



燕山大学
YANSHAN UNIVERSITY

硕士学位论文

MASTER'S DISSERTATION

论文题目 基于机器视觉的手写数字识别与目标追踪技术研究

作者姓名 于立强

学位类别 工程硕士

指导教师 史艳国 教授

2019 年 5 月

中图分类号：TP278

UDC：621

学校代码：10216

密级：公开

工程硕士学位论文

(技术研究类)

基于机器视觉的手写数字识别与目标追踪技术研究

硕 士 研 究 生：于立强

导 师：史艳国 教授

副 导 师：付 毅 高级工程师

申 请 学 位：工程硕士

工 程 领 域：机械工程

所 属 学 院：机械工程学院

答 辩 日 期：2019 年 5 月

授 予 学 位 单 位：燕山大学

A Dissertation in Mechanical Engineering

**RESEARCH ON HANDWRITTEN DIGIT
RECOGNITION AND TARGET TRACKING
TECHNOLOGY BASED ON MACHINE VISION**

by Yu Liqiang

Supervisor: Professor Shi Yanguo

Yanshan University

May, 2019

摘 要

机器视觉近年来发展迅速，受到各国研究人员的关注。基于视觉的手写数字识别与目标追踪技术一直是机器视觉领域的热门课题。本文以 RoboMaster 比赛为背景，在嵌入式微型处理器 NUC7i7BNH 的条件下，通过单目摄像头进行图像采集，对采集图像进行处理，完成对手写数字的识别和单目标的追踪。

利用棋盘标定法对图像采集系统中 USB 摄像头进行标定实验，并对采集的图像进行灰度化、滤波、增强对比度和图像二值化处理。其中，选用最大熵法对图像进行二值化处理，并根据数字面板的特殊性，对其进行改进，得到清晰的二值化图像。

通过研究经典的目标定位算法和数字面板的特征，改进经典的目标定位方法。提出基于边缘和几何关系相结合的目标定位算法，通过实验验证该算法对目标识别定位的准确性。

采用支持向量机的方法对手写数字进行识别，通过对样本进行一对一训练方法来解决多分类问题，并利用 5-CV 进行交叉验证。将不同核函数进行对比实验，采用径向基核函数，并对其进行参数修改，提高识别率。

在传统颜色识别算法的基础上，提出一种基于改进颜色识别与阈值分割相结合的方法，将运动目标检测并识别出来。在 STM32 单片机中进行 C 语言编写控制程序，利用 PID 算法控制电机，完成对目标的追踪任务。

关键词：嵌入式；手写数字识别；支持向量机；目标追踪

Abstract

Machine vision has developed rapidly in recent years and has attracted the attention of researchers from various countries. Vision-based handwritten digit recognition and target tracking technology has always been a hot topic in the field of machine vision. In this paper, based on the RoboMaster competition, under the condition of the embedded microprocessor NUC7i7BNH, the image is collected by the monocular camera, and the captured image is processed to complete the recognition of the handwritten digits and the tracking of the single target.

The calibration of the USB camera in the image acquisition system is carried out by using the checkerboard calibration method, and the acquired image is grayed, filtered, enhanced contrast and image binarized. Among them, the maximum entropy method is used to binarize the image, and according to the particularity of the digital panel, it is improved to obtain a clear binarized image.

The classic target location method is improved by studying the characteristics of the classical target localization algorithm and the digital panel. A target localization algorithm based on the combination of edge and geometric relations is proposed. The accuracy of the algorithm for target recognition and location is verified by experiments.

The method of support vector machine is used to identify handwritten digits, and the multi-classification problem is solved by one-to-one training method for samples, and cross-validation is performed by 5-CV. Different kernel functions were compared and tested, and the radial basis kernel function was used, and the parameters were modified to improve the recognition rate.

Based on the traditional color recognition algorithm, a method based on improved color recognition and threshold segmentation is proposed to detect and identify moving targets. In the STM32 MCU, the C language programming control program is used, and the PID algorithm is used to control the motor to complete the tracking task of the target.

Keywords: embedded; handwritten digit recognition; SVM; target trackin

目 录

摘 要.....	I
Abstract.....	II
第 1 章 绪 论.....	1
1.1 课题研究背景与意义.....	1
1.2 国内外研究现状.....	3
1.2.1 手写体数字识别技术研究现状.....	3
1.2.2 目标追踪研究现状.....	5
1.3 本文主要研究内容安排.....	6
第 2 章 图像采集系统与图像预处理.....	7
2.1 引言.....	7
2.2 图像采集系统.....	7
2.2.1 USB 摄像头.....	8
2.2.2 控制芯片选型.....	9
2.2.3 电机选型.....	9
2.3 坐标系转换.....	10
2.4 摄像头的标定与实验结果.....	12
2.5 图像预处理.....	13
2.5.1 图像灰度化.....	13
2.5.2 图像滤波.....	15
2.5.3 增强对比度.....	17
2.5.4 图像二值化.....	18
2.6 本章小结.....	22
第 3 章 字符区域识别与分割.....	23
3.1 引言.....	23
3.2 字符区域定位的研究.....	23
3.2.1 基于颜色特征目标定位算法.....	23
3.2.2 基于纹理特征的目标定位算法.....	24
3.2.3 基于边缘检测的目标定位算法.....	24
3.3 基于边缘信息和几何关系的研究.....	24
3.3.1 字符区域特征.....	24
3.3.2 本文改进定位算法.....	25
3.3.3 边缘检测.....	27

3.3.4 边缘提取	28
3.3.5 几何关系判定	29
3.4 图像质心	31
3.5 字符分割	32
3.5.1 常见的分割方法	32
3.5.2 基于字符轮廓的分割	33
3.6 本章小结	33
第 4 章 基于 SVM 的数字字符识别	34
4.1 引言	34
4.2 支持向量机概述	34
4.3 支持向量机原理	35
4.4 SVM 多类分类问题	37
4.5 基于 SVM 数字识别的研究	38
4.5.1 图像尺寸归一化	39
4.5.2 交叉验证	40
4.5.3 SVM 识别手写数字的实现	43
4.6 实验结果分析	43
4.7 本章小结	47
第 5 章 目标检测与控制的实现	48
5.1 引言	48
5.2 运动目标检测算法的研究	48
5.2.1 帧差法	48
5.2.2 背景差分法	49
5.2.3 光流法	50
5.3 基于改进颜色识别与阈值分割的研究	51
5.3.1 基于颜色识别	53
5.3.2 改进颜色识别	53
5.3.3 形态学操作	55
5.3.4 灯条识别	55
5.3.5 坐标输出	56
5.3.6 电机角度判断	56
5.4 运动控制与追踪策略的研究	57
5.4.1 PID 控制算法	58
5.4.2 云台系统	59
5.4.3 数字式 PID 的改进	60

目 录

5.5 开发环境	60
5.6 PID 控制与追踪实验	62
5.7 本章小结	64
结 论	65
参考文献	66
攻读硕士学位期间承担的科研任务与主要成果	70
致 谢	71

第1章 绪论

1.1 课题研究背景与意义

2013年4月的汉诺威工业博览会上，德国为了提高国家的竞争力，德国政府正式启动了“工业4.0”战略。从工业3.0时代到工业4.0时代的转变，实质上是将现有工业自动化转变为工业智能化和网络化，人工智能和人机交互技术将发挥至关重要的作用。机器视觉作为人工智能的一个重要分支，在生活和生产中发挥着重要作用，机器人通过机器视觉来帮助人们从事危险工作和人类无法运作的情况。

在我们生活中信息的主要来源是触觉、听觉和视觉。大部分信息主要是通过视觉传递到大脑，视觉相对于听觉和触觉获取信息的效率较高，所以视觉是人类感知客观世界的主要信息来源^[1]。机器视觉就是相当于人的眼睛，对外部环境进行采集、分析与处理。在1980年前后，Marr D提出视觉计算理论^[2]。他认为视觉是从获取信息，经过处理系统处理，最后储存起来供人们对图像信息分析理解的过程。该系统的提出很大程度上促进了机器视觉和图像分析技术的发展。由于现在计算机技术的快速发展，使得机器视觉也得到了相应的发展。机器视觉作为一门综合性学科，其应用了生物学、信息处理、计算机、模式识别等技术^[3,4]。现在该技术广泛应用于工业制造当中，通过视觉完成对目标的识别，帮助人们完成检测质量、分类等枯燥乏味的工作。

在机器视觉中，基于视觉的目标识别与追踪是其中较为有挑战性和创新性研究方向。视觉拍摄到的图像具有直观、实时、信息量大、无损等特点，因此吸引了很多研究人员对此进行深入研究^[5,6]。目标识别与追踪系统就是对采集到的图像序列进行处理，从图像序列中计算出目标的一些信息，如位置、速度、方向或位移量等。如图1-1所示维视智造智能分拣机器人，通过图像处理系统对拍摄到图像进行分析处理，识别出目标位置信息，机器人通过这些位置信息来对物料进行分拣，帮助人们完成乏味的分拣工作。图像处理信息的任务有两个：一是将环境信息转变为数字信息；二是利用数字信号进行运算处理得到想要的信息^[7]。视觉机器人在中国发展相对较为迅速，中国已将其应用到工业、交通、安全、航天、天文等领域。2010年我国机器视觉市场迎来了爆发式增长，机器视觉技术得到了较程度的发展^[8,9]。

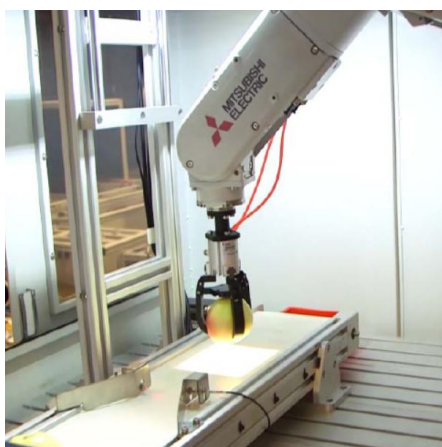


图 1-1 维视智造智能分拣机器人

从图 1-2 中的专利申请数量上进行分析,也可以看出机器视觉技术水平在 2010 年进入快速发展阶段。目标识别与追踪作为机器视觉的重要分支,所以研究目标识别与追踪技术有着重要的意义。

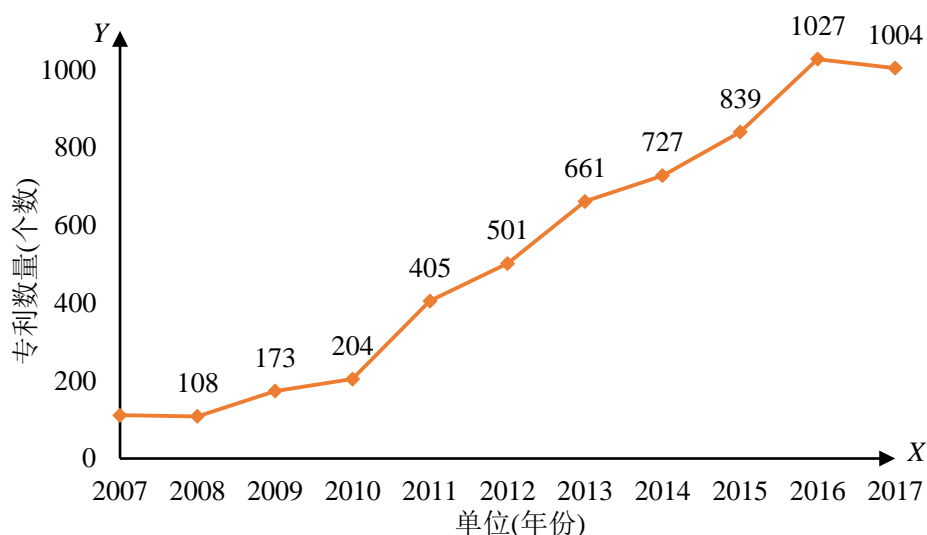


图 1-2 2007-2017 年中国机器视觉相关专利申请量变化图

机器视觉的高速发展,使目标识别与追踪同样得到了快速发展,目标识别与追踪技术快速应用到在很多领域中。在军事方面,经常应用于战场监控、无人飞行器、空中预警、精准制导中。在人类生活中同样有很多方面应用到了目标识别与追踪任务,比如经常使用的智能监控、智能交通系统、移动机器人、虚拟现实、人机交互等。因此目标追踪是机器视觉中一个比较有挑战性的问题。摄像头的普遍使用和日益剧增的需求使研究目标识别与追踪的人越来越多。

在目标识别中较为重要的字符识别应用同样较为广泛。字符识别的研究主要分两大部分,分别为手写体的识别和印刷体的识别。在手写体识别中又可分为联机手

写体识别和脱机手写体识别。脱机手写体识别占用一定空间，识别快，相对于联机手写识别错误率较高。联机手写体识别库是从云端中读取，本地占用空间少，相对于脱机识别略慢，错误率较低。为了提高机器人识别效率，所以研究如何提高脱机手写体识别率尤为重要。

在财务、金融领域，每天会处理大量财务和支票录入等工作。金融市场化进程日益加快，票据业务发展迅速，票据数量也越来越多。如今这些繁琐的工作仍由人工进行处理，如果实现手写字符的自动录入，大大提高传统工作方式的效率。所以手写数字识别系统具有重要的研究价值。

(1) 阿拉伯数字是世界上唯一通用性符号，具有较高的普遍性，所以研究数字识别意义十分重要。

(2) 手写体数字识别系统广泛应用于数据统计中。例如：报表的录入、支票的录入等应用中。

(3) 手写体数字识别系统还被应用于邮局的邮件分拣系统中，大大提高了邮件分拣速度。

(4) 手写数字识别一旦被成功掌握，易于在其它领域推广。

近年来，国内外研究学者通过不断对数字识别的研究，在手写数字识别方面同样取得了很大的成果。

1.2 国内外研究现状

1.2.1 手写体数字识别技术研究现状

大约在 20 世纪 20 年代末，字符识别的研究最早起源于欧美。1929 年 Tausheck 正式提出 OCR(Optical Character Recognition)概念。他是利用光学模板对字符进行识别^[10]。20 世纪 50 年代由于计算机的慢慢出现，使 OCR 开始步入正式研究阶段。越来越多的人开始对 OCR 的研究，1990 年进入发展期。1958 年 Crimsalpe 等人第一次使用计算机对光学字符进行识别，主要是通过将字符模板储存到计算机，通过其它机器对字符进行扫描，将得到的扫描字符与模板进行对比，来判断是否为字符。在 60 年代初，美国麻省理工学院 Murray Eden 认为 0-9 这几个数字，可以用有限的基本特征组成，为字符识别提供了一个重要的研究方向^[11]。针对字符的研究，20 世纪 70 年代 Parks 等人研究出来利用抽取拓扑结构特征的新方法。在 20 世纪 80 年代日

本进行了印刷体汉字的识别，并且公开演示。在 80 年代初，由于个人计算机的普遍使用，使得字符识别方面的得到了快速发展。在这期间提出了很多新型的技术。将人工神经网络、小波变换、分形引用到字符识别系统中，并取得了很好的成果。在 90 年代 Vapnik 等人在机器学习相关问题上得到了很大的进展，并建立了一整套完整的理论体系支持向量机(Support Vector Machines, SVM)，该方法成功应用到字符识别领域中^[12]。2013 年 Bengio 通过对不同环境下的 62 个大型手写数字集进行深入研究，利用统计学和大量仿真实验，最终证明了深层网络模型可以将手写数字中的本质特征提取出来。2017 年 Lee 提出了卷积深度置信网络，主要是在限制波尔兹曼机中添加了卷积操作，该方法可以有效的学习和提取环境中数字图像的特征。2017 年德国的 Friedhelm Schweriker 利用 SV_RBF40 的方法对手写数字进行识别，通过测试 10000 个样本数据，最终的识别率可以达到 95.56%。

早在 20 世纪 70 年代，我国开始对手写数字识别进行研究。中科院在 1974 成功将字符识别技术应用到邮件的分拣过程中。复旦大学成功研制出中国第一台印刷数字识别机。80 年代初，随着更多高校的研究，使数字识别技术得到较大的提高，并取得了很大的成就。黄巧巧在 Visual C++6.0 的编译环境下，利用 BP 神经网络实现对手写数字的识别，证明了 BP 神经网络可以应用于手写数字中^[13]。在自然场景的字符识别中，王强通过 PCA(Principal Component Analysis)和 CNN 相结合的新方法，并利用 SVHN 数据集进行了实验验证^[14]。曹丹提出基于 Hopfield 神经网络的识别模型，在数字识别方面较 BP 神经网络有较好的识别率^[15]。同样利用 BP 神经网络方面，上海交通大学通过压缩子特征的方法，最终使识别率达到 96.1%。对于银行票据中的手写数字识别，宋菀娟同样利用 BP 神经网络对 20000 个样本进行识别，得到的识别率为 95.39%^[16]。清华大学利用支持向量机的方法，完成对邮政单上面手写数字的识别，通过对 20000 个样本数据进行测试，最后得到的识别率为 92%。

通过上面的介绍，可以知道国内外的识别方法大体上可以分为：模板匹配方法、统计决策方法、模糊判断方法、逻辑判断方法、神经网络方法和支持向量机方法等。

在数字识别的实际应用中，希望对每个手写数字的准确识别，但是所有算法识别率都不能达到 100%。在实际环境中并不是只有单个数字，所以环境因素会干扰机器人对手写数字的识别，针对不同的复杂的环境将数字找出并准确识别出来，现有算法并不能准确识别。所以不论从技术方面还是实际应用上讲，研究复杂环境手写数字识别具有重要意义。

1.2.2 目标追踪研究现状

1937 年研究出第一部跟踪雷达 SCR-28, 使得追踪技术得到了人们的关注^[17]。机器视觉目标追踪方法相对起步较晚。机器视觉的发展非常快, 使机器视觉成为现代重要的研究方向^[18]。近些年来, 机器视觉的快速发展, 使得在视觉目标追踪方向已经研究出一些重要成果。比如在 1994 年, 美国斯坦福大学研究院研制的名为 Flakey 的移动机器人^[19]。该机器人可以对人进行识别, 并完成追踪任务。相对于国外, 我国也有一些成功案例。比如, 在上海交通大学研制的 Frontier-ITM 机器人, 其同样可以完成跟踪任务。由清华大学研究的 THIR 系统可以根据车道线检测完成跟踪任务。

视觉追踪经过不断地发展, 出现了很多比较新颖想法和算法。Yan 等人通过颜色直方图和 Mean Shift 相结合实现对目标的追踪^[20]。主要依据就是物体的颜色不变性、非刚性变形以及抗遮挡的共性, 但是在有物体遮挡时就会出现追踪目标丢失, 所以有一定的缺陷。针对复杂环境下的目标识别, Pere 与 Numimar 等人提出了基于颜色特征的粒子滤波观测模型^[21,22]。在目标追踪算法中除了对颜色的大量研究外, 还可以利用边缘信息提取的方法实现对目标追踪的任务^[23,24]。近年来, 人们通过对背景与目标进行分类, 使背景完全从图像中分离出来, 从而达到识别目标完成追踪任务。如 Avidan 提出支持向量机追踪算法, 利用向量机学习将目标与背景进行分类, 从而对目标进行追踪^[25]。

总体上说, 目标追踪技术发展非常迅速, 各方面也取得了很多研究成果。经过多年的发展, 主要形成以帧差法、光流法、背景差分法等为主流的运动目标检测方式。2008 年 Maddalena 等人提出自组织神经网络的背景减除法(SOBS), 该算法通过自组织方法建立背景的神经网络, 比较目标与模型之间的距离, 从而提取出运动目标^[26]。2014 年的 IEEE 变化检测研讨会(IEEE Change Detection Workshop)上, B.Wang 和 P.Dudek 提出一种新的快速算法自调整背景减法^[27]。但是并没有一个算法是通用的, 所以在实际应用中, 需要根据自己的实际环境背景和目标的特点, 选择或提出自己研究方法, 以达到自己的要求。

在目标追踪技术的实际应用中, 希望可以对视频图像序列进行长时间实时稳定的追踪。由于摄像机设备本身缺陷和实际环境的复杂性, 所以在图像采集中往往会引入一些噪声, 使图像的质量不高, 这样就使得追踪处理的难度大大提高。想要将实时性和准确性两个指标都达到实际的要求, 现有的算法不能实现, 所以不论从技

术研究方面还是从实际应用来讲本论文的研究都尤为重要。研究新型的目标追踪系统对于国防现代化建设和国民经济建设都有重大意义。

1.3 本文主要研究内容安排

本论文主要是利用摄像头采集图像，图像处理系统处理图像，完成对图像中手写数字的切割与识别，还有云台系统对特定运动目标的追踪。整个论文设计主要分为图像预处理、手写数字面板的定位与分割、手写数字识别和特定运动目标追踪。全文研究内容如下：

第 1 章：绪论。介绍了机器视觉的重要作用，阐述国内外对目标追踪与数字识别的研究现状，结合本文主要研究内容，确定本文研究路线。

第 2 章：目标识别的前处理。首先对研究对象的硬件条件介绍和根据摄像头标定理论完成摄像头的标定实验；然后对图像进行预处理，包括图像灰度化、图像滤波和对比度增强，得到较为清晰的灰度图像；最后进行阈值分割，针对数字面板的特点，本文对传统最大熵法进行改进，得到更加合理的二值图像。

第 3 章：字符区域识别与分割。首先对常用定位算法的研究，提出边缘检测与几何关系相结合的算法，准确的将数字面板识别与定位，最后通过分割技术得到所有手写数字图像。

第 4 章：基于 SVM 的数字识别。首先对支持向量机原理的介绍，其次对训练样本数据进行尺寸归一化处理，然后 SVM 对样本进行训练得到训练模型，最后通过反复试验，得到本文最优核函数与参数。

