



浙江工业大学

# 本科毕业设计（论文、创作）

题目：基于全局视觉的多机器人  
定位方法研究

作者姓名 张林语

指导教师 赵云波教授

专业班级 自动化 1404 班

学 院 信息工程学院

提交日期 2018 年 6 月 20 日

浙江工业大学本科毕业设计论文

基于全局视觉的多机器人定位方法研究

作者姓名：张林语

指导教师：赵云波教授

浙江工业大学信息工程学院

2018年6月

**Dissertation Submitted to Zhejiang University of Technology  
for the Degree of Bachelor**

**Research on Multi-robot Localization Method  
Based on Global Vision**

**Student: Zhang Linyu**

**Advisor: Professor Zhao Yunbo**

**College of Information Engineering  
Zhejiang University of Technology**

**June 2018**

# 浙江工业大学

## 本科生毕业设计(论文、创作)诚信承诺书

本人慎重承诺和声明：

1. 本人在毕业设计（论文、创作）撰写过程中，严格遵守学校有关规定，恪守学术规范，所提交的毕业设计（论文、创作）是在指导教师指导下独立完成的；

2. 毕业设计（论文、创作）中无抄袭、剽窃或不正当引用他人学术观点、思想和学术成果，无虚构、篡改试验结果、统计资料、伪造数据和运算程序等情况；

3. 若有违反学术纪律的行为，本人愿意承担一切责任，并接受学校按有关规定给予的处理。

学生（签名）：

年 月 日

# 浙江工业大学

## 本科生毕业设计（论文、创作）任务书

专业 自动化 班级 自动化 1404 班 学生姓名/学号 张林语/201403080725

一、设计（论文、创作）题目：基于全局视觉的多机器人定位方法研究

二、主要任务与目标：

1. 了解掌握多机器人定位常规算法 2. 设计定位系统，实时的获取视野内的每个机器人的二维坐标。3. 对系统进行分析和实现。

三、主要内容与基本要求：

1. 阅读相关文献，了解该领域基本研究现状； 2. 研究多机器人定位相关方法； 3. 提出并验证相关的定位系统； 4. 撰写毕业论文。

四、计划进度：

2018 开学前 收集相关资料文献，学习相关知识，完成外文翻译、文献综述；熟悉课题，做好开题准备。

第 1-3 周 完成开题报告，参加开题交流

第 4-8 周 设计系统并作验证等工作，接受中期检查

第 9-14 周 进行系统设计和实现等工作。撰写毕业论文初稿

第 15 周 论文修改，毕业答辩，提交相关文档资料

五、主要参考文献：

[1] 《与人共融的家庭服务机器人定位与导航》，段朋，山东大学，博士论文，2015 [2] 《基于视觉的多机器人协作 SLAM 研究》，苑全德，哈尔滨工业大学，博士论文，2016 [3] 《面向机器人的多目视觉检测算法研究及系统实现》，柯丰恺，华中科技大学，博士论文，2016

任务书下发日期 2017 年 12 月 20 日

设计（论文、创作）工作自 2017 年 12 月 20 日至 2018 年 6 月 20 日

设计（论文、创作）指导教师 赵云波

系主任（专业负责人） 杨马英

主管院长 张有兵

# 基于全局视觉的多机器人定位方法研究

## 摘 要

各式各样的机器人已经被广泛应用在多个领域中，机器人的移动作为一个极其重要的环节，涉及到了对机器人的定位。而无论是对单独的一个机器人，或是多机器人进行定位，选择方法的效果将直接影响到最终的结果。

旨在更好地利用视觉信息来进行多机器人的定位，本文提出了一种基于全局视觉的多机器人定位系统的设计方案。该定位系统在 MATLAB 上对固定于机器人活动区域上方的摄像头采集到的视觉信息进行相关处理后，能够实现多个机器人的定位，即输出每个机器人的二维坐标。

论文的主要工作如下：

1. 简述课题研究背景与意义及机器人定位方法的研究现状，介绍了一些定位方法并分析各自优缺点。

2. 建立一个基于 MATLAB 的开发环境上的视觉定位系统，对已有的几种检测与跟踪算法进行介绍和性能验证，对比后选取合适的方法并加以改进。

3. 将选择出来的检测和跟踪算法整合在一起，得到一种基于视觉的多目标跟踪算法的设计方案，介绍基本流程加以实现。然后由质心计算确定每个目标在图像坐标系中的二维坐标，根据透视原理确立转换关系，最终完成每个机器人的坐标输出。

**关键词：**视觉信息，多目标，检测跟踪，透视变换，定位

# Research on Multi-robot Localization Method Based on Global Vision

## ABSTRACT

Various kinds of robots have been widely used in many fields. The movement of robots is an extremely important part, involving the localization of robots. Regardless of whether it is a single robot or multi-robot, the effect and choice of the method will directly affect the final result.

In order to make better use of visual information for localization, this paper presents a global vision based mobile robot positioning system design. The localization system performs related processing on the visual information collected by a camera fixed above the active area of the robot on MATLAB, and could realize the localization of robots, that is, output the two-dimensional coordinates of each robot.

The mainly work of my study is as follows:

1. Introduced the background and significance of the research, including the research position of the localization method. Presented some methods and analyzed their advantages and disadvantages.

2. Based on MATLAB's development environment, establish a vision system, introduce several existing detection and tracking algorithms and verify performances. Select and improve the most suitable methods after comparison.

3. Combine the selected detection and tracking algorithms together, propose a vision-based multi-target tracking algorithm, and introduce the basic flow to achieve. Then the centroid calculation determines the two-dimensional coordinates of each target in the image coordinate system, establishes the conversion relationship according to the perspective principle, and finally completes the coordinate output of each robot.

**Key Words:** Visual information, Multi-target, Detection tracking, Perspective transformation, Localization

# 目 录

摘要 .....	I
ABSTRACT.....	II
目录 .....	III
<b>第 1 章 绪论 .....</b>	<b>1</b>
1.1 课题研究背景及意义.....	1
1.2 移动机器人的自定位.....	2
1.2.1 单传感器定位方法 .....	2
1.2.2 多传感器定位方法 .....	3
1.3 基于视觉的定位方法.....	4
1.4 主要研究内容.....	7
<b>第 2 章 定位系统的多目标检测 .....</b>	<b>8</b>
2.1 图像的预处理操作.....	8
2.2 两种传统检测算法的性能对比.....	9
2.2.1 帧间差分法 .....	10
2.2.2 背景差分法 .....	12
2.3 定位系统检测算法的选择与改进.....	15
2.3.1 混合高斯模型的基本原理.....	15
2.3.2 背景建模的基本流程 .....	16
2.3.3 算法的实现及改进 .....	17
<b>第 3 章 多目标跟踪算法的设计 .....</b>	<b>19</b>
3.1 Meanshift 应用于目标跟踪 .....	19
3.1.1 Meanshift 跟踪算法原理与流程 .....	19
3.1.2 Meanshift 算法的实现与结论 .....	21
3.2 基于卡尔曼滤波的目标跟踪算法.....	23
3.2.1 卡尔曼滤波基本原理 .....	23
3.2.2 卡尔曼滤波应用于目标跟踪.....	24
3.3 定位系统跟踪算法的改进与实现.....	26
3.3.1 多目标跟踪的两个难点.....	26



3.3.2	目标与轨迹的匹配过程.....	27
3.3.3	跟踪算法主程序及主要函数.....	29
3.4	跟踪结果展示.....	30
3.5	本章小节 .....	31
<b>第 4 章</b>	<b>基于视觉的多机器人定位.....</b>	<b>32</b>
4.1	坐标变换原理.....	32
4.2	多运动目标的坐标输出.....	34
4.3	应用于多机器人的定位.....	35
<b>第 5 章</b>	<b>总结 .....</b>	<b>37</b>
5.1	工作总结 .....	37
5.2	未来展望 .....	38
<b>参考文献</b>	<b>.....</b>	<b>39</b>
<b>致谢</b>	<b>.....</b>	<b>41</b>

# 第 1 章 绪 论

## 1.1 课题研究背景及意义

移动机器人技术的日趋成熟使其在各行各业被广泛地应用。无论是单独的一个机器人，还是涉及到分工合作的多机器人系统，机器人的移动问题都是必须考虑到的，而机器人在移动过程中的行为控制则需要根据实时的位置信息来进行调控，如何获取位置信息也就是我们所说的定位问题。

移动机器人的精准定位，是一个基本却又关键的问题。移动机器人在执行一项需要进行自主运动任务的过程中，首先要做的是确定自身位置，即定位问题。要想完成移动机器人的定位工作，首先需要为机器人配备一个由各种传感器构成的测量系统。视觉传感器因具有不需要接触测量、信息量丰富、成本低廉等优点，近年来在移动机器人的定位导航中结合以视觉测量、模式识别、图像处理等相关技术，已经得到了广泛应用。

根据定位出发点的不同，机器人的定位系统又被分为两种类型：外界定位系统，如基于顶级摄像机的视觉测量系统，作为一种全局定位系统，它可以对视野范围内的移动机器人进行跟踪以及位姿定位；自主定位系统指移动机器人结合对周围环境信息的评估与测量，通过自身传感器感知外界环境后来得到自身的相对位姿信息。

传统定位方法上获得机器人的位置信息的方式主要通过声纳或激光测距、电磁导航、GPS 等<sup>[1]</sup>。以上提到的几种方法目前的使用依然十分广泛，在结构化环境下依然能够保证一定的准确和有效性。可由于获取信息的局限性，此类方法目前只适用于相对简单且固定的场合。另一方面，随着要求逐渐的提高和应用范围的扩大、计算能力的提升，基于视觉的定位系统逐渐展现出它的优势。除此之外，视觉技术作为一项前沿技术，不仅可以应用在更加智能化的机器人系统上，还能对其他领域提供了一定的参考意义，包括了本文研究的基于视觉的定位方法。可见本次课题的研究的确是具有一定的实际应用意义的。

## 1.2 移动机器人的自定位

作为移动机器人所必须具备的功能之一，自定位功能也在随着时代的进步而被不断完善。在过去，人们多通过声纳或激光测距、电磁导航、GPS 等方式获得机器人的位置信息来完成定位的工作。而根据获取信息的传感器数量差异，移动机器人的自定位方法可分为以下两大类。

### 1.2.1 单传感器定位方法

利用单传感器来进行定位的方法主要分为基于里程计的定位方法以及基于视觉的定位方法这两种<sup>[2]</sup>。基于里程计的定位方法，由于里程计本身存在的累计误差，会导致当机器人运动距离较长时，会产生较大的误差，从而致使定位的不准确。我们需要注意到，视觉传感器因具有能够提供丰富的信息、不需要接触到环境、成本更低等优点，已被广泛应用于机器人的定位上，这一结果使得基于视觉的定位方法显得十分关键。机器人的视觉定位本质上就是基于视觉传感器获得的图像信息，通过鲁棒的估计算法对移动机器人的位姿参数进行相关运算，最终结合其他信息得到机器人的位姿信息<sup>[3]</sup>。

在视觉伺服领域里，目前已经有许多方法可以通过两视角来进行相关的运动估计，其中最典型的方法根据计算方法中的差别，可以将它们分为两种，基于本质矩阵的位姿估计方法和基于单应矩阵的位姿估计方法。

#### 基于本质矩阵的位姿估计方法

基于本质矩阵的位姿估计方法主要利用的是图像之间的对极约束，就是通过对对应关系得到预先给定的一幅图像上的某一特征点在另一幅图像上的匹配点。根据对极约束，我们知道它一定在对应的极线上<sup>[4]</sup>，之后我们可以得出相应的基本矩阵，利用已经标定的摄像机参数对本质矩阵进行计算，然后根据本质矩阵与有关参数之间的关系，利用算法得到最终的相对位姿参数。

在对这类方法进行的诸多研究中，较为经典的一种是八点算法，已经有工作将其应用在了一种机器人视觉导航中。需要注意的是，当视图之间的视差较小或周围环境的特征点是均分在一个平面上时，会有很大几率上发生类似于退化等情况，且得到的基本矩阵还可能出现病态，最终导致不能得出稳定的位姿估计解。

国外学者提出的一种五点法可以有效避免平面场景的奇异性，而由两幅视图估计出的多种可能的位姿解还需要由第三幅视图来确定解的唯一性。已经有研究表明，五点法已经被成功地应用到全景摄像机的视觉测距以及一些移动机器人的定位中，这种方法的问题在于增加了运算时间，不仅需要利用光束法来平差，其中的三角测量还需要额外的一些三维信息。

### 基于单应矩阵的位姿估计方法

单应矩阵描述了一种图像之间对应特征匹配点之间的映射关系。简单讲，就是在两幅图像中提取出至少四组特征匹配点，同时满足其中任意三点不在同一条直线上，然后再用线性和非线性优化的方法来对单应矩阵估算。因为在摄像机内的参数已知时，单应矩阵仅与相对位姿参数相关，所以可通过位姿分解方法来对单应矩阵进行分解计算，最终得到相对位姿参数。

