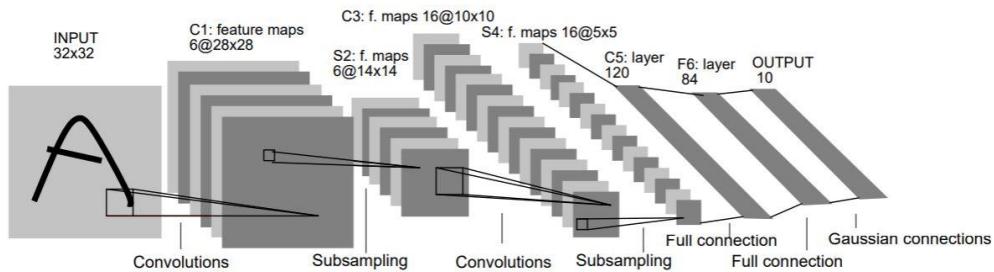


图2.2 LeNet-5 示意图^[20]

表2.2 视频行为识别常用 CNN 模型

	过程	特点
单帧 CNN	从视频中提取一帧，然后深入学习	仅适用一帧
CNN 扩展	从视频中提取多个帧，对这些帧进行卷积，并且卷积核为 $M \times N \times 3 \times T$ ，每个卷积层的 T 可以不同	不考虑同时进入网络的帧的次序
双流模型	空间流 CNN 和时间流 CNN 独立识别的结果，空间流是单帧的 CNN，时间流是处理叠加的光流帧的 CNN	可以处理时间的影响，然后与单帧处理相结合
LSTM	不使用全连接的层，而是使用 LSTM 在时间线上集成 CNN 最后一层的输出	可以有效地表达帧的顺序
C3D	卷积内核扩展到时域，在空间域和时域同时执行卷积	有很好地识别力

基于卷积神经网络的深度学习方法被广泛用于行为识别中。如 Makantasis 等^[52]利用深度学习对视频序列中的复杂事件进行检测和识别，其首先利用显著性映射对事件进行检测和定位，然后将深度学习应用于预先训练的特征识别与底层事件相对应的最重要的帧；Lin^[53]将一个半径边界作为正则化项加入到深度 CNN 模型中，有效地提高了 CNN 用于活动分类的泛化性能；Cao 等^[54]仔细研究了 CNN 作为联合特征提取和分类模型在适用于细粒活动中的适用性，他们发现由于存在较大的类内方差、较小的类间方差以及训练样本的可操作性有限等问题，在 SVM 分类器中直接使用从图像网络中学习到的深度特征的方法更为可取。

由于思考角度不同，卷积神经网络有多种的衍生模型以及改进思路。模型的适应性和传感器的可用性使融合多模态特征、融合多模态数据成为常用的行为识别方法，如在文献[55]中，为了识别从网络上采集的拥挤场景中的群体活动，作者混合了外观特征和运动特征。针对不同学习方式的组合，采用多任务深度学习能够提高计算效率，如 Tang 等^[56]的研究探讨了复杂事件识别的异类特征组合，并将此问题视为两个不同的任务：一个任务是估计识别事件的最有用的特征；另一个任务是使用与或图结构组合不同的特征。除了融合数据形式外，还有许多改进思路将多种类型的模型结合在一

起，如 Song 等^[57]使用视频和传感器数据提出了一个多模式多流的深度学习框架来解决以自我为中心的活动识别问题，Kavi 等^[58]提出了结合 CNN 和 LSTM 的多模型融合。

2.2.2 双流法及衍生方法

神经生物学认为在人脑初级视觉皮层中有两个通道可以输出信息：背侧流和腹侧流。背侧流主要涉及对运动信息和物体空间位置的处理，腹侧流主要涉及对物体识别的处理。基于此思想，在 2014 年的 NIPS (Advances in Neural Information Processing Systems) 会议上，Simonyan 等^[59]发表了一篇视频识别的文章，提出了基于两个独立识别流的双流法，即空间流从静止视频帧中动作识别，时间流从密集动作光流中识别行为，两个不同的流都通过 CNN 实现，最后进行信息融合。双流法的基本模型如图 2.3 所示。

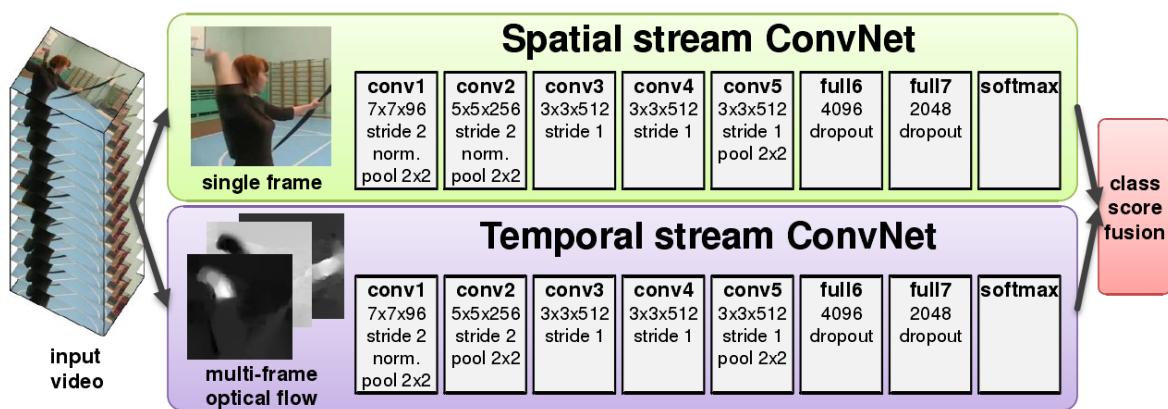


图2.3 双流法基本模型

Simonyan 等提出的双流法后来发展为更深网络的双流法，以及融合方式的双流法，甚至是除了双流还加入音频流的三流法。如 DeepMind 公司结合双流法和 3D 卷积提出了 Inflated 3D ConvNet 网络模型。

2.3 多视角行为识别

多视角下的行为识别可以分为两种，一种是多传感器的多视角，另外一种是基于多摄像机的多视角。多摄像机主要是通过多台摄像机从不同角度获取图像信息，融合不同角度的信息来弥补单台摄像机信息的不充分。本文的工作主要讨论多摄像机中多视角的情况。

单视角方法是使用一个摄像机来捕捉动作执行过程中的人体，但是，在不同的视角方向动作的视觉外观是完全不同的。单视角视频监控容易被摄像头自身角度影响或者是目标被其他物体遮挡，进而不能获取合适的判断特征。因此，单视角方法在训练

和测试过程中都设置了相同观测角度的基本假设，如果不满足这个假设，单视角方法的性能就会下降。多视角方法是近十年相对较新的研究方向，其利用多摄像机从不同角度进行全面的动作描述，实现独立于视角的动作识别。多视角下的行为识别可以分为基于视图的多视图识别、基于形状的多视图识别和主动识别。

基于视图的多视图识别目的是使用视点不变的局部描述符或具有变形公差的全局描述符为对象的 2D 图像添加视点公差。在给定训练多个视点图像中，可以通过跟踪多个视点之间共享的图像并相应地对图像进行聚类，或者学习它们之间相对二维位移寻找一组稳定的特征。为了进一步提高底层对象几何形状的精确度，这些 2D 图像元素也可以嵌入到隐式 3D 模型中。如果在测试中有多个视图可用，那么可以将图像组合起来将其视为单个更大的图像。近些年，CNN 结构得到了扩展，允许使用单个网络从图像序列中进行识别，方法包括跨所有视点的最大池，或将对象形状展开为全景图并跨每一行的最大池。如 Le 等^[60]提出了一种基于时空特征的方法，具体是通过无监督学习直接从视频中学习特征。当与堆叠和卷积的深度学习技术结合以学习分层表示时，该方法表现良好。

基于形状的多视图识别是指从重构或模型中学习一个显式三维形状，然后与深度图像、三维重构或形状补全的部分重构进行匹配。形状描述符包括局部表面属性分布，体素网格上的球面调和函数和三维局部不变特征。如 Weinland^[61]提出了一种在 3D 运动模板中使用运动历史体积进行多视图人体动作识别的方法。在该方法中，计算、对齐和比较由不同人在各种视点中执行的不同动作的运动历史体积。最近，CNN 被应用于三维形状，表现为三维占用网格，构建生成或判别网络。但是，从目前来看，基于二维视图方法的 CNN 的表现优于基于三维视图的方法。

主动识别的方法通常学习对象的生成模型，从未观察的视点预测对象外观，并根据熵减少的度量选择视图。Wu 等^[62]将对象建模为 SIFT 特征组成的三维云，将相机移动到可以显示尚未观察到的最大数量功能的视图。在文献[63]中提出了将主动识别纳入随机森林框架的方法，其中每个决策树都对对象分类和视点选择进行编码。最近，文献[64]中的 ShapeNets 框架提出将对象作为体素网格进行建模，并学习基于卷积深度信念网络的生成模型，以允许从未看到的视角进行视图合成。

2.4 公共数据集

在行为识别中，常用的公共数据集包括单视角数据集和多视角数据集，常见的单视图数据集包括 HCF101、KTH 等，如表 2.3 所示；常见的多视角数据集有 IXMAS、MuHAVi 等，如表 2.4 所示。

表2.3 单视角下常用行为识别数据集

数据集	样例 (张、 段)	类别 (个)	人数 (人)	摄像头个 数(个)	场景个数 (个)	格式	年份 (年)
KTH	2391	6	25	1	4	RGB	2004
Weizmann	90	10	9	1	1	RGB	2005
Hollywood	233	8	-	1	-	RGB	2008
UCF Sports	184	10	-	1	-	RGB	2008
UCF YouTube	3040	11	-	1	-	RGB	2008
Hollywood-2	3669	12	-	1	-	RGB	2009
MSR-Action3D	567	20	10	1	2	D+3D Joints	2010
Olympic sports dataset	783	16	-	1	-	RGB	2010
HMDB51	6849	51	-	1	-	RGB	2011
UCF50	6676	50	-	1	-	RGB	2012
MSR DailyActivity3D	320	16	10	1	1	RGB + D +3D Joints	2012
UCF101	13320	101	-	1	-	RGB	2013
UTD-MHAD	861	27	8	1	1	RGB + D +3D Joints	2015
Kinetics-600	500,000	600	-	1	-	RGB	2018

表2.4 多视角下常用行为识别数据集

数据集	样例 (张、 段)	类别 (个)	人数 (人)	摄像头 个数 (个)	场景个 数 (个)	格式	年份 (年)
IXMAS	2340	14	11	5	1	RGB	2006
MuHAVi	1904	17	14	8	-	RGB	2010
Northwestern- UCLA	1475	10	10	3	1	RGB、 RGB+D+3Djoints	2014
WVU-1	880	12	1	8	1	RGB	-

续表 2.4

Office Activity	1180	20	10	3	2	RGB+D	2014
UWA3D Multiview	900	30	10	5	1	RGB+D+3Djoints	2014
Act4 ²	6844	14	24	4	1	RGB+D	2014
CVS-MV-RGBD-Single	4320	10	18	3	1	RGB+D	2014
UWA3D Multiview II	1075	30	10	5	1	RGB+D+3Djoints	2015
M ² I	1784	22	-	2	1	RGB+D+3Djoints	2016
NTU+RGBD	56880	60	40	80	1	RGB+D+3Djoints	2016

2.5 本章小结

本章主要是为论文接下来的算法改进工作奠定理论基础。首先，介绍了传统行为识别算法研究点并概述了传统行为识别中常用的 HOG 与 SVM 结合方法和改进的密集轨迹算法；其次，介绍了基于深度学习的行为识别理论基础、研究点，并概述了常用的卷积神经结构双流模型及其改进模型；再次，介绍了对于多摄像头情况中多视角行为识别常用的研究方法；最后，分别介绍了行为识别中常用的单视角公共数据集与多视角公共数据集。

第三章 分布式智能监控系统设计

本章首先介绍面向老年人的智能监控系统的设计原则与设计需求，其次分析系统需要具备的功能，然后介绍系统的设计结构。在系统设计中分别介绍系统的四个组成部分，即监控平台、监控节点、监控终端和用户交互。

3.1 系统设计原则

面向老年人的智能监控系统的设计目标是通过技术手段实现高准确度、高便捷性的监护。因此，在进行智能监控系统设计时，应遵循以下原则：

1. 系统优化原则

在终端嵌入部分智能分析服务，一方面，可以充分利用终端设备富余的 CPU 资源，以实现内置的行为异常分析、数据简单分析等功能；另一方面，可以利用设备集成的 I/O 接口，可以在终端实现传感器采集、硬件控制、报警提示等综合功能，以满足监护全面化的要求。

2. 模块化设计原则

模块化设计是指将系统中的基本功能、智能化功能尽可能设计为独立的模块，以得到比较灵活的系统。这样在根据实际监控地点的需求适配硬件时，可以配合选取不同软件模块。

3. 操作与维护的原则

对于客户来说，使用操作肯定是越简单越好，所以智能化监控在设计的时候应尽可能的简单上手。系统的安装方式、交互界面、升级应简单明了，同时有详细易懂的操作提示。同时，在设计过程中要对各个功能模块进行规范化的合理分配，并要求具有良好的维护性^[65]。

4. 经济性原则

为了追求更高的实用性，在做系统设计时不能一味地思考更复杂更高级的解决方式，应在满足性能的前提下选择简单的方案。在进行设计之前，要充分的考虑各方面因素如研发成本、使用成本以及后期维护费用等。

3.2 系统设计需求

在分析老年人实际需求的基础上，同时为提升良好的用户体验，系统开发中除基本需求外还包含了部分衍生需求。总而言之，面向老年人的智能监控系统需求分析如下：

1. 全方位的防护能力

老年人居家养老过程中,因为日常生活范围是家庭居住范围而不仅限于单个房间,且老年人可能在任何监控区域发生意外,所以考虑到老年人实际的生活习惯,智能监控系统不能只适用于单个房间的监控或单一情况的监控。

2. 综合化的安全防范

为了全面保护老年人的生命财产安全,智能监控系统除了获取主要的视觉信息外,同时需要结合如生理参数、环境参数、活动轨迹等生活信息。而且,系统需要在具有智能分析的基础上具有较好的鲁棒性,能够实现在有遮挡物等复杂场景下的识别与分析功能。

3. 友好的交互界面

大多数智能系统功能复杂、专业性强、专业术语较多,普通用户很难快速熟悉其使用方式。因而可以使用图像化、操作提示、快捷工具等方法使用户的操作更加直观、提高系统的友好性。

4. 具有安全性

智能化监控系统也是基于计算机网络的监控系统,因而系统应对网络中的非法访问、非法入侵等应具有抵御能力,从而充分保证系统的数据安全性、访问安全性、传输安全性等。

3.3 系统功能分析

根据上文的设计原则和设计需求分析可知,设计的系统需要满足多功能要求,如监控功能、用户交互以及系统安全。因此,系统的功能划分如图 3.1 所示。

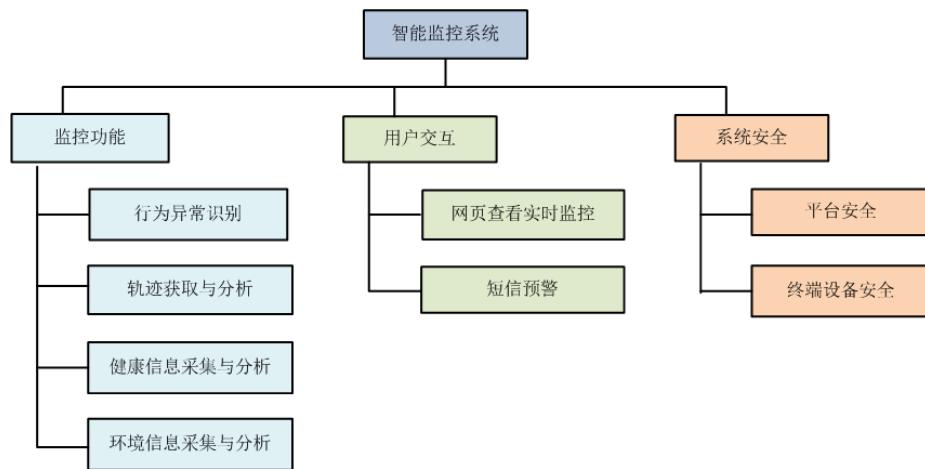


图3.1 智能监控系统功能图

1. 监护功能

由于老年人自身抵抗力较差很容易患上各种疾病,同时很多病都需要长期的监护

和治疗，因此需要一个监护功能去获取老年人的心率、睡眠以及运动指标等等。从而可以对这些参数指标进行分析处理，准确的提供对应医疗服务。目前，生理参数的获取可以通过可穿戴设备。

