

分类号：TP391.4

学校代码：10406

学号：

南昌航空大学  
硕 士 学 位 论 文  
(专业学位研究生)

双摄像机实时视频摘要系统

硕士研究生：

导 师： 副教授

申请学位级别： 硕士

学科、专业： 计算机技术

所在单位： 信息工程学院

答辩日期： 2018 年 6 月

授予学位单位： 南昌航空大学

# **Research on Real-time Video Summarization System With Dual Camera Products Assembly**

A Dissertation

Submitted for the Degree of Master

On the Computer Technology

**By**

Under the Supervision of

**A.Prof.**

School of Information Engineering

Nanchang Aeronautical University, Nanchang, China

June, 2018

## 摘要

场景监控已经成为信息时代必不可少的视频采集方式，并且发展出了多种形式。仅仅依靠人工方式对场景监控进行审核和查找已经很难满足目前爆发式增长的视频数据审核需求，因此，能够自动处理视频内容的视频摘要技术引起了人们极大的关注。目前传统的视频摘要技术在面对大场景的监控时，总是会存在生成的摘要中目标信息模糊不清、有效性差和高延迟等问题。本文针对以上问题并结合双摄像机的特点提出了一种双摄像机实时视频摘要系统。

本文提出的双摄像机实时视频摘要系统采用两个规格不同的摄像机进行高低搭配，两个不同规格的摄像机协同工作的基础是两个相机拥有统一的参考坐标系。本文以张正友教授的双摄像机标定为基础研究了标定方法，该方法使两个摄像机之间拥有统一的参考坐标系，并根据两个相机标定后的统一参考坐标进行目标在两个相机中同时定位，以达到一个相机中的目标可以在另一个相机中同步跟踪的目的。

结合标定的结果，使用清晰度一般的相机作为场景监控相机固定在监控区域上方，并对场景中的目标采用基于空间约束的混合高斯背景建模的方法进行目标的实时检测，再用改进的 Camshift 方法对检测到的目标进行实时跟踪。

针对大场景中的多个目标，制定合适的控制策略以实现场景中的每个目标进行标识、计算轨迹、估算速度、定位优先区域、优先级排序等操作，以实现场景中的每个目标都能有序、有效的进行信息采集。

通过设计云台摄像机的控制模块，使其作为目标抓拍相机，能够根据控制策略实现两个摄像机实时地协同工作，场景监控相机进行目标检测和跟踪采集目标在场景中的信息，目标抓拍相机进行目标高清信息的抓拍。最终，根据固定相机采集的视频信息和云台相机采集的目标高清信息进行整合，生成包含目标视频、轨迹、高清图像等多种形式的视频摘要。

**关键词：**视频摘要，双摄像机，标定，目标检测，目标跟踪

## **Abstract**

Scene monitoring has become an essential way of video collection in the information age ,and it has developed many forms. It is difficult to check the video by manual method only ,and also difficult to meet the demand of video data audit which is growing in the current explosive growth. Therefore, the video summarization technology that can automatically handle video content has aroused great concern. In the face of large scene monitoring, the traditional video summarization technology always has the problems of blurred target information, poor effectiveness and high delay. Based on the above problems and combined with the characteristics of dual cameras, a dual camera real-time video summarization system is proposed.

The dual camera real-time video summary system proposed in this paper it uses two different cameras to match the high and low levels. The base of the two different cameras working together is that two cameras have a unified reference coordinate system. This paper studies the calibration method based on the dual camera calibration of Professor Zhang Zhengyou. This method makes the two cameras have a unified reference coordinate system. According to the reference coordinates of the two cameras, the target is positioned at the same time in two cameras, so that the target in one phase machine can be tracked synchronously in another camera.

With the result of calibration, the camera is fixed above the monitoring area as the scene monitoring camera with the general camera, and the target is detected in real time by the hybrid Gauss background modeling algorithm based on space constraints in the scene, and the improved Camshift algorithm is used to track the detected target in real time.

In view of the multiple targets in the large scene, the appropriate control strategy is made to realize the operation of identifying each target in the scene, calculating track, estimating speed, locating priority area, prioritizing and so on. Each target in the scene can be collected in order and effectively.

The camera can be used as the target capture camera through the design of the control module of the cloud platform camera, and the two cameras can work together in

real time according to the control strategy. The scene monitoring camera can detect the target and track the information of the target in the scene, and the target capture the camera to get the target of the target high definition information. Finally, the video information collected by the fixed camera and the high definition information collected by the PTZ camera are integrated to generate a variety of video summaries including the target video, the trajectory, the high definition image and so on.