第 5 章：目标检测与追踪的实现。首先对常用运动物体检测算法的深入研究，可知常用算法并不满足本文追踪要求，然后通过对目标的研究，提出改进颜色识别与阈值分割相结合的办法，将运动目标检测出来，最后通过 PID 控制电机转动，完成对目标的追踪任务。

第2章 图像采集系统与图像预处理

2.1 引言

图像采集系统硬件主要包括控制芯片、电机和 USB 摄像头。工作时，电机带动 USB 摄像头转动拍摄不同方位的图像。通过分析采集到的图像，从而识别目标。为了保证图像的质量，需要对初始图像进行预处理操作。主要包括图像灰度化、图像滤波去噪、增强对比度和二值化图像。

2.2 图像采集系统

本文是以 RoboMaster 比赛为背景，实现对数字面板上的手写数字识别和对目标小车的追踪。在识别出目标后主要是通过控制二自由度云台的运动，实现对目标的追踪任务。云台系统由两个电机控制，一个控制云台左右转动称之为偏航轴(YAW)，另一个控制云台上下转动称之为俯仰轴(PITCH)。当云台处于水平方向时，俯仰轴出于零点位置，将转动的角度称之为俯仰角，在偏转轴处于零点时，云台指向正前方，左右偏转角度称之为偏航角。通过控制两个电机的转动使云台在一定的方位运动。云台的三维结构如图 2-1 所示，实物如图 2-2 所示。

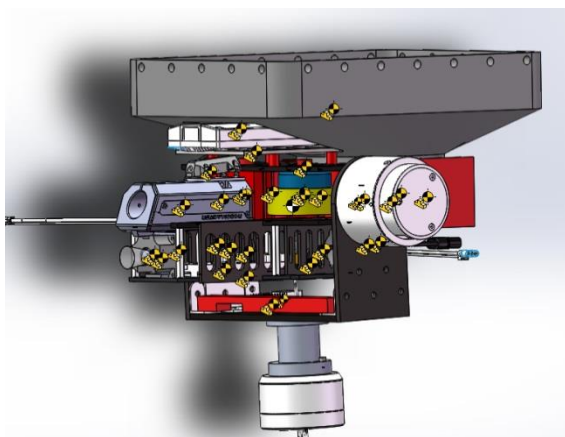


图 2-1 云台结构三维模型



图 2-2 云台实物图

由于机械结构的原因会使俯仰角受到一定的限制，由于偏航电机采用的是 24 路 2A 的导线滑环连接，所以可以进行 360° 旋转。导电滑环参数 24 路是指一共从滑环内部穿过了 24 跟导线，2A 是每根导线上可以承受的最大电流。云台的具体参数如表 2-1 所示。

表 2-1 云台参数

参数	PITCH	YAW
速度范围	-20dps~+20dps	-20dps~+20dps
角度范围	-30 °~+75 °	360 °

2.2.1 USB 摄像头

本文中的摄像头为简单的工业 USB 摄像头，可以将前方景象拍摄成一帧一帧图像，为了减少目标速度过快而产生的失帧现象，本文选用每秒传输 60 帧图像的高速摄像头。选择使用型号为 KS2A17 的摄像头，有效像素为 640×480，实物如图 2-3 所示。具体参数如表 2-2 所示。

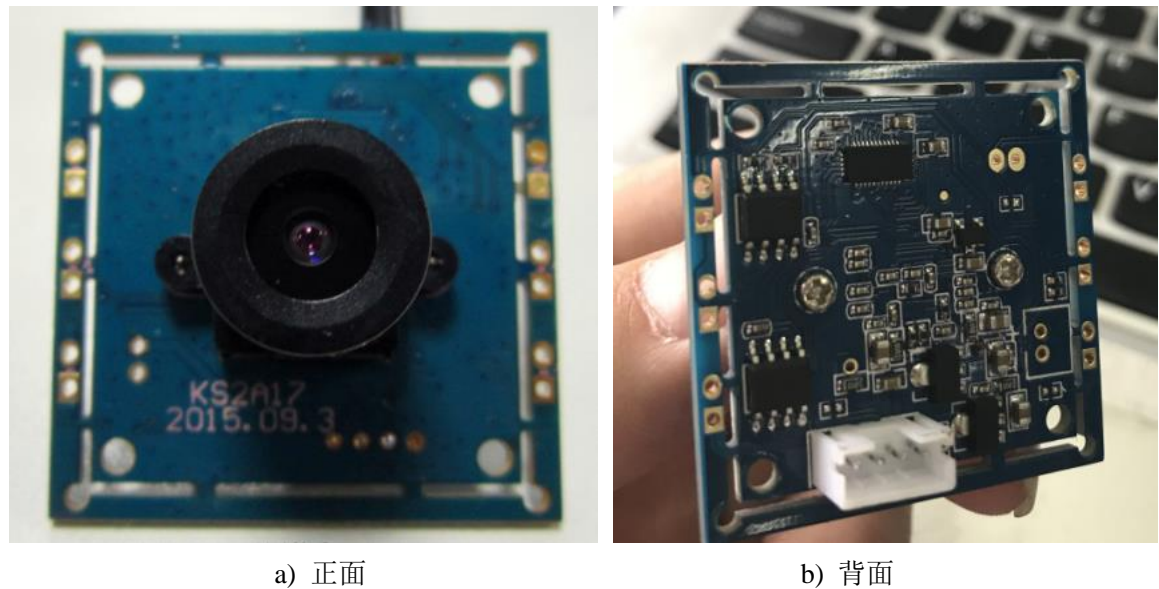


图 2-3 KS2A17 型号 USB 摄像头

表 2-2 KS2A17 参数

名称	参数
型号	KS2A17
传感器	OV2710
最大像素	210 万
USB 协议	USB2.0HS/FS
分辨率	640×480
工作电压	5V
支持的系统	WinXP/Vista/Win7/Win8; Linux with UVC (above linux-2.6.26) ; MAC-OS X 10.4.8 or later; Wince with UVC; Android 4.0 or above

2.2.2 控制芯片选型

云台控制模块采用 STM32F427 开发板作为主控制器。STM32F427 是一款基于 ARM Cortex-M4 内核 STM32 系列的 32 位的微控制器，相对 Cortex-M3 性能提高了 20%。STM32F427 可以对电机的精准控制、高频命令的收发和精准的姿态解算，将 STM32F427 作为主控制器。具体性能参数如表 2-3 所示。

表 2-3 性能参数

芯片型号	STM32F427
速度	180MHZ
程序储存容量	2MB
核心处理器	Cortex-M4
I/O 数	82
工作电源电压	3.6V
工作温度	-40℃~85℃
核心尺寸	32 位

2.2.3 电机选型

无刷电机广泛应用于航模、家电、电动车、智能小车、医疗等多个领域。无刷直流电机是典型的机电一体化产品。无刷电机结构比较简单，真正决定其使用性能的还是无刷电子调速器。无刷直流电机不仅含有直流有刷电机特性，而且还有频率变化装置，所以又名直流变频，国际通用名词为 BLDC。无刷直流电机相对于其他控制技术变频器的运转效率、低速转矩、转速精度要好，所以值得业界关注。下表 2-4 是有刷电机与无刷电机比较。

表 2-4 有刷电机与无刷电机比较

无刷电机	有刷电机
无电刷，低干扰	摩擦大，损耗大
噪音低，运转顺畅	发热大，寿命短
寿命长，低维护成本	效率低，输出功率小

由于本文对控制要求较高所以使用无刷电机。无刷电机是通过数字变频控制，可控制性强，从每分钟几转，到每分钟几万转都可以很容易实现。选择电机型号为大疆公司的 RM6623 和 RM6020 无刷电机。RM6623 作为俯仰旋转电机，RM6020 作为左右旋转电机。两个电机参数如表 2-5 所示。

表 2-5 RM6623 和 RM6020 电机参数

型号	RM6623	RM6020
工作电压	24V	24V
质量	605g	468g
转矩常数	0.38 N·m/A	741m N·m/A
转速常数	25RPM/V	13.33RPM/V
空载转速	600RPM	320RPM
堵转电流	5.3A	0.9A

2.3 坐标系转换

拍摄图像中某点在世界坐标系中位置的确定，就必须求得像素坐标系与世界坐标系之间的转换关系。假设空间有一点 P ，在图像中对应的像素点坐标为 (u, v) ，图像坐标系原点在成像表面与相机光轴的交点处，设 P 点在图像物理坐标系中坐标为 (x, y) ，在摄像头坐标系中坐标为 (X_c, Y_c, Z_c) ，在世界坐标系中坐标为 (X_w, Y_w, Z_w) ，具体的位置关系如图 2-4 所示。

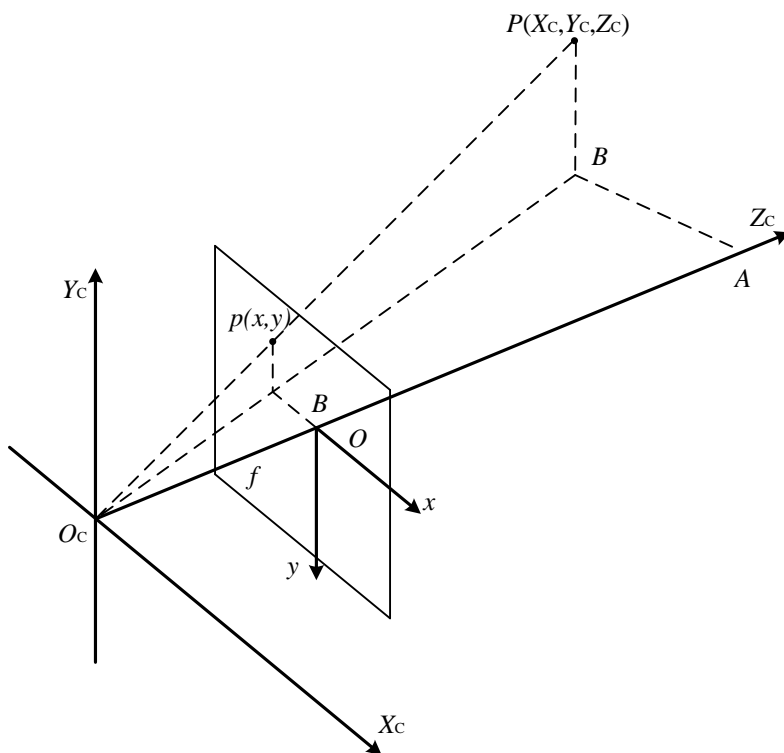


图 2-4 视觉系统各坐标系

(1) 图像物理坐标系与像素坐标系

像素坐标系的单位为像素，图像物理坐标系的单位为 mm，需要统一两者之间的

单位，在这里定义像素在横纵坐标上的物理尺寸为 d_x ， d_y 。通过位置平移与比例缩放实现两者之间的相互转化， (u_0, v_0) 透视主点坐标，具体的转换关系如下：

$$\begin{cases} u = \frac{x}{d_x} + u_0 \\ v = \frac{y}{d_y} + v_0 \end{cases} \quad (2-1)$$

进一步转化为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2-2)$$

(2) 摄像头坐标系与世界坐标系

世界坐标系通过旋转和平移得到相机坐标系，两者之间的转化关系式为：

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \mathbf{T} \quad (2-3)$$

其中， \mathbf{R} 和 \mathbf{T} 分别为旋转矩阵和平移矩阵。

(3) 图像物理坐标系与摄像头坐标系

由三角形相似原理可以得到两者之间的转化关系：

$$\begin{aligned} \frac{y}{f} &= \frac{Y_c}{Z_c} \\ \frac{x}{f} &= \frac{X_c}{Z_c} \end{aligned} \quad (2-4)$$

进一步转化为：

$$Z_c \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (2-5)$$

(4) 图像像素坐标系与世界坐标系

根据上面三部分之间的转化关系，可以推导出图像像素坐标系与世界坐标系的

转化关系:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2-6)$$

其中, f_x , f_y 为相机焦距, u_0 , v_0 为透视主点坐标。由于这些变量的值都是未知的, 为了得到求取参数矩阵, 就必须要进行摄像头的标定。

2.4 摄像头的标定与实验结果

OpenCV 标定应用场景较为广泛, 只需要读取相应的棋盘图像即可, 不需要过多的操作, 所以该方法简单易操作。在实际应用中, 为了使标定更加准确, 可以通过增加棋盘图片数量。本文采用的是平面棋盘为标定物的标定方法, 棋盘如图 2-5 所示。

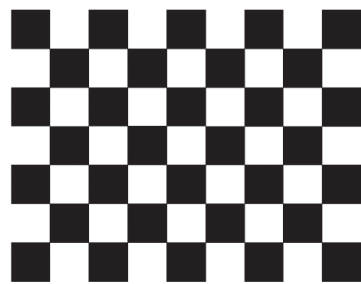


图 2-5 标定棋盘格

使用 KS2A17 摄像头拍摄不同距离和角度的棋盘图像, 如图 2-6 所示。

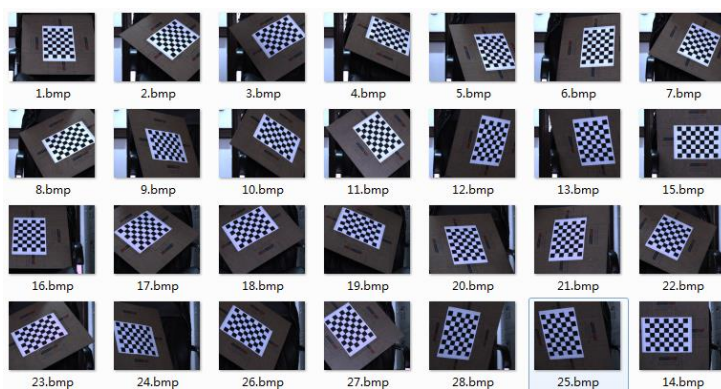


图 2-6 拍摄棋盘图片

OpenCV 函数库中包含两个基于棋盘标定的函数, 分别为单目标定和双目标定。本文系统是单目摄像头, 所以使用 OpenCV 函数库中的 `cvCalibrateCamera2()` 函数完成标定任务。首先读取拍摄的每张棋盘图片, 利用 `cvFindChessboardCorners()` 函数定

位图像中角点，然后紧接着调用 `cvFindCornerSubPix()` 函数进行亚像素分割，进一步优化细化角点，提高角点位置的准确性，最后调用 `cvCalibrateCamera2()` 函数进行大量计算，返回标定所需要的数据，最终得到的结果保存在 xml 格式文件中。表 2-6 为 OpenCV 标定结果。

表 2-6 OpenCV 标定结果

焦距 f_x (像素)	焦距 f_y (像素)	主点横坐标 u_0 (像素)	主点纵坐标 v_0 (像素)
450.35	453.42	320	240

2.5 图像预处理

图像预处理首先利用图像灰度化的方法降低图像处理数据；然后利用降噪技术来降低噪声对图像的影响；其次利用增强对比度的方法，降低噪声处理过程中对边缘信息的影响；最后通过图像二值化处理，使目标信息更加明显突出和进一步的将处理数据简化。预处理流程如图 2-7 所示。

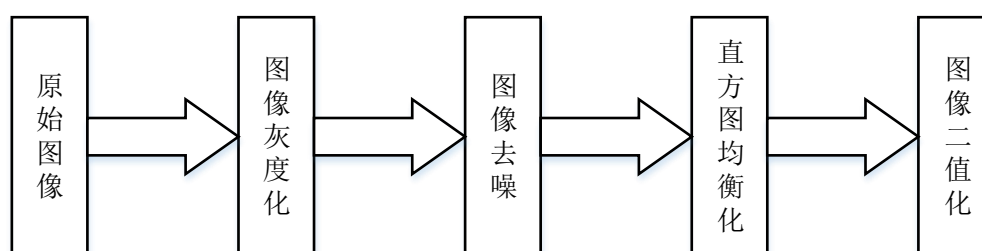


图 2-7 预处理流程

经过预处理过程，可以使后续在对目标信息特征进行提取、分割和识别过程更加简单和容易。

2.5.1 图像灰度化

本文摄像头采集到的图像都是由红、绿、蓝三通道组成。彩色图像相较于灰度图像含有更多的色彩信息^[28]。如果直接对图像进行处理，需要处理每个通道的信息，这样处理的数据就会非常大。图像灰度化得到的灰度图像只需处理单通道数据，减少两倍的处理数据。灰度图像并不影响对目标的提取、分割和识别。经过这样处理可以提高程序的处理速度，同时还可以节省储存资源。常见灰度化处理方法有以下几种：

- (1) 分量法。利用图像本身的通道的值，将三通道其中的一个通道的分量值作为

灰度图像的灰度值，可根据不同场合选择不同的通道作为灰度值。

$$\begin{cases} M_1(i, j) = R(i, j) \\ M_2(i, j) = G(i, j) \\ M_3(i, j) = B(i, j) \end{cases} \quad (2-7)$$

式中 $R(i, j)$ ——红色通道灰度值；

$G(i, j)$ ——绿色通道灰度值；

$B(i, j)$ ——蓝色通道灰度值；

$M_k(i, j)$ ——转换后灰度图像灰度值($k=1, 2, 3$)。

原图像采用分量法得到的三种灰度效果如图 2-8 所示。

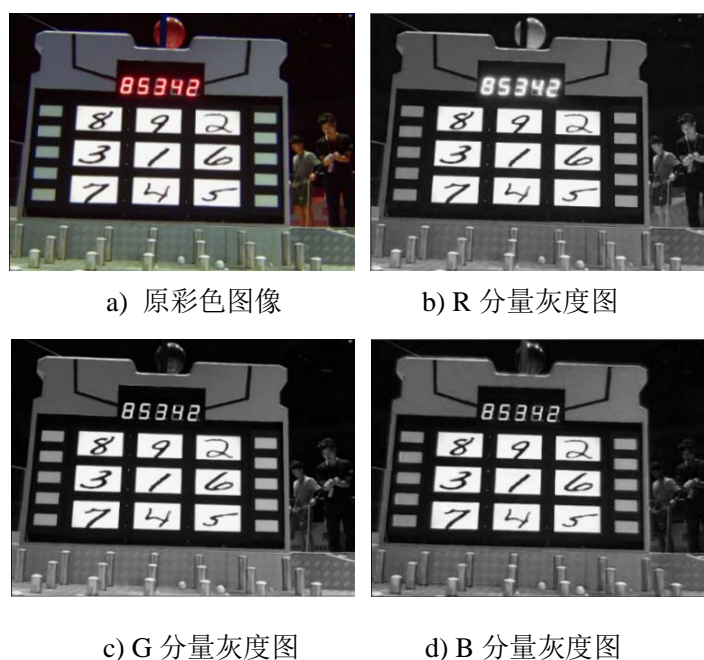


图 2-8 三色通道灰度图像对比效果图

(2) 最大值法。该方法是当前像素点的三个通道相对应数据的最大值作为灰度图像的灰度值。

$$M(i, j) = \max(R(i, j), G(i, j), B(i, j)) \quad (2-8)$$

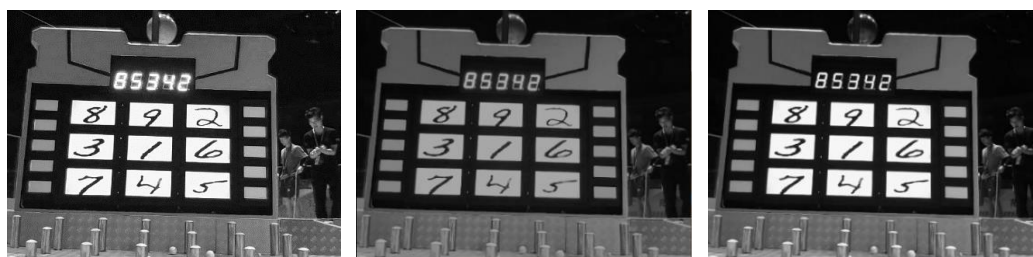
(3) 平均值法。将彩色图像中的三分量亮度相加并除以 3，得到的平均值作为灰度图像的灰度值。

$$M(i, j) = \frac{(R(i, j) + G(i, j) + B(i, j))}{3} \quad (2-9)$$

(4) 加权平均法。利用红绿蓝三部分不同的百分比，组合计算的结果作为灰度值。以人对三种颜色的敏感度不同为基础，选取不同的百分值，得到的运算公式：

$$M(i, j) = 0.30R(i, j) + 0.59G(i, j) + 0.11B(i, j) \quad (2-10)$$

利用最大值法、平均值法和加权平均值法处理效果如图 2-9 所示。



a) 最大灰度值图像

b) 平均值灰度图像

c) 加权灰度图像

图 2-9 其他灰度化处理比较图

分量法只需单通道数据，可能造成部分信息缺失；平均值法得到的灰度图像整体偏暗；最大值法得到的灰度图像部分区域过于明亮；加权法得到的灰度图像对本文数字面板处理效果较好，因此数字面板识别部分使用第四种方法。分量法在对于单一颜色分量的处理效果较好，由于在识别目标小车时需要对蓝色进行提取，所以该部分使用分量法进行处理。

2.5.2 图像滤波

图像噪声就是指图像数据中的不必要或多余的干扰信息。由于设备等的缺陷，数字图像在形成过程中易于被大量噪声污染。这些噪声通常表现为孤立的像素或像素块。对于数字图像信号，噪声表现为极大值或极小值，所以噪声降低了图像质量。内部噪音：一般来说，它可分为电路电流引起的噪声；设备运动引起的噪声；由光和电的基本性质引起的噪声和设备材料本身引起的噪声。