这一方法有效地避免了当极线约束失效时会出现的退化情况，同时加入的平面场景的约束能够使它更加的精确和稳定。已经有不少国内外研究工作者对这一方法做了大量研究，可是这些分解方法很多都需要进行繁琐的奇异值分解，还可能会出现多解现象。国内有学者就此给出了一种快速的单应矩阵分解方法，利用它解决了以上问题之后，单应矩阵的分解效率得到了明显的提高。

#### 1.2.2 多传感器定位方法

噪声存在的影响以及获取信息的有限性使单传感器定位方法很难一直保证输入信息的可靠性和准确性。多传感器定位通过对多传感器获取的信息进行融合在一定程度上可以有效解决信息量不足的问题。多传感器定位方法主要用到了用非线性滤波的方法，主要有以下三种：

- **扩展卡尔曼滤波方法**

作为卡尔曼滤波的一种拓展性应用，扩展卡尔曼滤波假设机器人在非线性系统中的系统状态是服从高斯分布的，同时它的运动观测模型也契合高斯噪声模型。这样的话，我们就能通过预测、观测和更新这三个传统步骤来对其位姿进行相关推算，最后得到最小二乘法意义下的最优位姿估计。文献<sup>[5]</sup>就有对该

定位方法进行了相关研究。它的缺点是，系统模型或噪声模型的误差较大时十分容易造成定位的失败，而且这一方法一般不适用于全局定位。

### ● 马尔可夫定位方法

马尔可夫定位方法将整个状态空间进行了离散化，与扩展卡尔曼滤波方法相比，它并没有局限于高斯假设分布，而且在表示机器人位姿的概率分布时，使用的是离散化后的空间单元的概率密度。目前已经分化出了很多不同的马尔可夫定位方法中最常用的两种是基于栅格的马尔可夫定位和拓扑马尔可夫定位。这一方法的缺点是效率较低，因为需要记录整个状态空间的概率分布，所以它会消耗大量的内存资源和计算时间<sup>[6]</sup>。

### ● 蒙特卡洛定位方法

蒙特卡洛定位方法的关键在于加权采样的过程，即指在机器人的位姿空间，我们能够根据实际概率分布赋予每一份粒子不同的权重，用一堆随机的采样粒子来模拟机器人位姿，最终就能够通过有限的离散样本近似得到机器人的位姿状态。该问题在国内外已经得到了大量充分的研究。与马尔可夫定位方法一样，蒙特卡洛定位方法并不局限于类似高斯分布假设的情况，但它的劣势是会有样本退化的情况出现<sup>[7]</sup>。

人们通常习惯用蒙特卡洛方法来处理地图定位问题（有时被称为绑架机器人问题）。在这里，马尔可夫假设被用来保持代理人姿态的基于粒子的后验表示，积分激光或视觉观察。通常情况下，这些方法只能在没有提供任何全球地理定位信息的小环境中运用。此外，最重要的是，他们依赖更具体的测量来源（如深度测量，声纳），并且只能在准确的二维平面图可用的小规模环境（如办公场景）中使用。虽然这里介绍的方法与这些方法相似，但是也提出了递归贝叶斯过滤算法，但这些方法使用了基于粒子的滤波技术，而由于众所周知的粒子损耗问题，这些滤波技术不能处理持续的不确定性<sup>[8]</sup>。

## 1.3 基于视觉的定位方法

根据摄像机数目的不同，可以将基于获取视觉信息的定位方法归纳为三种，双目视觉定位、多目视觉定位以及单目视觉定位：

## ● 双目视觉定位

双目视觉也被理解为立体视觉，正如人有一双眼睛一样，从两个视点来观测同一个环境，可以获取到更多的信息。利用三角测量原理，根据获取到的不同视角下的图像信息就可以通过计算得到两幅图像之间的像素偏差信息，最终得到目标的三维信息并完成定位。在双目视觉定位中，作为最关键的一部分，同时也是难度最大的环节，立体匹配存在着一些有待被更好地解决的问题，例如，具有特殊特征的目标可能会导致假匹配情况的出现；平行核线方向上的特征匹配经常会出现的歧义；随着摄像机之间距离增大，遮拦加重会导致可用来重建空间点的减少，最终增大匹配的难度。

针对以上的匹配问题，多种可应用于室内环境的视觉导航系统设计方案已经被提出，都可以有效解决相关的匹配问题，但它存在的劣势是算法的复杂性和偏高的计算成本。

## ● 多目视觉定位

多目视觉又称为全维视觉，利用多个摄像机来获取视觉信息，它可以用来获得范围更大的信息，包括同一目标的不同部分的信息。然而在得到更加丰富的信息后，不仅对被要求完成多幅图像之间匹配任务的视觉控制器提出了更高的要求，同样还容易导致图像畸变情况的增多。除此之外，系统的实时性和稳定性都难以得到保证。

## ● 单目视觉定位

单目视觉获取信息的方式相较于上面两种，显得信息量较少，与此同时也少去了一些处理上的麻烦。既不用考虑特征点的匹配问题和最优距离的设定，也不需要担心会出现类似于全维视觉中会产生较大畸变的现象。具有简单易用与适用范围广的特点使得它在可获得较多先验信息的环境中可以取得较好的效果，例如一种被应用广泛的使用到最小二乘互相关算法和贝叶斯算法的自定位方法。需要注意的是，由于最小二乘互相关算法的计算较为缓慢，在实时要求下能够处理的图像大小非常有限，而且不能很好的提取图像中的一些特征信息。鉴于此，有学者提出了一种基于环境特征的定位方法，它为基于视觉特征的定位方法提供了

一些更加高效的思路<sup>[9]</sup>。

实际上，机器人的视觉系统具体又分为两类，一种是通过安装在机器人身上的设备来采集视觉信息的，另一种则是由固定安装在机器人活动区域上方的设备来完成信息获取的<sup>[10]</sup>。然而安装在机器人自身上的设备进行获取视觉信息时，可能会出现自身晃动、视角局限或者物体遮拦这类会给后期信息处理带来麻烦和误差的问题<sup>[11]</sup>。后一种方法不用考虑这些问题的同时，还可以根据实际需求为机器人的定位提供更完善的信息。因此我们这次课题的视觉信息获取方法采用的是后一种，既呼应了基于全局视觉的这一主题，又具有一定的实际应用价值。

### ● 几种典型的视觉定位算法

在利用图像中角点的特征信息来进行三维环境重建的方法中，较为典型的一种是由 Harris 提出的一套 3D 视觉定位系统方案，具体用到卡尔曼滤波器来跟踪角点特征，而相机的运动轨迹和需要获取的三维信息是由跟踪结果来确定。

D.Murray 就移动机器人的地图创建和导航问题，提出过一种立体视觉算法，缺点是对机器人的自定位并未用到创建的地图，而且也未进行里程计的校正，导致误差会逐渐地增大。就此问题，有国内学者提出了一种将二维栅格地图结合三维特征来定位的方法，其中将平面物体的角点视作特征点与后一帧图像进行匹配，却也存在信息的丢失会导致定位精度的下降的问题。

前文提到的马尔可夫定位方法在视觉定位上也已经有了广泛的应用，缺点是不能自动学习地图，意味着需要人工的干预才能进行。对此，一种基于自然可视特征进行位姿估计的定位方法被提出，而其中被当作特征的是由机器人在自学习地图阶段检测到的一幅图像中的一部分缩略图的内容<sup>[12]</sup>。

而在即时视觉定位和地图构建方向，也就是 vSLAM 算法，一个基于局部不变特征的环境地图需要被建立起来完成机器人的定位，其中基于 SIFT 的目标识别和粒子滤波定位方法则可以有效解决“机器人绑架”这类特殊问题。与之相通的 MonoSLAM 算法，是根据单摄像头的轨迹来构建实时环境地图的，主要原理是先建立起一个摄像头的平滑运动模型，环境通过稀疏矩阵显示，最终用到概率估计的方法来跟踪特征点，完成摄像头的实时状态更新<sup>[13]</sup>。

## 1.4 主要研究内容

本课题研究的内容主要是对已有的相关算法进行对比分析之后,选取性能更好的方法加以改进并以此为基础,建立一套基于全局视觉的多运动目标定位系统。目标是能够利用获取到的单目视觉信息,即一段视频,在 MATLAB 上进行整体程序的设计与验证后,完成对视野内的多个机器人的跟踪和定位,结果为实时输出每个目标的二维坐标。本文五章的具体内容安排如下:

第一章作为绪论,主要概述了研究背景与意义,在介绍了几种典型的机器人定位方法并加以简单分析后,说明本文的主要研究方向和章节安排。

第二章,根据课题要求,对算法第一部分的构成,即运动目标的检测进行一些方法的介绍后,进行对应算法程序的设计与验证。最终根据检测结果的对比,选取更合适的一种方法并加以改进,完成整体设计的检测部分。

第三章,利用介绍的 `meanshift` 算法和卡尔曼滤波器的基本原理,将其分别应用于目标跟踪,对比跟踪效果加以分析。然后根据课题要求,选择一种基于卡尔曼滤波器的跟踪算法予以改进,结合第二章的检测部分算法完成多目标跟踪算法的设计,最后在 MATLAB 上完成程序编写。最后完成对不同的两段包含多运动目标的视频进行检测跟踪,并展示跟踪结果。

第四章在介绍了透视变换的基本原理后,将其拓展应用在前文提出跟踪算法中,同时在前文实现的跟踪算法中加入质心坐标输出的功能,然后展示结果。最后对算法整体结构进行介绍,将其与实际要求联系在一起。

第五章则对全文总结,分析不足,展望相关前景。



## 第 2 章 定位系统的多目标检测

本章主要工作是完成对多运动目标视觉定位系统中检测部分的算法选择。按照传统的视觉定位的步骤,先介绍了一种图像颜色模型及图像二值化处理的基本原理。然后对三种常用的目标检测方法进行了具体的分析和验证对比。根据分析对比实验结果,着重介绍了基于混合高斯模型做背景减法这一检测方法的应用,并对其进行改进,确定定位系统检测部分的算法。

### 2.1 图像的预处理操作

要想对一段视频进行处理,首先要了解的是对一幅图像的基本处理,所以接下来将介绍两个有关图像的基本概念。在计算机视觉中,有很多种对图像的颜色信息进行描述的方式,也可以说是存在多种颜色模型,即指一组可以通过数值来描述图像中所包含颜色信息的一个数学模型。对已经获取到的视觉信息进行处理的过程中,合适颜色模型的选择同样是十分重要的一项工作。而在图像的处理过程中,目标的检测经常会用到一种经典图像预处理方法,也将会在下面这一节进行简要介绍。

#### ● RGB 颜色模型

最典型的一种常用彩色模型就是多见于硬件设备的三基色模型,即为这里要介绍的 RGB 模型。它的形成主要是由于人眼中用来分辨颜色的锥状细胞可以将光谱的可见波段分成所谓的三个原色波段,即人眼对真实颜色形成的色觉,形式上都可以由红绿蓝三原色混合出来的一种复合光来形成。

国际上统一规定波长数值 (nm) 分别为 700, 546.1, 435.8 所对应的单色光作为红、绿、蓝三原色。由此可以建立起一个笛卡尔坐标系,其中红(R)、绿(G)、蓝(B)分别对应着三个坐标轴,之后根据三个原色分量的组合将不同的颜色用相同的方式出来。在 RGB 模型中,三个分量值都为零的零点代表的是黑色,而白色则是由三个分量都为最大值 255 的点来表示。值得注意的是,在这个形状为正方体的空间模型中,将原点与相隔距离最远的顶点相连,可以得到一条直观意

义上的对角线，也称为灰度线，而从黑到白的灰度值也正是分布在这条线上的。

RGB 模型一般多应用于显示器之类的硬件设备，但因为是基于光的叠加原理，所以它的缺点是不够直观，两种颜色之间的视觉差异不便于表达。

## ● 图像的二值化处理

作为一种基本的操作，二值化在所有的图像处理中都可能被用到。对于一幅图像我们需要关注的主要部分除去噪声，剩下的就是目标和背景了。二值化指的是将图像中的像素信息只用 0 和 1 两个值来表示，处理结果是图片能够只由黑和白两种颜色组成。这样在检测出移动目标后，就可以将背景与移动的目标的分离结果更加直观地表现出来，同时便于之后的处理。

图像的二值化处理需要先选定一个阈值，然后进行图像灰度值的分类，即将整个图像进行像素灰度值的筛选。其数学表达式为下式：

$$F(x, y) = \begin{cases} 1, & f(x, y) \geq T \\ 0, & f(x, y) < T \end{cases} \quad (2-1)$$

其中  $T$  为预先选定的阈值，当像素点的灰度值  $f(x, y)$  大或等于  $T$  时，就将输出的该点像素的灰度值  $F(x, y)$  设为 1，当输入的函数  $f(x, y)$  的值小于阈值  $T$  时，则将输出函数  $F(x, y)$  赋值为 0。1 对应白色，0 为黑色，这样就可以达到突出目标的效果。