老年人的室内活动轨迹在一定程度上反映了老年人的生活状态，对老年人的轨迹数据进行建模，可以分析判断老年人是否存在异常情况。对室内老年人进行轨迹定位、记录其停留时间，那么我们就可以通过这些数据分析老年人当前的活动情况。

随着老年人年龄的增长，身体机能会逐渐老化，这就导致了老年人相较其他年龄段的人更容易发生摔倒等严重后果的意外情况。除可能发生的意外事件，老年人心理脆弱、对亲人的依赖性增强，在情感方面老年人也可能会存在哭泣、呆滞时间过久等异常情况。因此系统需要对老年人实时状态进行查看和行为异常的识别。

对于人来说，周边环境往往能影响一个人的心情，而环境一般会包括温度、湿度以及空气质量等，所以对于室内环境的监测也是一个重要的功能模块。

2. 用户交互

监控系统与用户的交互主要体现在两方面，一方面是用户通过网页查询实时监控状态、对监控系统进行信息设置，实时监控状态获取包括获取环境数据、老年人健康状态、视频数据等信息，信息设置操作的内容包括监护对象个人、系统使用者等信息；另一方面是，监控系统分析到存在异常事件之后，通过短信向监护人发送监护对象的异常提醒。

3. 系统安全

智能监控系统运行的流程可以总结为首先通过硬件设备获取所需数据，其次在平台上对数据进行建模、分析，最后将数据呈现给用户、与用户进行交互。从运行流程来看，设备、平台都存在非法用户访问导致不安全的可能性。并且，全方位的监控系统监护老年人的方方面面，系统中的监护对象信息是较为隐私的信息。为保护使用者的个人隐私和保障系统的平稳运行，系统安全至为重要。

3.4 系统总体设计

综合上文分析，将系统设计为由监控平台、监控节点、监控终端、用户交互组成的分布式监控系统来满足系统需求和功能。其中，作为监控终端的传感器包括视觉传感器、温湿度传感器和气体传感器等，用于收集环境信息；监控节点分为两种类型，一种是由智能硬件树莓派实现，另一种是提供接口的可穿戴设备，监控节点提供简单的信息处理、数据传输等功能；监控平台包括云服务器、云对象存储、云数据库，分别主要提供程序分析、信息存储、图片存放等功能。智能化监控系统示意图如图 3.2 所示。

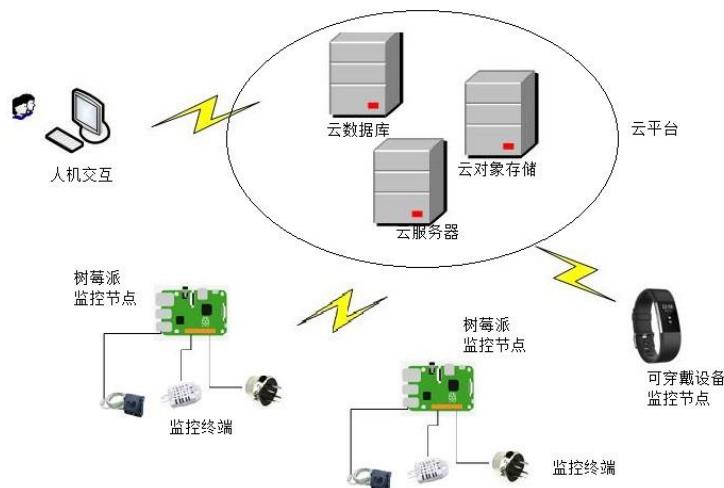


图3.2 智能监控系统组成图

由系统的总体设计得到智能化监控系统的整体框架如图 3.3 所示。数据从监控终端被采集之后，监控节点对其进行简要分析，监控平台进一步对数据进行分析和存储，最终将分析整理的数据呈现给用户。

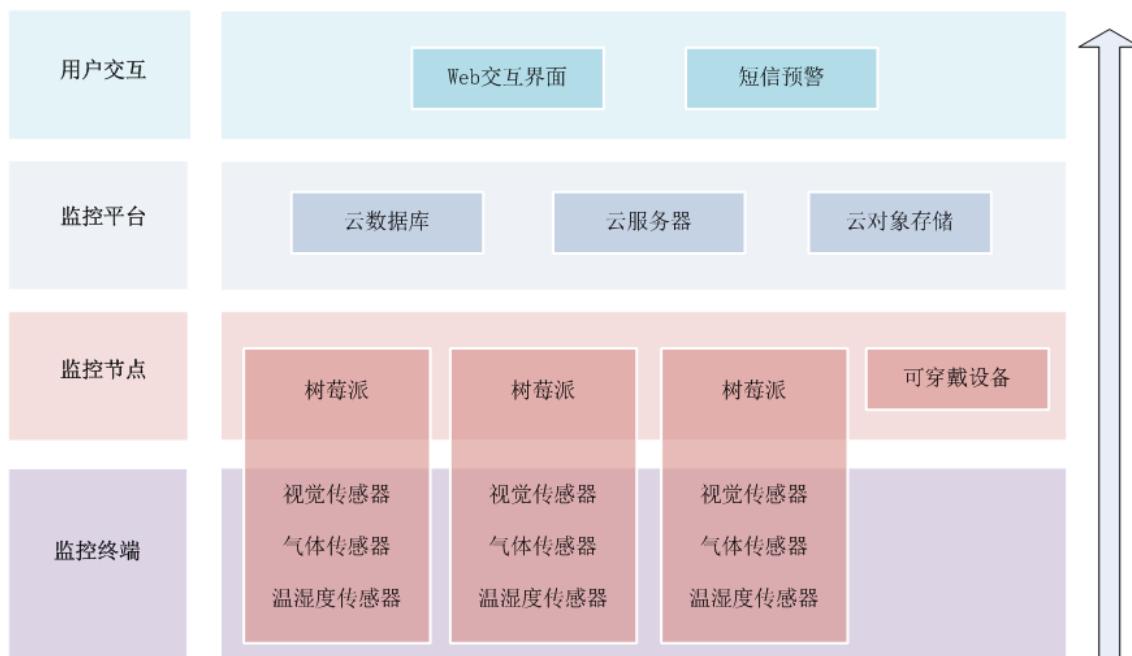


图3.3 智能监控系统框架图

监控节点会获取监控终端采集的信息，作为监控终端的多种传感器特别是视觉传感器会产生较大的数据量。系统选用分布式架构的目的是让监控节点提供基本的业务服务，通过过滤无用数据、处理简单数据来提高系统运行效率，以达到系统能更快地检测出异常情况的目的。

为了避免单摄像头存在的目标遮挡、角度不合适等问题，所设计的分布式智能监控系统提供采用多摄像头获取区域中多视角视觉信息的功能。换而言之，一个视野相

对宽阔的房间如厨房、厕所，可以采用单树莓派监控节点获取信息；而环境复杂的房间如客厅、卧室，可以采用多树莓派监控节点获取信息。监控系统的部署，可以采取如图 3.4 所示的方案，即一个房间内可以放置一个或两个树莓派监控节点。

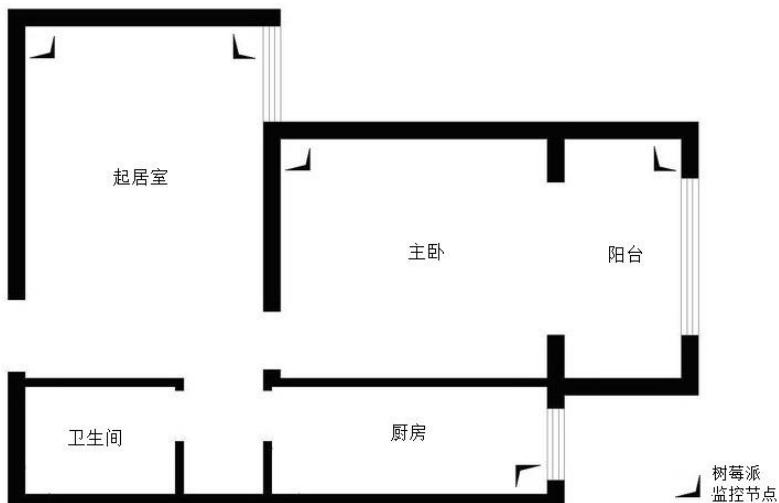


图3.4 树莓派监控节点部署示意图

3.4.1 监控平台设计

论文设计的系统监控平台由云服务器(Elastic Compute Service, ECS)、云数据库、云对象存储服务(Object Storage Service, OSS) 构成。云服务器是一种简单高效、具有强大处理能力的计算服务器，在使用过程中，随着业务的扩展或缩小，用户可以随时扩充或释放磁盘、带宽等资源。云对象存储是面向海量数据规模的分布式存储服务，其提供与平台无关的 Restful API 接口，可以实现在互联网任何位置存储和访问。云数据库是一种稳定、可靠、有弹性的在线数据库服务^[66]。

在论文的设计中云服务器承担系统、应用程序得运行平台角色，云对象存储负责的存储视频、图片类的监控信息，云数据库提供的功能是管理重要的数据信息。将监控平台设计为程序、图片、数据分离的方式是为了提高数据安全性、提升系统运行效率以及便于系统的扩展与维护。云服务器与云对象存储的搭配，将图片、视频、下载包保存在云对象存储中，可以降低存储费用。

目前主流云服务器有阿里云、腾讯云、百度云、华为云、亚马逊云等，对比各个服务的优缺点，系统最终选择的为阿里云服务器。分析来看，在技术方面上，阿里云支持更多的技术、安全性较高；从扩展性方面分析，阿里云集成了大部分主要的服务，同时提供了如云监控等技术服务。在安全性上，阿里云平台提供基础防护服务如分布式拒绝服务(Distributed Denial of Service, DDoS) 可以有效防止云服务器ECS实例受到恶意攻击，从而保证ECS系统的稳定，即当流入ECS实例的流量超出实例规格

对应的限制时，云盾就会帮助 ECS 实例限流，避免 ECS 系统出现问题。同时，在 Linux 实例中，阿里云安全外壳协议（Secure Shell，SSH）密钥对由公钥和私钥组成，是一种安全便捷的登录认证方式。相较于用户名和密码认证方式，SSH 密钥对有安全性、便捷性的优势。

阿里也提供了很多其他云产品，同一平台的产品会为开发提供便利性，因此云对象存储、云数据库选择的均为阿里产品。在监控平台中，云服务器、云对象存储、云数据库三者的关系如图 3.5 所示。

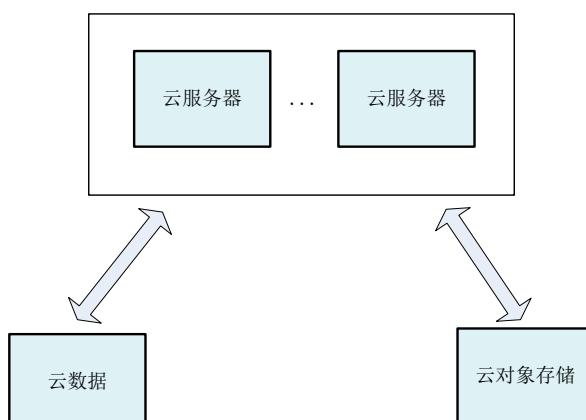


图3.5 云平台关系示意图

1. 云服务器设计

云服务器可以配置需要的操作系统类型以满足应用程序运行的平台要求和内存要求。作为监控平台、监控系统的核心，云服务器中的应用程序需要保证扩展性、安全性、灵活性。本文的云服务器主要由节点管理模块、资源管理模块、组合分析模块等模块组成，如图 3.6 所示。

节点管理模块是指设置系统内存在的监控节点设备信息和节点之间关联信息。根据设计的系统结构，一个目标对象拥有多个监控节点，为了区分不同的监控节点，给每个监控节点设置唯一的设备 ID 用来区分。因为多视角的监控系统可以提供更多的信息，所以将存在关系的树莓派监控节点绑定关联起来。

资源管理模块是指管理获取的信息数据将其存储云数据库或云对象存储。云服务器将接收到的环境数据存储到云数据库相应的数据表中，需要与用户交互显示时，通过程序查询云数据库获取数据。

组合分析模块根据监护系统采集到的监护数据，利用算法分析数据，并将分析结果推送给用户接口或通讯模块。因此，组合分析模块主要是由各种算法组成，提供的分析功能包括行为识别、轨迹分析、健康状态判断、生活环境判断。

数据交互模块是指与监控节点的数据交互，包括接收监控节点上传的数据、向监控节点传输文件、向监控节点发送操作指令等。

用户接口模块是实现系统使用者与系统信息的交互，如用户可以通过网页、移动终端的方式查看监控系统数据、配置监控系统信息等。

通讯模块是用来提供外部通信系统的信息交互，在系统自动判断发生意外时，通过程序读取云数据库中监护人联系方式，把报警消息经由通讯模块推送给监护人。

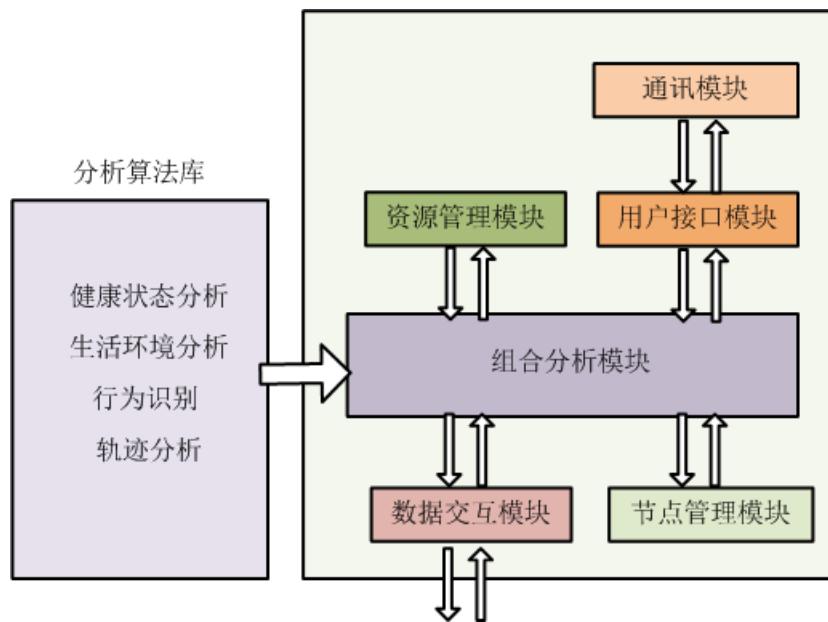


图3.6 云服务器软件模块图

2. 云数据库设计

课题设计中选取的阿里云数据库是关系型数据库（Relational Database Service, RDS），其支持 MySQL、SQL Server、PostgreSQL、MariaDB 等引擎。对于 Web 应用，MySQL 是应用最广泛的关系数据库管理系统，因此选用 MySQL 为数据库。为表示账户信息、监控对象、监控节点、获取数据的关系，分别设计 sys_user 表、gur_user 表、sys_node 表与 data 表。

3. 云对象存储设计

云服务器将接收到的用以判断的图片上传到云对象存储中，可以进一步地保证数据安全、分担云服务器的压力。云对象存储按照监控节点的设备 ID 编号构建顶层文件目录，各文件目录下按日期存放监控节点上传的图片，图片在以云服务器获取的时间命名，层次化的存储方式便于数据的访问。

3.4.2 监控节点设计

监控节点不仅拥有数据采集功能，还拥有数据分析功能。监控节点的分析功能分担了监控平台的压力，但其更侧重局部数据分析。监控平台的分析任务侧重宏观分析，比如监控平台会分析每一次定位点聚集形成的定位轨迹，同时会更加关注监控节点之