**Keywords:** video summarization,dual camera,calibration,target detection,object tracking

# 目 录

摘要.....	I
Abstract.....	II
目 录.....	IV
1 绪论.....	1
1.1 课题研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 视频摘要技术.....	2
1.2.2 双摄像机协同技术.....	3
1.3 本文的研究内容和组织结构.....	4
2 总体设计.....	6
2.1 总体结构设计.....	6
2.2 主要研究内容.....	8
2.3 主要内容设计依据.....	9
2.3.1 双摄像机标定.....	9
2.3.2 运动目标检测.....	10
2.3.3 运动目标跟踪.....	11
2.3.4 双摄像机协同工作和视频摘要生成.....	13
2.4 本章小结.....	14
3 双摄像机标定.....	16
3.1 双摄像机标定介绍.....	16
3.2 摄像机标定原理.....	16
3.3 双摄像机标定方法介绍.....	20
3.4 双摄像机标定的实验及分析.....	25
3.5 本章小结.....	29
4 运动目标检测和跟踪.....	30

4.1 运动目标检测技术.....	30
4.1.1 帧差法.....	30
4.1.2 背景减除法.....	32
4.2 运动目标检测应用及分析.....	36
4.3 基于 Camshift 的运动目标跟踪.....	39
4.3.1 Meanshift 算法原理分析.....	40
4.3.2 Camshift 算法搜索过程描述.....	41
4.3.3 改进的 Camshift 算法描述.....	42
4.4 运动目标跟踪应用及分析.....	43
4.5 本章小结.....	45
5 双摄像机协同工作及摘要系统实现.....	47
5.1 双摄像机协同工作.....	47
5.1.1 双摄像机分工.....	47
5.1.2 双摄像机协同工作策略.....	48
5.2 系统硬件及开发环境.....	50
5.2.1 系统硬件.....	50
5.2.2 开发环境.....	51
5.3 系统运行效果展示.....	52
5.3.1 系统界面.....	52
5.3.2 视频摘要生成.....	53
5.4 本章小结.....	55
6 总结与展望.....	56
6.1 总结.....	56
6.2 展望.....	56
参考文献.....	58
攻读硕士期间发表论文及参加科研情况.....	62
致 谢.....	63

## 第一章 绪论

### 1.1 课题研究背景及意义

随着时代的发展和科技的进步，人们记录数据和处理数据的方法正发生着巨大的改变。计算机出现后，随之出现的数据结构技术使得大家开始有了处理大量数据信息的能力，并且促使多媒体技术也飞速发展，图像、声音、视频等多媒体数据也逐渐成为主流的信息媒体传播形式。由于视频能给人视觉和听觉上的直观感受，并且能够记录和再现时间和空间上的各种信息，因此这种信息媒介在社会和军事上得到越来越广泛的应用<sup>[1, 2]</sup>。

信息处理技术的发展也带动了视频和各种形式多媒体的不断升级，视频等多种数据极速增长也在不断增加着人们处理视频数据的难度。因此，对视频以及各种多媒体数据进行有效组织和管理的各种技术也应运而生。例如，电子图书馆的视频检查搜索程序、电视台的广告、电影的精华摘要。然而大部分对视频的操作还停留在存储、编辑、播放这个层面上，计算机只能针对图像中的像素信息<sup>[3]</sup>进行处理，而其中包含的人类理解的内容难以用计算机识别和理解是计算机只能对图像像素信息处理的主要原因。

伴随着对视频语义信息处理的迫切需求，视频摘要技术也经过不断的发展慢慢的形成一门新的技术门类。视频摘要，就是以一种半自动或者全自动的方法，通过对视频的结构和内容的分析，从被分析的视频中摘要出有价值的信息，使这些有用信息以一种有效的方法拼接成精简的，能代表整段视频语义内容的视频概要<sup>[4]</sup>。

视频摘要技术在各研究机构的大力投入下已有了一定的发展，但是仍有许多难题需要克服。这些难题在机器对摘要视频语义的理解和摘要信息的准确性上表现的尤为突出<sup>[5]</sup>，即生成的摘要信息文不符题或者冗余复杂利用率低。其中视频语义理解的基础也是视频信息的准确，因此能否得到准确的详细的视频信息是解决其各种难题的第一关，也是关键的一关。在大场景及开阔地带的监控中，单独的摄像机所带来的视频精度已经难以满足视频摘要系统对视频处理的需求，多摄像机技术的独有的优势也引起了各研究机构<sup>[6]</sup>的注意。