二值化处理的过程中，合适的阈值设定至关重要。固定阈值法通常指人为设置一个固定的根据经验判断得到的阈值。自适应阈值法则是先对图像中的像素进行扫描，再根据扫描的情况对像素进行跟踪，根据像素周围的亮度自动设置一个阈值。多数时候，在经过多次试验后，采用固定阈值法设置阈值来进行目标检测是更为高效的。二值化之后图像的效果将在后续章节进行展现。

## 2.2 两种传统检测算法的性能对比

在对一段包含多目标运动信息的视频进行处理分析时，计算机和程序本身是无法确定我们感兴趣的目標的相关信息。所以我们只有先完成对运动目标的检测，包括提取目标的特征信息，构建合理的参考模型，之后才能进行对多目标的跟踪和定位。目前得到较广泛应用的检测方法主要为帧间差分法和背景差分法，接下来将这两种传统的运动目标检测算法进行简要的原理分析和性能对比。

### 2.2.1 帧间差分法

顾名思义，帧间差分法就是对一段视频中的相邻两帧图像作差分运算，然后设定合适的阈值，对差分后的图像进行二值化处理得到运动目标的轮廓。这种方法已经广泛应用在移动目标的检测识别中，虽然原理简单但是效果却还不错，不失为一种适用的目标检测算法。

实现帧间差分算法的具体步骤如下：

第一步，对一段视频里图像序列中提取连续到的两帧图像 $f_{k+1}(x, y)$ 、 $f_k(x, y)$ ，进行灰度化处理之后，将其进行减运算并取绝对值，得到中间图像 $D_k(x, y)$ ：

$$D_k(x, y) = |f_{k+1}(x, y) - f_k(x, y)| \quad (2-2)$$

第二步，设定合适的阈值，将 $D_k(x, y)$ 二值化，当其中某一点像素值大于阈值 $T$ 时，则将该像素视为目标像素，否则视为背景像素。二值化处理过程具体表现为，未变动的或者小于阈值的赋值 $0$ ，大于等于阈值的取 $1$ ，这样可以突出显示出检测到的移动目标，提取出对应的运动区域：

$$R_k(x, y) = \begin{cases} 1, & D_k(x, y) \geq T \\ 0, & \text{else} \end{cases} \quad (2-3)$$

这种方法在图像中包含多个运动目标或者是存在摄像机移动的情况下，可以得到不错的效果。具体的算法实现也较简单，提取运动目标区域的效果主要取决于对差分图像进行二值化时选取的阈值，在下面的算法实现中，选取的阈值是在尝试多次后得到的，图 2-1 是在 matlab 上编写算法后的运行结果：

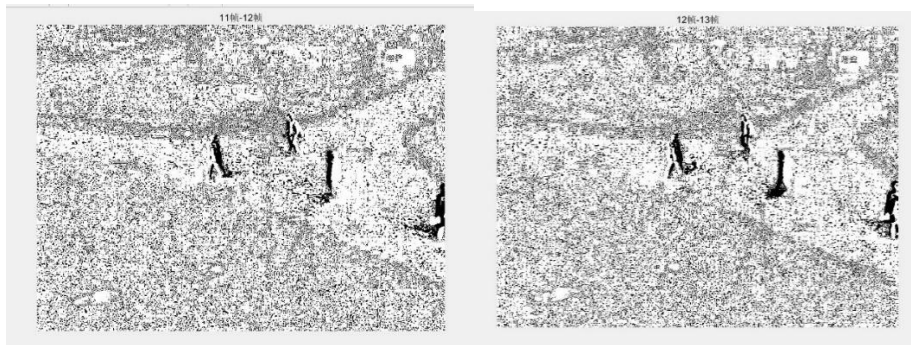


图 2-1 帧间差分法实验结果图

上面两幅图是经过简单的帧间差分之后得到的二值化图像，左边显示的是第 12 帧减去第 11 帧灰度化之后图像的结果，右边的是第 13 帧减去第 12 帧。结果显示，普通的帧间差分处理只能得到视频里运动目标区域的大致轮廓，同时存在着特别多的噪声，这对于要在多样化环境下进行多运动目标的检测无疑是不可接受的。

考虑到在运用上文介绍的帧间差分来提取运动轮廓这种包含了边缘检测的处理时，通常可以先进行一定程度的降噪，再进行下一步处理。所以针对以上出现的噪声情况，本文又采用了中值滤波的方法来进行处理。作为一种非线性数字滤波技术，中值滤波器经常被用来去除图像等信号中的噪声。它是通过检查和判断输入信号中的采样信息是否代表信号，然后利用由这些采样构成的观察窗来实现。首先完成观察窗口中的数值排序，并将位于中间的中值作为输出，最后将原先的值换成新的采样值并对其他信号重复以上操作。

在直接调用 MATLAB 里库函数中的中值滤波函数之后，这里完成了对之前帧间差分法的改进，将运动轮廓在原图像序列中对应的图像里标红显示，这里也利用到了上文中提及的 RGB 模型的原理。最终可以得到在原始图像中对运动目标的检测结果。在 MATLAB 上运行程序，结果如图 2-2 所示：



图 2-2 在第 12 帧检测到的运动目标的轮廓

可以看到,改进后的帧间差分法可以在视频的图像序列中将多运动目标检测出来,同时显示出轮廓,而且噪声过多的现象也有了明显的改善。但是在后续的验证中,我们发现当运动目标高速运动,或是静止不动以及被物体遮挡时,几乎得不到有效的检测结果。

上述问题也属于帧间差分法的缺点,能检测到运动对象的外部轮廓却不能对目标内部进行描述;检测效果较依赖帧间的间隔,当运动的目标在相邻帧中没有重叠时,检测会得到两个不同的目标轮廓;相对的,当目标在前后帧中近似于完全重叠时,结果将不会显示这一目标。因此,即使帧间差分法具有速率高,实时性强且不易受类似光照之类环境因素的影响,后文中,我们仍然不会使用帧间差分法来完成目标的检测。

### 2.2.2 背景差分法

背景差分法又称为背景相减法、减背景法,它的流程大致可以分为三步:建立一个背景模型、完成背景模型的更新以及前后景的分割。它的基本原理是通过特定方法获取到一段视频的背景图像,然后将每一帧的图像与背景图像做减法,从而得到输入图像与背景图像不同的区域,进而实现对视频中运动目标区域的检测。跟帧间差分法一样,表面上看起来原理简单,但背景模型构建方式的选取和更新方式都决定了最终的检测结果,因此选取合适又高效的背景方式是至关重要的。

本文根据模型建立的方式不同,将常用的背景差分法分为如下三类:简单背景建模法,基于统计信息的背景建模法和背景预测法。其中简单的背景建模法中包括均值法、中值法;复杂点的有,卡尔曼滤波器模型法、单高斯分布模型法以及下文中将着重介绍的混合高斯背景建模法。

首先本文根据背景差分算法的基本原理,在 MATLAB 上进行了相关算法的编写与功能实现,由于是测试算法的检测性能,所以只选取了起始前十帧的图像来完成背景模型的构建,用到的是均值法。背景图与检测结果如下两图所示:



图 2-3 构建的二值化背景图



图 2-4 对第 12 帧的差分结果

因为目的是为了对比检测算法的性能，所以这里用到的不是完整的背景差分算法，并没有加入背景的更新部分。从结果上来看，检测到的运动目标不同于帧间差分法只能检测到目标的轮廓，虽然不能断言背景差分法比帧间差分法好，但考虑到多目标跟踪过程中物体最常出现的静止情况，我们可以考虑用背景差分法来完成检测部分的设计。然而在运用背景差分法时，只是用均值法建模很明显是不能够应对实际需求中背景环境的复杂多变性的，同时背景的更新也是需要考虑到。

下面加入背景更新这一操作之后，试着在 MATLAB 上编写了基于背景信息统计的背景差分算法程序并将其实现，结果如图 2-5 所示：

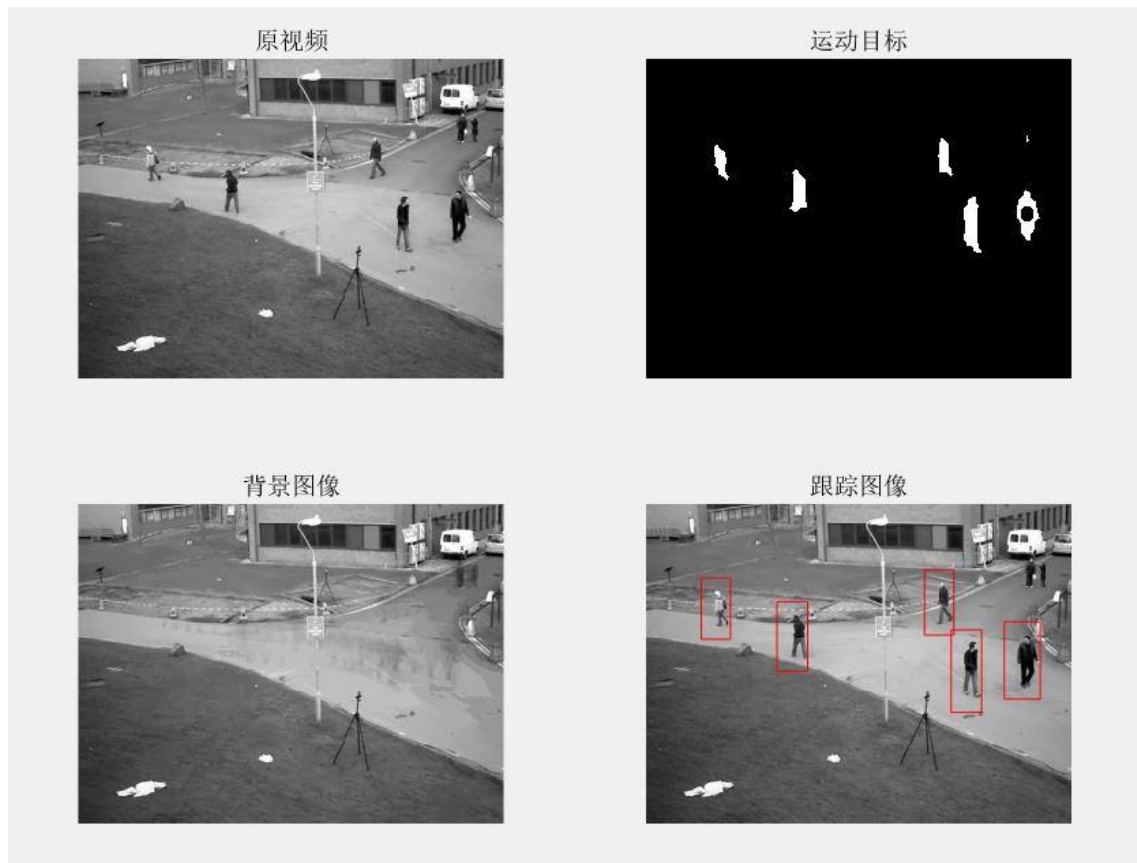


图 2-5 背景差分法的实现结果

可以看到，对视频每一帧图像进行灰度化处理之后，用相邻帧间差减的运动估计方法来构建背景，具体操作是根据阈值判断是否属于背景图像并加以统计，

得到左下角的背景图像。与当前帧做差分后得到右上角的检测效果，以二值图像显示。将检测到的运动目标框选出来，将其显示在原始帧图像上，如果对每一帧进行同样的处理后再按照图像序列还原成视频播放，即可完成某种意义上的目标跟踪，跟踪结果如右下角所示。但在实际算法的运行中，这种方法存在速率过于缓慢以及噪声干扰过大等问题。

## 2.3 定位系统检测算法的选择与改进

### 2.3.1 混合高斯模型的基本原理

混合高斯模型的背景建模是基于统计的样本信息来进行的。它将背景通过像素在一段时间内统计的样本值的概率密度等信息进行表示，并利用统计差分的方法完成对目标像素的判断，最终完成复杂背景的建模。

混合高斯背景模型首先假定的是，各像素之间的颜色信息是相互独立的，因此像素点的处理过程也是互不干扰的。同时将像素点的值在视频序列图像中的变化视作为一个不断随机生成像素点的值的过程，这样就可以用高斯模型来描述像素点的变化规律。