间数据的相关性。

1. 树莓派监控节点

节点以树莓派 3B+为硬件平台，具有成本低廉、便于连接传感器、易于安装的优势。树莓派监控节点提供 GPIO、I²C、串口等接口，并且支持 WiFi、蓝牙。树莓派属于智能硬件是将电脑集成到了一块电路板上的微型电脑，其兼容了完整的软硬件系统，可以直接开发。树莓派 3B+采用 64 位 4 核 ARM Cortex-A53 CPU，系统采用树莓派官方发布的 Ubuntu 16.04 系统，Ubuntu 是以桌面应用为主的开源 Linux 操作系统。

从层次上分析，监控节点向下采集监控终端数据，向上发送信息给监控平台；从功能上分析，除了承担信息传输中枢的任务，监控节点需要分担数据分析的任务。同时，树莓派监控节点作为一个相对独立的搭载操作系统的嵌入式平台，上面运行着各种各样的系统应用软件，因此树莓派监控节点需要具备设备安全的特性。

Linux 系统的一大优势在于它的安全性，Linux 的安全性得益于它的如下特点：具有卓越的补丁管理工具，可以在自动更新系统的同时升级系统中所有的软件；具有较为健壮的默认设置和权限管理，能将攻击带来的危害限定在一定的范围内；局域模块化设计，使得用户可以选择安装或删除任何一个系统组件；具有系统中默认整合的 SELinux 等安全机制，能够为系统提供保护；开放源代码框架，能更容易、更快速地发现和修复系统的安全问题等。

但是，Linux 是一个通用性操作系统，为了满足更广泛的用户需求，Linux 操作系统及应用软件都提供了多种配置参数，默认的设置安全性还不够。系统开发者应该根据系统安全需求，制定相应的安全策略，运用各种命令和工具进行合理的配置，并对系统进行监控。Linux 系统的安全性，可以从内核配置、文件系统安全、安全认证、网络安全以及 Linux 上的安全工具等方面设置。

对于采集到的环境信息或生理参数，数据分析的任务是对数据进行简要分析并决定是否上报数据。对于采集到的视觉传感器信息，数据分析的任务是指判断图像中是否存在明显或模糊的行为异常状态，若是模糊的行为异常，则向上发送数据，监控平台再通过多监控节点综合判断。对于采集到的环境信息，数据分析的任务是指数据是否超出正常范围，若数据不在正常范围，则上报数据、响应预警，若在正常范围，则保存数据。

树莓派监控节点与监控平台通信分三种形式，一是监控节点分析到数据异常，主动向监控平台上报数据；二是监控平台向监控节点请求信息；三是系统均未发生前两种情况，监控节点在固定时间间隔后上报数据。采用这种数据通信的方式，是为了在保证系统准确度的同时，保持系统效率。

树莓派监控节点与云服务器之间的通信传输可以选用 Socket 中的用户数据报协议（User Datagram Protocol， UDP）协议进行通信。Socket 套接字提供了三种类型，

即运用传输控制协议（Transmission Control Protocol, TCP）协议的流式套接字、运用 UDP 协议的数据报套接字、运用如控制报文协议（Internet Control Message Protocol, ICMP）等较低协议的原始数据报套接字。TCP 协议是面向连接的传输层协议，提供高效通信，能保证数据无误、数据无丢失、数据无失序以及数据无重复。UDP 协议是不可靠的无连接协议，在数据发送之前不需要连接，适用于小尺寸数据、应答困难、广播、组播、多媒体传输等。视频数据传输要求数据具有严格的实时性能，并且在一定程度上少量视频数据包的丢失不影响视频信息的提取，所以本文主采取 UDP 作为树莓派监控节点与云服务器之间的通信接口。

2. 可穿戴设备监控节点

智能手环、智能手表等可穿戴设备上拥有着大量的传感器，可以通过其采集到大量原始数据，对这些数据经过分析处理就可以获得监护对象的运动、健康状态和变化情况，这些都对分析用户自身身体健康状况有很大的意义。

目前，在用于生理数据收集的可穿戴设备中，以智能手环、智能手表居多。主流可穿戴设备厂商都拥有各自的云端数据存储服务并对用户开放使用，但大多数厂商是不提供二次开发接口的，这并不利于运用不同设备的数据进行统一管理。因为 Fitbit 品牌是提供了二次开发接口的，所以本文研究和开发过程中是使用 Fitbit Charge HR 手环作为可穿戴设备节点。

3.4.3 监控终端设计

监控系统需要完成全方位监护、用户交互、系统安全等功能，从技术上分析，为实现这些目标，监控终端需要采用基于生活线、基于传感器、基于可穿戴设备的老人监护技术的混合。在监控系统中作为监控终端的传感器总结如表 3.1 所示。

表3.1 监控终端所用传感器

监护功能	设备名称	接口方式	监控方式
行为轨迹	OV2710	USB	基于生活线
行为异常			基于视觉传感器
环境监测	DHT22/MQ-2	I2C/GPIO	基于环境传感器

1. OV2710

监控节点使用的树莓派平台支持通用串行总线（Universal Serial Bus, USB）接口、相机串行接口（CMOS Sensor Interface, CSI）的摄像头。USB 接口对应的视频数据传输方式为 UVC（USB video class）方式，CSI 是基于移动行业处理器接口（Mobile Industry Processor Interface, MIPI）的一种传输方式。节点中的 CSI 接口为树莓派官方中提供的完善接口，通过 CSI 接口的视频获取是通过命令行获取的，不利于系统综

合开发，因此选用 USB 接口。

选用的是 OV2710 型号的 CMOS 图像传感器，其可以通过串行摄像机控制总线（Serial Camera Control Bus, SCCB）接口或 MIPI 接口控制，传感器提供了全帧、子采样和窗口化 10 位图像。使用 OV2710 可以完全控制图像质量、格式和输出数据传输。所需的图像处理功能如曝光控制、缺陷像素取消、镜头阴影校正等可以通过 SCCB 接口进行编程。

互补金属氧化物半导体（Complementary Metal Oxide Semiconductor, CMOS）摄像机与电荷耦合器件（Charge Coupled Device, CCD）摄像机相比，结构简单、体积小，由于本身方便数字化，能更方便网络化、智能化，并且能容易与其它芯片结合，集成度高，可靠性高。

2. 环境传感器

选用的温湿度传感器为 DHT22 模块，其是一种数字型的温湿度复合传感器，产品具有良好的可靠性和稳定性。DHT22 温湿度传感器与 DHT11 温湿度传感器均为单总线驱动，但 DHT22 温湿度传感器的精度高于 DHT11 温湿度传感器，DHT22 温湿度传感器的温度测量范围为 -40~80 °C，温度测量精度 ±0.5 °C，湿度测量范围为 0~100%RH，湿度测量精度为 ±2%RH。除了 DHT 系列的温湿度传感器，利用 I²C(Inter Integrated Circuit) 通信的 SHT1X 系列的温湿度传感器精度也较高。

选用的气体传感器为 MQ-2 模块，其可以用来检测多种可燃性气体，在家庭环境中可以检测天然气的浓度是否过高。MQ-2 气体传感器可测的天然气浓度为 300pm~10000pm 之间，传感器在空气中无污染时模拟量输出电压值在 0.1~0.3v 之间；在浓度较高时，模拟量输出电压为 4v 左右。

3.4.4 Web 交互设计

对于独立的系统，个人可能会错过或忽略系统提示，并且假设有人偏离正常的期望则系统不能提供任何帮助。相反，具有用户交互功能的监控系统可以实时预警监护人员，监护人员可以在严重的健康损害行为之前进行帮助。系统提供 Web 界面交互与短信预警两种方式，其中 Web 交互需要在云服务器上搭建 Web 服务器。

表3.2 云服务器中网站部署方式

网站类型	WordPress	LNMP 环境	Java Web 环境	Node.js (CentOS)	Docker (CentOS7)	LAMP 环境
部署方式	镜像部署	ROS 模板部署	镜像部署	手动部署	手动部署	手动部署
	ROS 模板部署	镜像部署	手动部署			
		手动部署	插件部署			

阿里云服务提供适用多种系统的网站部署方式，其中适用于 Linux 系统的网站部署方式汇总如表 3.2 所示。

综合考虑开发语言、CPU 使用率，论文选用 LAMP 平台作为 Web 服务器，LAMP 分别代表 Linux、Apache、MySQL 和 PHP。Apache 是一个开源软件项目，特点是简单、速度快、性能稳定。超文本预处理器（Hypertext Preprocessor，PHP）是一种编程语言，支持多种操作系统，具有强大的可伸缩性、稳定高效运行、快速开发的特点，且支持和 JAVA 的交互等特性。MySQL 既是云数据库中选用的数据库类型。

PHP 语言负责后台与数据库的交互，执行在服务器端，网页访问 PHP 流程如图 3.7 所示。

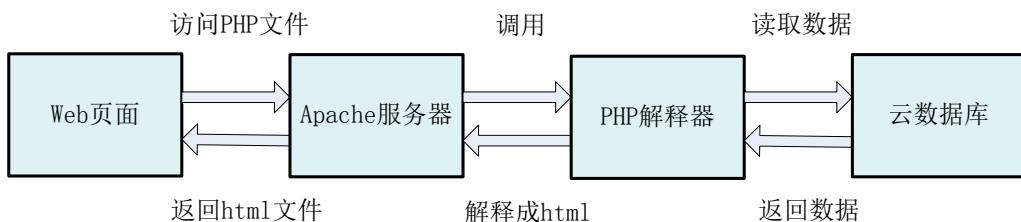


图3.7 网页通信流程图

由于浏览器是解析不了 PHP 代码的，所以当浏览器通过统一资源定位符去访问云服务器中 Apache 服务器对应路径下的 PHP 文件时，Apache 服务器会调用 PHP 解释器，PHP 解释器的功能是将 PHP 文件解释为 html，同样如果通过 PHP 去访问数据库时，也会将所得的数据解释为 html 代码的一部分，最终给浏览器返回的还是一个 html 文件。

因为选用的可穿戴设备为 Fitbit 智能手环，其提供的接口方式中只有 Web API 适用，因此需要在 Web 交互界面中调用其接口。

3.5 本章小结

本章以面向老人的家庭监护为背景基础，介绍了分布式智能监控系统的整体设计。首先，介绍了在老年人应用背景下系统的设计原则及需求；其次，分析出系统应具备全方位的监控功能、友好的交互功能以及安全性，其中全方位的监控功能可以判断居住环境、身体健康、日常行为中是否存在异常；最后介绍分布式智能监控系统的总体设计框架，并分别介绍了系统中监控平台、监控节点、监控终端、Web 交互四大组成部分的设计。

第四章 系统中行为识别算法的改进

本章主要介绍了监控平台算法分析模块中行为识别算法的改进，首先介绍了单视角下行为识别算法的改进，其次提出了一种多视角行为识别融合决策方法，最后在公共数据集上测试了改进方法的效果。

4.1 基于动态图像的 TSN 模型

基于深度学习的行为识别算法的改进可以体现在结构、输入、网络连接三方面，本文提出的改进的方法是输入层面的改进，是将动态图像作为网络的输入，以获取更多的视频帧间信息。

4.1.1 动态图像

在 2016 年 CVPR 会议上 Hakan Bilen 等^[67]首次提出了动态图像(Dynamic Image)的概念，动态图像是一种新的适用于卷积神经网络的视频压缩表示方法，可以捕捉视频中的时空动态演化信息。其是通过在视频的原始图像像素上直接应用 ranking pooling 的方法获取，每个视频可以生成一个 RGB 图像。

动态图像生成的思想是，将视频中的帧 I_1, I_2, \dots, I_t 用 ranking 函数表示， $\psi(I_t) \in R^d$ 表示从视频中每个独立帧 I_t 中提取的表示或特征集合， V_t 代表 t 时刻内特征的均值， V_t 可以表示为

$$V_t = \frac{1}{t} \sum_{\tau=1}^t \psi(I_\tau) \quad (4-1)$$

从式(4-1)可知每次的得分与 I_t 相关。排序函数与 t 时刻的分数相关联，t 时刻分数可以表示为

$$S(t|d) = \langle d, V_t \rangle \quad (4-2)$$

式(4-2)中 d 代表经过训练的参数向量，因此 t 时刻的分数可以反映视频帧中的排名。时间越往后，分数越高，假设 $q > t$ ，则有

$$S(q|d) > S(t|d) \quad (4-3)$$

利用 RankSVM 可以将 d 看为凸优化的问题，用公式表示为

$$\begin{aligned} d^* &= \rho(I_1, \dots, I_T; \psi) = \arg \min_d E(d), \\ E(d) &= \frac{\lambda}{2} \|d\|^2 + \frac{2}{T(T-1)} \\ &\quad \times \sum_{q>t} \max \{0, 1 - S(q|d) + S(t|d)\} \end{aligned} \quad (4-4)$$

$E(d)$ 中第一项是 SVM 中常用的二次正则化器, 第二项是计算有多少对 $q>t$ 得分函数排序不正确的情况。需要注意的是, 分数之间至少有一个单位距才被认为是正确的排序, 用公式表示为

$$S(q|d) > S(t|d) + 1 \quad (4-5)$$

在提出的动态图像中函数 $\psi(I_t)$ 是一个运算符负责将图像 I_t 的 RGB 分量叠加在一个更大的向量上。在所有情况下, 描述符 d 是一个实向量, 它具有与单个视频帧相同的元素数量。因此, d 可以被解释为一个标准的 RGB 图像。此外, 由于该图像是通过对视频帧进行排序得到的, 因此它总结了整个视频序列的信息。

计算动态图像的问题实质是计算公式(4-4), Hakan Bilen 提出 approximate rank pooling(ARP)的概念, 对动态图像生成进行优化。ARP 的推导是基于考虑等式(4-4)梯度优化中第一步的思想。从 $d = \bar{0}$ 开始, 通过梯度下降得到的第一个近似解是

$$d^* = \bar{0} - \eta \nabla E(d)|_{d=\bar{0}} \propto -\nabla E(d)|_{d=\bar{0}} \quad (4-6)$$

其中 $\eta > 0$, $\nabla E(\bar{0})$ 可以表示为

$$\begin{aligned} \nabla E(\bar{0}) &\propto \sum_{q>t} \nabla \max \{0, 1 - S(q|d) + S(t|d)\}|_{d=\bar{0}} \\ &= \sum_{q>t} \nabla \langle d, V_t - V_q \rangle = \sum_{q>t} V_t - V_q \end{aligned} \quad (4-7)$$

因而 d^* 可以近似为

$$d^* \propto \sum_{q>t} V_q - V_t = \sum_{t=1}^T \beta_t V_t \quad (4-8)$$

其中 β_t 为标量系数。展开加和为

$$\begin{aligned} \sum_{q>t} V_q - V_t &= (V_2 - V_1) \\ &\quad + (V_3 - V_1) + (V_3 - V_2) \\ &\quad + (V_T - V_1) + (V_T - V_2) + \dots + (V_T - V_{T-1}) \end{aligned} \quad (4-9)$$

利用 t 表示 β_t 为

$$\beta_t = (t-1) - (T-t) = 2t - T - 1 \quad (4-10)$$

V_t 用特征集合 ψ_t 表示，进而 d^* 可以表示成 ψ_t 的线性组合

$$d^* \propto \beta_t V_t = \alpha_t \psi(I_t) \quad (4-11)$$

其中 α_t 为

$$\alpha_t = 2(T-t+1) - (T+1)(H_t - H) \quad (4-12)$$

在式(4-12)中 H_t 为

$$H_t = \sum_{i=1}^t \frac{1}{t} \quad (4-13)$$

表示第 t 次谐波，并且 $H_0 = 0$ 。 α_t 的值对于给定的视频长度(t)是恒定的，因此不依赖于视频的内容。由公式(4-13)可知，在总和中 ψ_t 出现 $\sum_{i=t}^T \beta_i H_i$ 次。Ranking pooling 操作符降低为

$$\hat{\rho}(I_1, \dots, I_T; \psi) = \sum_{t=1}^T \alpha_t \psi(I_t) \quad (4-14)$$