图像滤波可以减少噪声对图像的影响。在尽量保留原有图像信息的情况下，减少噪声从而提高图像质量。由于采集到的图像都会被噪声污染，所以图像滤波是图像预处理必不可少的操作。当处理效果不好有可能在目标识别过程中产生误识别的情况。

图像滤波就是在尽量减少破坏图像中边缘信息的前提下，减少或消除图像中的噪声。其处理效果会影响到后期对图像分析与识别。下面是常用的一些滤波算法。

(1) 均值滤波

均值滤波采用的主要方法是邻域平均法^[29]。均值滤波通俗的讲就是，利用当前像素点与邻域像素点的平均值作为当前滤波后的像素值。可以一定程度上降低图像

的噪点，利用公式表示为：

$$g(i, j) = \frac{1}{m} \sum f(i, j) \quad (2-11)$$

式中 $f(i, j)$ ——当前像素点的灰度值；

$g(i, j)$ ——计算处理后像素点的灰度值；

m ——模板像素点个数。

对图 2-10a)原图添加椒盐噪声使处理效果更加明显，添加椒盐噪声得到如图 2-10b)所示，对其进行均值滤波操作得到如图 2-10c)所示，由于均值滤波使用像素附近的均值，所以该操作对图像细节破坏比较严重，因此图 2-10c)看起来较为模糊。



a) 原图 b) 添加椒盐噪声 c) 均值滤波

图 2-10 均值滤波效果图

(2) 高斯滤波

当噪点遵从正态分布式时，这样的噪声称之为高斯噪声^[30]。高斯白噪声包括热噪声和散粒噪声。对于正态分布的噪声，使用高斯平滑滤波器可以达到很好的滤波效果。高斯滤波同样也是基于邻域平均值的方法，但是需要对每个邻域像素进行加权处理，中心像素点权值最大，通过加权后像素相加取其平均值，将该值作为当前像素值。该方法可以最大限度的对原有数据保留，同时还可以一定程度上减弱图像噪声，但是该方法对于一些离散孤点噪声处理效果并不是很明显。对图 2-10b)进行高斯滤波处理，处理效果如图 2-11 所示。



图 2-11 高斯滤波图

(3) 中值滤波

在 1971 年 J.W.Tukey 发表中值滤波器的理论^[31]。经常应用于一维和二维信号的

处理，该技术应用十分广泛。

中值滤波理论基础是基于数学的排序统计，所以是一种非线性滤波。主要思想是对像素排序，利用区域中排序后的中间值作为当前的像素值。该算法在去除椒盐噪声和脉冲噪声有很好的效果，并且相对其他滤波操作可以很好的对边缘信息进行保留。由于是中间值代替当前值，所以较为接近该点的真实值。在傅里叶空间中中值滤波可以对其在高频分量进行减弱与消除，同样会改变低频分量。利用 3×3 模板进行中值滤波得到的效果如图 2-12 所示。



图 2-12 3×3 中值滤波图

对上述的三种方法进行分析，三种方法都可以将噪声进行一定程度上的减弱，中值滤波对椒盐噪声有很好的抑制作用，同时可以很好保留原始图像的边缘信息，但是均值滤波和高斯滤波相对处理效果并不是特别理想。因此，本文使用中值滤波进行图像处理。

2.5.3 增强对比度

增强对比度是将图像中的亮度值范围拉伸或压缩成显示系统指定的亮度显示范围，从而提高图像全部或局部的对比度。对于图像来说对比度越强，图像越清晰，对比度越小，图像越模糊。当图像背景和目标都比较暗淡或高亮时，可以通过增强对比度操作，使图像变得清晰。

(1) 直方图均衡化

直方图均衡化(Histogram Equalization)又叫做直方图平坦化，其主要是通过将图像进行非线性的拉伸，从而达到增强对比度的效果^[32]。对比度表示黑与白的比值，比值越大，从黑到白的渐变层次就越多。直方图均衡化是计算每级灰度值在图像中出现的频率，将图像中比较集中的灰度值区域将其拉升使其变得均匀。直方图均衡化是一个可逆的操作，但是在处理过程中没有选择性，有可能将不必要的噪点进行夸大，使其图像中 useful 信息降低。对图 2-13 中值滤波图进行直方图均衡化得到的效果如图 2-14 所示。

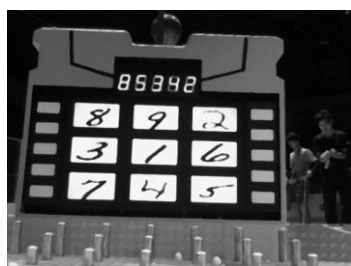


图 2-13 中值滤波图



图 2-14 直方图均衡化图

(2) 掩模操作

图像掩模操作和滤波操作相同，都是利用模板对原图进行卷积的过程，从而改变原始数据，得到想要的结果。从数学角度上说利用邻域像素值对当前像素值进行加权处理，从而改变该像素与邻域像素之间的影响关系，达到增强对比度的效果。

图像对比度增强所使用的函数为：

$$f(i, j) = 5 \times f(i, j) - [f(i+1, j) + f(i, j+1) + f(i-1, j) + f(i, j-1)] \quad (2-12)$$

对图 2-13 进行掩模操作，得到的图像对比度增强效果如图 2-15 所示。

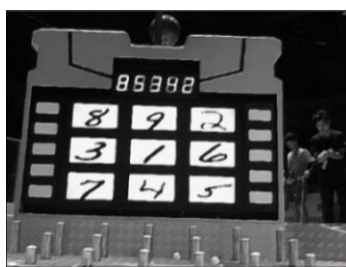


图 2-15 掩模操作处理图

通过两种方法的分析可知，直方图均衡化处理效果并不理想，由于本文研究对象的数字面板整体表现为白色，所以通过直方图均衡化会使白色面板区域四周变得明亮，更不利于对面板的提取。该方法多用于凸显整体偏暗或偏亮的情况。掩模操作可以很好的增强滤波处理后的对比度，因此本文使用掩模操作进行处理。

2.5.4 图像二值化

通过上面算法的处理，可以得到比较好的图像。图像中由于背景信息的存在会对后续的识别带来一定的困难，同时由于背景信息的存在会降低处理速度。本文使用二值化处理进一步将背景信息去除掉。

二值化处理，其表现形式是黑白图像，图像上点的灰度值只有 0 与 255。这样的操作不仅简化信息，而且可以进一步对信息的筛选。其主要原理就是利用阈值进行图像划分，当前像素值高于设定值将该像素值设定为 255，反之，当低于设定值将当

前像素值设定为 0。其二值化处理的公式为：

$$g(x,y)=\begin{cases}0 & ,f(x,y)<T\\255,& f(x,y)>T\end{cases}$$

(2-13)

式中 $f(x,y)$ ——原图像灰度值；

$g(x,y)$ ——处理后图像灰度值；

T ——设定阈值。

由上式可知，选取合适的阈值非常重要。设定阈值过小会使背景信息分割过少，阈值设定过高会使目标信息流失。通过设定不同的阈值进行阈值分割对比效果图，设定阈值分别为高、中和低三种阈值，得到效果如图 2-16 所示。

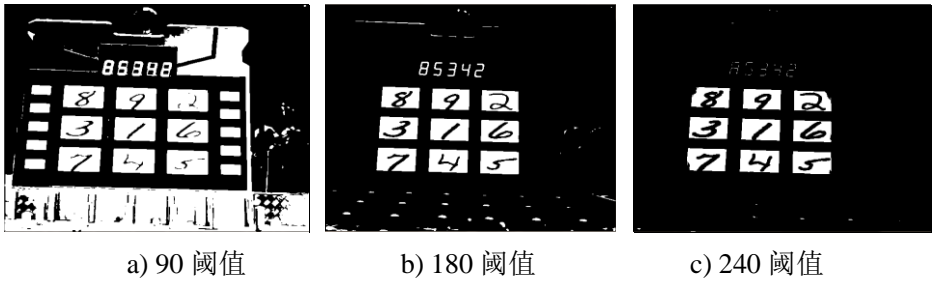


图 2-16 二值化

从上图的处理效果上看，在设定阈值为 90 时，得到的二值化图像中数字面板四周环境信息保留过多；设定阈值为 240 时，阈值设定过高使目标信息缺失；设定阈值 180 阈值时处理效果最好，所以一个阈值的选定，对分割结果影响很大。常见的阈值分割方法性能比较如表 2-7 所示。

表 2-7 常见阈值分割性能比较

阈值分割算法	适用图片	分割效果	计算量	处理速度	缺点
双峰法	背景和目標对比度强烈	较好	小	快	对比度较小不明显
最大熵法	低对比度的复杂图像	较好	小	较快	一维最大熵法抗噪性差
迭代法	所有图像	良好	一般	一般	细节难以区分，不稳定
Otsu 双阈值法	直方图呈三峰特性	良好	大	慢	算法速度慢

本程序需要实时进行检测，因此需要的处理速度越快越好。通过上面阈值分割方法的比较，处理速度较快的有双峰法、最大熵法。由于本文需要时效性，所以双峰法并不适用，因为双峰法需要人根据图像的直方图来确定阈值，而且需要通过

多个阈值比较才能得到较为理想的阈值。迭代法阈值分割是对双峰法的改进，可以根据不同的图像自动设定阈值，但是该方法相对于最大熵法处理速度较慢。所以综合考虑本文选用最大熵法进行阈值分割。

利用图像熵为准则进行图像分割是由 Kapure 提出来，现在仍然使用较广的一种图像熵分割方法^[34]。熵是来自于信息论中的信息熵，其表示系统中所含信息的一种度量。简单的说，信息熵越大，表示系统中的消息越加混乱，反之熵越小，信息越加有序。其数学表达式为：

$$H = \int_{-\infty}^{+\infty} p(x) \log p(x) dx \quad (2-14)$$

式中 $p(x)$ ——灰度 x 在系统中出现的频率；

H ——信息熵。

最大熵阈值算法思想：

- (1) 首先选定一个阈值 T ，以 T 为基础将图像分为两份，分别为高于阈值 T 的像素点作为目标类 M ，低于阈值 T 的像素点作为背景类 N ；
- (2) 然后分别计算 M 与 N 中每个灰度值所占的比例；
- (3) 分别计算背景和目标的信息熵；
- (4) 对图像中的每个灰度级进行步骤 1-3 运算， $\max(H_M + H_N)$ 时得到的 T 值作为最大熵的分割点。

经过实验，最大熵法没有达到预期的效果，如图 2-17a)所示。由于数字面板对比度较高，所以阈值分割的值会偏低。进行 10 组最大熵与传统分割比较如表 2-8 所示。

表 2-8 最大熵与传统分割比较

序号	最大熵阈值	经验阈值分割	阈值差
1	108	193	85
2	110	198	88
3	113	195	82
4	112	192	80
5	105	198	93
6	114	201	87
7	116	194	78
8	109	198	89
9	107	192	85
10	115	203	88

根据实验表明使用最大熵阈值法和经验阈值有一定的差别，将阈值偏差作为校正偏差值，最后得到校正公式为：

$$T = T_a + 85.5 \quad (2-15)$$

式中 T_a ——最大熵阈值分割阈值；

T ——最终阈值分割阈值。

利用最大熵法与改进最大熵法二值图像的效果如图 2-17 所示。改进算法流程图如图 2-18 所示。

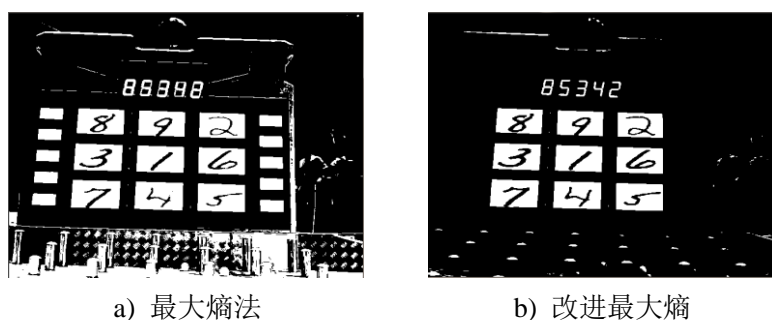


图 2-17 最大熵阈值分割

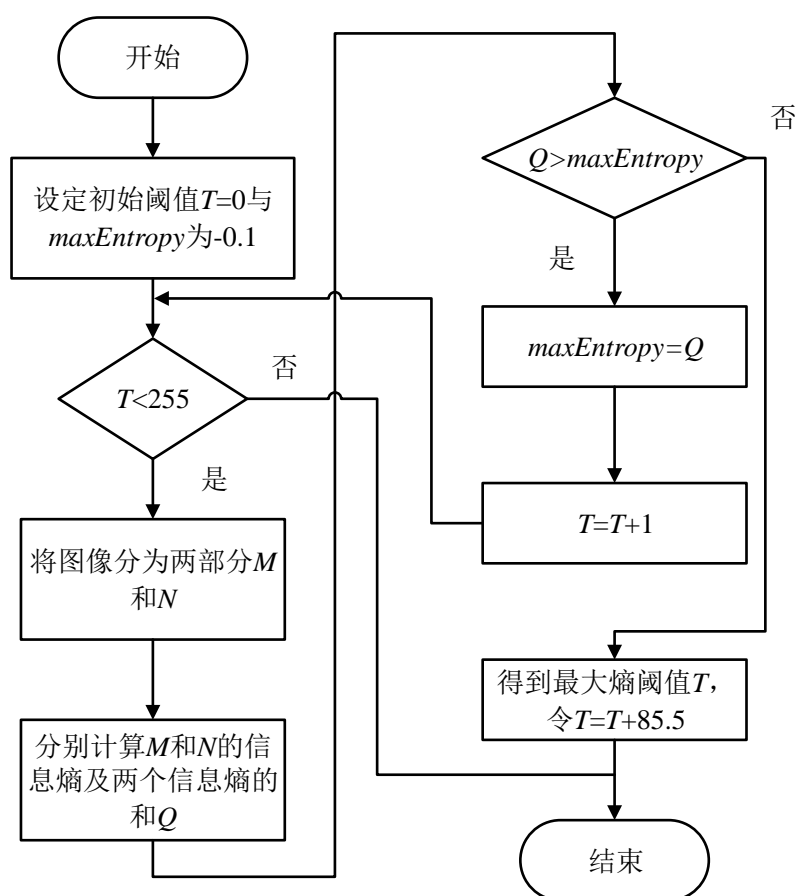


图 2-18 改进最大熵算法流程图

2.6 本章小结

本章首先对图像采集系统的设计，根据标定理论对摄像头进行标定实验；然后对采集图像进行图像灰度化，减少图像处理数据，加快处理速度，并通过常用滤波方法进行对比实验，观察滤波处理效果选用中值滤波进行降噪处理；其次利用掩模操作增强图像的对比度，降低图像滤波对图像边缘信息的影响；最后利用改进最大熵法对图像进行阈值分割，得到理想二值化图像。通过以上操作不仅仅减少了处理数据，而且提高了图像质量，为下一章目标定位与分割处理做准备。

第3章 字符区域识别与分割

3.1 引言

字符区域的识别与分割是本文的重点。定位字符的准确性，直接影响字符分割与识别的结果。在定位字符区域的位置后，进行下一步的字符分割和数字识别操作。较高的定位识别准确性，才能保障下一步数字识别的整体识别率。本章主要分为两部分内容：第一部分是对字符区域的识别与定位，主要是利用边缘检测与几何关系的相结合实现对字符区域的准确定位；另一部分是基于边缘特性实现对字符区域的二次分割，将图像中 9 个数字完整的分割提取出来。数字识别系统的整体流程如图 3-1 所示。

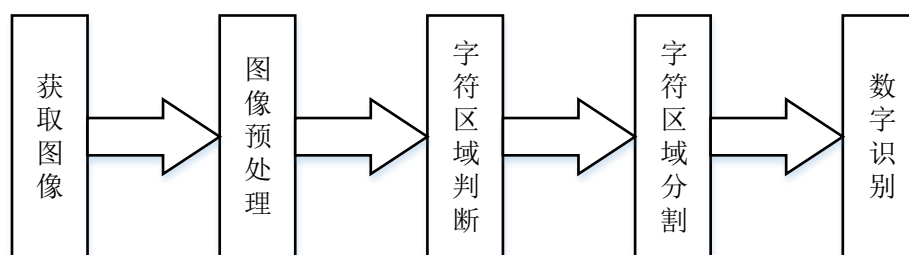


图 3-1 目标识别流程图

3.2 字符区域定位的研究

根据图像中信息，定位到所需要的位置信息，定位的准确性影响到对数字识别的准确率。因此字符区域定位在整识别系统中占有重要地位。下面是比较常见的定位算法。

3.2.1 基于颜色特征目标定位算法

有些物体颜色相对环境比较突出，通过颜色对目标进行定位较为容易实现，但是颜色阈值范围易受环境光照的影响。利用颜色特征对目标进行定位中，较为常见的是对汽车车牌的识别定位系统，车牌有着比较深的蓝色或黄色信息，相对环境较为突出。目标特征颜色比较明确的情况下，定位的准确率相对较高，如果当目标背景中含有和目标较为相似的颜色时，就会降低定位的准确性，这时需要利用目标的一些特性相结合的办法进行定位，比如说目标特征的矩形或圆形等。所以此方法适

合与一些其它方法相结合使用，进行相互取长补短^[35,36]。

3.2.2 基于纹理特征的目标定位算法

纹理特征相对较为突出时，可以使用纹理进行定位。纹理信息的好处是图像纹理不会因为环境的变化而改变，该方法具有旋转不变性，并且有较好的抗噪性^[37]。而目标的颜色阈值会因环境光照的变化而发生变化。

3.2.3 基于边缘检测的目标定位算法

该方法是利用边缘检测进行定位^[38]。算法相对比较简单，易实现，但是弊端也较为明显，当环境较为复杂时，检测出来的边缘信息较多使定位变得困难，所以该方法对图像质量有较高要求。

定位识别算法中，没有哪个算法适用于一切环境。需要根据不同的环境，选取不同的算法。实际的应用过程中，对于复杂环境单独使用某一个算法，并不能很好地将目标进行定位。

3.3 基于边缘信息和几何关系的研究

3.3.1 字符区域特征

数字识别对象是数字面板上的手写数字，所以需要将面板识别出来，具体的字符区域规格如图 3-2 所示。

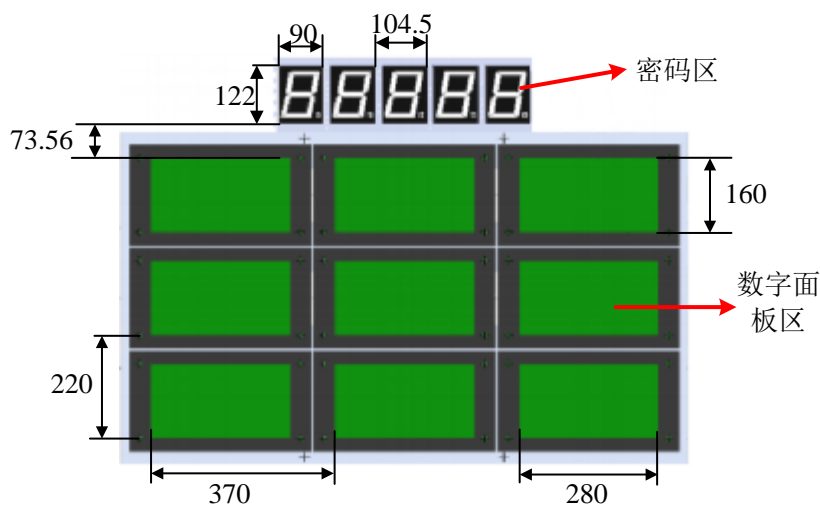


图 3-2 九宫格规格图

字符区域屏幕的表面为九宫格，每个格子都会显示一位手写数字，在九宫格屏

幕上部 LED 灯屏为密码显示区，随机显示 5 位不重复的阿拉伯数字密码，智能小车依次按照 5 位阿拉伯数字进行正确打击，每次打击完成之后数字面板的数字位置变化一次。开始时智能小车寻找字符区域位置，进行判断是否为所需要的目标，当为字符区域，智能小车对九宫格中任意一个格子进行弹丸射击，密码区将显示 5 位密码，九宫格随机出现 1 至 9 九个不重复的手写体阿拉伯数字，这时字符区域被激活。如果在 4.5 秒内没有受到继续击打，则密码区和九宫格关闭显示数字。具体的打击流程图如图 3-3 所示。