在常用对应多峰的多模态高斯分布模型中，图像像素点都是根据高斯分布的叠加建模的，不同的每一高斯分布对应一种不同的状态，对应状态下产生的像素点的颜色信息也不同。同时分布的权值和相关参数都是在不断更新的。假设在处理图像时，相互独立的像素点 R、G、B 三通道具有一样的方差，并将随机变量 X 的观测数据表示为 $\{x_1, x_2, \dots, x_N\}$ ， $x_t = (r_t, g_t, b_t)$ 对应 t 时刻的像素样本，则满足混合高斯分布采样点 $x_t$ 的概率密度函数如下所示：

$$p(x_t) = \sum_{i=1}^n w_{i,t} \times \eta(x_t, \mu_{i,t}, \tau_{i,t}) \quad (2-4)$$

$$\eta(x_t, \mu_{i,t}, \tau_{i,t}) = \frac{1}{|\tau_{i,t}|^{1/2}} e^{-\frac{1}{2}(x_t - \mu_{i,t})^T \tau_{i,t}^{-1} (x_t - \mu_{i,t})} \quad (2-5)$$

$$\tau_{i,t} = \delta_{i,t}^2 I \quad (2-6)$$



其中 $n$ 指包含的分布总数值。 $\eta(x_t, \mu_{i,t}, \tau_{i,t})$ 对应 $t$ 时刻的第 $i$ 个高斯分布。 $\mu_{i,t}$ 为均值， $\tau_{i,t}$ 为协方差矩阵， $\delta_{i,t}$ 为方差。 $I$ 是三维矩阵， $w_{i,t}$ 指的是 $t$ 时刻第 $i$ 个高斯分布的权重值。

### 2.3.2 背景建模的基本流程

1. 将 $K$ 个模型下每个点的像素值 $X_t$ 按下式完成比较，即可得到与新像素值匹配的分佈模型，且其均值偏差在 $2.5\sigma$ 内：

$$|X_t - \mu_{i,t-1}| \leq 2.5\sigma_{i,t-1} \quad (2-7)$$

2. 将像素值在匹配的模型符合要求时归为背景，否则归为前景。

3. 每个模型的权值更新如下式所示，其中 $\alpha$ 代表的是学习速率。当模型符合匹配时，令 $M_{k,t} = 1$ ，否则 $M_{k,t} = 0$ ，之后完成每个模型权重的归一化：

$$w_{k,t} = (1 - \alpha) * w_{k,t-1} + \alpha * M_{k,t} \quad (2-8)$$

4. 当未匹配模型的均值和标准差，即 $\mu$ ， $\sigma$ 保持不变时，则匹配模型的参数更新按下式进行：

$$\rho = \alpha * \eta(X_t | \mu_k, \sigma_k) \quad (2-9)$$

$$\mu_t = (1 - \rho) * \mu_{t-1} + \rho * X_t \quad (2-10)$$

$$\sigma_t^2 = (1 - \rho) * \sigma_{t-1}^2 + \rho * (X_t - \mu_t)^T (X_t - \mu_t) \quad (2-11)$$

5. 当第一步就没有出现匹配时，需要替换权重值最小的模型，令其像素值为该模型的均值，初始标准差取较大值，权重取较小值。

6. 将每个模型都根据 $w/a^2$ 进行降序排列，把权重大和标准差较小的模型排在前面。

7. 最后，将背景用前 $B$ 个模式表示，参数 $T$ 代表其中每个背景所占比例：

$$B = \arg(\min(\sum_{k=1}^b w_k > T)) \quad (2-12)$$

### 2.3.3 算法的实现及改进

根据混合高斯背景建模的原理和思路。在 MATLAB R2016a 上进行具体的代码实现，并对一段视频图像序列的前二十多帧进行统计背景信息建模处理，然后将每一帧图像与之差分，得到如图 2-6 所示结果：

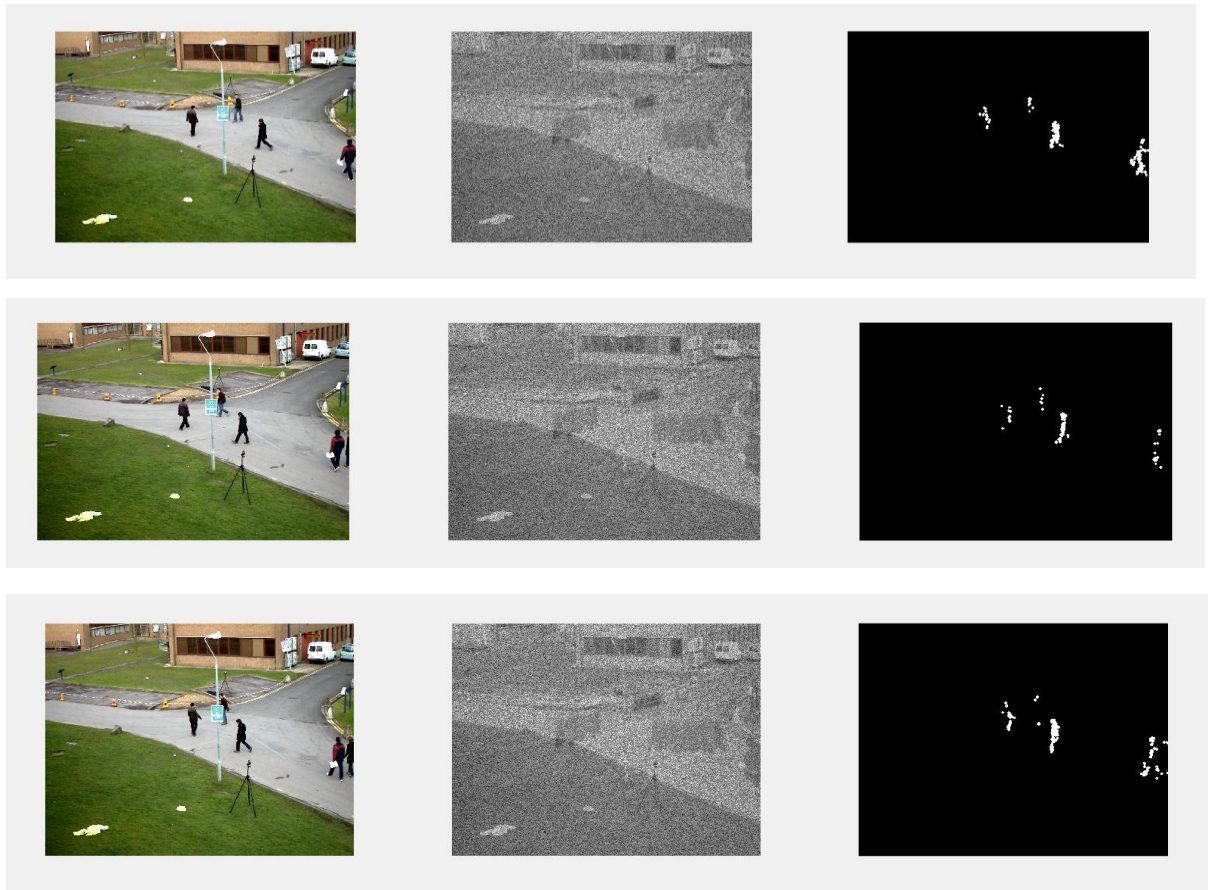


图 2-6 混合高斯背景建模实验结果

第一幅图像显示的是当前帧原图像，中间图像表示建立的背景图像，第三幅图显示了前景，表现为对运动目标的检测结果。对于在前后景分离之后，检测效果并不是很好，在这里的改进是利用形态学滤波操作来消除前景图中的噪声，并用 blob 分析子检测连通域，最终得到要跟踪的运动目标。

- 腐蚀与膨胀：腐蚀指通过消除边界点使边界向内部收缩的过程，经常被用来消除细小的噪声。膨胀则是将于目标接触的所有背景点合并到该目标中，使

边界向外部扩张的过程，可以用来填补物体中的空洞<sup>[14]</sup>。

- 开运算：一个先进行腐蚀后膨胀操作的过程，通常被用来消除小颗粒噪声，在细微处分离断开目标物之间的粘结，不明显改变面积的同时平滑目标物体的边界。其主要作用与腐蚀相似，与其相比，增加了操作过程中可以基本保证物体原有面积不变这一优点<sup>[15]</sup>。
- 闭运算：一种先进行膨胀后腐蚀操作的过程。常用于解决物体内出现细小空洞的现象，它还可以连接相邻物体，并且在平滑边界的同时并不明显改变目标物体的面积。

而 Blob 分析是指对从图像中获取的二值区域，进行面积、周长以及质心等特征的分析<sup>[16]</sup>。在程序中将会用 blob 分析得到所有连通域的中心以及跟踪矩形框的大小。在加入了以上的形态学操作和通过 blob 分析子检测连通域的方法，在消除噪声的同时，优化了检测效果，部分截图如下图所示：



图 2-7 改进后的检测结果

对比以上的检测结果，我们发现经过改进后的基于混合高斯背景建模的背景差分法检测效果基本符合后面跟踪部分的要求，所以在后文中将使用这种检测方法来检测运动目标并进行后续处理。

## 第 3 章 多目标跟踪算法的设计

本章将首先介绍两种目标跟踪方法，Meanshift 算法和基于卡尔曼滤波的目标跟踪算法，并在 MATLAB 上进行实验结果的对比，选取效果较好的并加以改进。然后在上一章的基础上，完成多目标跟踪算法的设计，展示成果。

### 3.1 Meanshift 应用于目标跟踪

#### 3.1.1 Meanshift 跟踪算法原理与流程

Meanshift 算法是基于梯度的无参数密度估计上的一种方法，也指一个迭代的步骤<sup>[17]</sup>。在将当前点移动到根据计算得到的它的偏移均值处后，将其视作为新的起始点，继续这种移动，迭代至达到一定要求才结束。实质上，它是一种核密度估计算法，每一次的迭代都指向密度函数的局部极大值处。

Meanshift 的基本原理如下：

假设在空间  $R^d$  内给定  $n$  个数据点， $x_i \in R^d$ ， $i=1, \dots, n$ ，在某一点  $z$  的核概率密度通过下式进行估计：

$$P(z) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{z-x_i}{h}\right) \quad (3-1)$$

其中， $h$  是带宽， $K(z)$  即为核函数。因为核函数需要满足收敛性，所以在这里采用的核函数是高斯核函数，表达式如下：

$$K(z) = \frac{1}{2\pi} \exp\left(-\frac{|z|^2}{2}\right) \quad (3-2)$$

引入核函数  $K(z)$  的轮廓函数  $g(z): [0, +\infty] \rightarrow R$  满足  $K(z) = g(|z|^2)$ 。

对公式 (3-1) 中  $P(z)$  求梯度，得到预测位置：

$$\nabla P(z) = -\frac{2}{h^2} \left[ \frac{1}{nh^d} \sum_{i=1}^n g\left(\left|\frac{z-x_i}{h}\right|^2\right) \right] \left[ \frac{\sum_{i=1}^n x_i g\left(\left|\frac{z-x_i}{h}\right|^2\right)}{\sum_{i=1}^n g\left(\left|\frac{z-x_i}{h}\right|^2\right)} - z \right] \quad (3-3)$$

其中 Meanshift 向量即为后面括号中的项，

定义  $z_{k+1} = \frac{\sum_{i=1}^n x_i g(\frac{z-x_i}{h})}{\sum_{i=1}^n g(\frac{z-x_i}{h})}$  的迭代过程，可以看出当核函数单调递减时，

Meanshift 向量是收敛的。化简得到标准迭代形式为，

$$z_{k+1} = z_k + \frac{\sum_{i=1}^n (x_i - z_k) g(\frac{z-x_i}{h})}{\sum_{i=1}^n g(\frac{z-x_i}{h})} \quad (3-4)$$