双摄像机和多摄像机技术在计算机视觉应用上因为有其使用灵活和用法多样



的优势,使得其在大场景监控、精确成像和协同应用等领域上有了飞速的发展。各种基于双摄像机和多摄像机的研究发现其在精确性上的提升,可以使得以前单摄像机效率和精度不高的问题得到极大的改观<sup>[7, 8]</sup>。因此,视频摘要领域中的精确性以及效率性问题能否结合双摄像机的优势得以解决是非常值得深入研究的课题。

双摄像机在拥有多样和灵活的使用优势的同时也伴随着摄像机增加带来的需要处理的数据增加的问题,这会对摄像机的控制和采集的成倍数据的数据处理带来新的难题。在实时性要求较高的场合对数据处理的速度和效率提出了更高的要求。

带着这些问题和思考,本文将就目前视频摘要领域最新的研究进行改进,结合双摄像机在大场景监控和灵活的监控方式,针对视频摘要系统固有的摘要信息准确性和时效性上进行研究和改进。

## 1.2 国内外研究现状

### 1.2.1 视频摘要技术

随着数字存储技术取得了巨大的飞跃,数字视频作为一种重要的媒体形式,越来越多地出现在人们的工作和生活中。数字视频技术发展的同时也涌现出了各种各样的视频库和多媒体资源管理系统,这些系统的核心就是如何利用和获取视频中所蕴含的信息,在这些需求和技术的基础上发展出了视频摘要技术。视频摘要(Video Summarization)<sup>[4]</sup>这一概念是在上世纪九十年代由美国的卡耐基梅隆大学的研究人员首先提出,它能帮助使用者高效的从视频中提取出有效的信息,供使用者不必花费大量时间和人力成本就能获取到视频中使用者感兴趣的部分。因此,视频摘要技术在早些年的时候就引起国内外研究者的广泛关注,经过国内外多媒体技术相关研究机构的大量投入,视频摘要技术取得了长足的发展。

目前国内外都有相对比较成熟的技术和研究,国外的有卡耐基梅隆大学<sup>[9]</sup>、哥伦比亚大学<sup>[10]</sup>、早稻田大学、布雷西亚大学等知名计算机和图像领域的高等学府在视频摘要技术领域展开研究。许多院校和机构的加入也完善了视频摘要技术的研究和应用,该领域也发展出多种不同风格的摘要技术和针对特殊场景的特殊摘要生成算法<sup>[12, 13]</sup>。国内的许多顶尖大学如清华大学、复旦大学、中国科学院计算所和国防科技大学等国内顶尖学府都对视频摘要技术做了大量的研究,在视频镜头的分割、视频检索、镜头探测、场景聚类、关键帧提取方面取得了相当的进展。

卡耐基梅隆大学是第一个开始对视频摘要技术开展研究的机构,其提出的

Infomedia 工程<sup>[9]</sup>主要是提取重要的视频和音频片段来创建视频摘要。通过对关键帧的提取,结合文字语义信息合成视频摘要,该系统有很好的可阅读性以及代表性,但是在摘要系统应用场景范围比较小,而且摘要信息清晰度有限。

国内方面,清华大学研究了一种对于人物之间的联系分析的自动提取摘要信息的算法,这种摘要算法主要针对有一定故事情节的视频设计。该算法针对个体进行分类,生成某个个体的摘要集合,但是摘要的时间分布比较散乱,不利于按时间检索和查阅,而且无法对细节信息进行检索。

### 1.2.2 双摄像机协同技术

在双摄像机协同跟踪领域,也有大量的机构和研究院对其进行了研究,研究的方向具体可分为基于活动和空间的协同跟踪、基于区块的双摄像机协同、以及基于特征匹配的多摄像机跟踪。问题和难点也主要集中在多摄像机非重叠区域中再识别以及接替跟踪的问题,重叠区域中多摄像机对目标的同步定位问题。

在多摄像机对非重叠区域中目标的交接跟踪也是一大热点,Zhao 等人<sup>[14]</sup>将各个摄像机监控范围内的目标进行检测和跟踪,然后将所有的结果和综合到一起分析出同一目标在所有摄像机范围内的信息。Tan 等人<sup>[15]</sup>将摄像机之间相对的信息,利用概率论和数理统计方面的研究构建贝叶斯概率模型,将轨迹和特征信息作为评估参数评估多个摄像机范围里的目标统一的可能性。这两种方法都存在计算量大,计算复杂度高的缺点,但是为此类研究开启了多摄像机协同跟踪的序幕。