Rank pooling 的另外一种构造是不计算中间平均值 V_t

$$V_t = \left(\frac{1}{t} \right) \sum_{q=1}^T \psi(I_q) \quad (4-15)$$

在排名分数中直接使用独立的视频特征 $\psi(I_t)$ ，在这种情况下，上面的推导结果的权重函数为

$$\alpha_t = 2t - T - 1 \quad (4-16)$$

动态图像有三个优势，第一个优势是生成的新的 RGB 图像可以由 CNN 架构处

理，其结构与静态图像的结构相同，仍旧可以使用标准的 CNN 架构从视频中学习合适的“动态”特性。第二个优势是动态图像的提取极其简单和高效，它可以让视频分类简化为使用标准 CNN 架构对单个图像进行分类。第三个优势在于压缩因子，因为整个视频是由相当于一帧的数据量来概括的。

动态图像的生成效果如图 4.1 所示，通过对效果图的观察可知，动态图像的焦点集中在动作对象上，背景往往被忽视掉了。这表明动态图像包含了进行动作识别所必需的信息。

同时，可以观察到动态图像对于不同速度的动作表现不同。对于缓慢的动作会在图像被模糊，比如图 4.1 中第一排的“吹干头发”；对于快速的动作，动态图像反映了动作中的关键步骤，如图 4.1 中第二行第一列中的准备挥杆和挥杆后停止的动作；对于速度更快的动作，动态图像反映了视频的不同部分，如图 4.1 中第三行第一列作为观看者出现的画面被叠加在马匹和骑马人的顶部。虽然动态图像可以捕捉到时间和运动的信息，但动态图像中的像素重排会呈现类似平移、模糊的效果。

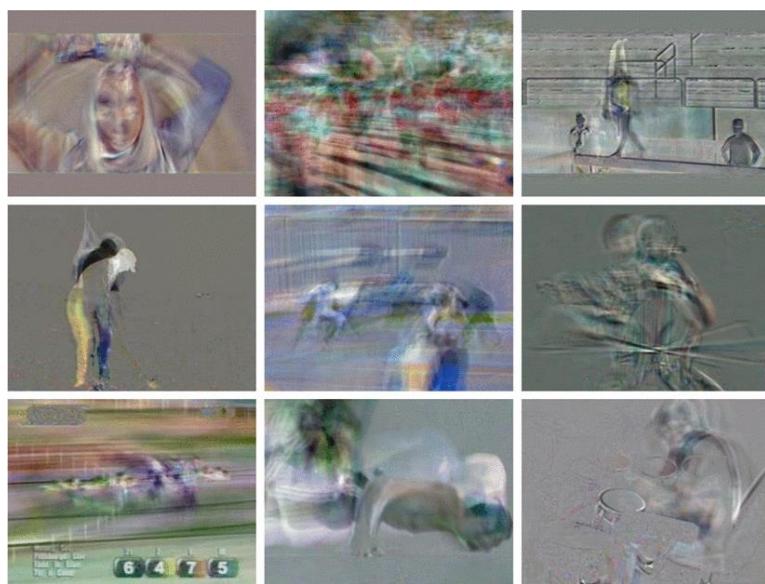
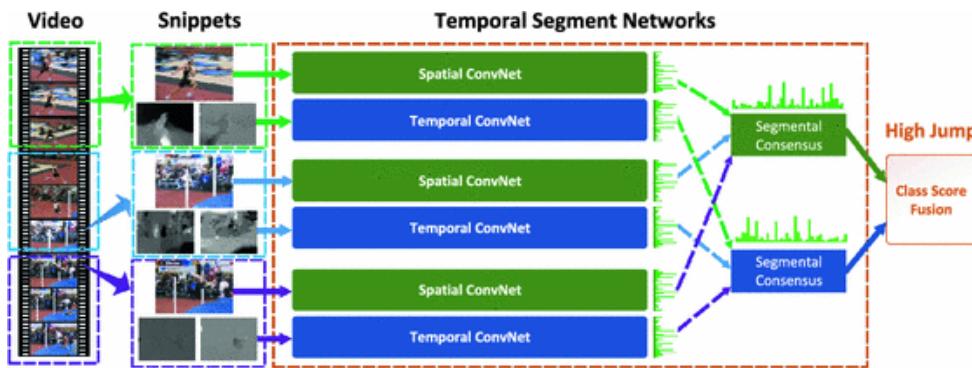


图4.1 动态图像效果图^[67]

4.1.2 TSN 模型

香港中文大学多媒体实验室 Wang 等在 2016 年提出 TSN 模型^[68]，TSN 模型是双流结构的改进，主要是解决长时间视频判断以及数据量少的问题。普通的双流结构是学习单帧或者短视频，而 TSN 模型输入的是两个光流特征图和一帧 RGB 图像，并且在做抽帧图像时运用的是稀疏的抽帧方式，运用稀疏采样是因为 TSN 模型考虑到了视频间的冗余情况。

图4.2 TSN 模型原理图^[68]

TSN 模型的原理如图 4.2 所示，具体实现过程是首先把视频分成 k 段，用 (S_1, S_2, \dots, S_k) 表示每个片段均匀地随机采样一个视频片段，用 (T_1, T_2, \dots, T_k) 表示输入包括 RGB 图像和两个光流特征图的 k 个片段；其次使用双流网络得到视频片段属于各类的得分；之后把不同片段得分取平均；最后通过 softmax 层输出。

TSN 模型可以用公式表达为

$$TSN(T_1, T_2, \dots, T_k) = H(G(F(T_1; W), F(T_2; W), \dots, F(T_k; W))) \quad (4-17)$$

式(4-17)中的 W 是指网络的参数， $F(T_i; W)$ 是指网络的输出代表该片段属于每个类的得分，即图 4.2 中空间流神经网络或光流神经网络的输出结果。 G 是一个段共识函数用以对所有片段中属于同一类别的得分做均值。预测函数 H 根据得分算概率，概率最高的类别就是该视频所属的类别，文献[48]中运用的是 softmax 函数。

在 TSN 模型中，网络损失函数用公式表达为

$$L(y, G) = -\sum_{i=1}^C y_i \left(G_i - \log \sum_{j=1}^C \exp G_j \right) \quad (4-18)$$

式(4-18)中 C 为行为总类别数， y_i 是 i 分类的基本事实标签。在 TSN 模型中段共识函数采用最简单的形式，用公式表达为

$$G_i = g(F(T_1), F(T_2), \dots, F(T_k)) \quad (4-19)$$

具体是采用聚合函数 g 从所有片段中相同类别的得分中推断出某个类别函数 G_i ，聚合函数 g 采用均匀平均法来表示最终识别精度。

TSN 模型允许利用多个片段与标准的反向传播算法共同优化模型参数 W 。在反向传播过程中，模型参数 W 关于损失值 L 的梯度用公式表示为

$$\frac{\partial L(y, G)}{\partial W} = \frac{\partial L}{\partial G} \sum_{k=1}^K \frac{\partial G}{\partial F(T_k)} \frac{\partial F(T_k)}{\partial W} \quad (4-20)$$

式(4-20) 中的 K 即为 TSN 模型中使用段数。

4.1.3 基于动态图像的 TSN 模型

文献[68]的训练结果显示 RGB 图像和 RGB 差异的结合可以将准确率提升，而动态图像可以包含更多的动作信息，因此本文提出基于动态图像的 TSN 模型 (Temporal segment networks based on Dynamic Image, DP-TSN)。

生成动态图像有两种方式，一种是将一整段视频生成一张动态图像，另一种是一段视频平均分成 k 段，每一段生成一张动态图像。DP-TSN 模型是将视频平均分成 K 段，每一段生成动态图像，将生成的动态图像作为空间流卷积网络的输入，光流卷积网络的输入是由 TVL1 光流算法提取的光流。

训练模型选择 ResNet，ResNet 运用残差的思想解决了随着网络加深准确率下降的问题。其比 VGG 网络深 8 倍，但也获得了更好的性能，自 2015 年被提出以来，广泛应用在检测、分割、识别等领域。DP-TSN 模型中，仍旧以 Softmax 函数作为预测函数。算法的框架如图 4.3 所示。

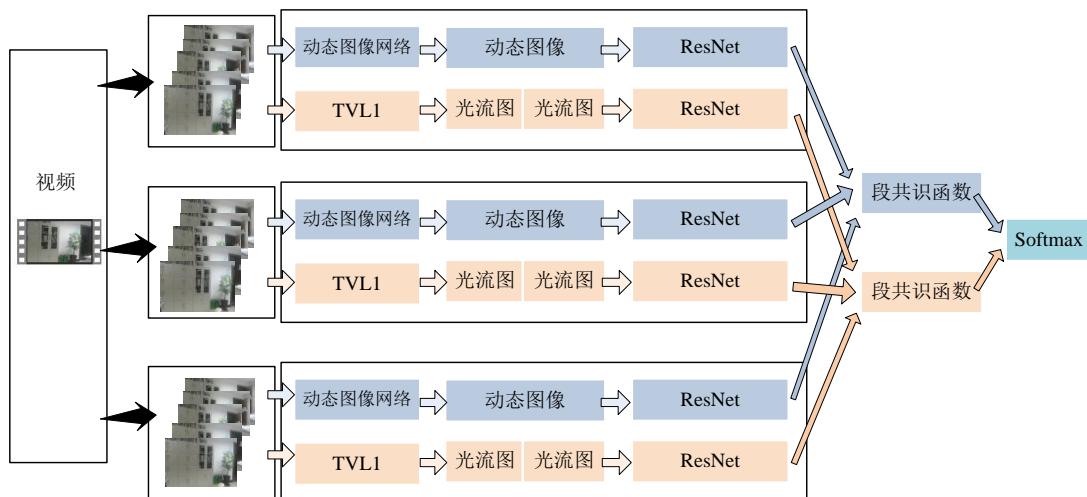


图4.3 DP-TSN 模型框架图

提出的 DP-TSN 模型相当于将之前在分段中随意选取一帧 RGB 图像的输入换成将分段信息压缩成动态图像的输入。算法明显的优势是动态图像获取更多的帧间信息，弥补了原始 RGB 图像无法包含帧间信息的不足。但是将动态图像替代原始图像 RGB 图像也损失了部分原始图像的纹理特征和外观特征。

4.2 多视角融合决策

提出的多视角下行为识别的算法是在提出的 DP-TSN 模型上做融合决策，既分别判断某个监控节点的行为异常概率，最后根据结果做融合决策。

当系统中的某个监控节点存在异常的可能性时，监控节点向监控平台发送指令与保存的视频数据，监控平台同时请求关联节点保存的视频数据，不同视角下的视频数据通过提出的 DP-TSN 模型判断出所属分类的得分之后再进行融合决策，对于综合决策本文采用最值融合策略来解决。

一般的投票方式即各部分求和统计概率最大的那个即可，这种方式简单明了，但是弊端明显。在模式识别中，各个部分求得的概率均不同，应该每个部分都参与投票考虑，但是更加侧重概率更高的结果。因此，结合相关文献，多视角融合算法当中决定采用的结合 Borda 投票策略。

4.2.1 Borda 投票

Borda 投票法是由法国数学家波达提出的，通常我们的投票方式是选民将自己的 一票投给某一候选人，但波达指出多数投票制度只适合两个候选者的情况，因此提出的波达技术法是对每一位候选人计数排序^[69]。Borda 投票通常的使用方式是每个投票者对不同的候选人进行打分并排序，排名靠前得高分并依次减少，然后计算所有投票者的 Borda 得分数，得分最高的候选者在竞选中中获胜。Borda 投票也可以采取排名靠前得低分，最后统计最低分的情况。

我们在对所有动作分类的时候，需要辨别的是多种动作，并且机器不会存在作弊的情况，所以适合运用 Borda 投票的方法。在多视角行为识别过程中首先获取每个视角下独立的状态概率，并将它们根据次序排好权重后输入到 Borda 算法，通过 Borda 算法规则计算出多视角下的综合结果。在本文设计的系统中是指被监控的对象的行为动作。Borda 算法的优点是快捷简便、易于实现，并且能最大化的体现最终结果。

4.2.2 结合 Broda 的融合决策

在 Broda 投票中每项的得分是确定最后确定排名的关键，同时每个选项的 Borda 分数是相对于所有可能的 Borda 分数确定的。在 Broda 投票中可以对所有的结果进行排序，然后在汇总之前对这些得分结果进行加权，某一结果获得的评价越高，该结果就会得到越多的分数，具体地说，在行为识别结果排名中，每一个排名低于其他结果就会获得一分，总分最多的选项成为最后的结果。

在 TSN 模型中，各个视角下通过 Softmax 函数都得到所有分类的概率值，然后对各个视角下的结果融合 Borda 投票机制进行决策。假设共有四个类别，首先，对每

个视角下得到的 Softmax 函数概率值结果进行排名。其次，对每个视角类别按照排名得到权重，则 1-4 名的权重依次为 4、3、2、1，并计算类别与权重的乘积。然后，对不同视角下相同类别的得分进行叠加。如在视角 1 中 A-D 四类的概率得分分别为 0.45、0.15、0.10、0.30，则根据 Borda 决策对应权重分别为 4、2、1、3；视角 2 下 A-D 四类的得分分别为 0.40、0.05、0.00、0.55，则根据 Borda 决策对应权重分别为 3、2、1、4。最后，求和 A-D 类的得分分别为 3、0.4、0.1、3.1，判断得分最高的为对应类别，在假设中即为类别 D。

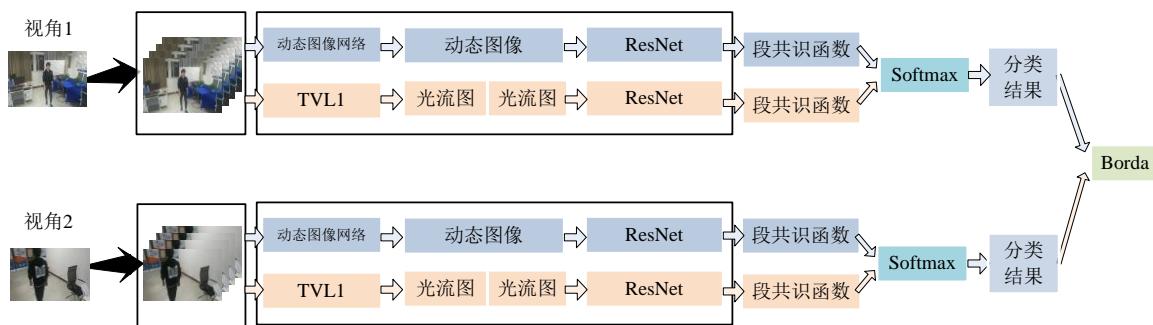


图4.4 结合 Borda 决策的多视角行为识别框架图

结合 Borda 的融合决策计算简单，易于实现，并且可以根据不同的权重值将得分之间的间距拉大，减少得分相同的情况。以假设来看，尽管在类别较少的情况下，也没有出现得分一致的情况，在类别较多情况下，出现的得分相同导致无法判断类别的情况会进一步降低。

4.3 实验和分析

1. 动态图像生成

动态图像以中山大学的 SYSU 数据集为示例，在数据集中选取三段相同动作的视频数据，将视频段分别通过动态网络生成的动态图像分别如图 4.4 所示。图 4.4 中的三幅图是三个不同的人完成扫地的动作。



图4.5 生成的动态图像示意图

2. 网络预训练

本文在对提出的 DP-TSN 模型进行训练时，选用的是公共数据集 UWA 3D Multiview Activity II Database（UWA3D II）中多视角数据的部分数据。UWA3D II 数据集是具有四个视角、30 种动作的数据集，但是训练和测试数据集选用了视频中的两个视角下的 20 种动作。

在 Windows 10 平台上进行 Open CV、Caffe 编译，编写网络配置文件，通过训练获得网络参数，生成 Caffe 模型。训练参数如表 4.1 所示。

表4.1 网络训练参数

参数名	初始学习率	学习策略	步系数	StepValue	最大迭代次数	梯度下降动量	权重衰减	训练方式
参数	0.001	Step	0.5	50000	20000	0.9	0.0005	GPU

3. 结果与分析

运用剩余的数据集在训练好的模型中测试提出的 DP-TSN 模型和结合 Borda 融合决策的多视角行为识别。测试提出的 DP-TSN 模型是将测试数据输入模型，得到准确率；测试结合 Borda 融合决策的多视角行为识别是将测试数据分两个视角分别通过网络再进行决策融合。测试得到准确率分别为 82.74% 和 86.44%，混淆矩阵如图 4.6、图 4.7 所示。