在介绍了 Meanshift 算法的基本原理之后，将其拓展应用到目标跟踪时，具体算法流程如下图所示：

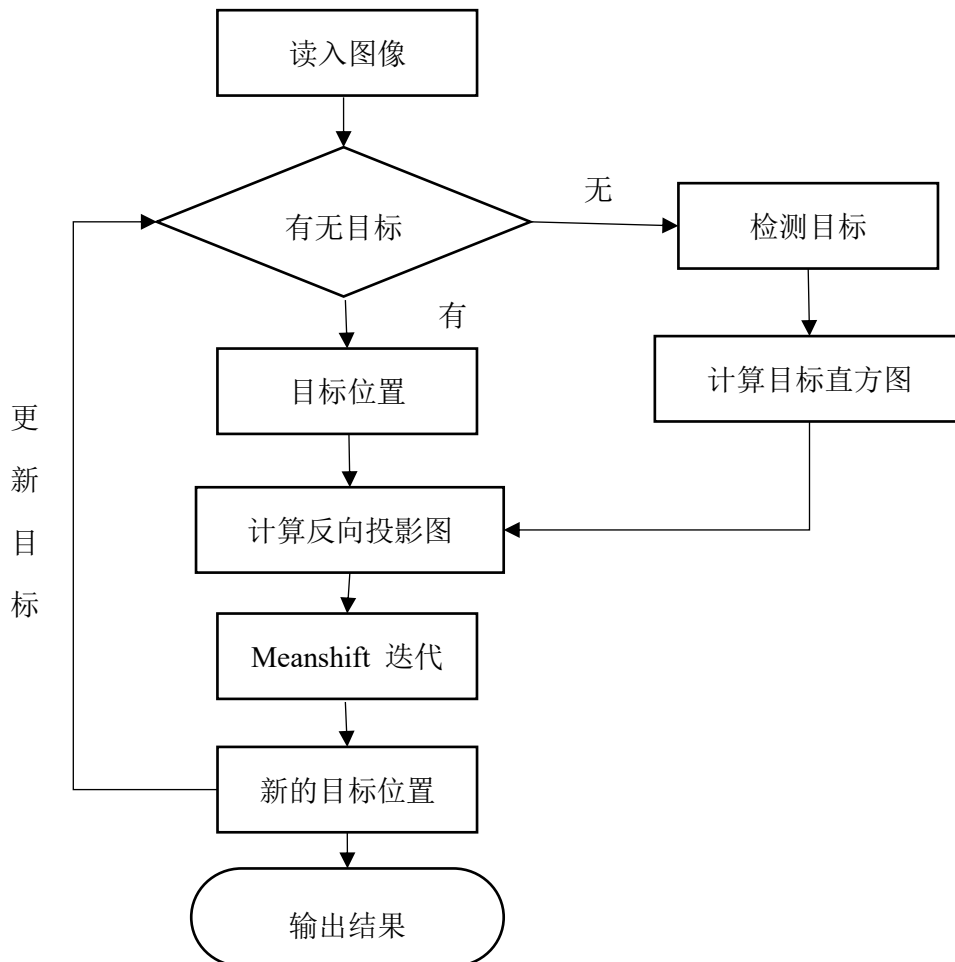


图 3-1 MeanShift 算法流程图

作为一种半自动跟踪方法，MeanShift 算法在起始跟踪帧通过手工确定搜索窗口来进行目标的选择；然后计算出核函数加权下的搜索窗口直方图分布，并以

相同方式得到第  $N$  帧对应窗口的直方图分布；最后根据分布的相似性最大原则，使目标沿密度增加最大的方向移动，即可完成对目标的跟踪<sup>[18]</sup>。这里还将通过更新窗口中心点的方式来显示目标的运动轨迹。

### 3.1.2 Meanshift 算法的实现与结论

根据上述流程在 MATLAB 上编程实现部分截图如下：



图 3-2 在起始帧手动框选出跟踪目标



图 3-3 目标的跟踪框及其运动轨迹的显示

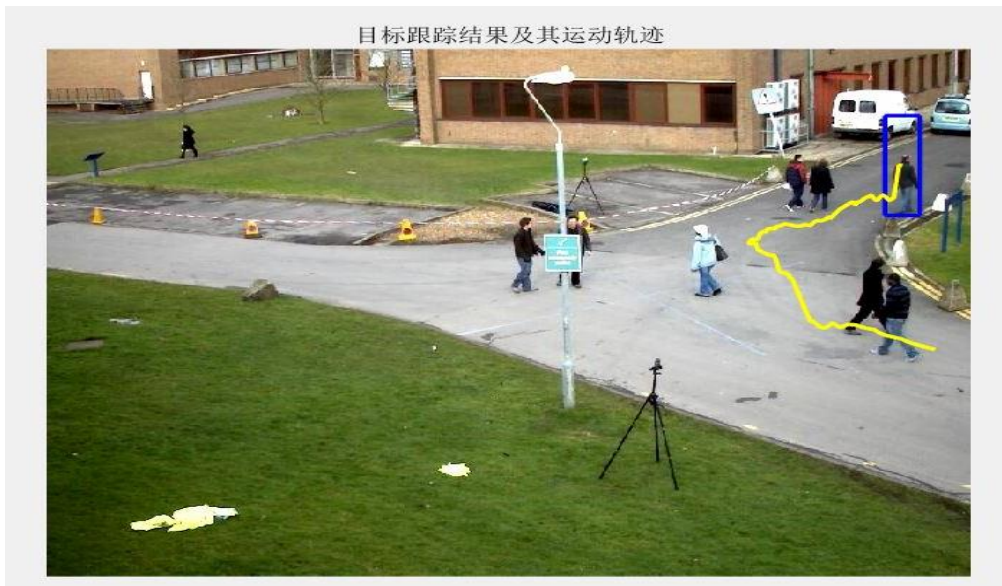


图 3-4 对不同目标的跟踪效果



图 3-5 问题：遮挡情况下的跟丢现象

实验结果表明，**meanshift** 算法对于单目标跟踪的效果良好，但是存在目标被遮挡后的跟丢问题。同时背景框的选取也很重要，如果框中背景部分过大，会产生过大的噪声。将它应用于多目标的跟踪困难较大，由此来看，**meanshift** 算法可以为我们提供有效的思路，但却不符合我们的要求。

## 3.2 基于卡尔曼滤波的目标跟踪算法

### 3.2.1 卡尔曼滤波基本原理

卡尔曼滤波是一种最优递推滤波方法，它的提出标志了滤波理论的建立。卡尔曼滤波器可以对动态系统中的状态序列进行计算得出线性最小方差误差估计，而在描述动态系统时需要用到系统动态状态方程以及系统观测方程<sup>[19]</sup>。

状态方程和观测方程通过根据系统当前测量值加上前一刻的估计值进行预测和修正不断迭代更新，目的是使观测噪声和系统噪声带来的误差逐渐减小到最理想的状态参数<sup>[20]</sup>。所以不用保存大量的测量数据，根据当前帧的观测值及前一刻的估计值，利用系统自身的状态转移方程、预测方程和校正方程通过预测当前时刻的估计值，就能够得到新的估计值，这一特性大大提升了它的实用价值。卡尔曼滤波算法还具备的优点，例如存在高斯白噪声时，它能够得到最优且无偏的最小方差估计。更为重要的是，即使是存在非高斯噪声，它的效果依然很好。具体数学模型的基本公式如下：

状态方程：

$$X(k+1) = A(k+1, k)X(k) + w(k) \quad (3-5)$$

观测方程：

$$Z(k) = H(k)X(k) + v(k) \quad (3-6)$$

其中， $X(k+1)$ 是  $k+1$  时刻  $n \times 1$  维的系统状态向量； $Z(k)$ 是  $k$  时刻  $m \times 1$  维观测向量； $A(k+1, k)$ 是  $n \times n$  维的系统状态转移矩阵； $H(k)$ 是  $m \times 1$  维系统观测矩阵； $w(k)$ 是  $k$  时刻过程的  $n \times 1$  维随机干扰噪声向量； $v(k)$ 是  $k$  时刻  $m \times 1$  维的系统观测噪声向量。 $w(k)$ ， $v(k)$ 通常假设为不相关的零均值高斯白噪声向量<sup>[21]</sup>。

卡尔曼滤波的整体结构如下，其中包含了预测和校正两个部分，一共五个基本公式，结构框图如图 3-6 所示：



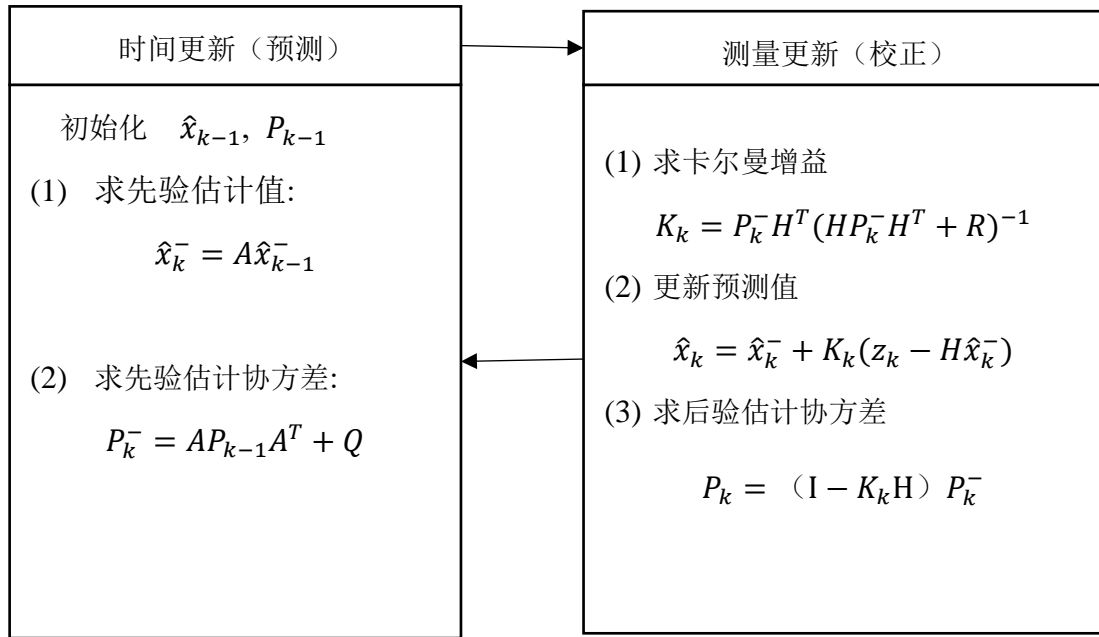


图 3-6 卡尔曼滤波器的基本结构

### 3.2.2 卡尔曼滤波应用于目标跟踪

在定位之前，需要先检测后跟踪，而跟踪效果关系到整个定位系统的结果，所以在跟踪算法的选择上需要结合实际需求，同时做出改进。由于卡尔曼滤波器同样具有预测目标轨迹的功能，同时它相较 Meanshift 来说，更为简洁，所接下来将通过设立简单运动模型的方式检验卡尔曼滤波用于目标跟踪的预测效果如何。

在上一小节中。我们已经大致了解了卡尔曼滤波的基本跟踪原理，接下来将通过两种简单的直线运动模型来检测卡尔曼滤波跟踪目标时的效果。考虑到机器人的运动方式，这里的运动模型选取的是匀速直线运动和匀加速直线运动，事先设定好速度与加速度的大小，通过原始运动轨迹与卡尔曼滤波跟踪轨迹的对比来检验卡尔曼滤波器的预测跟踪效果。如果对于简单模型的跟踪预测，卡尔曼滤波器的仿真表现也不错的话，说明可以考虑将其改进并用于定位系统的设计。