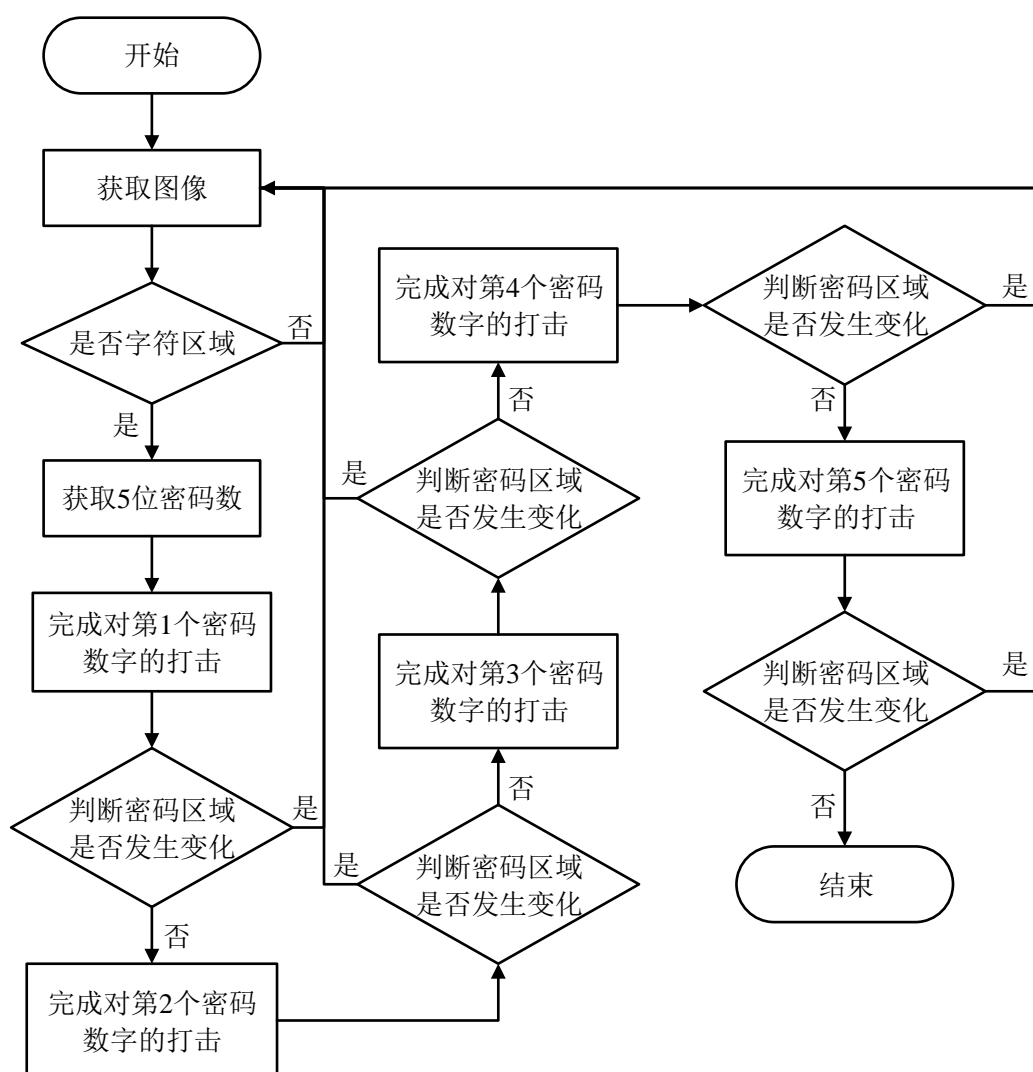


图 3-3 字符识别打击流程图

3.3.2 本文改进定位算法

由于环境的复杂性，只根据单个特征并不能将字符区域识别出来。所以根据本

文的研究对象的纹理和特征,提出基于边缘检测和几何关系相结合的目标定位算法。传统边缘定位算法对图像进行边缘检测,检测出来的边缘信息过多,如果只利用边缘信息,并不能将目标精确定位,这就会对后期图像分割造成很大的影响。本文利用字符的几何关系进行筛选,从而达到进行一步确定目标位置。具体的流程如图 3-4 所示。

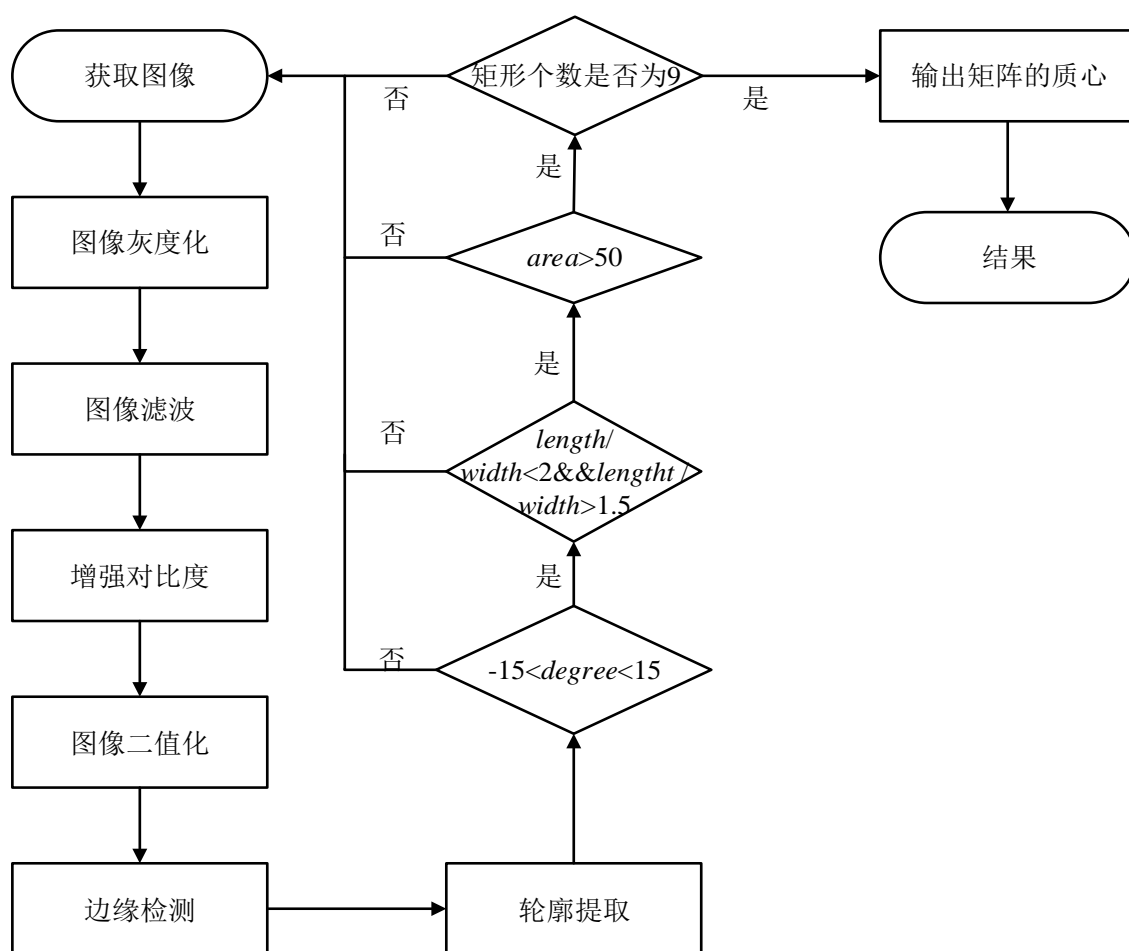


图 3-4 九宫格定位流程图

具体步骤是:

- (1) 在获取到图像后,利用图像灰度化降低图像处理数据,然后通过图像滤波降低图像噪点;
- (2) 经过掩模操作增强图像对比度,减少滤波对边缘信息的损害;
- (3) 利用改进最大熵法对图像二值化处理,使信息进一步简单化;
- (4) 通过图像边缘检测,检测出图像中边缘信息;
- (5) 利用最小包边矩形,存储每个连通区域的边缘信息;
- (6) 通过判断矩形特征(长宽比、矩形倾斜角度范围)进一步对目标信息筛选,最

后通过所有满足条件的矩形数量来判断是否为目标。

3.3.3 边缘检测

边缘就是图像中像素点发生剧变的位置。边缘检测的主要目标是突出字符区域而非字符区域边缘信息要尽可能的减少^[33]。边缘检测首先进行滤波操作，边缘检测对噪声比较敏感，主要是算法的核心就是对图像强度的导数，导数对噪声是比较敏感的，所以滤波是必备的操作；然后增强图像对比度，为了使边缘信息突显出来，需要进行图像对比度增强，将边缘信息处像素值拉大；最后对边缘进行检测。下面简单的介绍 Sobel 算子和 Canny 算子两种较为常见的边缘算法。

(1) Sobel 算子

该算子是基于离散微分的算子。根据微分求导，求出像素点的梯度矢量或法矢量。Sobel 有两个核函数分别为水平方向和垂直方向。

水平方向：将原始图像 A 与一个奇数的核函数进行卷积操作，如下面核大小为 3 时，计算结果为：

$$\mathbf{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \cdot \mathbf{A} \quad (3-1)$$

垂直方向：将图像 $f(x, y)$ 与一个奇数的核函数进行卷积操作，如下面核大小为 3 时，计算结果为：

$$\mathbf{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \cdot \mathbf{A} \quad (3-2)$$

图像的每一个像素的横向及纵向梯度近似值可用以下的公式结合，来计算梯度的大小：

$$G = \sqrt{G_x^2 + G_y^2} \quad (3-3)$$

为了提高效率，可以简化算法为：

$$G = |G_x| + |G_y| \quad (3-4)$$

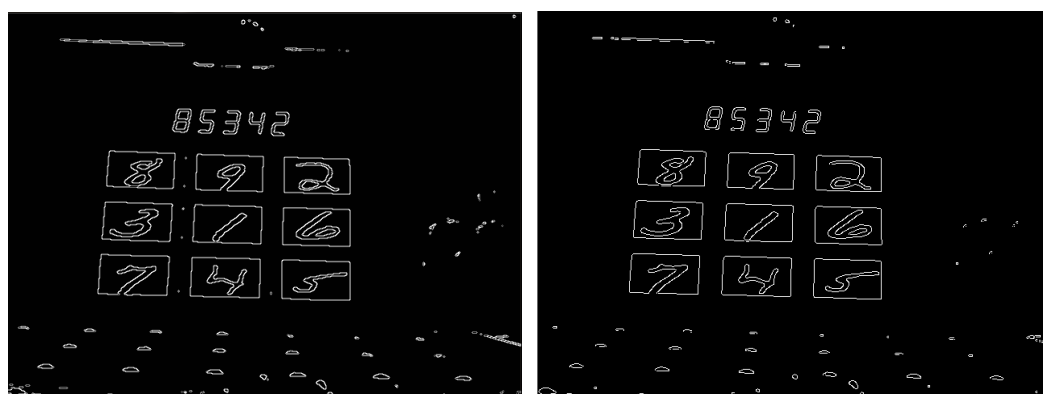
(2) Canny 算子

Canny 算子是常用的边缘检测中较为强大检测算法之一。该算法由 John F.Canny

研究出来。Canny 算子首先滤波降低噪声，然后计算梯度幅值和方向，这个步骤和 Sobel 算子相同使用水平和垂直方向上的两个核，然后计算幅值和方向，其梯度方向一般取 0° 、 45° 、 90° 、 135° ；其次进行非极大值抑制，这一步排除非边缘像素，保留一些细线条，最后进行滞后阈值，设定两个阈值，将图像分为三段，低于低阈值去掉，中间阈值与高阈值相连保留，没有相连的去除，高阈值的保留。

通过上面的步骤可以看出 Canny 算子是一个多级边缘检测算法。最优边缘算法的标准就是要求低错误率、高定位性和最小响应。

利用两种方法对第 2 章中图 2-17b) 进行边缘检测得到的效果如图 3-5 所示。



a) Soble 算子边缘检测

b) Canny 算子边缘检测

图 3-5 两种边缘检测效果图

根据两种效果图和要求，本文最终选用 Canny 算子进行边缘检测。

3.3.4 边缘提取

Canny 算子将边缘信息检测出来，本文定位算法需要对边缘信息进行判断，然而 Canny 算子检测出来的曲线是由一些单独孤立的像素点组成，并没有将曲线轮廓作为一个整体。所以下一步工作就是将检测出来的每条曲线中所有像素点组装成轮廓。OpenCV 函数库中 findContours 函数^[39]可以将所需要的边缘信息提取出来，数组集 contours 可以将每一条曲线中像素点组装成一个整体存储起来。最后通过 drawContours 函数将保存起来的曲线画出来，查看是否为所需要曲线。

findContours 检测出来每条曲线之间的关系如图 3-6 所示。图中有 5 条曲线轮廓，将其标号为 0-4，图中 0 号在最外层，没有同级别曲线，1 号和 2 号为 0 号子级曲线，1 号和 2 号为同级区域，1 号是 2 号前一个曲线，没有子级曲线，2 号没有后一曲线，没有后一曲线，2 号子级曲线为 3 号和 4 号。

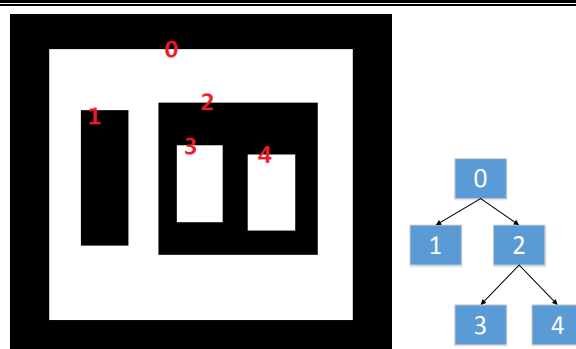


图 3-6 边缘提取轮廓关系图

`findContours` 函数在寻找曲线时，需要检测出来的曲线模式如表 3-1 所示。

表 3-1 `findContours` 函数可选的轮廓检索模式

标识符	含义
<code>RETR_EXTERNAL</code>	表示只检测最外层轮廓。对所有轮廓，设置 <code>hierarchy[i][2]=hierarchy[i][3]=-1</code>
<code>RETR_LIST</code>	提取所有轮廓，并放置在 <code>List</code> 中。检测轮廓不建立等级关系
<code>RETR_CCOMP</code>	提取所有轮廓，并且将其组织双层结构(顶层为连通域的外围边界，次层为空的内层边界)
<code>RETR_TREE</code>	提取，立网状的轮廓结构

本文只要知道外轮廓信息就行，所以使用 `RETR_EXTERNAL` 轮廓模式。利用 `drawContours` 函数将每个轮廓画出来，得到的效果如图 3-7 所示。



图 3-7 边缘轮廓线图

3.3.5 几何关系判定

通过上述方法得到的边缘信息，可以看出图像中包含许多背景边缘信息，所以需要进一步处理将所有背景边缘信息去除掉。

(1) 最小包边矩形

几何关系判定主要是判断矩形倾斜角度、长宽比和面积是否在设定值范围内，所以需要将每个边缘信息用最小包边矩形框选出来。利用 `OpenCV` 函数库自带的函数最小包边矩形，得到效果如图 3-8 所示。

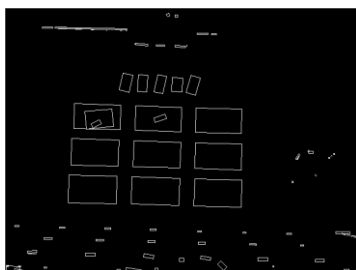


图 3-8 最小包边矩形

从图中可以看出将 9 个矩形边框完好框选出来,同时图中还含有好多错误矩形,所以需要通过下面的判断条件进行判断,将错误的信息去除掉。

(2) 角度判断

从图 3-8 中可以看出图像中每个矩形都有不同程度的倾斜角度,由于字符区域的矩形由于拍摄角度的问题,都会出现不同程度的畸变,主要的畸变情况如图 3-9 所示。

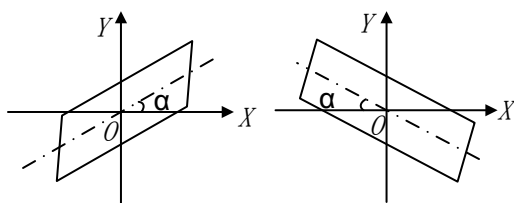


图 3-9 字符倾斜方式

数字显示屏可能发生倾斜现象,但其倾角并不会特别大,因此通过设定矩形倾斜角绝对值在 15° 范围内,可以将部分较大倾角的矩形去除掉。对图 3-8 进行角度判断后得到效果如图 3-10 所示。

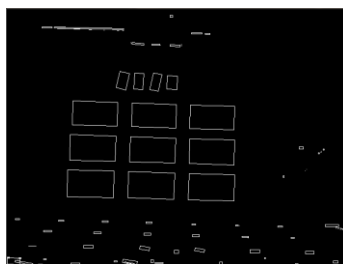


图 3-10 角度判断

通过角度判断后,与图 3-8 相比较,图 3-10 明显将一些不要矩形去除掉。

(3) 长宽比判定

由图 3-2 可知 9 个数字面板固定长宽比为 1.75,由于拍摄角度,矩形会发生较小的变形,因此在长宽比判断范围为:

$$1.5 < \frac{length}{width} < 2 \quad (3-5)$$

对图 3-10 进行长宽比得到效果如图 3-11 所示。

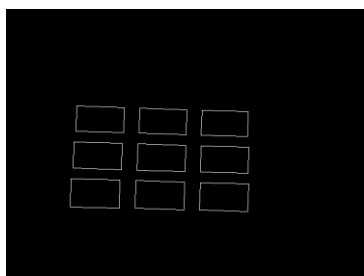


图 3-11 长宽判断

(4) 面积判定

为了保证数字面板定位的准确性，利用面积进一步确定是否为数字区域。面积的判断可以将目标与摄像头距离较远的情况滤除掉。

3.4 图像质心

图像矩具有平移、旋转、尺度等特征不发生改变的特点，因此也可以称为不变矩。经常用来计算几何特征，如质心、方向等。将图像看成二维薄片物体，每点像素值作为该点的密度，得到的密度函数为 $f(x, y)$ ，二维离散数学表达式为：

$$M_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q f(x, y) \quad (3-6)$$

其中， p, q 表示矩的阶。

M_{00} 零阶矩阵表示所有像素点的质量。用上式的一阶矩可以将图像的质心计算出来，质心表达式为：

$$\begin{cases} x_0 = \frac{M_{10}}{M_{00}} \\ y_0 = \frac{M_{01}}{M_{00}} \end{cases} \quad (3-7)$$

将图 3-11 识别出来的 9 个矩形，使用式(3-7)的质心算法公式得到矩形中心坐标，运行结果如图 3-12 所示。

矩形的宽:48.6843	矩形的长: 84.5689	矩形的中心: [383.491, 332.274]
矩形的宽:48.7049	矩形的长: 86.1423	矩形的中心: [270.637, 328.989]
矩形的宽:49.3268	矩形的长: 86.5596	矩形的中心: [156.513, 325.952]
矩形的宽:46.3988	矩形的长: 84.162	矩形的中心: [384.539, 266.112]
矩形的宽:46.3727	矩形的长: 84.4461	矩形的中心: [273.402, 262.724]
矩形的宽:46.9771	矩形的长: 85.3646	矩形的中心: [161.099, 259.347]
矩形的宽:43.9222	矩形的长: 81.7	矩形的中心: [384.78, 201.964]
矩形的宽:44.0963	矩形的长: 82.909	矩形的中心: [275.829, 198.481]
矩形的宽:44.4563	矩形的长: 84.019	矩形的中心: [165.49, 194.775]

图 3-12 运行结果

矩形中心坐标如表 3-2 所示。矩形的序号是从左到右，从上到下依次为 1-9。

表 3-2 质心坐标

序号	像素坐标(像素)
1	(165.49,194.775)
2	(275.829,198.481)
3	(384.78,201.964)
4	(161.099,259.347)
5	(273.402,262.724)
6	(384.539,266.112)
7	(156.513,325.952)
8	(271.637,328.989)
9	(383.491,332.274)

3.5 字符分割

简单的说就是将图像分割成互不相关的几个区域。一般的分割方法有阈值分割、边缘分割、颜色分割、区域分割等。

3.5.1 常见的分割方法

(1) 灰度阈值分割法，灰度阈值分割法是最为常见的分割方法^[40]。只需设定阈值就可将背景和目标分割开来，其关键点就是阈值的设定。将高于阈值的为目标，低于阈值的为背景，或高阈值的为背景低阈值的为目标。

(2) 颜色分割法，当图像中物体与背景有明显颜色差异时，可以利用颜色阈值进行图像分割。该方法最大缺点是不同的环境光强，会对颜色阈值造成影响，如果还用原来设定颜色阈值进行分割，有可能会出现分割不完整的现象。

(3) 边缘分割法，在图像中物体有明显边缘轮廓的情况下，而且环境较为简单可以使用边缘分割法。边缘检测的图像分割方法的基本思路是先确定图像中的边缘像素，然后再把这些像素连接在一起就构成所需的区域边界。常见的边缘检测算法已经在第 3 章第 2 小节讲过。

(4) 区域分割法，传统的区域分割方法有区域生长、区域分裂和两种方法的结合，其中最基础的是区域生长法。区域生长是根据事先定义的准则将像素或者子区域聚合成更大区域的过程。区域生长首先选择合适的生长点；然后确定相似性准则即生长准则；最后确定生长停止条件。一般来说，在无像素或者区域满足加入生长区域

的条件时，区域生长就会停止。

本文需要将每个数字分割出来，上面几种方法在对于简单的图像的分割有很好的效果，但是不能将区域分割出来。

3.5.2 基于字符轮廓的分割

字符分割的主要目的就是为数字识别做准备。OpenCV 函数库中的 ROI 函数可以利用表 3-2 得到的每个矩形的质心和矩形边长数据将 9 个矩形分割出来。利用 ROI 函数将九个数字都分割出来的效果如图 3-13 所示。

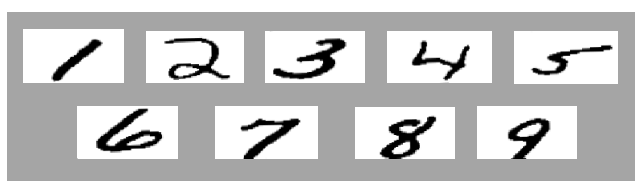


图 3-13 分割 9 个白色面板

在分割出面板后，为了分割图像和训练数字图像相同需要进一步处理，同时由于训练数字图像是黑底白字，bitwise_not 函数对分割图像进行颜色反转操作，最终效果如图 3-14 所示。