对比传统方法 80.50% 的正确率，由于 DP-TSN 模型运用动态图像可以获取视频帧间信息，因此在正确率上有 2.24% 提升。对比图 4.6、图 4.7 可知，在多视角情况下对 DP-TSN 模型结合 Borda 投票进行融合决策，可以得到比单视角情况下更高的正确率。

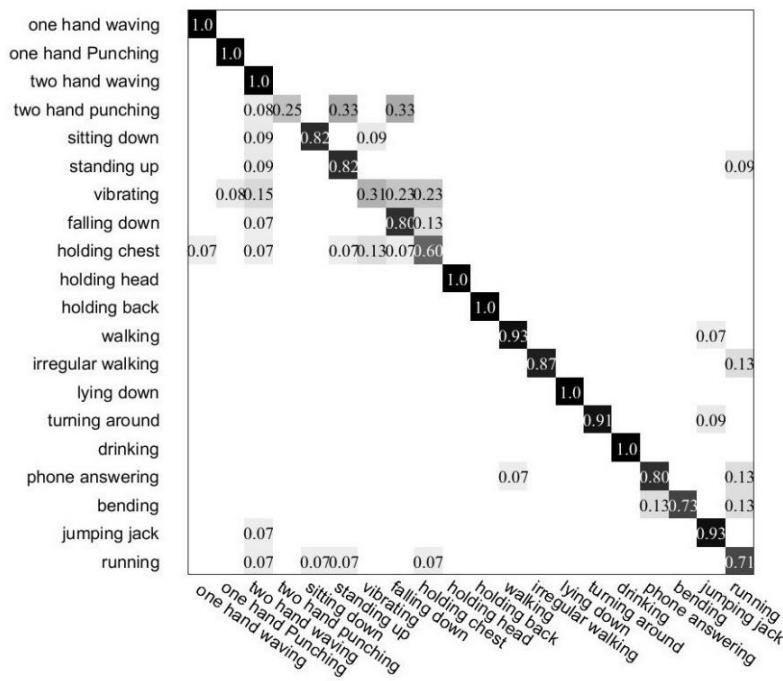


图4.6 DP-TSN 模型数据测试图

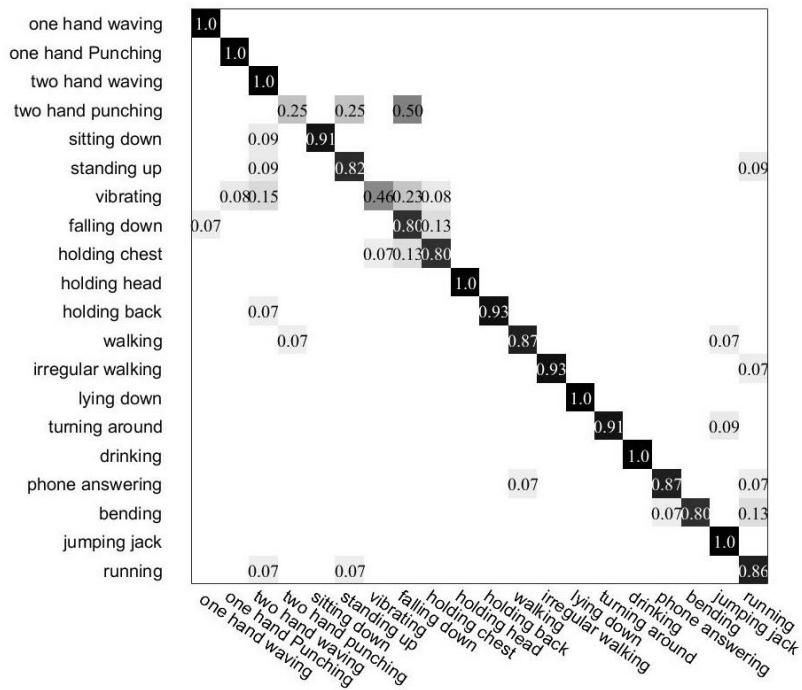


图4.7 结合 Borda 决策的多视角行为识别数据测试图

4.4 本章小结

本章主要是对云服务器算法分析模块中的行为识别算法进行改进。首先，提出了基于 TSN 模型改进的行为识别方法，改进模型融合了动态图像的概念，即将生成的

动态图像作为 TSN 模型中的 RGB 图像输入，弥补原模型中帧间信息不足；其次，介绍了提出的多视角行为识别算法，具体是提出了融合 Broda 决策的多视角行为识别，提出的融合决策具有快捷简单、易于实现的优势；最后，在 Windows 平台上测试了提出模型的效果，运用公共数据数据集 UWA3D II 分别测试了提出的 DP-TSN 模型和结合 Borda 融合决策的多视角行为识别方法。

第五章 分布式智能监控系统实现

本章介绍的主要内容是智能监控系统中各部分的实现。首先分别描述了监控终端、监控节点、监控平台、Web 页面的实现过程，然后是对系统进行了数据采集测试和功能测试。

5.1 监控终端实现

监控终端部分的开发是在树莓派平台上对传感器进行驱动开发，并封装传感器接口。

1. 视觉传感器驱动开发

通常，摄像头通过 V4L2 (Video for Linux two) 框架编写驱动代码，V4L2 框架主要的作用是使程序有发现设备和操作设备的能力，其通过回调函数来实现数据采集过程中设置的功能，常用的设置功能包括摄像头的频率、视频压缩格式、图像参数、帧频等。然而，系统实现过程中运用的摄像头为免驱摄像头，则不需要编写摄像头驱动程序。

2. 环境传感器开发

温湿度传感器采用的 DHT22 模块，DHT22 温湿度传感器为单总线驱动，驱动函数作用一方面是对其时序的控制，另一方面是对外提供读温度、读湿度的接口。

烟雾传感器采用的是 MQ-2 模块，通过 MQ-2 烟雾传感器的 DO 接口可以获取是否检测到烟雾。环境传感器与树莓派监控节点连接的硬件电路图如图 5.1 所示。

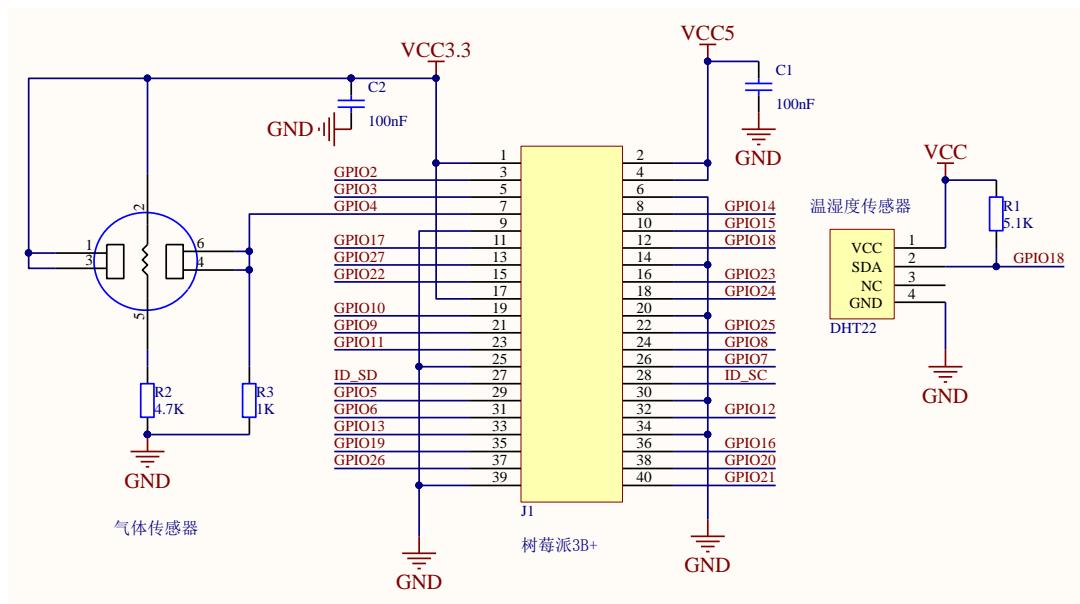


图5.1 环境传感器硬件连接图

为了测试不同的 linux 驱动框架，其中温湿度传感器选用的是杂项设备框架，烟雾传感器选用的是字符设备框架，杂项设备框架实质上字符设备框架的封装，不同之处在于杂项设备框架中将主设备号固定为 10。

5.2 监控节点实现

5.2.1 树莓派监控节点实现

1. 节点安全实现

安全策略的目标是保护数据的同时维护用户隐私的完好，策略围绕对系统的访问能力、对系统上安装软件的权限、数据的访问权限、从故障中恢复的方法进行制定。监控节点使用数据安全性策略、访问安全性策略实现安全保障^[70]。

数据安全性策略通过文件访问权限和数据加密两种方式实现。文件访问权限是利用访问控制列表实现，系统存在 Root 账户和使用者账户，Root 账户访问系统文件以及使用者账户文件，而使用者账户仅可以访问数据文件夹；数据加密是使用 Linux 统一密钥设置（Linux Unified Key Setup, LUKS）加密磁盘。

访问安全性策略是通过基于密钥加强 SSH 远程访问、防止地址伪装实现。基于密钥加强 SSH 远程访问是将公钥文件复制到远程 SSH 服务器上，客户端使用 SSH 连接服务器时将采用基于密钥的认证方法，如果任一密钥文件缺失或不匹配会拒绝登录 SSH 服务器；树莓派监控节点在设置防火墙的基础上阻止地址伪装，地址伪装是攻击者使用虚假 IP 地址向服务器发送恶意数据包时经常使用的一种技术。

2. 数据采集

环境数据采集是设备安装传感器驱动之后，在监控节点主程序中，以每分钟 1 次的频率调用温湿度传感器、气体传感器数据读取接口，并把获取到的数据写入文件“/home/usr_data/data.txt”。

采集视频数据可以根据底层 V4L2 驱动框架的流程来写视频数据获取流程，也可以采用开源项目 MJPG-streamer 获取视频数据，MJPG-streamer 项目中也是根据 V4L2 框架获取视频数据。为了方便与主程序的结合，课题选择的是前一种方案，并以每秒 25 帧的频率获取的图片信息存储在目录“/home/usr_data/image”下，保存的视频时长为 2 分钟。当树莓派节点判断可能发生意外情况时，先将目录“/home/usr_data/image”下的数据发送给云平台，再发送实时视频流。

3. 数据传输

为了实现更高效、更准确的系统效果，监控节点发送精简的有用信息给监控平台，因此采用简单的机制控制数据传输流程。针对非异常信息采用时间机制，当数据与上次数据传输时间间隔超过 5 分钟，监控节点主动向监控平台发送时间段内的环境和图

像数据，上传的压缩图像数据是每 10 秒间隔截取一帧。针对异常信息采用事件机制，即监控节点判断获取数据为可能异常数据后，将可能异常数据传送给监控平台，并且持续传输数据直到异常事件得到监护人的响应或者监控平台判断数据为正常时停止传输。采用响应机制，有系统用户通过客户端查询数据时，监控平台向监控节点请求数据，监控节点向上层传输视频数据，直到客户端断开链接。数据传输流程具体如图 5.2 所示。

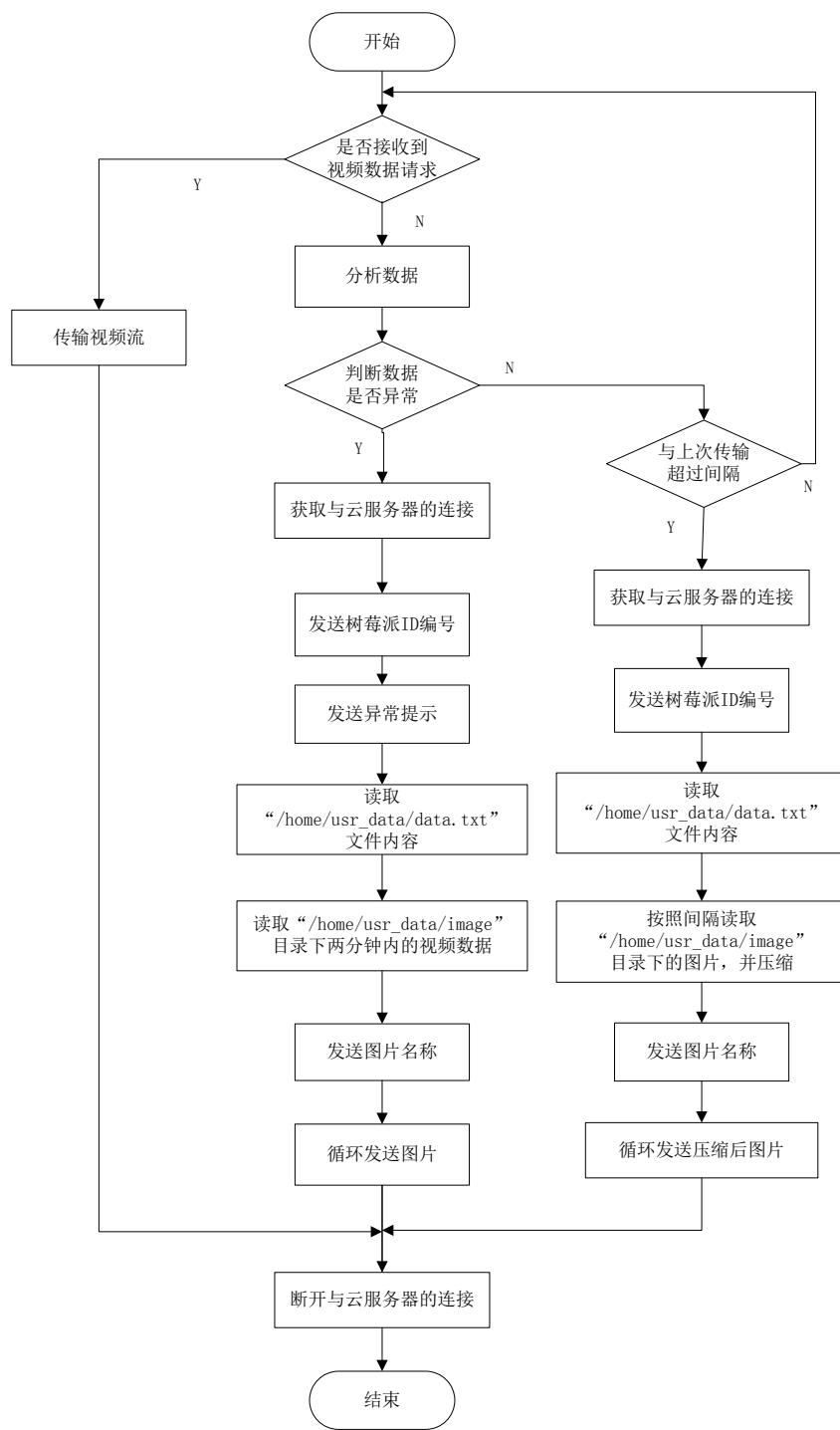


图5.2 监控节点数据上传流程图

树莓派监控节点与云服务器之间采用 Socket 通信，客户端代码运行在监控节点中，服务端代码运行在服务器端。客户端与服务端连接之后一次完整的数据发送流程是，首先向监控平台发送监控节点的 ID 编号，发送 ID 编号的目的是让监控平台区别数据的来源；其次读取 “/home/usr_data/data.txt” 文件中的温湿度信息和烟雾信息，读取完成之后依次发送给监控平台；最后监控平台读取目录 “/home/usr_data/image” 下的图片，依次传输图片名字与图片内容。

4. 生理参数分析

由智能手环获取到的关键生理参数分别为心率、睡眠质量、卡路里消耗。安静状态下，成年人正常心率为 60~100 次/分钟，理想心率应为 55~70 次/分钟。一般来说，年龄越小，心率越快，老年人心跳比年轻人慢。根据常识，在监控节点中将设置心率范围为 50~100 次/分钟，若不在这个范围内，向监控平台上报生理数据异常。人老了之后，睡眠时间会越来越少，老年人一天的总体睡眠时间是在 5~6 小时范围内都是合理的，如果某天时间少于四个小时，则睡眠数据作为异常数据上报。

5. 环境信息分析

环境信息采集的是温湿度数值和烟雾状态，温湿度的高低主要是影响居住体验，并非导致意外的主要因素，但烟雾浓度过高时会导致中毒或爆炸。因此在监控节点阶段，只将烟雾浓度作为是否意外的标注，若烟雾为存在的状态，上报异常数据。