在 MATLAB 上编写程序，得到的实验结果如下两图所示：

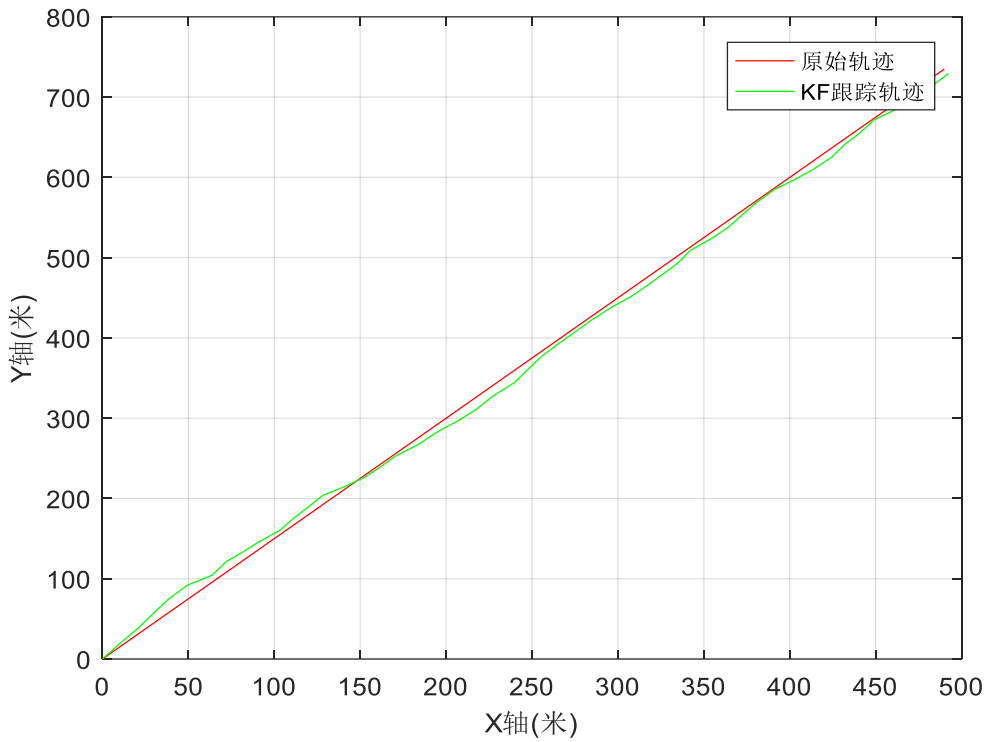


图 3-7 匀速直线运动模型的轨迹对比

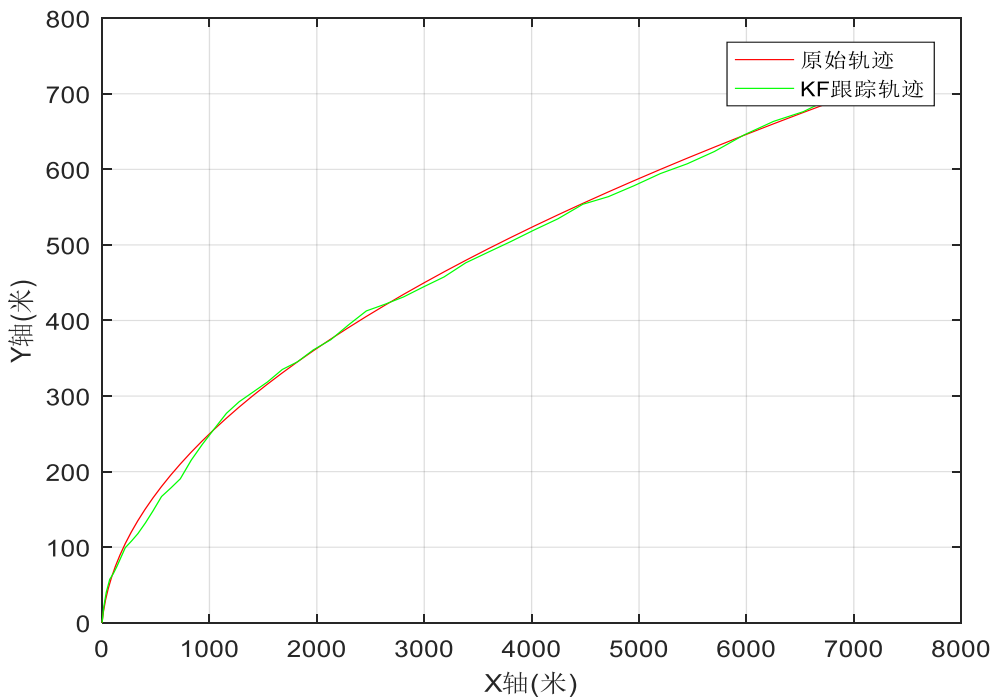


图 3-8 匀加速运动模型的轨迹对比

实验结果表明，卡尔曼滤波器应用于运动目标的跟踪结果存在一定的误差，但总体跟踪的预测性能良好，适用于对运动目标的跟踪预测。

### 3.3 定位系统跟踪算法的改进与实现

在对比了两种典型的目标跟踪算法并验证性能之后，结合题目的要求，接下来完成的工作是完成对多运动目标的跟踪算法设计。

基于全局视觉的多目标跟踪算法主要由两部分组成：

1. 将每帧中的运动目标检测出来。
2. 将检测目标与之前正在跟踪的同一目标关联起来。

在第二章中已经完成了对运动目标检测部分算法的设计，所以这里可以直接根据检测到的结果完成对目标的跟踪算法设计。在进行多目标跟踪算法的设计之前，先对多目标跟踪的两个难点进行介绍：

#### 3.3.1 多目标跟踪的两个难点

##### ● 遮拦问题：

遮拦在多目标跟踪中较为常见，因为目标被背景中物体或其他目标所遮挡的情形并不少见。具体可以将其分两个阶段：目标进入遮挡的过程中信息的丢失情况会加重；目标出离开遮挡的过程中目标信息得到恢复对应信息的增多。然而跟踪算法需要保证获取到的目标信息的足量和稳定才能保证跟踪的效果。所以遮挡问题十分容易降低跟踪的稳定性甚至于丢失目标。如何有效处理尤其是严重的遮挡问题，始终是多目标跟踪的一大难点。当前多数算法无法很好地处理这一问题，同时当目标丢失后，没有特别有效的重新获取目标的方法。

##### ● 目标数目的随机性：

目标数量的变化也是多目标跟踪的一大难点。变化主要源于新目标的出现，同时会出现的目标消失及目标的分裂与合并现象。将其拆分开来看，数量的变化分为两种情形：总数的增加，包括目标的增加和分裂；总数的减少，包括目标的消失和合并<sup>[22]</sup>。如何快速且准确地得出当前图像中包含的目标个数，同样也是多目标跟踪的一大有待解决的问题。

在完成了对运动目标的检测之后，我们通过矩形框的方式对目标进行标识并将其显示。为了解决上述的遮拦问题，这里通过使用卡尔曼滤波器进行预测得到预测中心，将其调整后同样作为轨迹的矩形框输出。而在分配轨迹给新检测到的

目标时，本文中用到了卡尔曼滤波器来对下一帧图像里的目标区域进行预测，并且在其出现的预测区域内完成目标的匹配。

在匹配过程中将运用到了匈牙利匹配算法，这样就可以将新一帧图像中检测到的多个目标匹配到各自对应的轨迹。具体实现过程中，它将预测得到与检测到的质心之间的距离之和进行了最小化求解。

### 3.3.2 目标与轨迹的匹配过程

#### ● 匈牙利匹配算法基本原理：

匈牙利匹配算法通俗来讲是一种用增广路的算法思想求二分图中的最大匹配的算法<sup>[23]</sup>。如下图所示，我们可以到图上的点分为两种，左边是数字，右边是字母，并且数字与数字、字母与字母之间是没有互相连接的边的。这种点被分开成两边且每边的点之间不存在相连现象的图就叫做二分图，示例图如下：

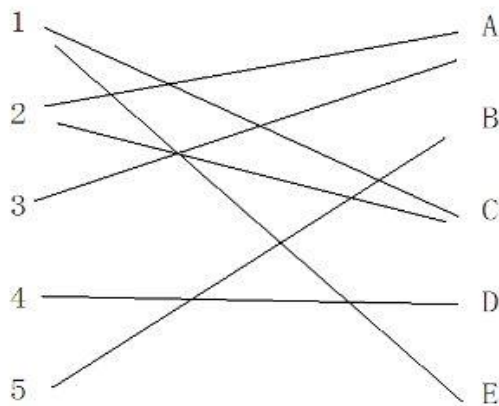


图 3-9 一种简单的二分图示例

匈牙利算法就是用来处理二分图匹配问题的，它求的是二分图的最大匹配，就是指在二分图中连出尽可能多的边，同时使它们遵从二分图匹配的规则<sup>[24]</sup>。增广路算法流程如下图所示：

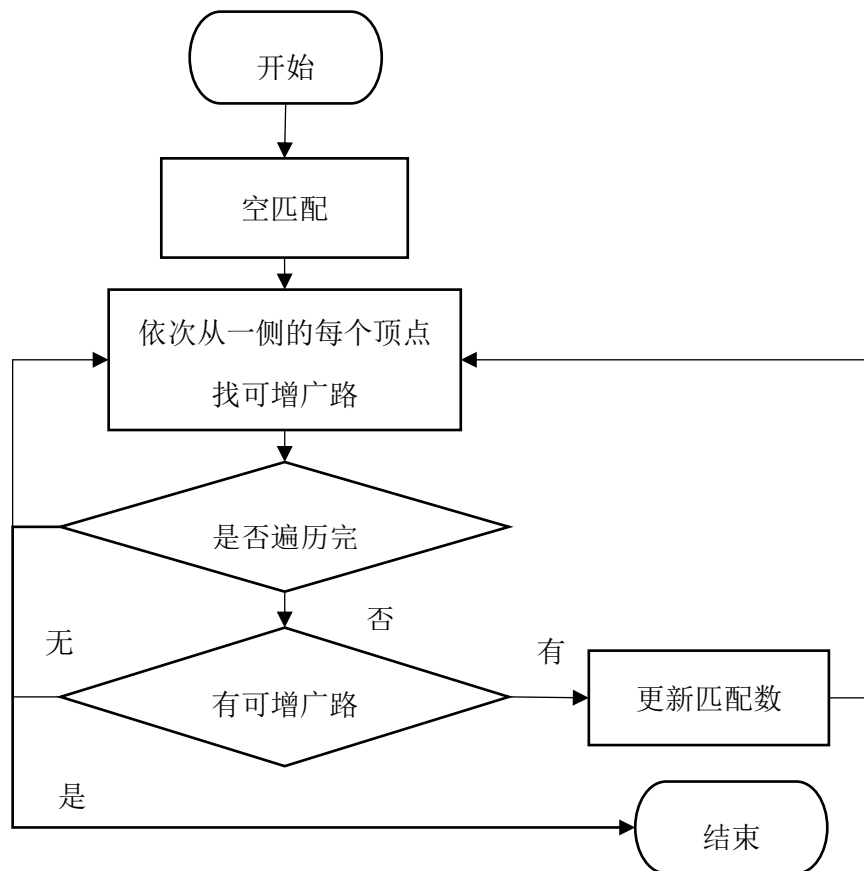


图 3-10 增广路算法流程图

匹配的过程实质上就是一直在找一条路径，能够把二分图的左半部分的其中一个未匹配节点与右半部分的其中一个未匹配节点相连，并加入到已经匹配的节点中。将其应用于跟踪部分具体可以分为两步：第一步，计算每种匹配的损失矩阵 $[M N]$ ，其中 $M$ 表示的是轨迹数目， $N$ 对应检测到的目标数目。第二步，求解损失矩阵，设定阈值，进行分配轨迹与检测目标。

### ● 更新跟踪轨迹和检测目标

更新跟踪轨迹和检测目标的过程如下，将能匹配到检测目标的轨迹更新至当前这一帧中的已检测目标的位置。将达到一定阈值的匹配不到检测目标的轨迹删除，即该目标已经连续多帧未被检测到。同时将新的未分配的检测目标添加到一个新建的轨迹中。

### 3.3.3 跟踪算法主程序及主要函数

根据前文的思路完成多目标跟踪算法的编程，主程序部分截图如下图所示：

```
function bishe()
% 创建用于阅读视频的系统对象，检测移动对象，并显示结果
obj = setupSystemObjects(); %初始化函数
tracks = initializeTracks(); %初始化轨迹对象
nextId = 1; % 下一个轨迹的编号

while ~isDone(obj.reader) % 检测移动物体，并在视频帧中跟踪它们
    frame = readFrame(); %读取一帧
    [centroids, bboxes, mask] = detectObjects(frame); %前景检测
    predictNewLocationsOfTracks(); %根据位置进行卡尔曼预测
    [assignments, unassignedTracks, unassignedDetections] = ...
        detectionToTrackAssignment(); %匈牙利匹配算法进行匹配

    updateAssignedTracks(); %分配好的轨迹更新
    updateUnassignedTracks(); %未分配的轨迹更新
    deleteLostTracks(); %删除丢掉的轨迹
    createNewTracks(); %创建新轨迹

    displayTrackingResults(); %结果展示
end
```

图 3-11 多目标跟踪算法主程序截图

主要用到的函数及其功能如下表所示：

表一 程序中调用到主要函数的功能

函数名称	功能简介
obj = setupSystemObjects()	初始化系统对象函数
[centroids, bboxes, mask] = detectObjects(frame)	前景检测函数
predictNewLocationsOfTracks()	卡尔曼预测函数
detectionToTrackAssignment()	检测轨迹分配函数
update(un)AssignedTracks()	轨迹更新函数
displayTrackingResults()	显示跟踪结果函数

### 3.4 跟踪结果展示

根据上文的设计思路，在 MATLAB 上编写程序进行调试相关参数后，对两段包含多运动目标的视频进行检测跟踪的部分结果截图如下两图：



图 3-12 包含多个行人视频的跟踪结果

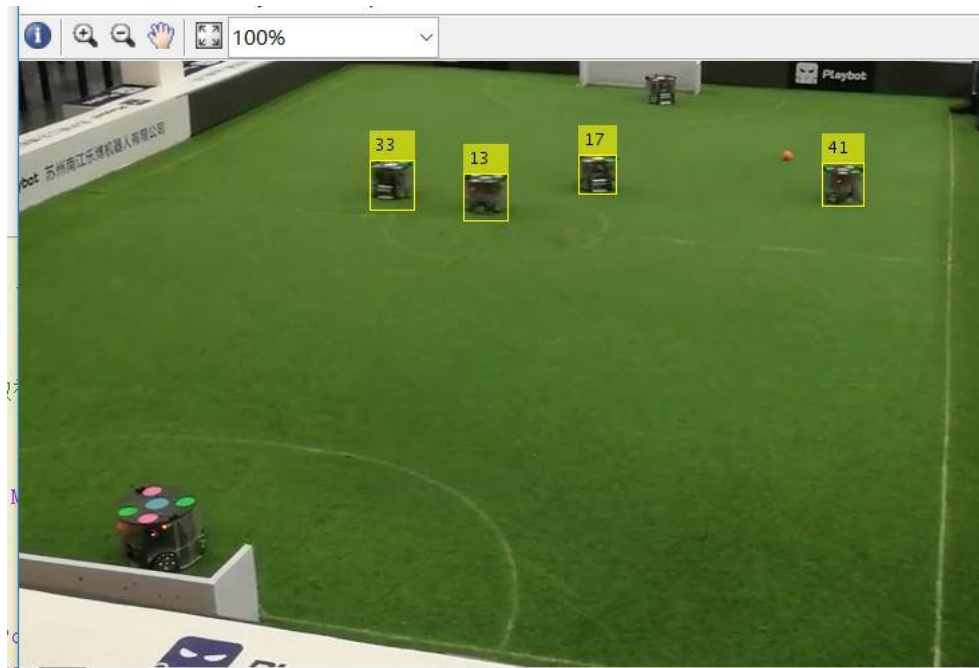


图 3-13 足球机器人比赛视频中的跟踪结果

### 3.5 本章小节

在第二章的工作基础之上,本章对两种常用跟踪算法进行了实验测试,在对比分析了两者各自的优缺点之后,选择用卡尔曼滤波器来完成之后对多目标的跟踪处理。对定位系统进行了整体的算法设计,检测部分用到的第二章中的改进后的基于混合高斯背景建模的背景差分法,跟踪部分主要是基于卡尔曼滤波的轨迹匹配方法,而多目标跟踪中的轨迹匹配是根据匈牙利匹配算法的原理来完成的。加入了更新轨迹和删除失效轨迹的功能后,最终完成对多运动目标的跟踪,显示出来的效果基本达到要求。