图 3-14 数字分割图像

经过多次实验，随机选择对不同数字面板定位运行时间如表 3-3 所示。

表 3-3 定位系统运行时间

序号	1	2	3	4	5	6	7	8	9	10
运行时间(s)	0.234	0.245	0.259	0.247	0.239	0.273	0.246	0.251	0.274	0.269

由于计算机自身状态，多次试验结果产生一定的差异，但是定位系统程序运行时间都保证在 0.28s 时间范围内，完全满足实验要求。

3.6 本章小结

本章重点研究目标定位算法，首先对传统定位算法进行比对分析；然后通过研究文中目标的特征，提出了一种基于边缘和几何关系相结合的定位算法，完成数字面板的精确定位；最后通过 ROI 函数和数字位置信息将数字面板上数字分割出来。

第4章 基于SVM的数字字符识别

4.1 引言

在二十世纪九十年代中期 Vapnik 和 Corinna Cortes 等人提出支持向量机(Support Vector Machine)的机器学习方法,主要的理论依据是 VC 维理论和统计中最小化原则。SVM 处理一些非线性和小样本分类等问题有很好的效果^[41,42]。在机器学习中, SVM 比传统的机器学习要好, SVM 在模式识别或回归问题等方面得到了较为普遍的应用。支持向量机就是在统计学习理论的基础上发展出来的,它在解决小样本、非线性以及高维问题表现出特有的优势。尽管目前统计学习理论和支持向量机方法还有很多问题需要进一步研究,但它们正在成为继神经网络之后的机器学习领域新的研究热点,也将推动机器学习理论和技术的发展。

4.2 支持向量机概述

SVM 主要是通过寻找一个超平面将训练样本分割开来,利用超平面来进行分类,将问题变成为二次规划问题。核函数可以解决线性不可分的问题,核函数的作用就是将样本从不可分低维度空间变成可分高维度空间,再利用超平面将训练样本特征分割开来。

SVM 算法中,选择一组特征子集(Support Vector, 简称 SV),核函数根据特征子集将样本整体数据进行划分。经过这样的处理不仅可以大大提高目标识别的准确度,同时还可以降低算法在计算时的计算量和程序总的占用量。支持向量机在实际工程中就是完成一些分类问题,它的好处就是在保证完成分类之后,还可以保证分类的准确性和快速性。

(1) SVM 算法只是通过一组样本的特征子集进行分类的,但并不是靠采集到的样本的每个样本, SVM 算法一系列的计算仅仅围绕采集样本的整体特征子集。所以最终的计算结果对样本空间的维数影响并不是很大,这样不仅减少了计算量,同时还避免了由于数据量多出现的错误,从而提高数字识别的准确性。

(2) SVM 算法是通过特征向量来决定最终的分类结果,而整个算法过程中并不是每个特征向量都会用到,所以在所有特征向量中会出现多余的特征向量,可以对这

部分向量删除掉，从而降低训练时间和减少运行内存。

(3) 通过参阅资料可知支持向量机算法有很强的推广能力，它的适用范围相对比较广泛，可以应用到很多领域实践中，而且最终的结果还是比较有说服力的。

(4) SVM 算法模型建立相对来说比较简单，不用太多人为的干扰，可以应用于多分类情况，所以 SVM 算法具有普遍性。

(5) 由于 SVM 算法模型简单和人为干扰少，所以相对实验结果就比较客观，因此 SVM 算法进行分类可以很大程度上反应实际数据的分类情况。

(6) SVM 算法在非线性不可分的情况下，可以将样本从低维空间映射到高维空间，因此 SVM 算法可以解决一些高维数据问题^[43]。

4.3 支持向量机原理

现代统计学理论是支持向量机算法的理论依据，SVM 算法主要是去解决有限的样本容量的分类和识别问题。SVM 算法通过寻找一个最优超平面将样本数据分成两份，因此 SVM 算法分类问题转变成寻找一个超平面问题，如图 4-1 中 L_1 所示。所谓的最优分类超平面不仅将样本数据准确的分割成两份，并且样本到平面的距离最大，这个距离叫分类间隔。前面准确无误的将样本分成两类是保证经验风险达到最小值，而尽量寻找距离最大的面是使算法数据更加接近数据本身，降低实际风险。最优超平面公式：

$$\vec{w} \cdot \vec{x} + b = 0 \quad (4-1)$$

式中 $\vec{w} \in R^d$ ——分类超平面的法向量；

$b \in R$ ——阈值。

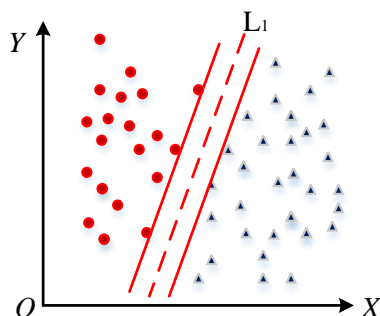


图 4-1 线性可分情况下最优超平面

当给定随机样本数据对其进行分类时，使用最优分类超平面进行分割，该样本被分类到正确的一面的几率就越高，同时被误分类的概率就会越小，准确率就会越

高，最终的预测准确率也会越高。通过平行于 L_1 的两个红线直线的样本点称之为支持向量。所以支持向量的确定也就意味着超平面的确定。

(1) 在线性可分的情况下，在 N 维空间中，必定可以找到超平面。超平面中 b 为分类面的偏置项，该值表示了分类面到原点的间距^[44]。当找到正确超平面将样本分隔开，就会对样本中任意的样本点都有一定不等式成立：

$$\begin{cases} \vec{w}^T \vec{x}_i + b \geq +1, y_i = +1 \\ \vec{w}^T \vec{x}_i + b \leq -1, y_i = -1 \end{cases} \quad (4-2)$$

(2) 当出现线性不可分情况，再用上面优化问题的方法来解决现在的问题，就会找不到可行的解。为了更好的解决线性不可分的问题，Vapnik 提出了新的概念软间隔^[45]。也就是说通过超平面分类样本数据后使被错误分类的样本数据变少，如图 4-2 所示。引入新变量 $\xi_i > 0$ (松弛变量)，那么约束条件变为：

$$\begin{cases} \vec{w}^T \vec{x}_i + b \geq +1 - \xi_i, y_i = +1 \\ \vec{w}^T \vec{x}_i + b \leq -1 + \xi_i, y_i = -1 \end{cases} \quad (4-3)$$

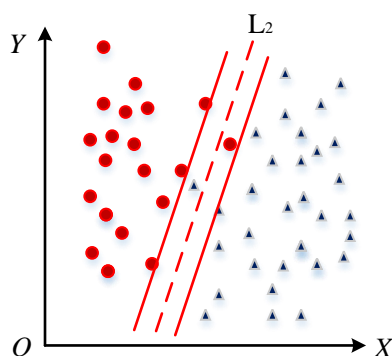


图 4-2 非线性可分情况下最优超平面

(3) 实际应用中样本数据大部分是相互掺杂在一起，在这种情况下不能寻找到超平面将样本数据分隔开，Vapnik 解决这个问题的办法是通过构造一个核函数，将样本数据从低维空间映射到高维空间中，这使得样本在低维不可分变到高维可分^[46]。从而在高维空间中寻找一个超平面将样本分割开。

SVM 的核心内容之一就是核函数^[47-49]，引用核函数就是解决从低维不可分到高维可分问题，但是这种操作将计算复杂程度大大增加。核函数一般形式为：

$$K(x_1, x_2) = \varphi(x_1) \cdot \varphi(x_2) \quad (4-4)$$

ϕ 为映射函数。在满足 Mercer 定理情况下,核函数 $K(x_1, x_2)$ 才有效。经常应用的核函数如表 4-1 所示。

表 4-1 核函数类型

核函数类型	核函数公式
线性核函数	$K(x, y) = x \cdot y + c$
多项式核函数	$K(x, y) = (ax \cdot y + c)^d$
径向基核函数	$K(x, y) = \exp(-\gamma \ x - y\ ^2)$
高斯径向基核函数	$K(x, y) = \exp(-\frac{\ x - y\ ^2}{\sigma^2})$
Sigmoid 核函数	$K(x, y) = \tanh(\alpha x + c)$

4.4 SVM 多类分类问题

本文需要对 9 种数字进行分类,而 SVM 的本质是解决二分类问题。SVM 在解决多分类的问题^[50,51]主要由两方面的办法:第一种比较直接,通过更改 SVM 目标函数,优化超平面参数,找到一个可以实现多分类目标函数,这种方法看似简单但是实际需要大量的计算和构造,实现起来非常困难。第二种方法利用 SVM 本身的两类分类方法,可以通过二分类组合的方式进行多分类问题。目前较为普遍的是以下三种组合方式:

(1) 一对多法(one-versus-rest)

单独分类不组合,有几类训练几个分类器。例如训练数字 1,就将 1 的样本作为正样本,2-9 数字的样本作为负样本,进行正负样本分类。本文有 9 个数字,需要训练 9 次得到 9 个分类器。

(2) 一对一法(One-Versus-One)

一对一方法是将每个数字两两进行二分类,比如对 1-9 进行分类,需要 1 和 2-9 每个数字进行二分类,然后 2 在和 1 与 3-9 这 8 个数字分别进行二分类以此类推,最后得到 $9(9-1)/2=36$ 个分类器。当一个未知样本输入进来时,样本数据依次通过每个分类器,将匹配度高的,作为最终结果。

(3) DAG-SVM

DAG-SVM 思想来源于 Platt 所研究的 DDAG(决策导向循环图)。OVO-SVM 可能会出现拒分或误分等问题,DAG-SVM 可以很好解决这类问题。这种算法开始时和 OVO-SVM 两种算法都需要进行两两分类,最终得到的分类器个数为 $K(K-1)/2$ 。

不同之处在于，OVO-SVM 需要未知待测样本逐个通过每个分类器根据投票输出最终结果，而 DAG-SVM 不是将样本逐个输入到每个分类器中，而是将所有的二分类器构造成了有向环图的结构，通过这样的操作，得到的分类器有 $K-1$ 个，因此这种操作是效率大大提升。对本文中 9 种数字的 DAG-SVM 决策分类整体过程如图 4-3 所示。

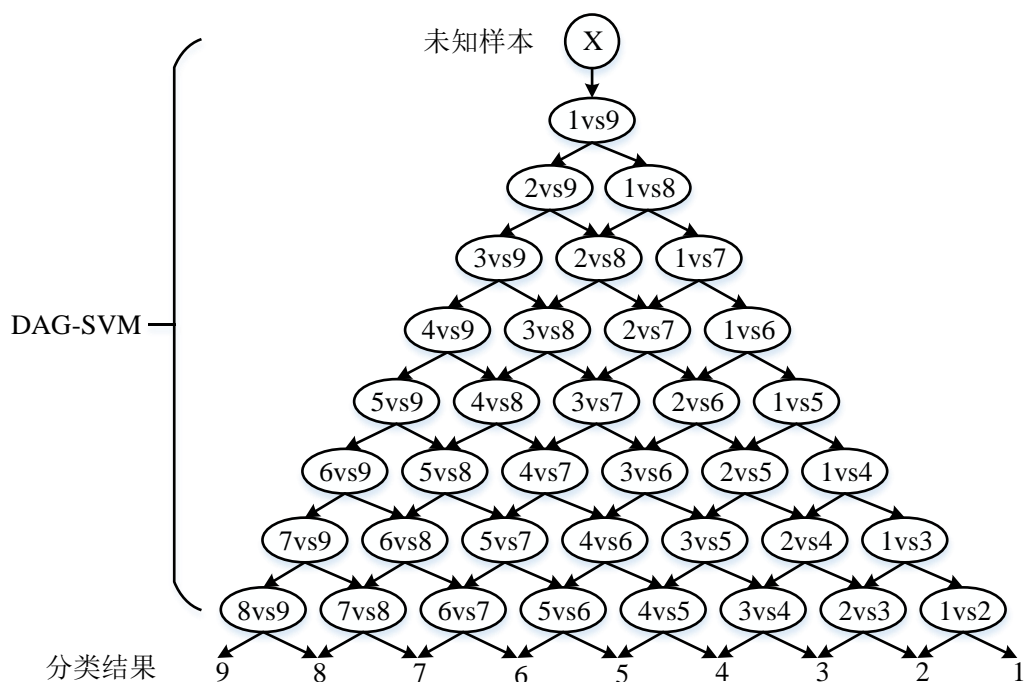


图 4-3 未知样本 DAG-SVM 分类过程

4.5 基于 SVM 数字识别的研究

首先准备样本集；然后训练训练模型；最后通过的到的训练模型识别手写数字。具体的训练过程如图 4-4 所示。

具体的步骤为：

- (1) 准备样本图像；
- (2) 样本图像预处理；
- (3) 图片尺寸归一化；
- (4) 字符图片特征提取；
- (5) 核函数选定：本系统采用 RBF；
- (6) 选取最佳参数 C ，训练样本模型；
- (7) 利用样本数据测试模型的优劣。

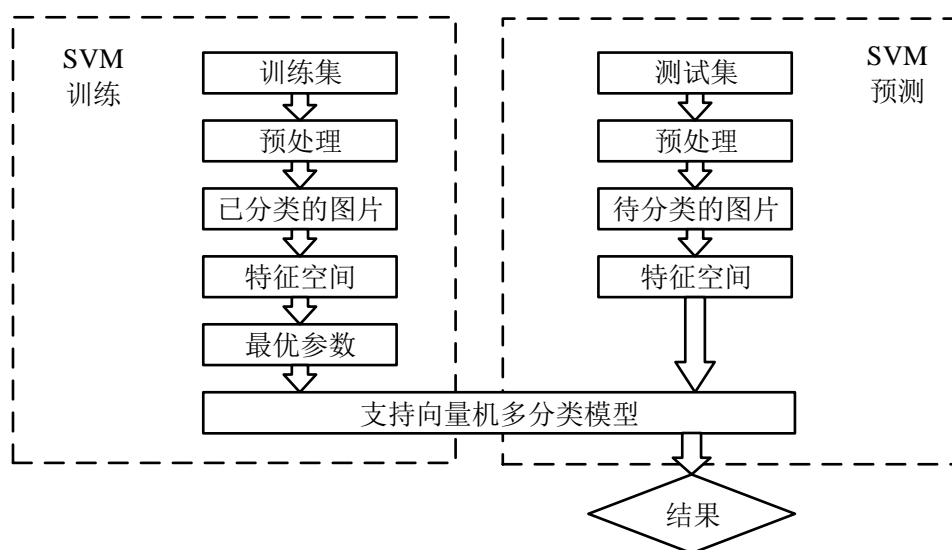


图 4-4 SVM 训练和预算整体流程图

4.5.1 图像尺寸归一化

通过第3章已经知道了字符区域，并将9个数字区域都分割出来，但是分割出来的图像和训练的数字图片大小不相同，利用SVM训练模型识别图像上数字必须要求图片格式与训练图片格式相同。第3章中分割出的数字图片大小和训练图片大小都不相同，图像像素值大小如表4-2所示。

表 4-2 图像尺寸

数字值	1	2	3
尺寸(像素)	84×44	82×44	82×44
数字值	4	5	6
尺寸(像素)	85×47	84×46	84×46
数字值	7	8	9
尺寸(像素)	87×46	86×48	85×49

在训练时，需要样本数据的大小格式相同，所以需要进行尺寸归一化处理，将每个图像的尺寸大小都变成28×28，不仅将图像尺寸进行统一，而且还能减小数据量，加快处理速度。

(1) 尺寸缩放

如果将尺寸为 $M \times N$ 图像缩小到图像尺寸为 $m \times n$ 时，需要将图像进行划分 $(M \times N)/(m \times n)$ 大小互不相交的小块，将每个小块中数字的平均值，作为缩小后的

像素值。对第3章分割得到的图像进行缩放,得到的手写数字缩放效果如图4-5所示。



图 4-5 传统尺寸改变

明显可以看出缩放后数字对比原来的数字变得模糊。主要是该算法是采用周围的均值进行计算的,二值图像只有两个值,所以取得均值后的数值肯定小于或等于255,图像变得模糊。

(2) 本文尺寸缩放

本文数字图像缩放的具体步骤是:首先将图片预处理得到二值化图像,再经边缘检测提取出数字外部边缘信息,利用 ROI 分割出数字区域;然后通过均值法进行缩放到 26×26 大小;其次通过补填图像背景,使数字图像在整个图像的中心位置;最后通过图像二值化将数字变得清晰。具体的算法流程图如图4-6所示,并以数字8的图像为例用该算法流程进行处理效果如图4-7所示。

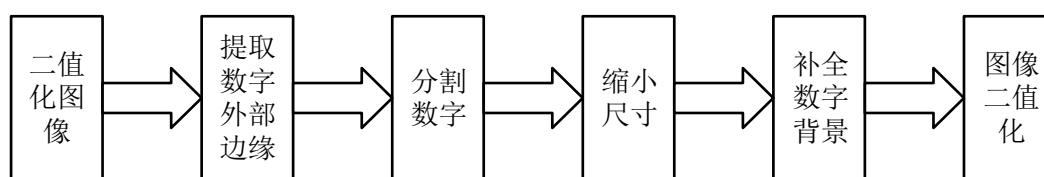


图 4-6 改进尺寸缩放流程图

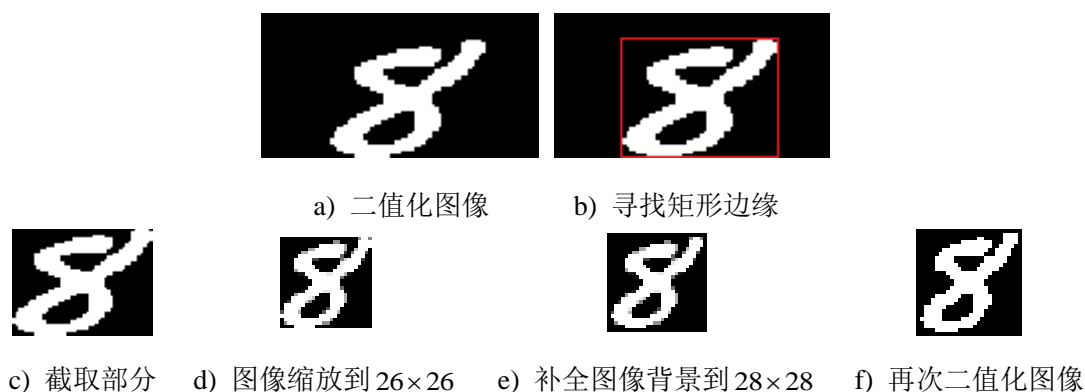


图 4-7 图像尺寸归一化过程

该算法减少了图像部分无用的背景信息,优化了尺寸归一化算法,使图像中数字的信息占比量增加,利用二值化进一步减少了数据处理量。

4.5.2 交叉验证

交叉验证^[52](Cross Validation, CV)主要是利用统计分析的方法,分析分类器的好坏。基本思想就是将样本数据分成两份,分别用于训练分类器的训练数据和作为判

别手写数字识别率多少的测试数据。

(1) Hold-Out Method

将样本分成两份，取其中一份作为训练数据，这部分叫做训练集；另一份样本数据集用来验证训练好的分类器或模型，用于分析训练数据的可靠性，这部分叫做测试集，这部分数据样本应该少于总样本数据的三分之一，训练样本数据达到百分之八十最优。测试集测试分类器的测试结果作为判断训练模型好坏的依据。

优点：这种办法简单，这要将样本分为不等的两份就行，没有过多的复杂计算。

缺点：由于这种方法是在训练之前就将样本数据分为两份，最后并没有进行交叉训练和验证。样本数据随机分为两份对测试结果有很大影响，可能出现不同的分类产生有不同结果，而且有可能很大差别。因此 Hold-Out Method 验证方法严格来说没有很大说服力。

(2) K-fold Cross Validation(简称 K-CV)

该方法得到 Hold-Out Method 验证方法的启发，样本数据分成 K 份数据集，将 K 份中的一份作为测试集，另外数据作为训练集。 K 的取值不小于 2，通常情况时大于等于 3。以 10 组为例，具体的算法流程如图 4-8 所示。

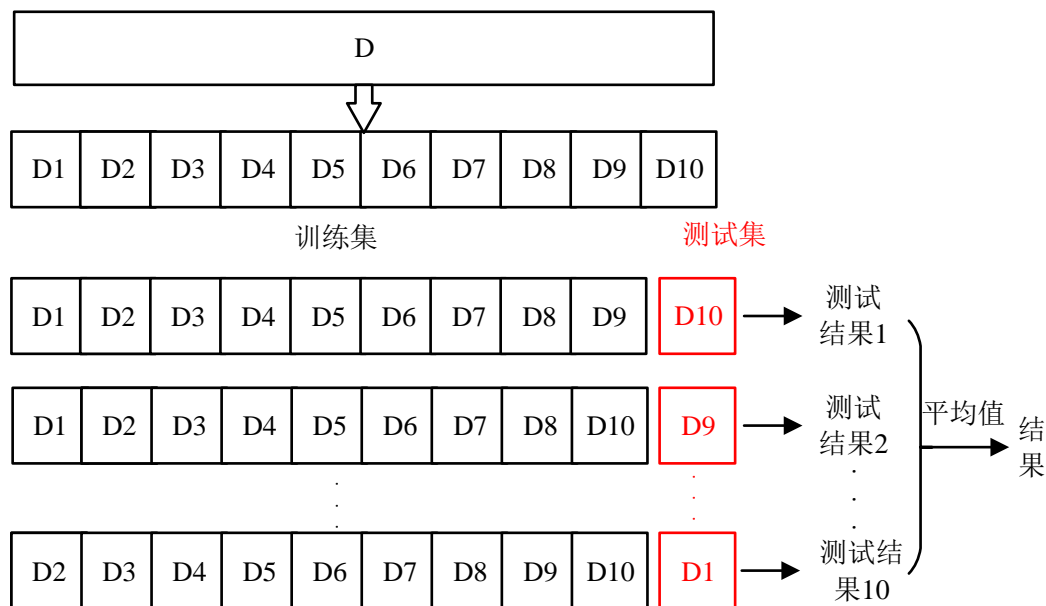


图 4-8 10 组 K-CV 流程

首先将样本数据 D 随机分成 $D1$ 到 $D10$ 的 10 份数据。特殊情况下，可以将样本分为不均等的样本子集，可以按照需求或遵循一种规则划分样本数据；然后将 $D10$ 作为测试集，将 $D1$ 到 $D9$ 数据集作为训练集，在训练完后，用 $D10$ 测试集测试得到的识别率，作为第一组测试下的准确率；其次和第二步相同，不一样就是将 $D10$ 换