多摄像机同步定位可以区分为区域同步定位和相对位置同步定位,基于区域的定位是将区域作为分割将多个摄像机在不同区域中的目标重新匹配。Mittal 等人的研究是基于高斯颜色模型来处理多个摄像机中同以物体的匹配定位<sup>[16]</sup>,实现了多个摄像机中区分同一目标并同步位置信息。这类方法的优势在于相对简单和迅捷,但是可靠性上存在诸多问题。在几何约束的基础上,普遍将约束种类分为二维和三维的约束方法。三维的约束方法是将目标还原到三位环境中,并且获取目标三维的特征信息用作同步定位,这种方法在实现难度上比较大而且存在计算量巨大无法满足实时性要求的缺陷。二维平面上的可以克服三维定位的问题,可以满足比较高的实时性要求<sup>[17]</sup>,其中 Kran 和 Shah<sup>[18]</sup>使用的是基于平面的单应性约束原理,即二维平面的目标到摄像机投影图像上的成像,通过实现在不同摄像机采集的目标物的特征来匹配得到目标的位置信息达到同步的效果。这种方法需要处理的数据量比较小,实时性也比较好,能较好的满足对实时性要求较高的的要求。

## 1.3 本文的研究内容和组织结构

本文研究的主要内容是在对大场景监控的时候使用双摄像机实时的生成视频摘要，基于以上目的本论文研究分为以下几个部分：

(1) 对两个摄像机进行标定，优化标定算法，实现两个摄像机坐标系的统一，根据标定出的统一坐标系实现场景监控相机对目标抓拍摄像机球面极坐标系的映射，使得标定出的坐标信息与目标抓拍摄像机的水平角度和俯仰角度，以便目标抓拍摄像机抓拍运动目标。

(2) 研究运动目标检测和运动目标跟踪算法，遴选出一种或者改进一种满足本课题实时性要求的运动目标检测跟踪算法，利用固定的场景监控摄像机对场景中的运动目标进行检测和跟踪，并且将图像的坐标信息转换成标定后的公共坐标系中的坐标信息。

(3) 设计两个相机协同工作的控制模块，使系统可以根据两个摄像机标定的公共位置信息，有效的控制目标抓拍摄像机运动到坐标系中目标相对应的俯仰角度对应的位置。

(4) 设计视频摘要生成方法，使两个摄像机采集的数据生成由多种表现形式信息组成并且能够代表目标主要信息的视频摘要。该摘要根据两个摄像机采集的不同清晰度的视频和图像信息，完成两个摄像机采集的数据进行融合，生成有效的视频摘要。

论文结构安排如下：

第一章绪论，本章内容主要为课题的研究目的和意义，介绍了目前双摄像机的技术现状和视频摘要技术的发展情况。

第二章总体设计，本章介绍双摄像机实时视频摘要系统的具体构成和设计原理，将课题各部分内容的设计思路和原理进行逐一描述。

第三章双摄像机标定的研究，本章介绍了双摄像机标定的概况和实现原理，实现标定出两个摄像机彼此图像间映射的转换矩阵。

第四章运动目标检测和跟踪方法的介绍和实现，使用了基于空间约束混合高斯背景建模的方法建立背景模型，利用背景差分法进行运动检测，并使用 Camshift 方法对检测到的运动目标进行实时跟踪。

第五章设计双摄像机协同工作的方法并将其运用在双摄像机实时视频摘要系统中，整合标定、检测和跟踪内容实现两个摄像机协同对场景中的目标进行高清抓拍操作，生成能供查阅的多种形式信息的视频摘要系统。

---

第六章总结和展望，总结课题的完成情况，以及课题存在的不足之处和有待改进的地方，对以后此类课题发展方向进行展望。

## 第二章 总体设计

本文研究的课题是双摄像机实时视频摘要系统，目的是针对当前在大场景中视频摘要系统生成的视频摘要所获取的摘要信息不够清晰的问题，这种问题产生的原因在于大场景中采集的图像和视频要兼顾到监控的大面积区域，因此必然会导致目标的清晰度的不足。场景的面积和清晰度是一对相互制衡的两个指标，要想进行场景的大面积监控多数情况下会以损失清晰度为代价。