## 第 4 章 基于视觉的多机器人定位

在前面两章已经中完成了对多运动目标跟踪算法的设计工作，而实际应用到多机器人的定位中，还需要考虑到不同场景下具体要求。例如，机器人所在环境的坐标系的建立以及坐标转换等问题。本章就其中的坐标转换进行了研究并在程序中加以实现。

### 4.1 坐标变换原理

为了完成对机器人进行实际空间位置的确定，需要建立起图像坐标系与世界坐标系的对应关系<sup>[25]</sup>。通过对图像模型进行逆透视变换，实现图像坐标系向世界坐标系的转换，从而完成对机器人的定位。其坐标变换模型如下图所示：

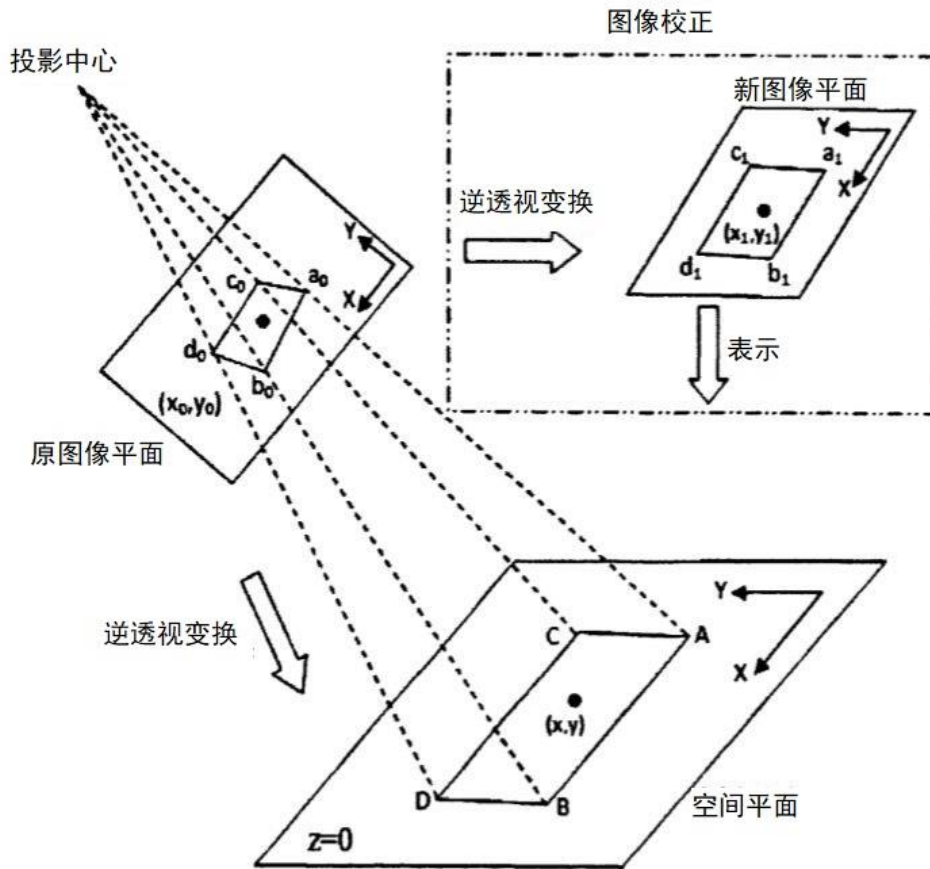


图 4-1 坐标变换模型

透视变换的本质是将图像投影到一个新的视平面，其通用变换公式为：

$$U_0 = M_{3 \times 3} X_{3 \times 1} \quad (4-1)$$

其中  $U_0 = [x_0, y_0, w_0]^T$ ； $X_{3 \times 1} = [u, v, w]^T$ 。(u,v) 表示原始图像像素坐标，(x,y) 为变换之后的图像像素坐标。 $x = x_0/w_0$ ， $y = y_0/w_0$

转换矩阵  $M_{3 \times 3}$  具体构成如下：

$$M_{3 \times 3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} T_1 & T_2 \\ T_3 & a_{33} \end{bmatrix} \quad (4-2)$$

其中  $T_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ ，表示图像的线性变换矩阵； $T_2 = [a_{13}, a_{23}]^T$ ，用于产生图像的透视变换矩阵； $T_3 = [a_{31} \ a_{32}]$ ，表示了图像平移的过程矩阵。

对于一个点横纵坐标的透视变换，有数学表达式：

$$x = \frac{x_0}{w_0} = \frac{a_{11}u + a_{21}v + a_{31}}{a_{13}u + a_{23}v + a_{33}} \quad (4-3)$$

$$y = \frac{y_0}{w_0} = \frac{a_{12}u + a_{22}v + a_{32}}{a_{13}u + a_{23}v + a_{33}} \quad (4-4)$$

在 MATLAB 上根据坐标转换原理编写程序，对一幅图像进行透视转换处理，结果如下图所示：

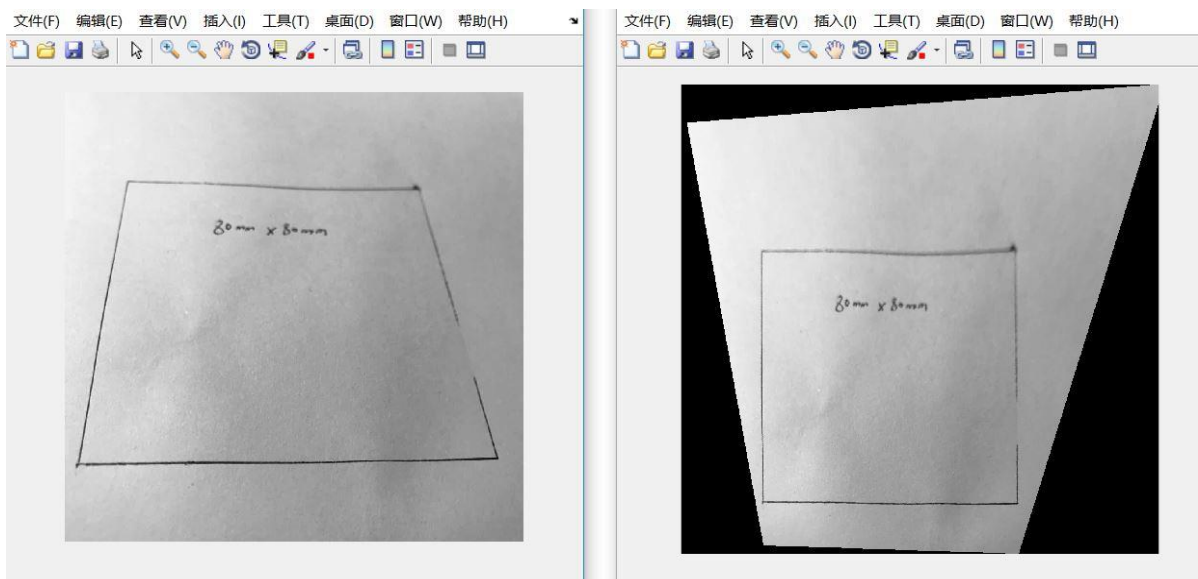


图 4-2 透视转换效果图

## 4.2 多运动目标的坐标输出

在跟踪部分，由于我们已经得到了多个跟踪目标的跟踪框的矩阵  $bbox$ ，所以在进行对目标跟踪的同时，我们只需要先由质心计算得到目标的中心点的位置，然后利用预先根据背景图像求出的坐标转换矩阵，就能够完成对目标质心的二维坐标的转换和输出。流程框图如下：