成 D9, D9 作为测试集, D1 到 D8 和 D10 作为测试集, 得出第二组的测试结果; 最后将 10 组得到测试结果相加除以 10 得到的平均值作为训练模型的性能指标。

优点: K-CV 可以将样本分为 K 份, 但是每一份都能作为测试集, 可以很好的增加算法的寻优能力和较高的可信度。

缺点: 将数据分成 K 份, 这样操作增加了算法的复杂性和不稳定性。因为 K 值得选取有很大的差别, 所以当 K 值较大时可以提高分类准确性, 伴随着就是算法的复杂性, 有可能出现过学习现象, 同样当 K 的取值太小时, 准确性会下降, 有可能出现欠学习现象。

(3) Double Cross Validation(2-fold Cross Validation, 记为 2-CV)

Double Cross Validation 方法和 Hold-Out Method 有很大的相同都是将原本样本分为两份, 但是不同之处就是将样本分为两份大小相同的样本。将样本随机分两份, 分别为训练集和测试集。开始利用训练集训练模型, 并用测试集测试得出识别率, 然后将训练和测试集对调, 再次训练模型, 并得到测试结果, 最后将两个结果取其平均值, 作为训练模型的好坏标准。

优点: 同样和 Hold-Out Method 验证一样, 简单易懂, 计算简单。

缺点: Double Cross Validation 验证方法并不常用, 主要是将样本一分为两份使训练样本数太少, 所以训练集不能完全代表整体的样本数据。这个方法训练的模型和整体的训练模型率有明显的差别。当被分的子集差别大时, 得到的结果往往达不到实验要求。

(4) Leave-One-Out Cross Validation(简称 LOO-CV)

该方法是每个样本单独作为测试集, 比如样本数据有 N 个样本, 那么 LOO-CV 就是 N-CV, 将其中一个作为测试集, 剩余的样本数据作为训练集。将这 N 组的准确率的平均值作为性能指标。

优点: 这个方法可以无限接近原始数据分布规则, 因为能够将每个训练样本点作为训练对象, 所以这个方法训练的结果值得信赖; 实验过程中没有随机因素会影响实验数据, 确保实验过程是可以被复制的; 将交叉验证的寻优能力发挥到了极致, 大大增加了全局最优参数的寻优能力^[52]。

缺点: 训练模型无限接近原本样本数据, 所以研究成本高, 同时当样本数据较多时, 训练时间会比较长; 这种方法无限接近样本数据, 所以准确率可以大大的提高, 同样使用这样的方法算法复杂程度大大增加。

4.5.3 SVM 识别手写数字的实现

SVM 是一个二分类器，但是样本类型有 9 种，所以不能直接使用 SVM，如果想要直接使用 SVM 就必须改变二次规划的目标函数，但是更改二次规划函数太过于复杂，所以很少使用。通过分析 SVM 多分类器方法的优缺点。由于本文中样本有 9 类，为了避免正负样本数据差别太大而出现的不均衡的问题，本文采用一对一的方法对样本进行训练，同时还保证了每个样本的正确率。

本文最终使用 5 折交叉验证，就是 K-CV 方法。将数据分成 5 份，得到 5 组数据的平均值作为最后的测验结果。本文利用训练好的模型对第 3 章的 9 个图片上数字进行识别，识别流程如图 4-9 所示。

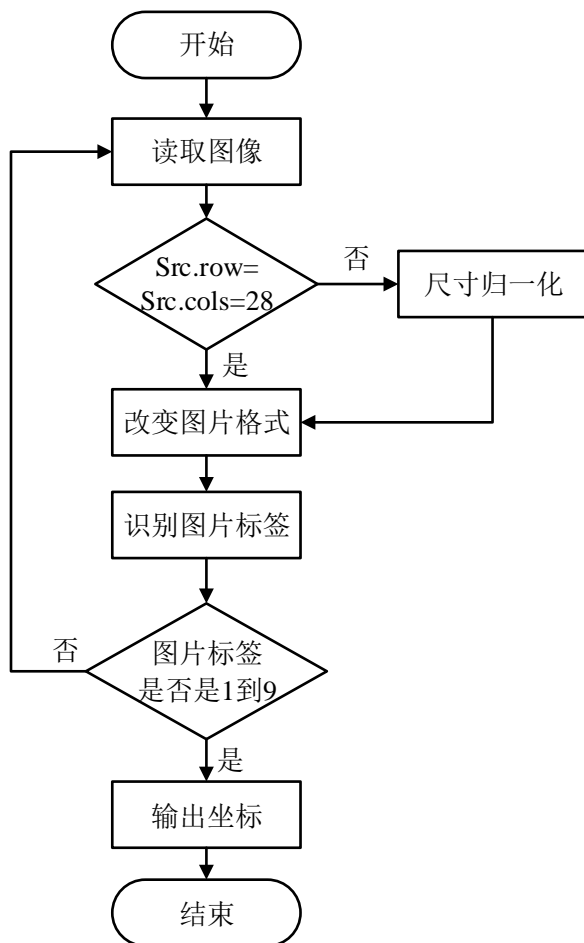


图 4-9 数字识别流程图

4.6 实验结果分析

基于 SVM 的数字识别实验全部过程都是在 Windows 系统环境下采用 c++ 的语言

编译环境下实现的，本文主要是通过 OpenCV 函数库下的 ml.cpp 文件，这里面都是机器学习的算法，提供了正太贝叶斯、K 最近邻、支持向量机、决策树、ADA Boost、梯度提升决策树、随机森林、人工神经网络、EM 算法。

利用支持向量机对数字识别，在 OpenCV 中的 SVM 的类型如下表 4-3 所示。

表 4-3 SVM 的类型

类型符号	含义	作用
C_SVC	C 类支持向量分类机	n 类分组($n \geq 2$)，允许用异常值惩罚因子 C 进行不完全分类
NU_SVC	ν 类支持向量分类机	n 类类似然不完全分类的分类器。参数为 ν 取代 C (其值在区间[0, 1]中， ν 越大，决策边界越平滑)
ONE_CLASS	单分类器	所有的训练数据提取自同一个类里，然后 SVM 建立了一个分界线以分割该类在特征空间中所占区域和其它类在特征空间中所占区域
EPS_SVR	ϵ 类支持向量回归机	训练集中的特征向量和拟合出来的超平面的距离需要小于 p 。异常值惩罚因子 C 被采用
NU_SVR	ν 类支持向量回归机	ν 代替了 p

根据 Vapnik 等人的研究表明，核函数对训练结果有一定的影响，为了选择最好的核函数，利用本文数字样本进行训练的到测试结果。径向基函数 $C=10$ ， $\sigma^2=100$ 测试样本运行结果如图 4-10 所示。不同核函数对 SVM 的性能影响对比实验如表 4-4 所示。

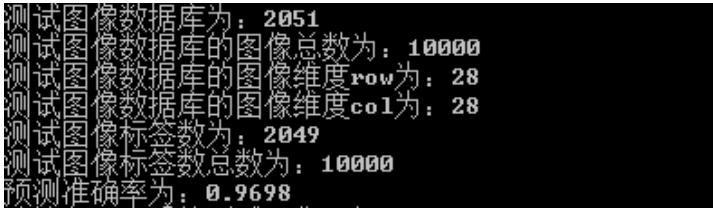


图 4-10 径向基函数 $C=10$ ， $\sigma^2=100$ 运行结果

表 4-4 不同核函数对 SVM 的性能影响

核函数	对应的识别率(%)	参数选择
多项式核函数	92.12	$d=1$
线性核函数	92.28	
径向基函数	96.98	$C=10$ ， $\sigma^2=100$

经过实验可以知道在三种核函数中，径向基函数相对好些。最终使用径向基函数训练样本。

径向基函数在 OpenCV 中需要设定两个比较重要的参数，即 C 与 γ 。在两

个参数中 C 是惩罚因子，其表示出现错误的几率，当 C 越大时，错误率越小。另一个参数 $gamma$ 是径向基函数自带的一个参数，其表示数据映射到高维空间后的分布。 $gamma$ 值越大表示支持向量越少。

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right) = \exp(-gamma \cdot \|x - y\|^2) \quad (4-5)$$

$$gamma = \frac{1}{\sigma^2} \quad (4-6)$$

通过实验优化参数数据。开始通过锁定惩罚因子 C 的值，改变 $gamma$ 的值，从而求得实验中 $gamma$ 的最优值。本次初次实验通过固定惩罚因子 C 为 100，测试 $gamma$ 的实验数据如表 4-5 所示。

表 4-5 参数 $gamma$ 选取结果

惩罚因子	参数 $gamma$	准确率(%)
100	0.1	97.23
100	0.01	96.60
100	0.001	86.15
100	0.0001	80.32

经过上述的实验，可以得到参数 $gamma=0.1$ 时训练模型结果最好。想要得到加好的数据，采用减小数据跨度的方法，进行了下一步实验如表 4-6 所示。

表 4-6 参数 $gamma$ 最优选取实验结果

惩罚因子	参数 $gamma$	准确率(%)
100	0.15	93.13
100	0.1	97.23
100	0.06	98.04
100	0.05	98.2
100	0.04	98.32
100	0.03	98.22
100	0.02	97.98
100	0.01	96.6

在减小参数数据跨度后，取其比较有代表的数据得到上述的实验数据如表 4-6 所示。参数 $gamma$ 是体现模型复杂程度，数据走向是先升高后降低的趋势，参数 $gamma$ 为 0.04 时，实验结果最好。

为了得到惩罚因子的最优参数,将参数 γ 固定,改变惩罚因子。通过实验寻找惩罚因子的最优参数。首先经过上面的实验可以得到 γ 最优参数为 0.04。在寻找最优参数 C 时,参数 γ 为 0.04,进行实验得到结果如表 4-7 所示。

表 4-7 参数惩罚因子 C 实验结果

惩罚因子	参数 γ	准确率(%)
1	0.04	98.12
10	0.04	98.32
100	0.04	98.32
1000	0.04	98.32

通过最优参数 γ 确定后得到数据 C 为 10 左右时,模型的识别率最高。上表中 C 的跨度较大,为了进一步精确,缩小跨度,进一步得到较好的数据参数。实验结果如表 4-8 所示。

表 4-8 参数 C 最优选取实验结果

惩罚因子	参数 γ	准确率(%)
1	0.04	98.12
2	0.04	98.27
3	0.04	98.33
4	0.04	98.32%
5	0.04	98.32%
6	0.04	98.32%

通过上式表明当惩罚因子 C 为 3 时,训练结果最好。 C 的越大与越小都会使泛化能力变差。训练模型的读取时间较长,所以在程序开始时就将训练模型进行读取,从而减少在字符区域识别后进行数字识别使得时间,10 组不同数字运行时间结果数据如表 4-9 所示。

表 4-9 运行时间

序号	时间/s	序号	时间(s)
1	0.012931	6	0.013987
2	0.015006	7	0.014257
3	0.013921	8	0.014675
4	0.014765	9	0.015018
5	0.014961	10	0.014586

经过上面的实验可以得出单个识别时间可以达到 0.01550s 以下,可以完全达到

本文识别时间要求。

4.7 本章小结

本章首先介绍了支持向量机的一些原理，主要包括最优分类超平面、线性向量机、非线性向量机以及核函数；然后针对本文数字图像大小不同，对图像尺寸归一化处理，最后通过本文的数据进行训练和测试。经过反复的实验与参数优化调节，得到较高的识别率 98.33%，对数字识别进行识别时间验证，得出 SVM 可以对手写数字进行快速准确的识别。

第 5 章 目标检测与控制的实现

5.1 引言

在视觉追踪系统中，关键步骤是对移动目标的检测。如果检测不到运动目标，就无法完成对目标追踪的任务。目前常用的运动目标检测算法有帧差法^[53]、背景差分法^[54]和光流法^[55]。根据本文追踪目标的复杂性，提出一种利用改进颜色识别与阈值分割相结合的方法，完成复杂目标的快速追踪。

5.2 运动目标检测算法的研究

5.2.1 帧差法

帧差法利用相邻两帧或三帧图像进行差分运算，取其绝对值，得到两帧图像的亮度差，对差分图像进行二值化处理，如果两幅图像的背景是完全一样的情况下，通过作差可以完全将背景消除^[56]。相邻两张图像拍摄时间较短，所以前后两张图像的环境变化不大，变化最大的为图像中运动物体，将两张图像进行做差，通过设定阈值，规定小于阈值的像素点为环境，高于阈值的像素点就是追踪目标，可以将环境像素点过滤掉。帧差法的流程图如图 5-1 所示。

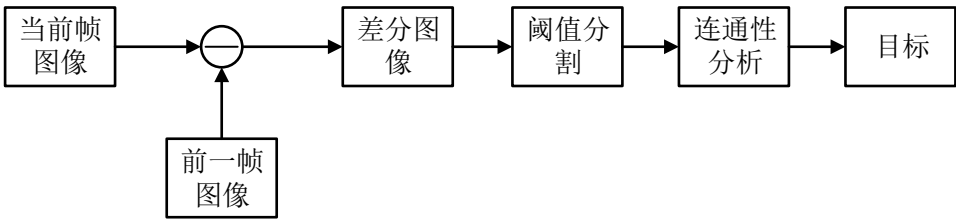


图 5-1 差分法流程图

前后两帧图像相减，后一帧图像与前一帧图像灰度值不发生变化的部分就会被减掉，减掉的部分包含大部分背景图像和少量目标图像。在运动目标识别中，该方法可以实现对多目标的识别。

帧差法的公式为：

$$D_k(x, y) = |f_k(x, y) - f_{k-1}(x, y)| \tag{5-1}$$

$$g_k(x, y) = \begin{cases} 255, & D_k(x, y) \geq T_0 \\ 0, & D_k(x, y) \leq T_0 \end{cases} \tag{5-2}$$

式中 $D_k(x, y)$ ——差分图像;

T_0 ——设定阈值。

T_0 值的选择对识别有很大影响, 当 T_0 值设定过小时, 处理后的图像就会含有背景信息, 不能将目标分割出来; 当 T_0 值设定过大, 目标信息丢失过多出现大量空洞和漏检, 检测出来的运动目标位置不准确。

帧差法对均匀移动的物体识别效果较好, 但是实际情况目标并不一定匀速运动, 有可能加速或减速运动, 这时使用帧差法进行运动目标识别, 检测出结果会出现漏检和多检的情况。

相邻的两帧图像拍摄时间非常短, 可以将前一帧图像作为下一帧图像的背景, 有良好的实时性。背景图像都是前一帧图像, 因此背景图像没有积累。帧差法优点是算法简单和计算量小。缺点是对噪声较为敏感。当目标体积大或整体颜色较为一致时, 识别过程中内部会产生空洞, 不能完整的将目标提取出来。

5.2.2 背景差分法

背景差分法是利用当前帧图像与背景参考模型比较来检测运动物体的一种方法。该算法识别的好坏与背景图像有较大关系。经常应用于视频监控系统中。当目标进入拍摄视野中, 将当前帧图像和背景图像做差, 通过设定阈值, 大于设定阈值的像素点, 作为是目标信息, 小于设定阈值的像素点, 作为背景信息。相对于帧差法, 背景差分法可以识别出视野范围内不动的目标。

背景差分法的公式为:

$$D_k(x, y) = |f_k(x, y) - b_0(x, y)| \quad (5-3)$$

$$g_k(x, y) = \begin{cases} 255, & D_k(x, y) \geq T_0 \\ 0, & D_k(x, y) \leq T_0 \end{cases} \quad (5-4)$$

其中, $b_0(x, y)$ 为我们背景图像, $f_k(x, y)$ 为当前帧图像, T_0 为设定阈值。

背景图像的准确度, 直接影响背景差分法对运动目标的检测结果。由于环境的复杂性和不可预测性, 环境会发生一定变化, 使输入背景图像与实际环境信息有一定的差别, 检测出来的目标不准确。比如环境光照的变化、摄像机的抖动和环境物体的改变等。

背景差分法的优点是检测速度快、准确率高、易于实现。该算法的关键就是对背景图像的获取。

5.2.3 光流法

1950 年 Gibson 提出光流的概念。光流是由于目标运动、相机运动或两者共同运动所产生的。给图像中每个像素点赋予一个速度矢量，从而形成光流场，在没有运行物体时，光流场表现为连续均匀，在出现运动物体时，光流场变得不再连续均匀。从而检测出运动物体。

该方法的假设条件是：

- (1) 相邻图像序列之间的亮度是恒定；
- (2) 物体在相邻的图像中的运动识别叫“微小”的；
- (3) 保持空间一致性；即，同一子图像的像素点具有相同的运动。

现实三维中的点与图像中的像素点是相对应的。当时域中发生变化，图像中每个像素点就会产生相应的速度变化。二维变化是由三维空间速度变化经过投影而来，其原理图如图 5-2 所示。可以通过分析每点的速度变化来寻找目标。

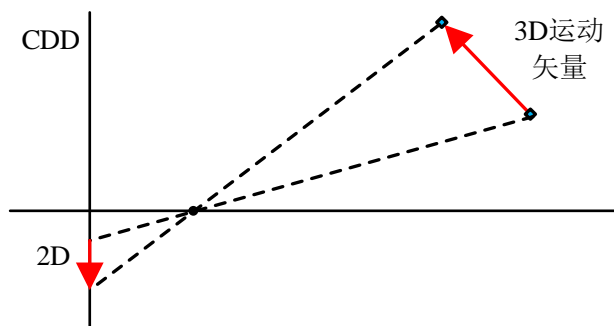


图 5-2 光流法原理

根据光流法的前提假设，目标或相机运动非常短暂的时间 Δt ，分别在 x 和 y 方向上移动的距离分别为 Δx 和 Δy 。 Δt 无限接近 0 时，得到 5-5 公式：

$$F(x, y, t) = F(x + \Delta x, y + \Delta y, t + \Delta t) \quad (5-5)$$

对其上面的式子进行一级泰勒级数展开，得到式 5-6：

$$F(x + \Delta x, y + \Delta y, t + \Delta t) = F(x, y, t) + \frac{\partial F}{\partial x} \Delta x + \frac{\partial F}{\partial y} \Delta y + \frac{\partial F}{\partial t} \Delta t + \varepsilon \quad (5-6)$$

忽略式中的二阶无穷小，可以得到光流约束方程：

$$F_x dx + F_y dy + F_t dt = 0 \quad (5-7)$$

图像在运动场中水平和垂直方向上面的偏移量分别为 u , v ：

$$\begin{cases} u = \frac{\Delta x}{\Delta t} = \frac{dx}{dt} \\ v = \frac{\Delta y}{\Delta t} = \frac{dy}{dt} \end{cases} \quad (5-8)$$

可以将光流约束方程改写为:

$$F_x u + F_y v + F_t = 0 \quad (5-9)$$

在实际应用中,将研究对象看成刚体,所以目标的所有像素点都具有相同的速度矢量。光流场的计算是视觉运动检测中一个研究重点,根据不同的约束条件会有不同的计算方法。其中较为经典的是L.K(Lucas&Kanada)法和H.S(Horn&Schunk)法^[57]。

光流法并不需要摄像头固定,在背景信息不需要输入的情况下,还可以很好的将目标检测出来。最大的缺点是在阴影、透明性、遮挡、多光源和噪声等情况下,使用光流法来检测运动物体时,光流法的计算量非常大,没有办法使光流法达到实用性和实时性,很难应用于实时检测系统中,这种方法对噪声是非常敏感的,噪声使计算结果精度降低,难以检测到运动目标的精确边界。

对三种算法进行对比,得到表5-1所示。

表5-1 三种运动目标检测算法的比较

运动目标检测技术	优点	缺点
帧差法	对视觉场内的光线逐步变化不敏感;算法简单易实现。	对静止的物体无法检测;对目标运动方式敏感;容易产生空洞的检测。
背景差分法	能检测出静止物体;算法简单易实现。	对长时间光线变化敏感;背景更新模型难建立。
光流法	直接计算运动信息;对目标的颜色不敏感;能容忍尺度的变化。	计算量很大;实时性差;需要特定的硬件支持。

常见的三种方法都有其的优缺点,本文运动目标并不是一直运动,帧差法容易产生空洞和多检的缺点;背景差分法要求摄像头固定,同样不符合本文要求;光流法计算量非常大不满足实时性要求,所以传统的三种运动目标识别方法都不满足本文实验要求。

5.3 基于改进颜色识别与阈值分割的研究

本文需要再识别出敌方小车后,对敌方小车上的装甲板进行打击直至将目标消灭。利用传统的运动目标检测算法都不能将装甲板准确的检测出来,所以传统算法

不满足本文使用要求。小车整体结构比较复杂,如图 5-3 所示,通过对整车的检测来完成对装甲板的打击难度比较大,而且最终打击准确率会很低。通过研究装甲板特点,发现装甲板较为突出的特点是在左右两侧有比较明显的灯条。追踪的最终任务就是完成对装甲板的准确打击,所以本文直接对装甲板进行识别。