针对这些问题，本课题对大场景下的高清晰度的实时视频摘要技术做出相应研究，利用云台摄像机的灵活性和两个摄像机协同工作的优势来弥补在监控大场景的普通监控系统在清晰度的目标信息采集方面的弱点。本文将针对这一目的设计出能够利用双摄像机（一个场景监控摄像机，一个目标抓拍摄像机）的灵活性，一个监控场景中的目标，并根据两个摄像机经过双摄像机标定确立的场景的坐标系来共享位置信息以实现目标抓拍摄像机捕捉目标的高清信息，再根据两个摄像机采集的不同的信息，将数据融合成视频摘要。

### 2.1 总体结构设计

本课题研究的双摄像机实时视频摘要系统与传统的视频摘要系统相比，核心优势在于采用了双摄像机，能够从两个方面（大场景监控和高清信息采集）对场景中的目标进行信息采集，采集的信息中既包含了大场景中目标的运动、位置、轨迹和高清图像等信息。为了控制云台根据实时检测和跟踪到的目标位置信息对目标进行抓拍，本文设计了两个摄像机协同工作的方法并以两个摄像机标定为基础产生相同的坐标系供两个摄像机共享要抓拍的目標的位置信息。那么整体分为四个部分双摄像机标定、运动目标检测、运动目标跟踪、双摄像机协同工作并生成视频摘要，根据个部分的设计将各部分整体集成为一个两个摄像机能够协同工作的实时视频摘要系具体集成图如图 2-1 所示：