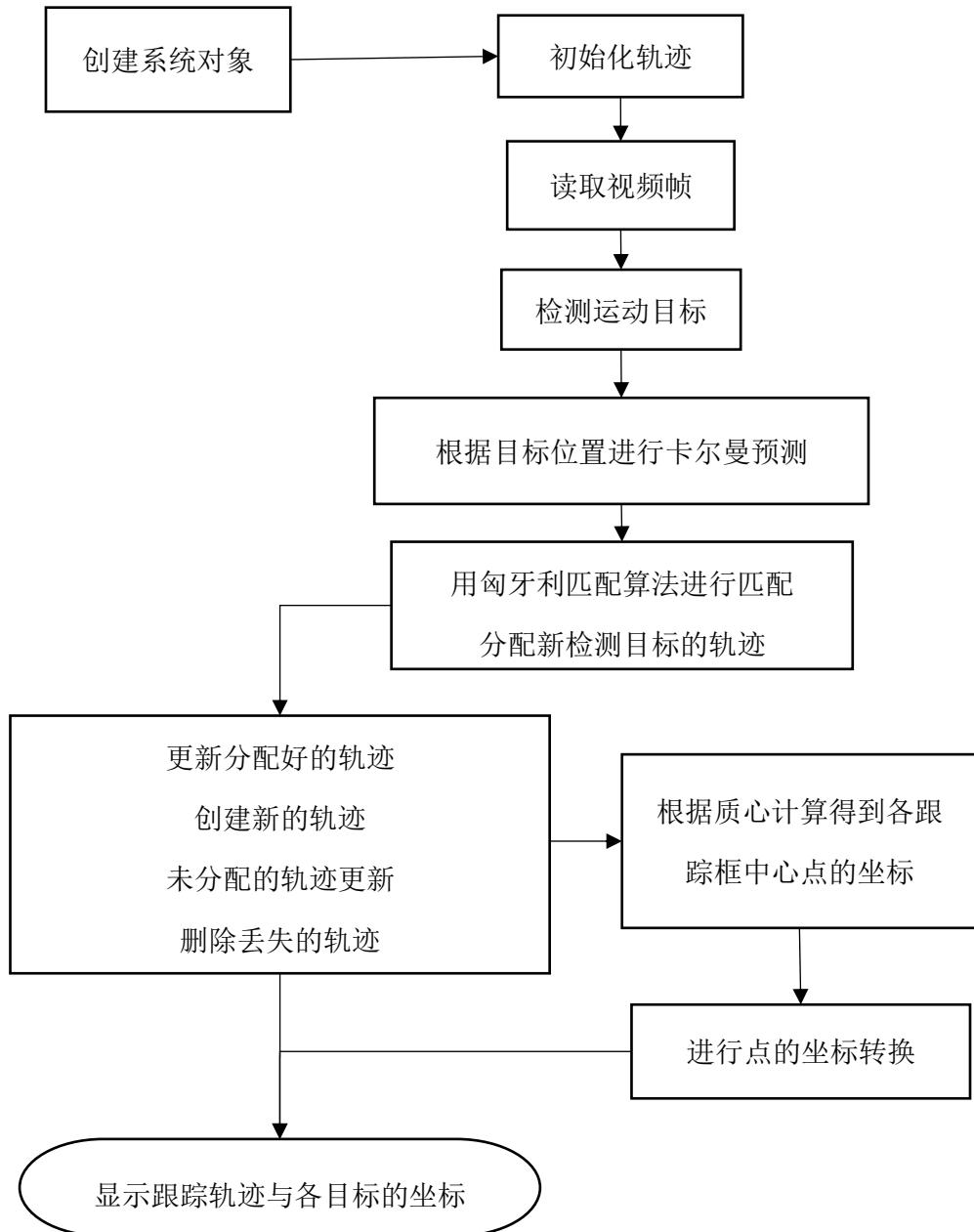


图 4-3 定位系统的整体流程框图

在 MATLAB 上进行程序的功能添加和修改后得到的结果部分截图如下图：

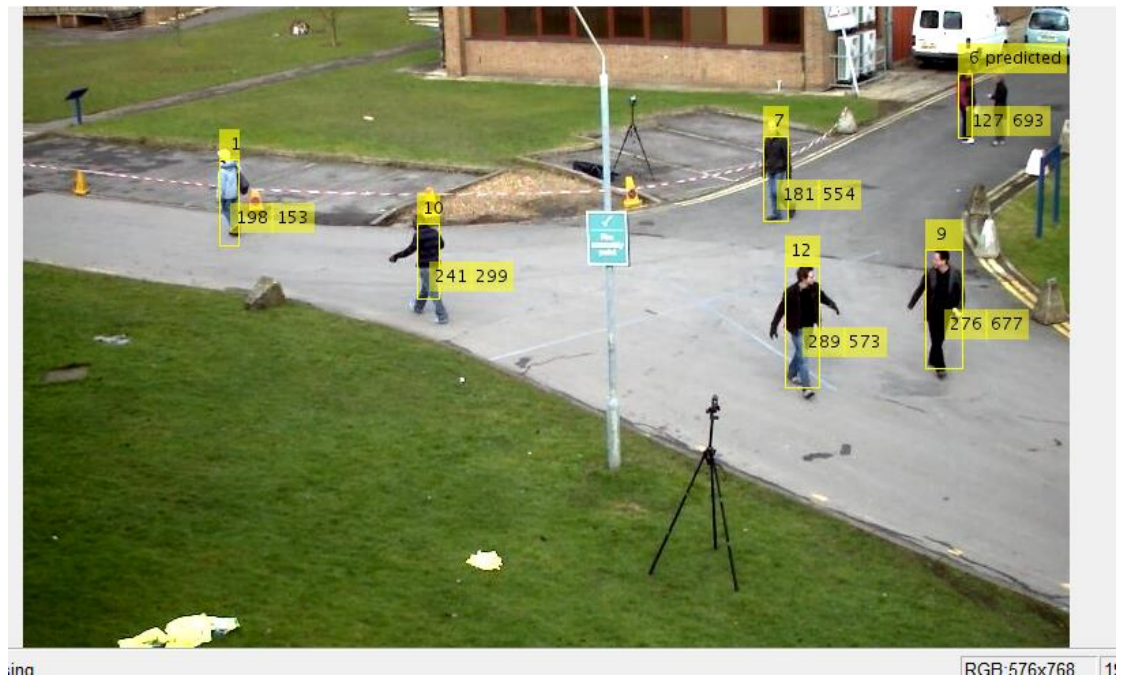


图 4-4 对视野内的多目标进行坐标的实时输出结果截图

由对一段包含多个行人(运动目标)的视频跟踪并显示坐标的结果可以看到,当出现在检测区域(视野)的目标保持运动或者短时间的静止时,都能够将其检测且用跟踪框突出。同时根据相关计算,可以将多个目标的中心点坐标输出,之后再经过坐标转换就能够得到目标在实际坐标系(俯视视角)中的二维坐标。

### 4.3 应用于多机器人的定位

在完成了多目标定位系统的整体设计之后,将其应用于多机器人定位时,假设的情况下,是对根据固定于机器人活动区域上方的摄像头采集到的视觉信息进行相关处理后,达到多个机器人定位的目标,即输出每个机器人的二维坐标。将该定位系统应用于实时定位时,只需要将 USB 摄像头获取的视频进行实时传输并处理即可完成。虽然可能会存在视频传输过程中的延时和丢帧现象,但整体算法在对多目标的检测到坐标输出过程中不存在明显的延时,符合同步输出二维坐标的课题要求。

在利用本文中提出的定位系统来对一段足球机器人的比赛视频进行处理时,我们可以得到如下图所示的结果截图:



图 4-5 对多个足球机器人的定位效果截图

由结果可见，本文设计的定位系统应用于多机器人定位时的效果明显是不错的，保持着一定的稳定性、准确性和及时性。

基于视觉的机器人定位，实质上就是先检测，再跟踪，最后输出坐标这一位置信息。系统的难点就是在于检测和跟踪的精度，这些都是需要经常对比试验才能得到对应特殊情况下的最优解的。本文的贡献在于经过对比分析多种检测和跟踪的算法后，将其有机结合，并加入了坐标输出这一功能，最终实现了多目标（机器人）的定位目标。然而整体框架实际应用到基于全局视觉的多机器人的定位中，仍然存在一些具体细节上的问题，例如：

1. 在对多机器人的实际定位中，如果摄像头的设置不是如同本文中的假定，固定于活动区域上方，那么可能会出现更多的遮拦问题。
2. 机器人的形状对于算法提出的具体要求。本文中由于跟踪的对象是行人，具有高度明显大于宽度，上肢躯干无缝隙等特征，所以在算法中的一些参数上的设定具有一定的针对性。而实际应用于不同形状的机器人的跟踪时，这些参数都是需要进行相应的修改的。
3. 本文中目标跟踪之后输出的二维坐标是基于质心计算得到的质心的图像中的像素坐标透视转换得到的。应用于实际定位中，不同情形下要求输出的坐标可能不是质心坐标，相对具体坐标系的转换也有待进一步解决。

## 第5章 总结

### 5.1 工作总结

无论是单独的一个机器人，还是涉及到分工合作的多机器人系统，考虑到机器人的移动问题时，如何获取机器人的具体位置信息是至关重要的。基于全局视觉的多机器人定位方法，是指利用顶端摄像机获取到的视觉信息来完成对视野内的机器人的跟踪定位，具有一定的实用意义，且可以拓展应用到车辆检测以及目标识别等相关领域。本文的研究主题是基于全局视觉的多机器人定位系统，实际内容主要是基于视觉信息处理的多目标检测跟踪与坐标输出算法设计。贡献在于对比分析多种检测和跟踪的算法并将其有机结合，在加入了坐标输出这一功能后，实现了多目标（机器人）定位的目标。现将所做工作总结如下：

（1）在查阅了许多相关文献后，完成课题研究背景与意义的综述，总结机器人定位方法的研究现状，介绍了部分定位方法并简要分析各自优缺点。

（2）在介绍了需要用到的视觉处理基础原理后，对两种运动目标检测方法进行了简介和性能的验证，对混合高斯背景建模进行实现与改进，检测效果有了一定的提升，为整体定位系统提供了良好的基础。

（3）对 `meanshift` 算法和卡尔曼滤波器进行简单的理论推导，在 `MATLAB` 上编写程序并对比跟踪效果，选择了卡尔曼滤波的方法，并根据匈牙利匹配算法的思想进行改进并完成算法的主体部分的设计。将第二章与第三章的跟踪算法整合在一起，达到多目标的检测和跟踪并显示结果。

（4）介绍坐标转换原理并进行验证，将其添加到跟踪算法中，根据质心计算的方法完成对多目标的同步坐标输出。然后就应用于多机器人定位展开了一定的思考和分析。

本次课题的研究工作由于条件受限，主要对象是包括多行人的一段视频，在此基础上完成了整体定位算法的设计，但其实用性证明了应用于多机器人定位上的可行性，从某种程度上基本达到了课题的要求。

## 5.2 未来展望

视觉信息的处理与应用技术随着时代发展的要求显得愈加炙手可热，将视觉技术与机器人结合在一起，更是预示了人工智能的无限前景。基于视觉的机器人定位方法众多，更多巧妙高效的设计思想令人不仅感叹知识就是力量。本文的工作虽然具有一定的实用性，但同时存在更多的是局限性，实际应用起来仍然需要进行改进。

需要改进的地方例如，遮拦物以及系统噪声对跟踪结果的影响仍然有待进一步减小；如果摄像头不是固定的，该怎么进行视觉信息的处理；实用于实时获取坐标时，计算速率对于算法结构的优化提出的更高要求等。

随着研究的深入，自身能力的有限加上算法设计过程中依然存在着的不少问题，使我更为深切地体会到自身研究深度的不足与知识上的短缺，同时也不断感受到这门学科的魅力所在。如果能够将课题进一步研究下去，可以为其设计一个用户可视化界面，这样不仅可以应用于机器人的定位，甚至可以用于智能家居系统或者交通检测系统等实用性领域。路漫漫其修远兮，虽然这些都是后话了，我还是希望能够看到相关技术更多被实用于生活中。

## 参考文献

- [1] 段朋. 与人共融的家庭服务机器人定位与导航[D]. 济南:山东大学, 2015:2-6.
- [2] 于仕琪, 刘瑞祯. 学习 OpenCV[M]. 清华大学出版社, 2011: 456-464.
- [3] 徐田来, 游文虎, 崔平远. 基于模糊自适应卡尔曼滤波的 INS/GPS 组合导航系统算法研究[J]. 宇航学报, 2005, 26(5): 571-575.
- [4] 柯丰恺.面向机器人的多目视觉检测算法研究及系统实现[D]. 武汉:华中科技大学, 2016: 12-26.
- [5] LI J, DING Q X, WANG P. Monocular vision measurement method of object position and orientation based on quaternions[J]. Aeronautical Computing Technique, 2011: 2-5.
- [6] ZHANG Z F, HANSON A R. 3D reconstruction based on homography mapping[C]. Proceedings of ARPA Image Understanding Workshop, 1996: 1007-1012.
- [7] 迟健男, 徐心和. 移动机器人即时定位与地图创建问题研究[J]. 机器人, 2004,26 (1) : 92-96.
- [8] 苑全德. 基于视觉的多机器人协作 SLAM 研究[D]. 哈尔滨: 哈尔滨工业大学, 2016: 22-36.
- [9] 鲁书山. 移动目标检测与跟踪技术的研究及 Matlab 实现[D]. 太原:中北大学, 2017: 13-23.
- [10] FISCHLER M A, BOLLES R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. Comm of the Acm, 1981, 24(6): 726-740.
- [11] NISTER D, NARODITSKY O, BERGEN J. Visual odometry for ground vehicle applications[J]. Journal of Field Robotics, 2006, 23(1):3-20.
- [12] 肖江,陈想,丁亮. 一种改进的 Mean Shift 运动目标跟踪算法[J]. 信息技术. 2017(1): 2-4.
- [13] 孙同同, 孙首群. 基于视觉检测与跟踪技术在机器人中的应用[J]. 组合机床



- 与自动化加工技术. 2016(2): 2-3.
- [14] MENG C, ZHOU N, XUE X L, et al. Homography-based depth recovery with descent images[J]. Machine Vision and Applications, 2013, 24(5): 1093-1106.
- [15] SIMMONS R, KOENIG S. Probabilistic robot navigation in partially observable environments[J]. Proceedings of the International Joint Conference on Artificial Intelligence, 1995:1080-1087.
- [16] THRUN S, FOX D, BURGARD W, et al. Robust Monte Carlo localization for mobile robots[J]. Artificial Intelligence, 2001, 128(12):99-141.
- [17] 张娟,毛晓波,陈铁军. 运动目标跟踪算法研究综述[J]. 计算机应用研究. 2009(12): 2-5.
- [18] 贾艳丽. 基于视频图像序列的运动目标检测与跟踪[D]. 哈尔滨:哈尔滨工程大学, 2012: 23-26.
- [19] NISTER D. An efficient solution to the five-point relative pose problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6):756-770.
- [20] 杨阳. 基于视频的运动目标检测和跟踪技术[D]. 杭州:浙江大学, 2014: 12-16.
- [21] 王瑞. 智能视频监控系统的研究与开发[D]. 济南:山东大学, 2012: 5-10.
- [22] 郭文亮. 基于全局视觉的移动机器人导航系统研究[D]. 北京:北京交通大学, 2011: 43-45.
- [23] CHEN C, SCHONFELD D. Pose estimation from multiple cameras based on Sylvester's equation[J]. Computer Vision and Image Understanding, 2010, 114:652-666.
- [24] KNORR M, NIEHSEN W, STILLER C. Online extrinsic multi-camera calibration using ground plane induced homographies[C]. Intelligent Vehicles Symposium, 2013: 236-241.
- [25] Jirawimut R, Prakoonwit S, Cecelja F, et al. Visual odometer for pedestrian navigation[J].IEEE Transactions on Instrumentation and Measurement, 2004, 52(4): 1166-1173

## 致 谢

大学四年真就印证了一句从小就学会的谚语，光阴似箭，日月如梭。回首这段时光，所有得到与失去，留下来的都是自己的经历。在这告别之际，我想对陪伴我走过这美好四年的师长和亲友表达诚挚的谢意。

首先，我要感谢的是见证了我完整四年的母校，浙江工业大学，她为我所提供的环境和关怀都是无比宝贵的，感谢您的孕育之恩。

然后，我想感谢我的指导老师，赵云波教授。在选题之前，赵老师在平时课上表现出来的思维缜密、严谨治学和和蔼可亲就已经给我留下了十分深刻的印象。而在完成毕业设计的整个过程中，赵老师总能够在我困惑的时候提供有效的帮助，每次与老师交流之后，都会有许多收获。师恩难忘，赵老师对我的帮助是不可衡量的，感谢您的指导之恩。

最后，我要谢谢大学四年里遇到的所有老师，同学和朋友，还有实验室的吕文君等学长们，与你们的相处都是我应该感激的经历。还有永远都应该感恩的父母，你们是我整个人生中最重要的组成，也是我必须努力的理由。

转眼就到了说再见的时候，大学四年的收获将是我接下来人生道路上的永远推动力。雄关漫道真如铁，而今迈步从头越。希望自己怀着感恩的心，不忘过去，不惧未来，今后付出更多的努力去收获精彩。