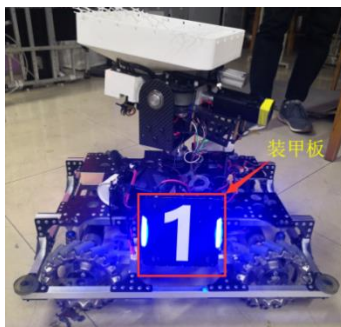


图 5-3 目标小车

装甲板是一个 90×90 的黑色方框,在其左右两侧有高亮蓝色的灯条,对于装甲板的识别主要是对蓝色灯条的识别,经过研究发现灯条附近表现为蓝色,而在其中心表现为白色,所以装甲板的识别流程图如图 5-4 所示。

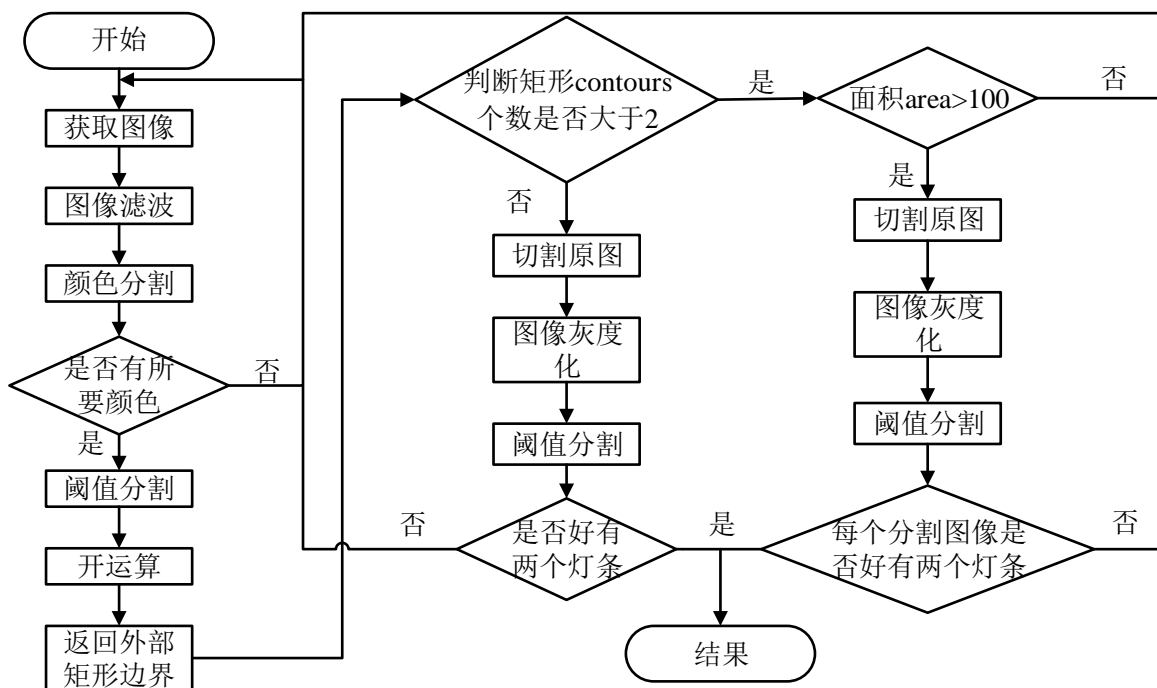


图 5-4 装甲板识别流程图

本文追踪小车是利用了两次阈值分割,第一次将蓝色区域分割出来,第二次检测蓝色区域是否有灯条,通过判断是否有灯条,来判断是否为检测目标。

5.3.1 基于颜色识别

通过观察首先想到的方法是通过颜色进行识别。颜色识别是比较简单的算法，且易于实现。采集到的图像是 RGB 三通道组成的，RGB 通道并不能很好地反映出物体具体的颜色信息，人类直观上看，HSV 符合人们对颜色的理解，同时 RGB 的颜色阈值受光照的影响。所以在颜色识别过程中，需要将 RGB 的颜色空间转换到 HSV 空间中。首先通过颜色空间转化，再通过滤波处理减少噪声，通过颜色阈值分割最后通过二值化处理，得到目标颜色区域。利用颜色识别算法对图 5-3 进行识别，得到的实验效果如图 5-5 所示。

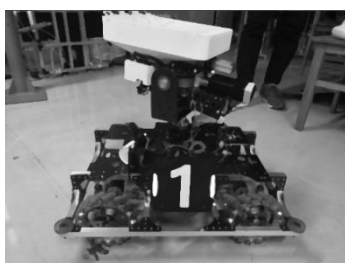


图 5-5 颜色分割二值化图像

本文需要对目标中装甲板进行打击，基于颜色识别不能过准确的将装甲板识别出来，只是将装甲板附近的区域识别出来。在视野中出现相同颜色时，基于颜色识别就会出现误检测的情况。

5.3.2 改进颜色识别

研究图像灰度化时，发现灰度化中的分量法，其红、绿、蓝每个通道得到的灰度图像有所不同。对图 5-3 原图进行蓝色分量法灰度处理和权值灰度图像得到效果如图 5-6 所示，在灯条附近蓝色空间灰度图相对于权值灰度图更加亮。



a) 权值灰度图



b) 蓝色分量灰度图

图 5-6 灰度图像对比

基于颜色识别需要对每个通道都进行阈值分割。改进算法仅对蓝与绿两个通道

进行计算，并且只是进行简单的加减计算，这样的操作可以减少运算时间。改进颜色分割的算法流程如图 5-7 所示。

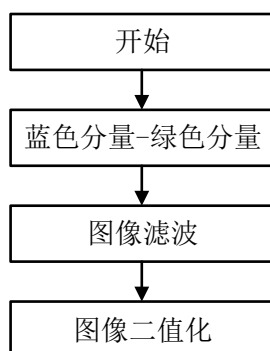


图 5-7 改进颜色分割流程图

通过蓝色分量法减去绿色分量法得到的效果如图 5-8 所示。可以看出部分背景已经去除了，在灯条的附近相对比较亮，最后通过图像二值化得到效果如图 5-9 所示。



5-8 相减灰度图

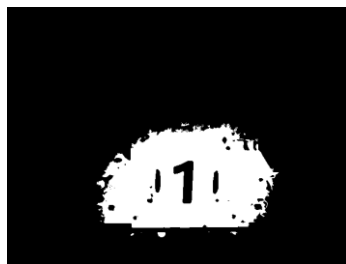


图 5-9 图像二值化

改进颜色识别算法和传统颜色识别算法都能得到很好的二值化图像。取得 10 次运行程序时间如表 5-2 所示。

表 5-2 算法时间对比表

序号	颜色识别算法时间(s)	改进颜色识别算法时间(s)
1	0.017420	0.004178
2	0.017567	0.004874
3	0.018124	0.005124
4	0.018542	0.005457
5	0.017987	0.004578
6	0.016987	0.004975
7	0.018247	0.005412
8	0.016587	0.005148
9	0.017345	0.004984
10	0.017843	0.005648
平均值	0.017665	0.005038

通过实验可知改进后颜色识别算法运行速度更快，明显提高了 3.5 倍。

5.3.3 形态学操作

图 5-10 得到的二值图像，其边缘附近有好多的毛刺和一些孤立的白点。可以利用形态学操作来减少这些毛刺和孤立点。形态学中包括膨胀和腐蚀。膨胀和腐蚀都是对白色部分所说的，膨胀可以使高亮部分对周围进行膨胀，使原来白色部分区域变大，腐蚀是膨胀的反运动，使白色部分减少。本文利用开运算减少这些毛刺，使其变得润滑，同时过滤掉一些较小的白点。该操作将边缘润滑的同时并不明显改变其面积。开运算是先腐蚀后膨胀的过程。开运算的计算公式为：

$$A \circ B = (A \ominus B) \oplus B \quad (5-10)$$

上式中，其中 \ominus 和 \oplus 分别表示腐蚀和膨胀。对上面图 5-9 的二值化图像进行开运算得到的效果如图 5-10 所示。



图 5-10 开运算

5.3.4 灯条识别

通过颜色识别将蓝色区域分割出来，但是机器人不知道是否为装甲板区域，需要进一步处理。首先利用最小包边矩形将颜色识别区域框选出来，得到最大矩形区域的坐标值，通过坐标值对图 5-6a) 进行分割，然后通过阈值分割将灯条分割出来。

利用最小包边矩形将这部分标注出来如图 5-11a) 所示，通过面积进行筛选得到如图 5-11b) 所示。

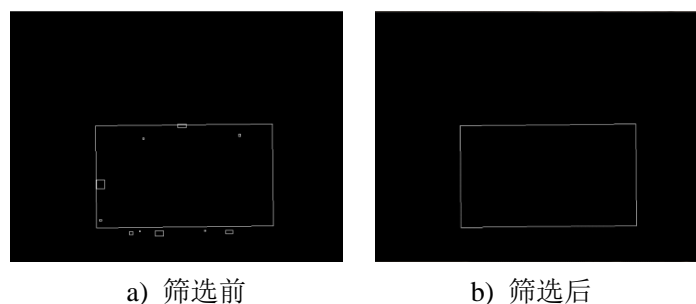


图 5-11 最小包边矩形

通过最小包边矩形检测出蓝色区域，在根据矩形的左上角坐标和矩形的长宽，

通过 OpenCV 中函数 ROI 截取滤波图像对应的图像如图 5-12 所示。



图 5-12 ROI 截取部分

对灯条的识别，首先进行图像阈值分割，为了防止有其他白色区域，进行图像开运算操作如图 5-13 所示，最终得到的效果如图 5-14 所示。

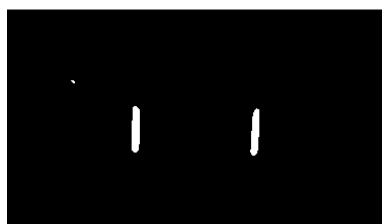


图 5-13 开运算

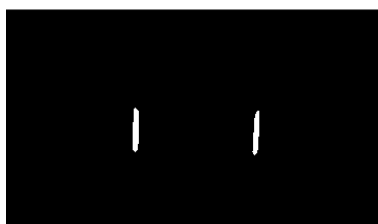


图 5-14 灯条识别结果图

5.3.5 坐标输出

经过两侧阈值分割完全可以将装甲板上的灯条识别出来，然后利用第三章中的质心求法，求出两个灯条质心，即 (x_1, y_1) 和 (x_2, y_2) ，截取滤波图像的位置坐标是 (x_m, y_m) ，最终装甲板的中心坐标为式(5-11)所示，根据坐标控制摄像头运动。

$$\begin{cases} x_0 = x_m + \frac{x_1 + x_2}{2} \\ y_0 = y_m + \frac{y_1 + y_2}{2} \end{cases} \quad (5-11)$$

5.3.6 电机角度判断

本文需要完成对目标的追踪任务，所以需要给定电机转动的角度，但是转动的角度需要目标与摄像头之间的距离。单目测距根据小孔成像原理，具体的测距原理模型如图 5-15 所示。

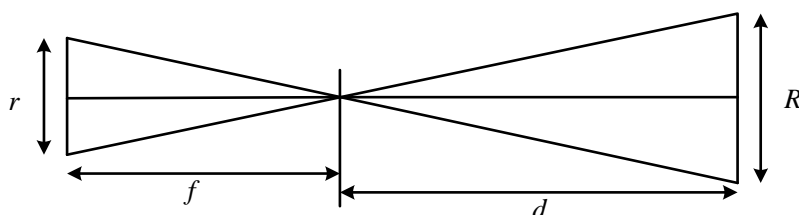


图 5-15 测距模型

根据小孔成像原理可以得到公式：

$$d = f \frac{R}{r} \quad (5-12)$$

其中， d 表示目标距离摄像头的距离， f 为摄像头的焦距， R 为目标实际距离， r 为目标在图像中的像素距离。

利用第 2 章得标定结果可以得到摄像头内参数，所以摄像头每个单元沿横轴和纵轴的距离为：

$$s_x = \frac{f_x}{f} \quad (5-13)$$

$$s_y = \frac{f_y}{f} \quad (5-14)$$

通过上面两个式子，可以得到像素的尺寸：

$$d_x = \frac{f}{\sqrt{f_x^2 + f_y^2}} \quad (5-15)$$

根据图像可以得到两点的像素差为 s ，得到图像像素距离：

$$r = s \times d_x \quad (5-16)$$

结合上式可以得到最终的距离为：

$$d = f \frac{L}{l} = \frac{R \sqrt{f_x^2 + f_y^2}}{s} \quad (5-17)$$

由于本文装甲板上两个灯条之间的中心距是已知，所以可以根据灯条位置将目标与摄像头的距离计算出。目标在摄像头横纵坐标的偏移量分别为 x ， y ，通过计算可以知道云台的俯仰角 θ 和偏航角 α 为：

$$\begin{cases} \alpha = \tan^{-1} \frac{x}{\sqrt{y^2 + d^2}} \\ \theta = \tan^{-1} \frac{y}{d} \end{cases} \quad (5-18)$$

5.4 运动控制与追踪策略的研究

云台整个控制系统如图 5-16 所示。云台系统主要是由主控模块、CAN 总线连接电机模块和视觉处理模块。

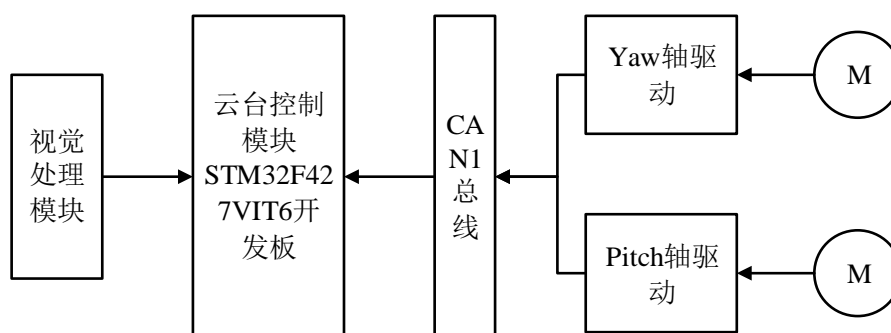


图 5-16 系统框图

追踪系统性能好坏和云台运动性能的好坏有着直接关系。云台运动性能的好坏取决于电机控制性能，所以电机的控制算法非常重要。在实际工业中利用最多的控制方式就是 PID 控制，所以本文选用 PID 算法对电机控制。

5.4.1 PID 控制算法

本文摄像头的电机用 PID 算法控制。PID 控制是最为经典的控制理论^[58]。PID 算法结构的简单、控制性能稳定可靠、计算量小、参数易于调整的特点，经过多年的发展，PID 是现在较为普遍的控制方法，并在很多领域都广泛使用。PID 可以控制温度、流量、压力等。因此经常应用在工业控制中。只要对参数进行调节就行，不用考虑数学模型，就可以达到预期的效果。PID 控制由控制器、调节器还有被控制器三部分组成。如下图 5-17 为 PID 基本原理图。

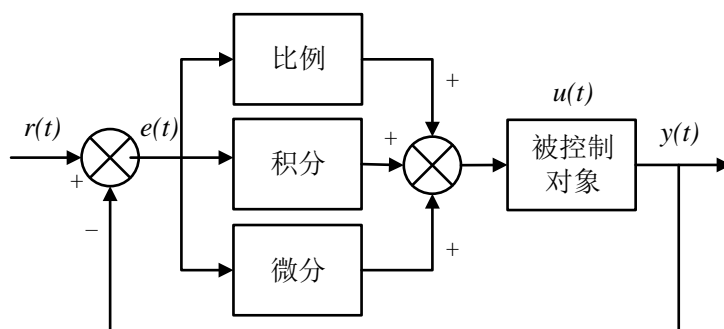


图 5-17 PID 控制模型图

图中， $r(t)$ 为输入量， $y(t)$ 为输出量， $e(t)$ 为误差。

模拟 PID 控制原理：

$$u(t) = K_p \left[e(t) + \frac{1}{T_i} \int_0^t e(t) dt + T_d \frac{de(t)}{dt} \right] \quad (5-19)$$

式中 K_p ——比例系数；

T_I ——时间积分常数；

T_D ——时间微分常数。

计算机只能处理离散的信号，所以需要将上面的方程进行离散化，得到新的方程式如下：

$$u(k) = K_p \left\{ e(k) + \frac{T}{T_I} \sum_{i=1}^k e(k) + \frac{T_D}{T} [e(k) - e(k-1)] \right\} \quad (5-20)$$

其中， $k=0, 1, 2, \dots$ (采样序号)， $u(k)$ 为第 k 次的输出量， T 为时间变量， $e(k)$ 为第 k 次的偏差。

PID 的控制器由三部分组成：比例，积分和微分，其制约系数分别为 K_p 、 T_I 和 T_D 。这三部分所起的作用分别为：

(1) 比例环节：按比例反应系统偏差。当系统产生偏差比例环节就会产生响应，按照比例来减小误差。系统的响应快慢是由 K_p 所决定。当增大 K_p 值时，系统的响应速度就会变快，当值过大时，系统的稳定性就会下降，系统产生震动。

(2) 积分环节：其主要的作用是用来消除系统误差。积分作用的强弱取决于积分时间常量。为了加快系统趋于稳定速度，就要增大积分作用，从而需要减小 T_I 值。当 T_I 过小会使系统震荡。 T_I 的作用反映到 PID 控制的性能指标上有：增大 T_I ，系统的超调量减小，调节时间变长，反之调节时间变短。

(3) 微分环节：微分环节能够对误差信号的变化率做出响应，其主要作用是当系统产生较大的误差变化趋势时提前做出反应并进行适当的调整。 T_D 越小，微分作用越强，系统的响应速度越快，但 T_D 值过小会因被控对象时间延迟产生问题。将 T_D 的作用反映到 PID 控制的性能指标上有：增大 T_D ，系统的上升时间变长，调节时间变长，反之时间变短。

5.4.2 云台系统

本文是通过视觉进行目标检测，所以其反馈信息系和的视觉传感器相连接的，通过视觉常感器判断目标所在位置，将位置发送到控制器，然后控制器控制直流电机的转动。本文的 PID 控制算法是在 KEIL5 软件中进行编写的，编写完成后在下载主控制器 STM32 上。STM32 单片机通过控制 PWM 波控制电机的转动角度。为了保证图像采集系统电机的控制精度，电机的控制采用三闭环控制。三闭环控制流程图如图 5-18 所示。

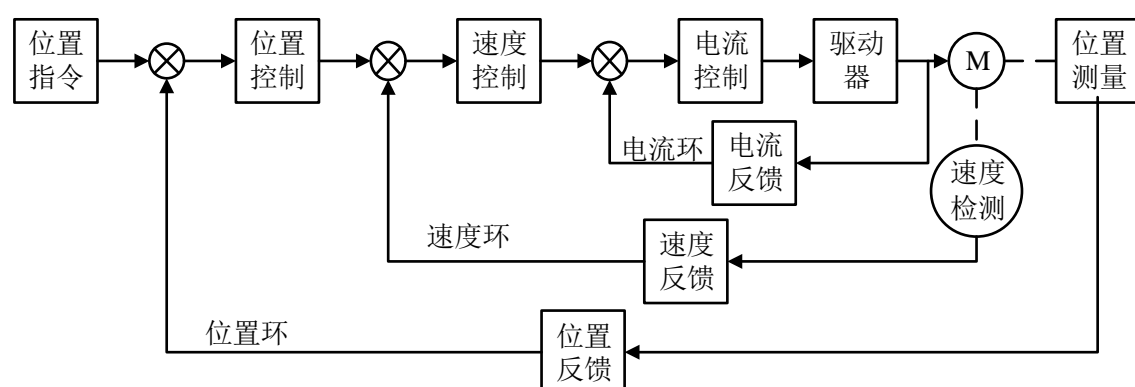


图 5-18 三闭环 PID 控制

5.4.3 数字式 PID 的改进

(1) 带死区的 PID 算式，在实际电机控制过程中，当控制动作过于频繁时，会引起系统的振荡。为了避免系统的振荡，引入带死区 PID 算式。公式如下：

$$e(k) = \begin{cases} e(k), & |e(k)| > B \\ 0, & |e(k)| < B \end{cases} \quad (5-21)$$

其中， B 表示死区的大小，具体数据可根据实验控制数据确定。当 B 值设定过小时，就会使调节动作过于频繁，不能达到稳定调节的目的；当 B 值过大时，会使系统产生滞后。

(2) 积分分离的 PID 算式，PID 算法中积分就是控制系统的稳态偏差，但是当系统开始、停止或大幅度改变值时，就会积分饱和现象。为了将系统由于积分饱和，导致系统出现超调和振荡现象，在 PID 控制中引入积分分离 PID 算式。具体的算法是，给定一个判断阈值，当偏差 $e(k)$ 大于判断阈值时，取消 PID 中积分作用；当偏差 $e(k)$ 小于判断阈值时，才启用积分作用。

(3) 微分项的改进算法，当有阶跃信号输入时，会增加微分项的输出，控制系统容易产生振荡。同时微分作用容易引进高频干扰。为了避免这种情况发生，往往在 PID 控制中串联一个低通滤波来抑制高频干扰。