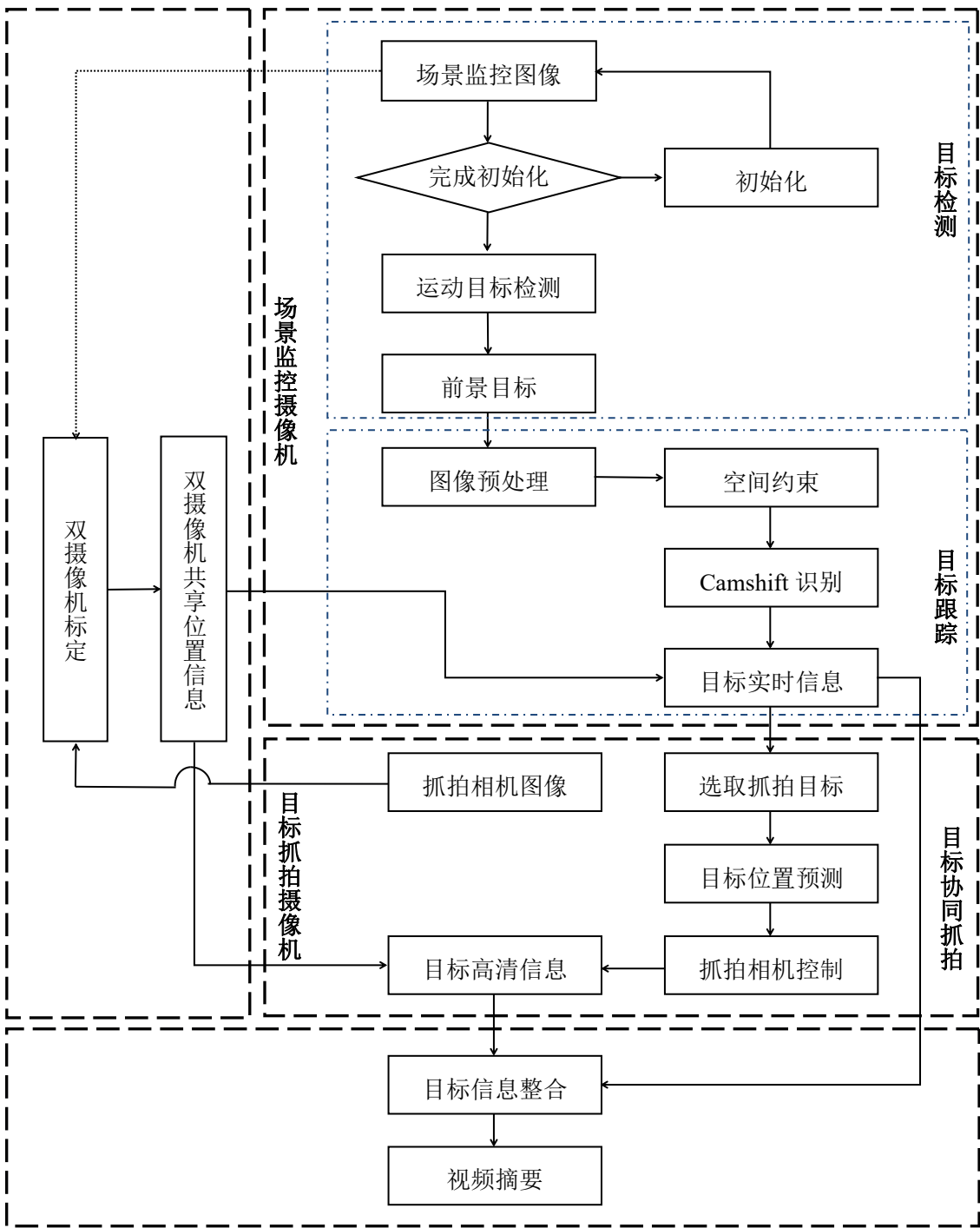


图 2-1 总体设计集成示意图

由图 2-1 可以从相机功能上将系统分为两个部分：场景监控部分和目标抓拍部分。这两部分共同的基础是以两个摄像机采集的图像为依据进行标定后两个相机存在的共同的位置坐标系，这部分使得两个不相关的功能产生紧密联系，互相提升了效率。在场景监控部分可以按照对场景中目标不同操作将这部分分为目标检测和目标跟踪，这部分作为双摄像机实时视频摘要系统的主要部分之一对目标

进行发现并持续跟踪，将采集的信息作为视频摘要重要的一部分。双摄像机协同工作是集合了标定对两个摄像机建立合适的关联，并依据目标检测和跟踪采集到的数据控制云台摄像机进行目标高清信息抓拍，将抓拍的高清信息作为视频摘要的补充。

## 2.2 主要研究内容

本文在设计双摄像机实时视频摘要系统的时候采用了模块化的思想，以便达到高内聚、低耦合的效果，使各部分之间既能相互配合，又尽量少的影响彼此的运转。双摄像机实时视频摘要系统划分为四个部分，双摄像机标定部分、场景监控和运动跟踪部分、双摄像机协同工作、视频摘要生成部分。各部分功能如下：

### (1) 双摄像机标定

双摄像机标定部分主要任务是负责构建系统所需要的参考坐标系，分别先标出场景监控摄像机和目标抓拍摄像机的内参数运用于对外参数的标定。设计出坐标转换方法分别将两个相机图像坐标和相机本身坐标建立映射，将相机坐标和世界坐标系建立映射，这样图像坐标和世界坐标之间就存在相应的方法进行相互转换。有了坐标相互转换方法就可以根据场景摄像机中跟踪到的运动目标求解出目标在世界坐标的位置，来操控目标抓拍摄像机运动到指定的位置，来抓拍目标的清晰图像信息。

目前双摄像机标定还是存在两个明显缺陷，一个是标定方法一般要求条件比较苛刻，比较难以操作，一个是标定后精度难以保证，很容易出现误差过大的情况，还有可能经过多次映射的过程导致误差会逐渐放大，到最后误差大到无法正常使用。这些问题都是这个部分面对的主要难题和需要解决的问题。

### (2) 场景监控和运动跟踪

场景监控和运动跟踪部分主要功能是对整个大场景进行监控，对场景中的运动目标先识别再跟踪。场景监控摄像机要求的像素没有那么高，只要求能对场景进行建模，使用差分法对运动目标进行检测，使用 Camshift 或者其他运动跟踪方法对运动目标进行实时跟踪，并将运动目标的坐标信息实时反馈成世界坐标系中的坐标，以供目标抓拍摄像机对识别和跟踪到的目标进行实时抓拍。

### (3) 双摄像机协同工作

双摄像机协同工作部分作为整个系统的控制核心，它综合了场景监控和运动跟踪部分采集的运动目标的坐标信息和移动信息，来控制目标抓拍摄像机对目标

进行实时跟踪，获取到目标的高清图像信息。本部分根据场景摄像机检测出的运动目标的位置信息和运动信息，将位置信息和运动信息进行整合，对需要目标抓拍摄像机进行跟踪的先后进行排序，然后根据排序对目标抓拍摄像机发出需要被跟踪的目标的位置信息和运动信息进行实时跟踪抓拍。

#### (4) 视频摘要生成

根据时间对运动目标进行跟踪记录目标在场景中的时间和轨迹信息，对跟踪成功的目标进行抓拍其高清图像信息，并对抓拍成功的目标进行事件记录，这样在时间维度上就形成了相对有序的、有效的目标信息检索。在场景中各种信息产生的同时对运动目标的位置和运动信息进行记录，记录成图片或者以短视频形式进行存储再综合目标信息生成视频摘要。这样就可以使在不对完整视频进行查看的前提下，通过检索生成的视频摘要快速地检索出运动目标在场景中存在的关键的、有效的、清晰的信息。