6. 视觉信息分析实现

在监控节点端视觉分析的实现是通过传统行为识别算法 iDT 算法，将 iDT 算法移植在树莓派上，并利用获取到的视频测试。IDT 算法可以标注显示行为动作，若存在异常或判断模糊的情况则进一步上报数据。

5.2.2 可穿戴设备监控节点实现

可穿戴设备节点的访问是通过 RESTful Web 服务接口，接口是一组由统一资源标识符（Uniform Resource Identifier, URI）标识的资源组成。识别出资源之后可以应用它的操作集，并且接口为给定资源提供多种表示。可穿戴设备节点的数据获取过程根据 Fitbit 官方开发文档可以整理为以下四步。

1. 注册 Fitbit 账户，通过 Fitbit 官网注册要用的开发账号和使用的应用程序类型，考虑需求本文选用个人应用的类型。注册完成后可以获得用于后续鉴权的一个 Key 和一个 Secret。
2. 鉴权，使用 OAuth2.0 鉴权接口，验证是否有访问数据的权限。
3. 读写数据，认证之后可以调用不同的 API 接口，根据获得的权限访问用户数据。
4. 请求数据，Fitbit 以资源的形式管理数据，不同的数据存放在不同的资源目录

中。因此，读写不同类型数据的差别在对 http 发送的 get 请求中 URL 不同。

读取可穿戴设备节点的流程如图 5.3 所示。

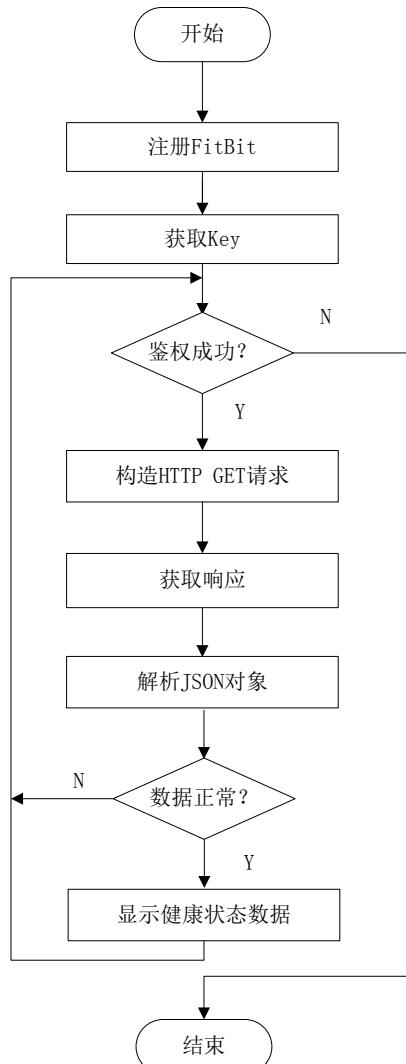


图5.3 可穿戴设备数据获取流程图

5.3 监控平台实现

5.3.1 云服务器实现

云服务器采用组件设计的监控系统软件能够具有较强的灵活性，即可满足小型视频监控应用的简单需求，又可轻松地升级为大中型以及以上级别的视频监控应用，各个模块的实现为如下内容。

1. 数据交互模块实现

服务器中代码每一个进程使用一个 UDP 连接，线程总数通过线程池进行管理和限制，通过心跳机制来保证 UDP 连接不断开。

接收监控节点上传的环境信息以及视频监控图片，监控节点有数据发送给交互模块时，服务端执行接收函数。一次数据接收流程涵盖监控节点编号接收、环境信息接收、图片数据接收。

云服务器向树莓派发送指令数据，当云服务器向监控节点发出信息采集指令或其他指令时，通过 socket 通信中的 send 函数。

服务端心跳机制，当 UDP 在设定的时长内都没有进行数据的读写时，这时候需要对其进行监测，查看客户端是否已经掉线。本文采用的方法是向树莓发送一个保存 TCP 连接的能够判断调度是否被丢弃的心跳包，从而确定连接是否正在进行。系统包含字符串“kAlive”即为心跳包。

远程升级监控节点程序，监控节点的主程序通过执行 C 语言的动态库，该动态库的名称为“client.so”。当需要升级监控节点程序时，将修改好的代码生成动态库，通过 SSH 传输给监控节点。

2. 节点管理模块实现

智能监控系统中节点之间的关联可以借鉴 ZigBee 绑定的思想，Zigbee 中绑定是指在两个节点的应用层上建立起来一条逻辑链路。Zigbee 中四种绑定方式分别是：两个节点分别通过按键机制调用函数、Match 方式、ZDP_BindReq 方式、手工管理绑定表。智能监控系统的节点关联借鉴其中的 Match 方式和手工管理绑定方式。

手工管理绑定是指事先知道绑定的节点信息通过应用程序调用绑定函数来实现手工绑定表管理，优点是自由度大、实现简单，缺点是需要知道监控节点信息、应用程序复杂。

Match 方式是一个节点通过函数允许或禁止本节点被 Match，然后另外一个节点在一定的时间内发起匹配请求，允许被 Match 的节点会响应请求，发起请求的节点在接收到请求回复之后自动处理绑定。

智能监控系统使用这两种方式进行绑定，即在测试或订制的设备中运用手工管理绑定，在普通设备中采用 Match 方式。用户可以通过 Web 交互界面操作，云服务器程序接收 Web 命令后检测关联设备的合法性，然后云服务器将关联信息写入云数据库。

节点管理模块会自动匹配同一监护对象下的不同监控节点，若是判断到会有不同关联节点同时出现同一目标，则会自动匹配节点。

3. 资源管理模块实现

云数据库提供了 IP 白名单来实现网络安全访问控制，在云数据库中将云服务器 IP 地址设置在 IP 白名单中，云服务器通过内网或外网访问云数据库，内网访问的稳定性和安全性较强。云服务器访问云对象存储可以借助于云服务器实例 RAM (Resource Access Management) 角色，把 RAM 角色 和 ECS 实例关联起来，实例

内部的应用程序就能通过 STS (Security Token Service) 临时凭证访问其他云产品。

4. 用户接口模块实现

使用云服务器搭建 LAMP 平台的操作步骤如下：准备编译环境、安装 Apache、安装 PHP 环境、安装 PHPMyAdmin，阿里平台上有具体的流程。

5. 分析算法库实现

算法部分主要实现的是与视觉传感器相关的算法。行为识别算法是移植上一章节的多视角行为算法，多视角行为算法是基于卷积网络实现的，需要首先移植深度学习框架，再将训练好的模型放置进去。

6. 通讯模块实现

通讯模块借助阿里平台提供的短信服务，支持国内和国际快速发送验证码、短信通知。通过 API/SDK 方式调用短信发送能力，将信息发送至指定手机号码。

5.3.2 云数据库实现

阿里云的云存储内核为的 MySQL 内核，数据库类型选择关系型数据库设计的关系表^[71]，为表示监护人员、监护对象、监控系统之间的关系，设计了 sys_user 表、gur_user 表、sys_node 表、data 表。

1. sys_user 表

该表主要用于用户登录系统时的身份验证，并在用户注册时保存新用户的个人信息。后台程序通过在云数据库中验证输入的账号、密码信息，确定用户是否有权限进入系统。Sys_user 表的具体设计如表 5.1 所示。

表5.1 sys_user 表设计

字段名	类型	长度	字段说明	备注
countName	varchar	40	账户名称	主键
passWord	varchar	40	密码	
phNum	varchar	11	联系方式	

2. gur_user 表

将每个账户下监护的目标对象存储在 gur_user 表中，以此可以达到存储不同用户对应的监控对象的目的。Gur_user 表的具体设计如表 5.2 所示。

表5.2 gur_user 表设计

字段名	类型	长度	字段说明	备注
countName	varchar	40	账户名称	主键
objName	varchar	30	监护目标对象	

3. sys_node 表

在监控系统中一个被监护对象有不同监控节点，每个监控对象拥有的监控节点所关系保存在 sys_node 表。Sys_node 表中相同的 key 值，表示节点关联，具体设计如表 5.3 所示。

表5.3 sys_node 表设计

字段名	类型	长度	字段说明	备注
objName	varchar	40	监护目标对象	主键
nodeID	varchar	15	监控节点编号 ID	
key	int	4	节点匹配关系	

4. data 表

Data 表用来存储树莓派监控节点上传的信息，其中图片存储的是图片名称，需要使用图片时从云对象存储服务器中获取。Data 表具体设计如表 5.4 所示。

表5.4 data 表设计

字段名	类型	长度	字段说明	备注
nodeID	varchar	15	监控节点编号 ID	主键
tem	float	4	温度数值	
hum	float	4	湿度数值	
gasCon	varchar	4	气体状态	
photo	varchar	30	图片名	
date	datetime	6	数据上传时间	

5.3.3 云对象存储实现

通过阿里云控制台，创建一个特定名称的 Bucket 列表，每一个列表相当于是一个容器，在存储容器中对每一个监控节点新建一个文件夹，即每一个监控节点拥有对应的文件夹。

通过阿里云官方接口，可以实现图片在云对象存储中的上传和下载。上传的核心代码为 oss_put_object_from_file(oss_request_options_t *, aos_string_t*, aos_string_t*, aos_string_t* , aos_table_t * , aos_table_t *)；下载的核心代码为 oss_get_object_to_file(oss_client_options, &bucket, &object, headers, params, &file, &resp_headers);

5.4 Web 页面实现

Web 页面开发语言为 html5、JavaScript、CSS。Html5 嵌入了视频、图形的特性符合监控系统显示视频监控、图表分析的需求；CSS 是一种描述 html 文档样式的语言，负责 html 元素的显示；JavaScript 是一种灵活的动态脚本语言，可在客户端浏览器中执行。网页可以分为结构层、表示层、行为层三个层次实现，在设计的智能监控系统中结构层是由 html5 负责创建，主要用于描述网页内容；表示层由 CSS 负责创建，主要是控制如何显示有关内容；JavaScript 连接前端和后端服务器，负责对事件做出反应。

网页与后台的通信选用基于 Ajax 的长轮询，Ajax 全称为 Asynchronous Javascript And XML，其是指一种创建交互式网页应用的网页开发技术。智能监控系统使用者从客户端发起数据请求，云服务器接受客户端请求会检查是否存在新数据需要响应，若不存在 Ajax 进入阻塞等待状态；若存在，Ajax 通过回调函数处理信息并给云服务器发送下一个请求。通信方式如图 5.4 所示。运用 Ajax 实现长轮询的优势是数据实时性高、可以减少数据请求资源。

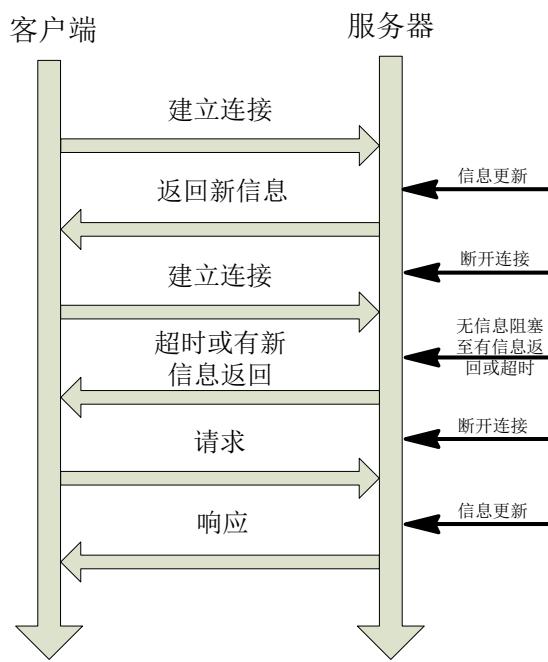


图5.4 客户端与服务器交互示意图

根据系统功能与交互需求，网页由监控界面、监控节点操作、数据分析、信息管理四个组成部分，集中展示了系统中监控平台、监控节点、监控终端的功能。

监控界面是系统整体实时数据的显示，显示的实时数据包括视频数据、传感器数据、健康状态数据。其中视频数据可以根据按钮选择不同监控节点的视频数据；其中传感器数据以不同的监控节点划分显示，主要有温湿度数据、气体状态数据、天气；

其中健康状态数据通过可穿戴设备节点获取，主要有心率和睡眠时间。界面左边为系统的界面导航。

监控节点操作页面是通过用户交互实现监控系统中树莓派监控节点的绑定和升级功能。

数据分析界面是显示环境数据、老年人生理状态数据、行为数据变化趋势，以及现状综合分析结果。

信息管理页面是通过用户交互显示用户信息和监控对象信息，用户信息即为系统使用者的信息，包括系统登录必备信息、个人联系信息、监护对象信息。监护对象信息为被监护老年人的个人信息和使用的监控节点。

5.5 系统测试和分析

将智能化功能分别加于监控平台、监控节点来实现，即监控系统上下层相结合的实现方式。对比将智能化功能加于系统前端实现或系统后端实现，这是一种折中的方式，也是较为合理和平衡的智能化监控系统实现方式。一方面，可利用监控节点富余的 CPU 资源实现部分简单而基本的智能功能；另一方面，监控平台集中资源实现更高端的或者更面向高层应用的智能化功能。将智能化功能均匀分布在监控系统的各个环节，才能达到既高效又平衡的高性价比的效果，才能真正将智能化监控的优势提高到一个新的层次。

5.5.1 系统数据测试

1. 环境数据测试

环境数据测试包括温湿度测试与气体浓度测试。温湿度通过高精度温湿度计作为标准值，对比得到采集数据的准确率。表 5.5、表 5.6 分别表示了采集到的温度数据、湿度数据。

表5.5 温度数据

测试数据	1	2	3	4	5	6	7	8	9	10
测试值(°C)	17.5	14.4	16.6	16	13.5	17	18	18.2	20.3	25
标准值(°C)	17.6	14	16.7	16	13.4	17.3	18	18.1	20.5	24.5
误差(%)	0.5	2.8	0.6	0	0.7	1.7	0	0.5	0.9	2.0

表5.6 湿度数据

测试数据	1	2	3	4	5	6	7	8	9	10
测试值(%RH)	61.2	70.1	45.7	44.8	54.1	52.1	50.5	62.7	58.9	44.8

续表 5.6

标准值(%RH)	60.4	68.6	48.2	46.1	54.4	52.6	51.1	59.7	58.4	48.5
误差(%)	1	2	5	3	1	1	1	5	1	8

通过监控节点采集的测试值与标准值对比可知，温度误差在 0%~2.8% 之间，湿度误差在 1%~8% 之间。

不同于温湿度测试的定量分析，烟雾传感器通过在传感器周围释放少量气体，进行定性分析。气体测试数据如表 5.7 所示。

表5.7 烟雾传感器测试数据

测试数据	1	2	3	4	5	6	7	8	9	10
烟雾测试值	有	有	无	无	有	有	有	无	无	无
真实情况	有	有	无	无	有	有	有	无	无	无
误差	无	无	无	无	无	无	无	无	无	无

综合上文分析，通过监控节点测试的环境数据与标准值相差较小，因此认为监控节点的环境采集功能正常；

2. 健康数据测试

可穿戴设备的健康数据测试，主要是通过对比获取数据与可穿戴设备自带 APP，得到获取数据是否存在有误的情况。因为没有真实的设备，健康数据的测试来自于 Fitbit 提供的模拟器，得到的数据格式如图 5.5 所示。

```
{
  "badges": [
    {
      "badgeType": "DAILY_FLOORS",
      "dateTime": "2019-01-27",
      "image50px": "http://www.fitbit.com/images/dash/badge_daily_floors10.png",
      "image75px": "http://www.fitbit.com/images/dash/75px/badge_daily_floors10.png",
      "timesAchieved": 3,
      "value": 10
    },
    {
      "badgeType": "DAILY_STEPS",
      "dateTime": "2019-01-27",
      "image50px": "http://www.fitbit.com/images/dash/badge_daily_steps5k.png",
      "image75px": "http://www.fitbit.com/images/dash/75px/badge_daily_steps5k.png",
      "timesAchieved": 6,
      "value": 5000
    },
    ...
  ],
  1
}
```