5.5 开发环境

(1) 编译环境

上文算法与程序在 Windows 平台的实现采用的编译环境是 Visual Studio 2013，使用的函数库为 OpenCV3.3.0。

Linux 操作系统发布于 1991 年。由芬兰著名电脑黑客 Linux Torvalds(林拉兹, 托瓦斯)开发。其主要特点是开放、多用户、多任务, 还有安全、稳定、高性能。所以深受广大用户喜爱。该系统设计思想来源于 Unix 的核心思想。该系统支持 32 位与 64 位硬件。其广泛应用于服务器上面的操作系统。

Ubuntu 系统是基于 Linux 上的一种免费操作系统。该系统是由 Mark Shuttleworth 在 2002 年创造出来。Ubuntu 是基于 Debian 发行版和 GNOME 桌面环境的操作系统。所以说 Ubuntu 是以桌面应用为主的 Linux 操作系统, 并且 Ubuntu 版本是 Linux 系统当中较为受欢迎的版本。该系统可以为所有用户提供一个有安全有可靠的操作系统。本文的编程环境为 Ubuntu 14.04 操作系统, 图像处理函数库为 OpenCV3.3.0。

(2) 嵌入式平台的介绍

本文使用的是英特尔公司最新研发的 NUC7i7BNH 的微计算机如图 5-19 所示, 可以在该计算机上直接进行编程和运行调试, 相对于一些不支持 OpenCV 或者 CUDA 的单片机与嵌入式平台, 它省去了大量移植过程, 使开发变得快捷、容易, 只需要在其上的 Linux 系统中进行编程即可。



a) 正面

b) 背面

图 5-19 NUC7i7BNH 处理器

该处理器的大小为 $115\text{mm} \times 111\text{mm} \times 51\text{mm}$, 所以本文的处理器非常的小, 其使用的是第七代智能英特尔®酷睿™i7-8650CPU@1.90GHz×8。

(3) Ubuntu 程序的编译运行

OpenCV 的 Linux 版本, 采用 Cmake 编译, 不同于 PC 的架构, 所以需要进行交叉编译。将编写好的 C++ 程序, 采用下面命令即可完成编译:

```
g++ main.cpp -lopencv_core -lopencv_imgproc -lopencv_highgui -o main
```

使用的编译器为 g++, 执行文件为 main.cpp, 最后生成可执行文件。在执行时只需要输入命令: /main 即可, 方便简单。

5.6 PID 控制与追踪实验

(1) 电机控制射击实验

本文研究对象是通过 PID 控制电机的转动，在识别出目标后都要实现对目标的打击任务，所以进行打击实验。打击目标为黑色箱子，打击距离为 300cm。首先在小车上进行装弹如图 5-20 所示。



图 5-20 进行装弹

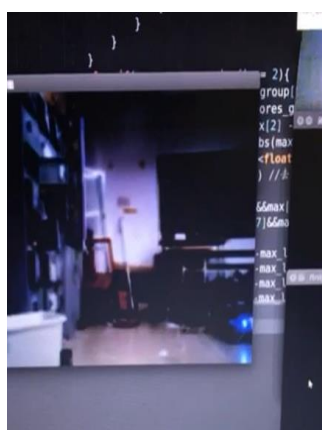
准备工作完成以后进行射击实验如图 5-21 所示。



图 5-21 射击实验

(2) 视觉追踪实验

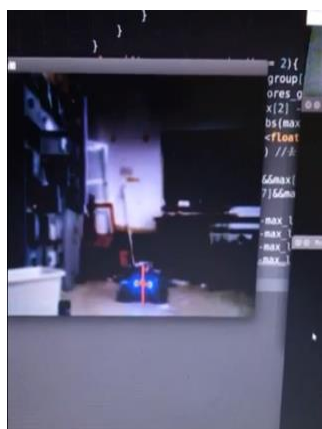
为了检测本文提出的基于特性目标追踪方法的有效性，实验采用 Ubuntu14.04 操作系统，使用 OpenCV 函数库，g++编译环境作为实验平台。该摄像头的分辨率为 640×480 。当目标出现在视野中后，程序自动进行识别追踪。并将装甲板用红色“十”字标记出来，表示识别出目标。对识别追踪实验进行连续帧的实验，截取部分图像如图 5-22 所示。



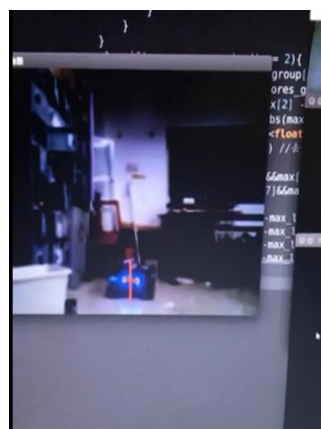
a) 第 1 帧图像



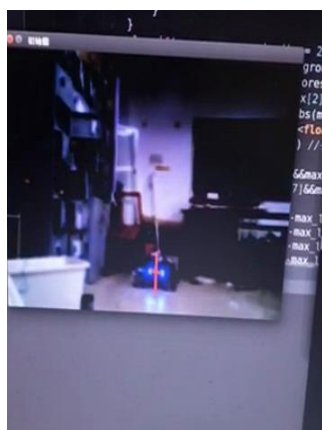
b) 第 86 帧图像



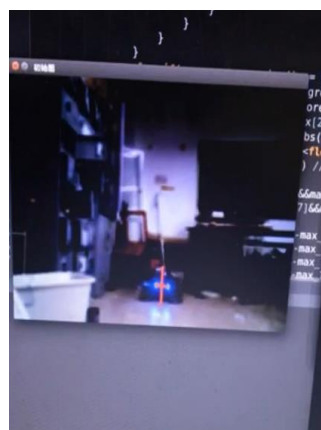
c) 第 123 帧图像



d) 第 156 帧图像



e) 第 179 帧图像



f) 第 213 帧图像

图 5-22 目标追踪实验

通过实验表明，本文算法准确的将目标小车识别出来，并完成追踪任务。输出目标运动的位置信息如表 5-3 所示。并利用表 5-3 中数据完成运动物体的位移曲线如

图 5-23 所示。

表 5-3 目标位置

说明	X 轴距离(cm)	Z 轴距离(cm)
第 1 帧图像	233.83	154.36
第 86 帧图像	168.34	167.29
第 123 帧图像	83.38	152.86
第 156 帧图像	48.92	124.67
第 179 帧图像	83.76	183.48
第 213 帧图像	103.54	245.23

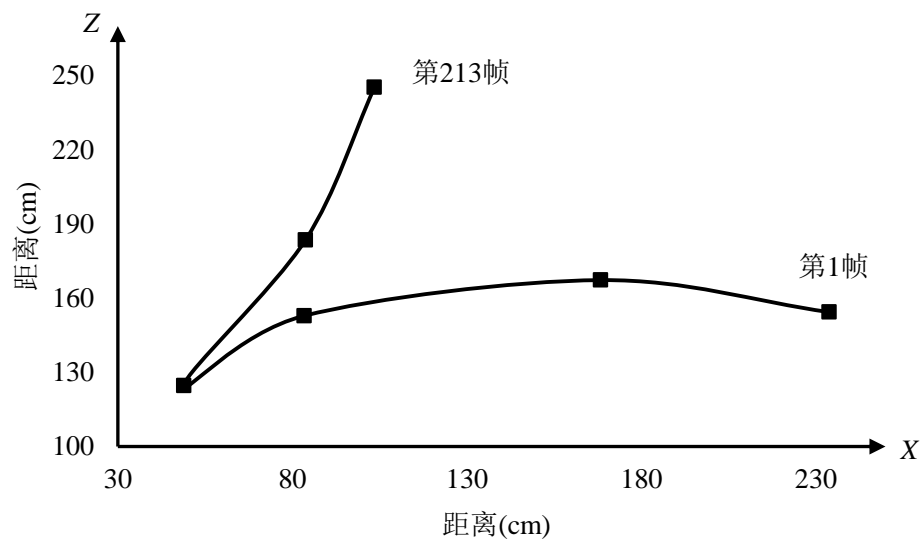


图 5-23 运动目标位移曲线图

5.7 本章小结

本文首先介绍经常使用的动态目标识别的方法，根据这些方法的不足和研究对象的特征，提出基于特性目标追踪方法；然后利用 PID 算法对云台进行控制；最后进行云台控制射击实验和小车追踪实验，验证了本文对目标小车追踪算法的准确性和稳定性。

结 论

本文以 RoboMaster 小车为研究对象,该小车集成了微型处理器 NUC7i7BNH 和单目摄像头模块,首先通过单目相机采集图像,然后经过在处理器上运行图像处理程序,实现对手写数字的识别,并完成对特定目标的识别与追踪。课题研究成果总结如下:

(1) 针对传统最大熵法对数字面板进行二值化处理的设定阈值偏小问题,本文对最大熵法进行了改进。通过多次实验,得到最大熵法与经验阈值之间的关系式,得到清晰的二值化图像。

(2) 针对数字面板矩形定位精确度低的问题,本文提出一种基于边缘检测与几何关系相结合的方法,实现了对数字面板的精确定位。

(3) 针对手写数字较难识别的问题,本文利用支持向量机对手写数字识别。通过边缘信息对数字进行尺寸归一化,信息进一步减小。通过实验得到最优参数,在时间满足技术要求下得到最终识别率为 98.33%,为以后支持向量机对数字识别作参考。

(4) 针对复杂目标追踪问题,本文提出了一种基于改进颜色识别算法与阈值分割相结合的方法,实现了对目标的快速追踪。

在机器视觉领域,本文主要研究手写数字和视觉追踪,对视觉领域发展有一定的促进作用,但是还存在一定的问题,仍需要改进的方向是:

(1) 基于 SVM 的手写数字识别问题,其主要是利用支持向量,并不是每个样本数据都作用于 SVM,可以对样本数据进行改进,找到样本中所有支持向量的样本数据,不仅可以增加训练速度,而且减少储存数据。

(2) 本文只是将小车的部分区域识别出来,不能将整个小车完整检测出来。

参考文献

- [1] 文家昌. 基于 LabVIEW 机器视觉的产品检测平台设计与应用[D]. 广州: 华南理工大学硕士学位论文, 2012: 2-3.
- [2] Marr D, Nishihara H K. Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes[J]. Proceedings of the Royal Society B: Biological Sciences, 1978, 200(1140): 269-294.
- [3] 朱朔. 基于机器视觉的无人平台目标识别与跟踪系统设计[D]. 南京: 南京理工大学硕士学位论文, 2018: 3-4.
- [4] ADI 公司. 高性能 DSP 的嵌入式视觉应用: 借机器一双慧眼[J]. 电子产品世界, 2013(2): 21-24.
- [5] Priya C T, Praveen K, Srividya. Monitoring of Pest Insect Traps Using Image Sensors & Dspic[J]. International Journal of Engineering Trends & Technology, 2013, 4(9): 89-90.
- [6] 殷建军, 张铁民, 潘春华. 分辨率实时可调的无线图像传感器节点设计与试验[J]. 农业工程学报, 2017, 10(03): 190-197.
- [7] 熊智华. 基于 DSP 平台的机器视觉系统优化研究[D]. 重庆: 重庆大学硕士学位论文. 2005: 3-5.
- [8] 王亚鹏, 许玮, 丁辰龙. 共话机器视觉技术发展[J]. 自动化博览, 2009, 26(8): 40-42.
- [9] 胥磊. 机器视觉技术的发展现状与展望[J]. 设备管理与维修, 2016, 24(9): 7-9.
- [10] 苏云, 张庆来. 面向海量文献的数字化系统研究[J]. 图书与情报, 2010, 26(2): 85-89.
- [11] 闫江宝, 张长胜, 周雷. 手写体数字的识别方法研究[J]. 数据通信, 2018, 186(05): 43-45.
- [12] 王晓娟, 白艳萍. 基于 BP 神经网络的手写体数字的识别方法研究[J]. 数学的实践与认识, 2014, 44(7): 112-116.
- [13] 黄巧巧. 基于 BP 神经网络的手写数字识别系统研究[D]. 湖北省: 华中师范大学硕士学位论文. 2009: 40-45.
- [14] 王强. 基于 CNN 的字符识别方法研究[D]. 天津: 天津师范大学硕士学位论文. 2014: 34-36.
- [15] 曹丹. 基于 Hopfield 神经网络的脱机手写数字识别[D]. 湖南省: 中南大学硕士学位论文. 2009: 41-42.
- [16] 宋婉娟. 基于支持向量机的数字串识别[J]. 湖北第二师范学院学报, 2010, 27(2): 84-86.

- [17] 郑敏, 赵玉芹, 田红伟. 基于多传感器数据融合的机动目标跟踪算法研究[J]. 探测与控制学报, 2006, 28(5): 43-45.
- [18] 陶鹏. 基于视觉的机器人障碍物识别与目标追踪[D]. 南京: 南京师范大学硕士学位论文, 2013: 3-5.
- [19] 王继玉. ROS 移动机器人跟踪系统的设计与实现[D]. 郑州: 郑州大学硕士学位论文, 2018: 2-3.
- [20] Yang C, Duraiswami R, Davis L. Efficient Mean-shift Tracking Via a New Similarity Measure[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE, 2005, 1, 176-183.
- [21] Perez P, Hue C, Vermaak J, et al. Color-based probabilistic tracking[J]. Eccv, 2002, 1: 661-675.
- [22] Nummiaro K. An Adaptive Color-based Particle Filter[J]. Image & Vision Computing, 2003, 21(1): 99-110.
- [23] Deutscher J, Reid I. Articulated Body Motion Capture by Stochastic Search[J]. International Journal of Computer Vision, 2005, 61(2): 185-205.
- [24] Shen C, Hengel A and Dick A. Probabilistic Multiple Tracking[C]// In: International Conference on Digital Applications, Sydney 2003, 1, 399-408. Cue Integration for Particle Filter Based Image Computing: Techniques and Applications, Sydney 2003, 1, 399-408.
- [25] Avidan S. Support vector Tracking[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004, 26(8): 1064-1072.
- [26] Maddalena L, Petrosino A. A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications[J]. IEEE Transactions on Image Processing, 2008, 17(7): 1168-1177.
- [27] Wang B, Dudek P. [IEEE 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)-Columbus, OH, USA (2014.6.23-2014.6.28)] 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops-A Fast Self-Tuning Background Subtraction Algorithm[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. IEEE, 2014: 401-404.
- [28] 孙水发, 夏平, 龚国强. 冈萨雷斯版“数字图像处理”教材的教学方法研究[J]. 中国电力教育, 2013(16): 88-89.
- [29] 李晓梅. 基于图像处理的钢轨表面缺陷检测与定位研究[D]. 兰州: 兰州交通大硕士学位论文, 2018: 19-20.

- [30] Cumani A, Guiducci A, Grattoni P. Image Description of Dynamic Scenes[J]. Pattern Recognition, 1991, 24(7): 661-673.
- [31] 王立. 基于中值滤波和小波变换的改进型图像去噪研究[D]. 北京: 北京邮电大学硕士学位论文, 2009: 7-9.
- [32] 王永鑫, 刁鸣, 韩闯. 基于迭代直方图均衡化的常规光源下水下成像增强算法[J]. 光子学报, 2018, 47(11): 97-107.
- [33] 唐泉, 樊晓平, 黎燕. 基于分块聚类的车牌图像二值化方法[J]. 计算机仿真, 2010, 27(3): 286-289.
- [34] Pun T. A New Method for Grey-Level Picture Thresholding Using the Entropy of the Histogram[J]. Signal Processing, 1980, 2(3): 223-237.
- [35] 马秋芳. 面向工业应用的图像颜色识别方法研究[D]. 哈尔滨: 哈尔滨工业大学硕士学位论文, 2016: 28-34.
- [36] 陈海霞, 崔茜. 复杂背景下基于颜色的车牌定位方法研究[J]. 现代电子技术, 2016(23): 60-62.
- [37] 邹丽晖. 基于纹理特征的木材表面缺陷识别方法的研究[D]. 哈尔滨: 东北林业大学硕士学位论文, 2007: 20-25.
- [38] 罗文婷, 李中轶, 李林, 等. 基于改进 Canny 边缘检测算法的道路标线自动识别及定位[J]. 西南交通大学学报, 2018, 53(6): 1-10.
- [39] 毛星云, 冷雪飞, 王碧辉, 等. OpenCV3 编程入门[M]. 北京: 电子工业出版社. 2015: 304-306.
- [40] 颜微. 改进的二维阈值图像分割方法[D]. 湘潭: 湘潭大学硕士学位论文, 2016: 15-19.
- [41] Adandkon M M, Cheriet M. Support Vector Machine .Pattern Recognition Letters, 2009, 1(3): 1303-1308.
- [42] 朱亚辉, 黄襄念. SVM 方法在模式识别应用领域中的发展与研究[J]. 现代计算机, 2015(4): 20-24.
- [43] 余承波. 基于支持向量机(SVM)的车牌识别[D]. 淮南: 安徽理工大学硕士学位论文, 2018: 38-41.
- [44] 王昕. 基于 EEMD 和支持向量机的天然气管道泄漏诊断[D]. 大庆: 东北石油大学硕士学位论文, 2015: 15-18.
- [45] 王至超. 支持向量机方法及应用研究[D]. 济南: 山东师范大学硕士学位论文, 2012: 21-23.

- [46] 薛洁, 王剑平, 张果. SVM 分类器在继发性干燥综合征诊断中的应用[J]. 价值工程, 2016, 35(26): 231-234.
- [47] Ha M, Wang C, Chen J. The Support Vector Machine Based on Intuitionistic Fuzzy Number and Kernel Function[J]. Soft Computing, 2013, 17(4): 635-641.
- [48] Song H, Ding Z, Guo C, et al. Research on Combination Kernel Function of Support Vector Machine[C]// International Conference on Computer Science & Software Engineering. IEEE, 2008: 838-841.
- [49] 张倩, 杨耀权. 基于支持向量机核函数的研究[J]. 电力科学与工程, 2012, 28(5): 42-45.
- [50] 周涛丽. 基于支持向量机的多分类方法研究[D]. 成都: 电子科技大学硕士学位论文, 2015: 27-29.
- [51] 刘宏, 李锦涛, 崔国勤. 基于 SVM 和纹理的笔迹鉴别方法[J]. 计算机辅助设计与图形学学报, 2003, 15(12): 1479-1484.
- [52] 薛洁. 基于参数优化的 SVM 分类器在继发性干燥综合征诊断中的应用[D], 昆明: 昆明理工大学硕士学位论文, 2017: 27-29.
- [53] Meier T, Ngan K N. Video Segmentation for Content-Based Coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8): 1190-1203.
- [54] Gupte S, Masoud O, Martin R F K, et al. Detection and Classification of Vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2002, 3(1): 37-47.
- [55] Barron J L, Fleet D J, Beauchemin S S. Performance of Optical Flow Techniques[J]. International Journal of Computer Vision, 1994, 12(1): 43-77.
- [56] 张赛钰, 朱小玲, 汪衍广, 等. 基于帧差法与 Mean-shift 算法相结合的运动熔滴识别与跟踪方法[J]. 上海交通大学学报, 2016, 50(10): 1605-1608.
- [57] 郭向威. 基于改进光流法的运动目标检测与跟踪研究[D]. 南宁: 广西大学硕士学位论文, 2018: 23-25.
- [58] 卓凌波. 智能 PID 算法在矿山充填自动控制中的研究与应用[D]. 长沙: 湖南师范大学硕士学位论文, 2016: 29-31.

攻读硕士学位期间承担的科研任务与主要成果

(一) 参与的科研项目

- [1] 史艳国, 于立强, 申胜宇, 刘明余. 205-F 和 215-F 轿车玻璃检测与分析, 企业合作项目.
- [2] 史艳国, 于立强, 申胜宇, 刘明余. 旭硝子汽车玻璃 DBO 工装设计, 企业合作项目.
- [3] 史艳国, 刘明余, 于立强, 申胜宇. 威卡威汽车车窗亮条检测.
- [4] 史小华, 于立强, 杨超儒. 融合 CPS 的《单片机原理及应用》课程建设, 单片机课程教改. 课题编号. 2018CPS13.

(二) 科研获奖

- [1] 史小华, 于立强. 助老/助残可如厕智能电动轮椅. 第六届全国“TRIZ”杯大学生创新方法大赛二等奖, 2018.
- [2] 史小华, 于立强. 《椅行天下—养老/康复可如厕智能轮椅》. “创青春”河北省大学生创业大赛一等奖, 2018.
- [3] 林赛, 于立强. 第四届河北省“互联网+”大学生创新创业大赛金奖, 2018.

致 谢

转眼间已经过了三年的研究生生活，研究生生活将结束。在这三年中无论是生活还是学习上帮助过我的老师与同学，在这里我都致以最真诚的感谢。从开始对研究生的憧憬到对以后工作的憧憬，这三年中不仅在学习上让我更进一步，而且对自己以后的生活与工作方式更加清晰。

衷心感谢导师史艳国教授对本人的精心指导。他的言传身教将使我终生受益。史老师在学习上对我们的严格要求和对我们生活的热情关心。在学术上史老师有着对科学的严谨思维，对教育的认真和执着，给我留下深刻的印象，对我以后的研究工作起很大指导作用。史老师从开始到论文的完成过程中，感谢老师的严格把关，使我顺利的完成本文。

史小华老师在我研究生期间，带领我们参加各种科研比赛，增加我们动手能力与扩展我们眼界，同时感谢对我论文的帮助与指导。

在这期间课题室的同学给与我的关心与帮助，在这里我由衷的感谢你们对我的帮助与关心，和你们在一起使我的生活变得精彩。

另外，非常感谢我的父母，没有你们对我的支持和鼓励，就没有我的今天。衷心的感谢你们在物质与精神上的帮助与关心，是我坚强的后盾。