## 2.3 主要内容设计依据

### 2.3.1 双摄像机标定

双摄像机标定部分是本课题的重点也是双摄像机实时视频摘要的核心内容之一，双摄像机标定的精度决定了两个相机协同工作的效率和效果。针对标定的精度问题和标定难易程度问题，对双摄像机标定进行学习和研究并对比了当前几种相机标定难易程度对精度的影响，决定以张正友教授对标定的研究为基础实现对本课题中双摄像机进行标定。

本文的双摄像机标定采用了张正友教授的两步标定法对相机进行标定，过程如图 2-2 所示，通过反复迭代的方式对相机的内外参数不停的进行逼近，直到满足迭代终止条件，得到的结果收敛到一定的范围，可以得到满足精度要求的参数。



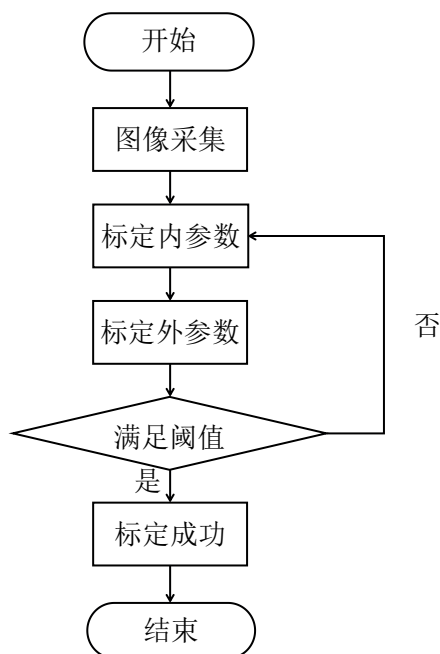


图 2-2 双摄像机标定流程图

### 2.3.2 运动目标检测

本课题对双摄像机进行明确的分工，以实现既采集场景中目标的运动信息又采集目标的高清信息。实现场景中运动目标的实时检测并且稳定跟踪是目标抓拍摄像机抓拍目标高清图像信息的基础之一。场景监控相机作为目标抓拍摄像机来获取目标运动信息和位置信息，可以使用清晰度一般的低成本摄像机作为场景监控摄像机，运动目标检测部分的作用就是快速的在场景中找到运动目标的位置，实时地反馈运动目标的位置和大小，以供跟踪模块对检测到的运动目标进行后续工作。

根据场景监控摄像机对监控区域的实时采样，建立稳定的背景模型，实时地区分出相对于背景模型的运动的目标，即对场景中的运动目标进行每一帧的检测，以实现快速地稳定地检测出运动目标的目的，流程如图 2-3 所示。