图5.5 可穿戴设备获取的部分数据图

5.5.2 系统功能测试

1. 用户注册测试

用户注册的对象是老年人的监护人，注册界面如图 5.6 所示。通过测试可知，系统注册正常。

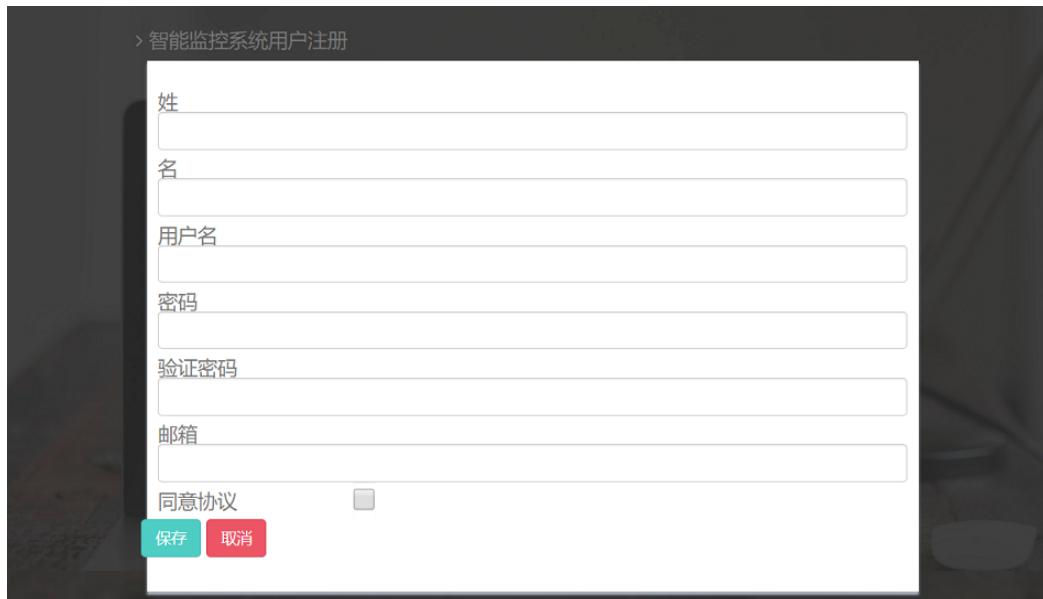


图5.6 系统注册界面

2. 用户登录测试

用户登录主要用以老年人的监护人登录系统，界面如图 5.7 所示，测试可知用户登录功能正常。

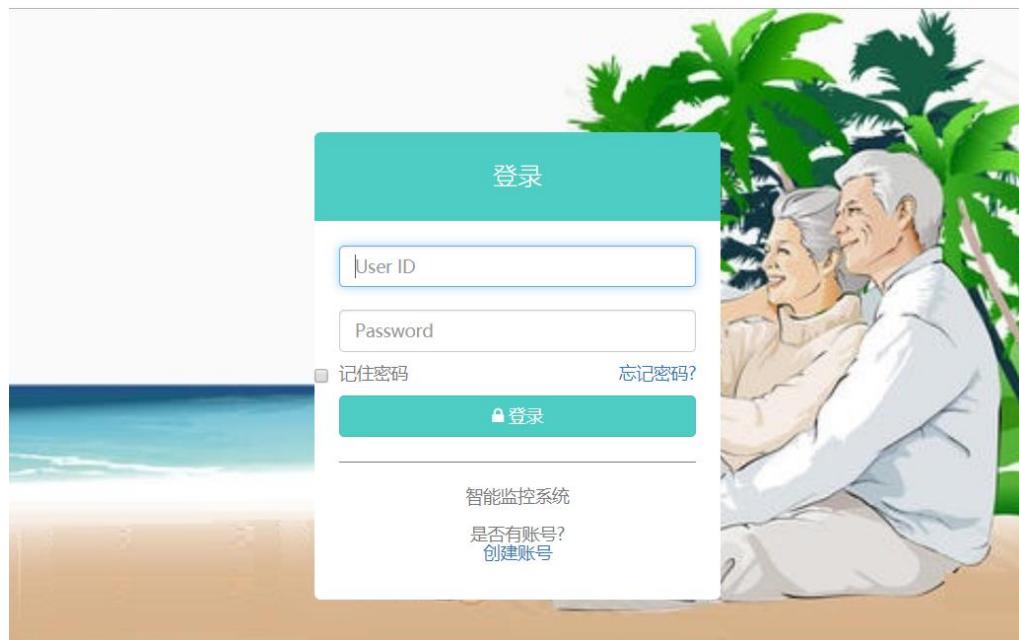


图5.7 系统登录界面

3. 监控界面测试

监控界面测试是测试数据在网页上的显示是否正确，测试示意图如图 5.8 所示。

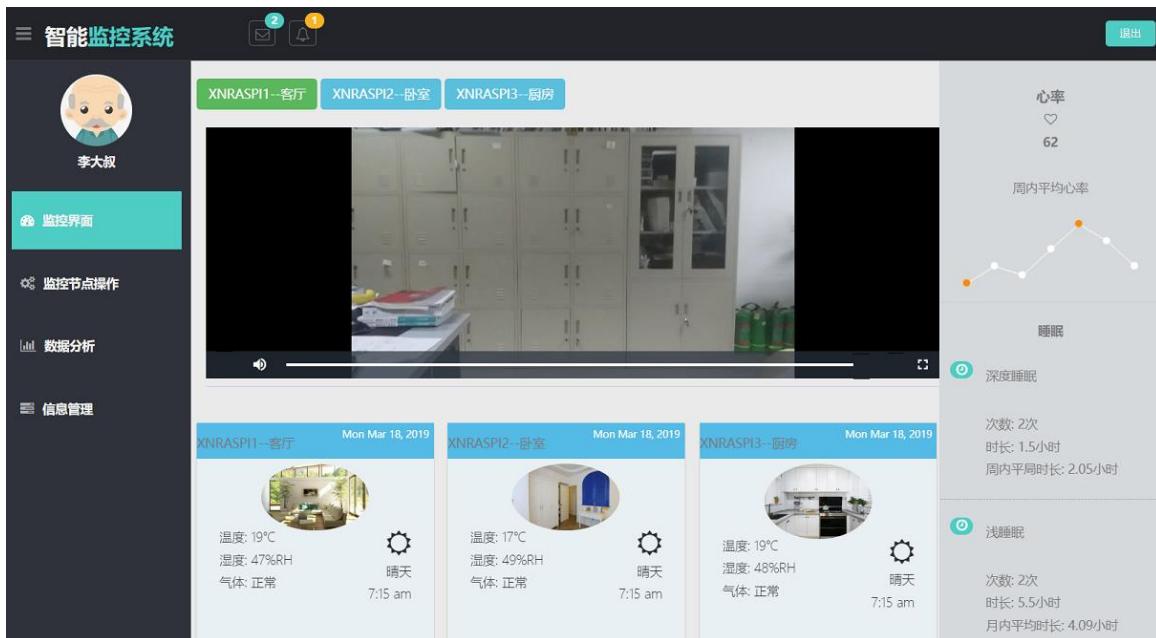


图5.8 监控界面示意图

4. 监护功能测试

监护功能测试是测试老人发生意外时是否有警报提示，以视频监控到跌倒为例，



图5.9 异常发生时监控视频图



图5.10 系统短信预警图

监控界面显示图如图 5.9 所示，短信预警信息如图 5.10 所示。

5. 监控节点操作测试

监控节点操作主要是针对树莓派监控节点的节点绑定和节点升级操作，节点绑定操作如图 5.11 所示，节点升级操作如图 5.12 所示。

设备ID	位置	存在的绑定关系	增加绑定节点	删除绑定节点
XNRASPI1	客厅	< XNRASPI2 >	<button>添加</button>	<button>删除</button>
XNRASPI2	卧室	< XNRASPI1 >	<button>添加</button>	<button>删除</button>
XNRASPI3	厨房	< >	<button>添加</button>	<button>删除</button>

图5.11 系统节点绑定图

图5.12 系统节点升级图

6. 数据分析测试

数据分析图示显示的分别是一周内温湿度数据、深浅睡眠时间、心率数据，在实际页面显示中这三种数据由一个变化的图表，横坐标是以星期标注的时间轴。周内温湿度数据变化如图 5.13 所示，蓝色为温度线，橙色为湿度线。

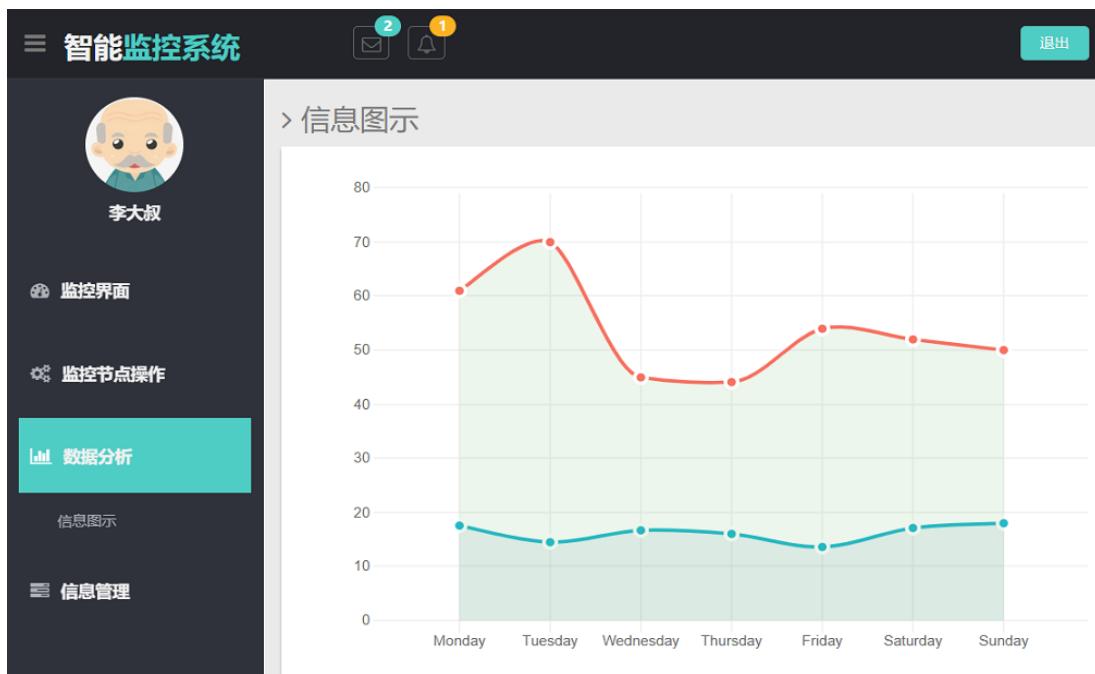


图5.13 周内温湿度变化图

周内睡眠数据如图 5.13 所示，其中蓝色为深睡眠数据，橙色为浅睡眠数据。

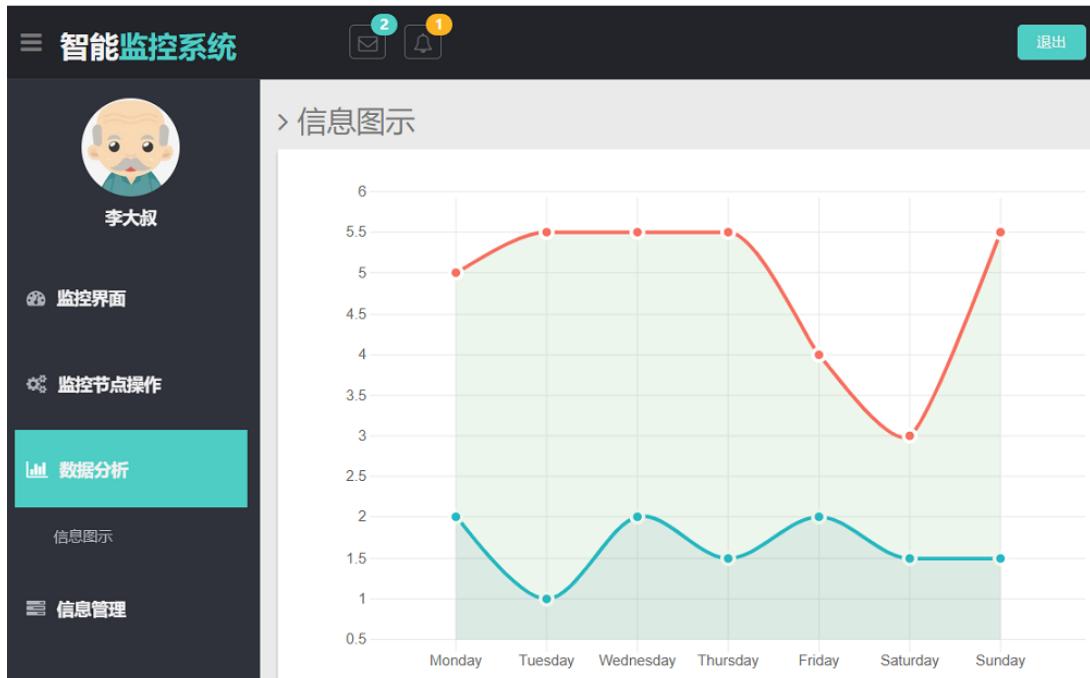


图5.14 周内睡眠时长变化图

周内心率数据如图 5.15 所示，其中纵轴数据为每天的平均心率。



图5.15 周内心率变化图

7. 用户管理测试

用户账号管理与监护对象管理都是用户管理的子功能。用户账号管理主要用以修改用户数据、增删监护对象等，如图 5.16 所示。



图5.16 用户账号管理图

监护对象管理主要是设置监护对象基本信息、增删监控节点等，如图 5.17 所示。



图5.17 监护对象管理图

5.6 本章小结

本章从组成部分介绍了面向老年人设计的分布式智能监控系统的实现细节。首先，介绍了视觉传感器、温湿度传感器、气体传感器等监控终端的开发流程；其次，介绍了树莓派监控节点数据采集、数据传输、节点安全实现的过程，并介绍了可穿戴设备监控节点信息获取的过程；再次，介绍了监控平台中三大云产品的实现过程，其中云服务器中实现了数据交互、节点管理、资源管理等六个模块；从次，介绍了基于 html5 开发的人机交互界面实现过程；最后，从系统数据和功能两方面对实现的监控系统进行测试。

第六章 总结与展望

6.1 研究内容总结

论文以老年人监护为核心，针对这一目标分别在系统和算法两方面进行研究。设计的智能监护系统从居家、养老机构等应用背景出发，采用具有良好伸缩性的分布式架构。同时，智能监护系统从实际场景出发，给予老年人在居住环境、身体状态、日常行为等方面全方位保护。为了充分利用智能监控系统中的视频信息，提升监护系统异常判断的准确程度，论文从单视角、多视角两个方面改进了行为识别算法。本文的具体研究内容总结如下。

1. 设计和实现监控系统全方位防护、用户交互的功能，并且考虑了系统安全性。其中提供的全方位防护具体体现在可以监测老年人居家活动所有区域、可能发生意外的主要因素，而不是只考虑老年人在某一部分区域下的某种意外；其中提供的用户交互功能主要通过 Web 界面交互、短信预警两种方式实现；其中提供的系统安全功能主要通过监控平台和监控节点两个层面实现。

2. 设计和实现监控系统的资源合理性、处理高效性。论文设计的智能监控系统采用分布式架构，由监控终端、监控节点、监控平台、用户交互四部分组成。设计的分布式架构可以使系统运行在居家养老等小型应用环境下，也可以使系统运行在养老中心等大型应用环境下。其中监控节点依托智能硬件或可穿戴设备承担简单的数据分析功能，节点在运用富余 CPU 资源的同时减轻了监控平台的数据分析压力。

3. 论文改进了基于 TSN 模型的卷积神经网络模型。在分析目前基于深度学习的行为识别算法基础上，思考了当前卷积神经网络模型中对视频时序信息运用不充分的问题，并提出了运用 DP-TSN 模型的解决方案。因为多帧视频图像构成的动态图像保留了较好帧间信息，所以对常用的 TSN 模型在输入层面上改进，将原始模型输入的 RGB 图像替代为动态图像。

4. 论文在多视角场景下提出了基于 Borda 的融合决策。提出的多视角下融合决策的实现是依据多个单视角的分类结果，利用 Borda 投票规则再次融合判断，在 Borda 投票中每个视角中排名靠前的分类具有更高的权重。采用 Borda 融合决策的多视角行为识别比单视角行为识别具有更高准确度。并且，提出的基于 Borda 的融合决策复杂度低、易于实现。

本文设计、实现的面向老年人的智能监控系统，满足了老年人在居家养老、机构养老环境下对监护的需求。为了充分利用采集的视频数据，论文同时改进了行为识别算法，提升了系统监控的准确度。

6.2 研究工作展望

论文实现的分布式智能监控系统，虽然满足了在不同场景下对老年人全方位监护的功能，但是系统从智能化、软件优化角度仍有可以完善的地方。

1. 轨迹定位在本文的设计方式是采用视觉定位方案，因此在不知道房屋模型情况下，无法做到多房间之间的全局定位。论文可以进一步对比蓝牙定位方案与视觉定位方案的精确度，蓝牙定位方案可以考虑在带蓝牙的监控节点中增加 iBeacon 功能，树莓派监控节点与可穿戴设备节点分别担任发射者、接收者的角色。

2. 监控平台与监控节点的通信线程通过线程池的进行管理，虽然线程池实现拥有优点，但是每个线程都占用一定的系统资源会导致系统资源浪费，并且会有某个线程维护的 TCP 连接没有进行网络 I/O 读写事件处理的情况。论文下一步优化可以选用 Socket 通信框架代替线程池以弥补不足。

3. 目前系统中由监控平台承担大部分分析工作，一方面是因为尚不确定运用深度学习的算法是否可以运行在树莓派监控节点上；另一方面是分布式的实现并不是类似于自组织网络，导致监控节点之间无法自己交换信息。论文下一步工作考虑完成系统智能优化和算法优化，使监控系统更加智能化，实现即插即用的效果。

本文的研究内容主要分为系统和算法两大部分，每一部分的内容都需要一定的精力去认真研究与实现。仅从系统部分来说，系统的设计和实现包含系统架构选择、监控节点实现、Web 服务器部署、交互界面实现等大量内容。因此论文在构思之初可以将重心选择在系统上，去探索更多的通用框架、隐私保护、性能优化等方面内容，进而实现更完善、更智能的系统。

参考文献

- [1] United Nations population division. 2017 Revision of World Population Prospects [EB/OL].
<https://population.un.org/wpp/>
- [2] 王桂新, 干一慧. 中国的人口老龄化与区域经济增长[J]. 中国人口科学, 2017(03):32-44.
- [3] Hudson J, Orviska M, Hunady J. People's attitudes to robots in caring for the elderly[J]. International journal of social robotics, 2017, 9(2): 199-210.
- [4] Hitchcock K. Dear life: on caring for the elderly[M]. Black Inc, 2016.
- [5] Romanou A. The necessity of the implementation of Privacy by Design in sectors where data protection concerns arise[J]. Computer law & security review, 2018, 34(1): 99-110.
- [6] Kidd C D, Orr R, Abowd G D, et al. The aware home:A living laboratory for ubiquitous computing research[C]. International Workshop on Cooperative Buildings. Springer, 1999:191-198.
- [7] Costa Â, Castillo J C, Novais P, et al. Sensor-driven agenda for intelligent home care of the elderly[J]. Expert Systems with Applications, 2012, 39(15): 12192-12204.
- [8] Dawadi P N, Cook D J, Schmitter-Edgecombe M. Automated Cognitive Health Assessment Using Smart Home Monitoring of Complex Tasks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2013, 43(6):1302-1313.
- [9] Suryadevara N K, Mukhopadhyay S C. Determining Wellness through an Ambient Assisted Living Environment[J]. IEEE Intelligent Systems, 2014, 29(3):30-37.
- [10] Do H M, Pham M, Sheng W, et al. RiSH: A robot-integrated smart home for elderly care[J]. Robotics and Autonomous Systems, 2018, 101:74-92.
- [11] Ghorbel M, Betg éBrezetz S, Dupont M P, et al. Multimodal notification framework for elderly and professional in a smart nursing home[J]. Journal on multimodal user interfaces, 2013, 7(4): 281-297.
- [12] Armentia A, Gangoiti U, Priego R, et al. Flexibility support for homecare applications based on models and multi-agent technology[J]. Sensors, 2015, 15(12): 31939-31964.
- [13] Tran D D, Le T L. Abnormal event detection using multimedia information for monitoring system[C]. 2014 IEEE Fifth International Conference on Communications and Electronics (ICCE). IEEE, 2014:490-495.
- [14] Stone E E, Skubic M. Fall Detection in Homes of Older Adults Using the Microsoft Kinect[J]. IEEE Journal of Biomedical and Health Informatics, 2015, 19(1):290-301.
- [15] Wiliem A, Madasu V, Boles W, et al. A suspicious behaviour detection using a context space

- model for smart surveillance systems[J]. Computer Vision and Image Understanding, 2012, 116(2):194-209.
- [16] Bian Z P, Hou J, Chau L P, et al. Fall Detection Based on Body Part Tracking Using a Depth Camera[J]. IEEE Journal of Biomedical and Health Informatics, 2015, 19(2):430-439.
- [17] Johansson G. Visual perception of biological motion and a model for its analysis[J]. Perception and Psychophysics, 1973, 14(2): 201-211.
- [18] Song Y, Morency L P, Davis R. Multimodal human behavior analysis: learning correlation and interaction across modalities[C]. Proceedings of the 14th ACM international conference on Multimodal interaction. ACM, 2012:27-30.
- [19] Roach J W, Aggarwal J K. Determining the movement of objects from a sequence of images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1980, PAMI-2(6):554-562.
- [20] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [21] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in neural information processing systems. NIPS, 2012:1097-1105.
- [22] Feichtenhofer C, Pinz A, Wildes R. Spatiotemporal residual networks for video action recognition[C]. Advances in neural information processing systems. NIPS, 2016:3468-3476.
- [23] Girdhar R, Ramanan D. Attentional pooling for action recognition[C]. Advances in Neural Information Processing Systems. NIPS, 2017:34-45.
- [24] 贾立平. 老人健康监护系统的设计与实现[D]. 北京航空航天大学, 2014.
- [25] 吕莎莎. 老年人手腕式血压仪设计[D]. 湖北工业大学, 2016.
- [26] 贾国忠. 面向老年人的智能音箱语音交互设计研究[D]. 华南理工大学, 2018.
- [27] Huang B, Tian G, Wu H, et al. A method of abnormal habits recognition in intelligent space[J]. Engineering Applications of Artificial Intelligence, 2014, 29:125-133.
- [28] 马宝庆. 基于全方位计算机视觉的独居老人监护系统[D]. 浙江工业大学, 2014.
- [29] 武传胜, 武林. 一种物联网声音监护老人的方法及其系统: 中国, 201511018590.1[P/OL]. 2016-04-13[2019-03-19].
- [30] 余文明. 基于物联网的空巢老人远程室内监护系统的设计与实现[D]. 华中科技大学, 2016.
- [31] Qiu Z, Yao T, Mei T. Learning spatio-temporal representation with pseudo-3d residual networks[C]. Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2017:5533-5541.
- [32] Zhao Y, Xiong Y, Lin D. Recognize actions by disentangling components of dynamics[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE,

- 2018:6566-6575.
- [33] Wang Y, Zhou L, Qiao Y. Temporal hallucinating for action recognition with few still images[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018:5314-5322.
- [34] Wang Y, Long M, Wang J, et al. Spatiotemporal pyramid network for video action recognition[C]. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. IEEE, 2017:1539-1538..
- [35] Zhao L, Wang J, Li X, et al. Deep convolutional neural networks with merge-and-run mappings[J]. arXiv preprint arXiv:1611.07718, 2016.
- [36] Wei Y, Xia W, Lin M, et al. HCP: A flexible CNN framework for multi-label image classification[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(9): 1901-1907.
- [37] Li Y, Wang S, Tian Q, et al. Feature representation for statistical-learning-based object detection: A review[J]. Pattern Recognition, 2015, 48(11): 3542-3559.
- [38] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. International Conference on computer vision & Pattern Recognition. IEEE, 2005:886-893.
- [39] Porikli F. Integral histogram: A fast way to extract histograms in cartesian spaces[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2005:829-836.
- [40] Wang X, Han T X, Yan S. An HOG-LBP human detector with partial occlusion handling[C]. IEEE International Conference on Computer Vision. IEEE, 2010:32-39.
- [41] Paisitkriangkrai S, Shen C, Zhang J. Performance evaluation of local features in human classification and detection[J]. IET Computer Vision, 2008, 2(4):236-246.
- [42] Cao X, Wu C, Yan P, et al. Linear SVM classification using boosting HOG features for vehicle detection in low-altitude airborne videos[C]. IEEE International Conference on Image Processing. IEEE, 2011:2421-2424.
- [43] Ye Q, Jiao J, Zhang B. Fast pedestrian detection with multi-scale orientation features and two-stage classifiers[C]. IEEE International Conference on Image Processing. IEEE, 2010:881-884.
- [44] Liu W, Lv J, Gao H, et al. An efficient real-time speed limit signs recognition based on rotation invariant feature[C]. 2011 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2011:1000-1005.
- [45] Boi F, Gagliardini L. A Support Vector Machines network for traffic sign recognition[C]. International Joint Conference on Neural Networks. IEEE, 2011:2210-2216.
- [46] Wang H, Schmid C. Action Recognition with Improved Trajectories[C]. 2013 IEEE International Conference on Computer Vision. IEEE, 2013:3551-3558.

- [47] Afshar S, Salah A A. Facial Expression Recognition in the Wild Using Improved Dense Trajectories and Fisher Vector Encoding[C]. Computer Vision & Pattern Recognition Workshops. IEEE, 2016:66-74.
- [48] Qingwu L, Haisu C, Yan Z, et al. Human Action Recognition Using Improved Salient Dense Trajectories[J]. Computational Intelligence and Neuroscience, 2016, 2016:1-11.
- [49] Shu Z, Yun K, Samaras D. Action detection with improved dense trajectories and sliding window[C]. Workshop at the European Conference on Computer Vision. Springer, 2014:541-551.
- [50] Russell S J, Norvig P. Artificial intelligence: a modern approach[M]. Malaysia; Pearson Education Limited, 2016.
- [51] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation[C]. Proceedings of the IEEE international conference on computer vision. IEEE, 2015:1520-1528.
- [52] Makantasis K, Doulamis A, Doulamis N, et al. Deep learning based human behavior recognition in industrial workflows[C]. 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016:1609-1613.
- [53] Lin L, Wang K, Zuo W, et al. A Deep Structured Model with Radius-Margin Bound for 3D Human Activity Recognition[J]. International Journal of Computer Vision, 2016, 118(2):256-273.
- [54] Cao S, Nevatia R. Exploring deep learning based solutions in fine grained activity recognition in the wild[C]. International Conference on Pattern Recognition. IEEE, 2017:384-389.
- [55] Shao J, Loy C C, Kang K, et al. Crowded Scene Understanding by Deeply Learned Volumetric Slices[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(3):613-623.
- [56] Tang K, Yao B, Fei-Fei L, et al. Combining the Right Features for Complex Event Recognition[C]. Proceedings of the 2013 IEEE International Conference on Computer Vision. IEEE, 2013:2696-2703.
- [57] Song S, Chandrasekhar V, Mandal B, et al. Multimodal Multi-Stream Deep Learning for Egocentric Activity Recognition[C]. Computer Vision & Pattern Recognition Workshops. IEEE, 2016:24-31.
- [58] Kavi R, Kulathumani V, Rohit F, et al. Multiview fusion for activity recognition using deep neural networks[J]. Journal of Electronic Imaging, 2016, 25(4):043010.
- [59] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[C]. Advances in neural information processing systems. NIPS, 2014:568-576.
- [60] Le Q V, Zou W Y, Yeung S Y, et al. Learning hierarchical invariant spatio-temporal features for

- action recognition with independent subspace analysis[C]. Computer Vision & Pattern Recognition. IEEE, 2011:3361-3368.
- [61] Weinland D, Ronfard R, Boyer E. Free viewpoint action recognition using motion history volumes[J]. Computer Vision and Image Understanding, 2006, 104(2-3):249-257.
- [62] Wu K, Ranasinghe R, Dissanayake G. Active recognition and pose estimation of household objects in clutter[C]. IEEE International Conference on Robotics & Automation. IEEE, 2015:4230-4237.
- [63] Doumanoglou A, Kim T K, Zhao X, et al. Active Random Forests: An Application to Autonomous Unfolding of Clothes[C]. European Conference on Computer Vision. Springer, 2014:644-658.
- [64] Wu N Z, Song S, Khosla A, et al. 3D ShapeNets: A Deep Representation for Volumetric Shape Modeling[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2015:1912-1920.
- [65] 雷玉堂. 现代安防视频监控系统: 设备剖析与解读[M]. 电子工业出版社, 2017.
- [66] 范贵生, 虞慧群, 陈丽琼, 刘冬梅. 分布式嵌入式系统的自适应能耗管理和分析[J]. 软件学报, 2014, 25(02):419-438.
- [67] Bilen H, Fernando B, Gavves E, et al. Dynamic Image Networks for Action Recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2016:3034-3042.
- [68] Wang L, Xiong Y, Wang Z, et al. Temporal segment networks: Towards good practices for deep action recognition[C]. European conference on computer vision. Springer, 2016:20-36.
- [69] Zahid M A, De Swart H. The borda majority count[J]. Information Sciences, 2015, 295: 429-440.
- [70] 王超. 基于树莓派的图像去雾算法的实现[D]. 安徽理工大学, 2018.
- [71] 周辰彦. 基于树莓派的虫害图像测报系统[D]. 浙江理工大学, 2018.

致谢

时光飞逝，三年的硕士研究生生活已渐入尾声。回首这段时光，一切都历历在目，让我难忘。在这段时间的学习和生活中，老师、同学、朋友和亲人的无私关心与支持都给了我克服困难的动力与勇气，是我人生路上宝贵的精神财富。在本文即将完成之际，谨向身边给我支持与鼓励的所有人表示由衷的感谢！

首先要感谢千博教授、杨刚教授等老师。从步入研究生阶段开始，老师们就在学习与生活中给予我无微不至的关心与照顾，无论我遇到什么困难，老师们都会孜孜不倦的给我指导解决困难，帮助我很多。本文从选题、开题、查找资料、研究实现等各个环节都是在千老师和杨老师的细致指导与严格要求下完成的。老师们对待工作的责任感，认真负责的人格魅力以及渊博的知识内涵，都使我受益良多，将在我以后的工作与生活中产生深远的影响。

感谢实验室的同届同学、师弟师妹以及所有同学朋友们，感谢你们一直以来对我的帮助与陪伴，很高兴与你们度过了一段美好的时光，愿我们友谊长存。

感谢 612 宿舍的所有同学，感谢你们生活中的陪伴与帮助。在此，向你们表示最真诚的谢意！同时也祝福你们一切顺利，前程似锦！

特别要感谢我的家人，是你们最无私的爱一直鼓舞着我，支持着我，成为我不断进取的动力，没有你们的支持，就没有我现在的成绩。你们为我付出了极大地心血，是我最坚实的后盾。

最后，感谢在百忙之中审阅本文的专家们！谢谢您！

作者简介

1. 基本情况

李冬月，女，河北邯郸人，1993年11月出生，西安电子科技大学机电工程学院控制理论与控制工程专业2016级硕士研究生。

2. 教育背景

2012.08~2016.06 河北农业大学，本科，专业：物联网工程
2016.08~ 西安电子科技大学，硕士研究生，专业：控制理论与控制工程

3. 攻读硕士学位期间的研究成果

3.1 发表学术论文

[1] 李冬月, 杨刚, 千博. 物联网架构研究综述. 计算机科学, 2018, 45(11): 27-31.

3.2 申请（授权）专利

[1] 杨刚, 李冬月, 千博等. 一种药品视频说明书及说明书生成方法: 中国, CN201710291254.7. 2017.

3.3 参与科研项目及获奖

- [1] 企业合作项目, 基于 Nellcor PMB05N 模块血氧采集系统设计, 2017.9~2018.4, 已完成, 负责信号处理算法编写.
- [2] 自研项目, 基于视频的老人异常行为监护系统设计, 2018.09~至今, 项目持续进行中, 负责软件实现.



西安电子科技大学
XIDIAN UNIVERSITY

地址：西安市太白南路2号

邮编：710071

网址：www.xidian.edu.cn