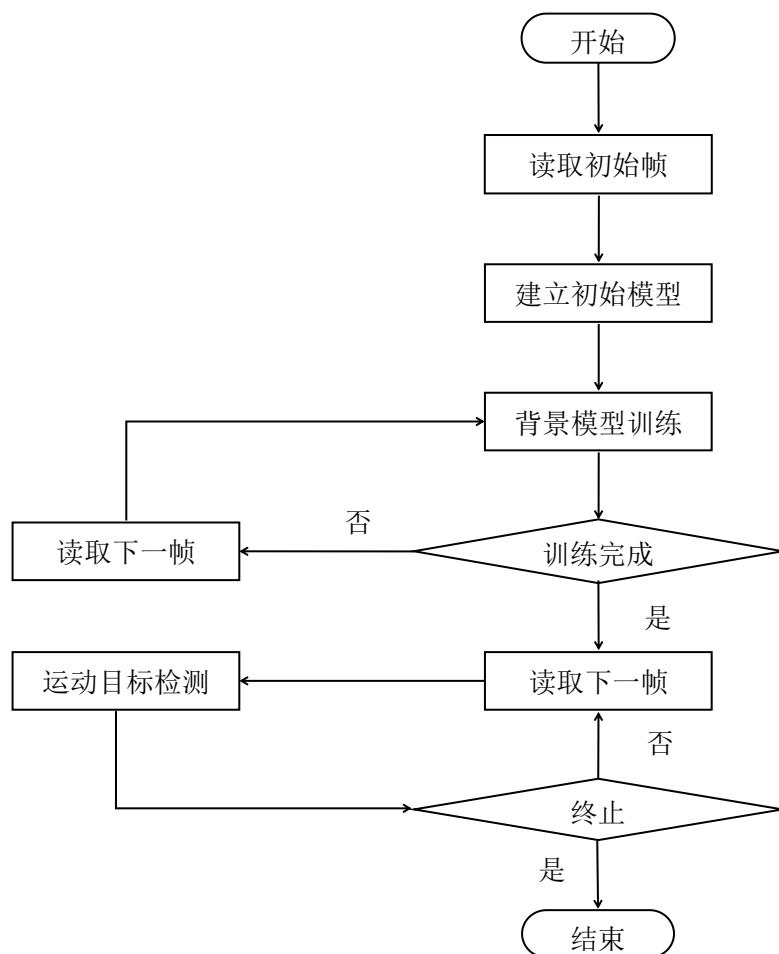


图 2-3 运动目标检测流程图

### 2.3.3 运动目标跟踪

目标跟踪是指在视频的图像序列中分析和查找出相应的待跟踪目标的位置和轮廓信息。通常典型的目标跟踪系统包含了以下几个方面：目标的外部信息模型、目标运动信息模型和目标的搜索策略。根据目标外部信息的采集和抽象方法，目标跟踪算法通常可以分为如下两类：生成模型法和判别模型法。

通过生成模型法在一些特征空间如颜色空间、色域空间等抽象出能够代表出目标的特征模型，对建立的搜索区域进行搜索，通过对比搜索区域中所有相似目标的相似度来确定跟踪目标的结果。基于生成模型法的目标跟踪技术，基本流程如图 2-4 所示。

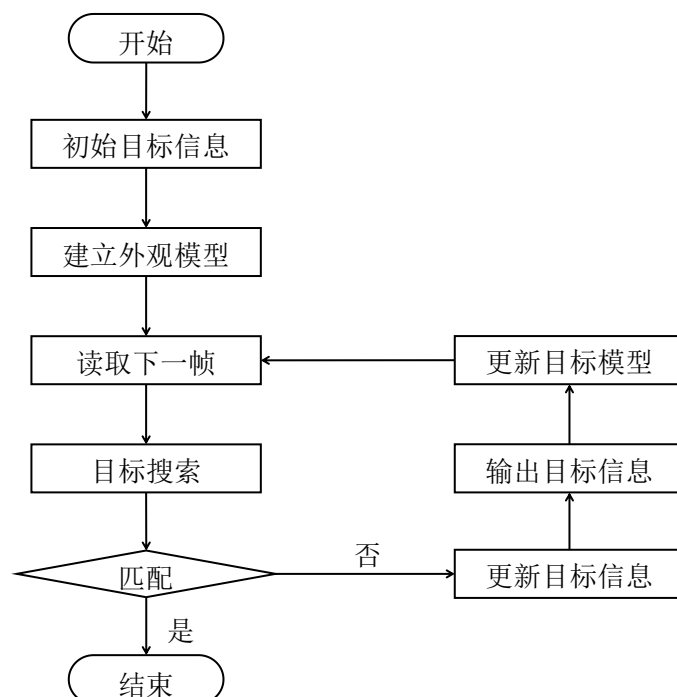


图 2-4 基于生成模型的目标跟踪流程图

基于判别模型法的目标跟踪方法，其基本流程如图 2-5 所示。判别模型法的关键点在于设定合理的阈值，将运动目标和背景模型通过阈值将二者区分出来，达到实时确定运动目标位置的目的。基于这种方法的目标跟踪性能的优劣主要体现在分类的过程中阈值是否能很好的将前景和背景割离。

目前常见的目标跟踪算法有 Camshift、TLD、CT 等。其中比较流行的跟踪算法是基于生成模型法的 Camshift 算法。这种跟踪方法主要是基于 meanshift 模型对目标的颜色直方图进行统计得到其概率分布，求解出颜色概率的特征向量，根据这些特征生成目标模型信息在搜索区域中进行相似性匹配。另外两种跟踪算法是基于判别模型法的运动跟踪算法，TLD 和 CT 主要基于两种不同的分类器进行场景和前景的分离，其精度和速度都比较依赖于对这两种分类器效果，而且训练分类器又需要相当的数据基础，实现起来也有一定的难度。Camshift 的运算相对以上两种基于判别法的跟踪方法相对快速且可以选择合适的特征分布方法进行建模，本课题采用基于 Camshift 的目标跟踪方法为基础采集场景中运动目标的实时信息。

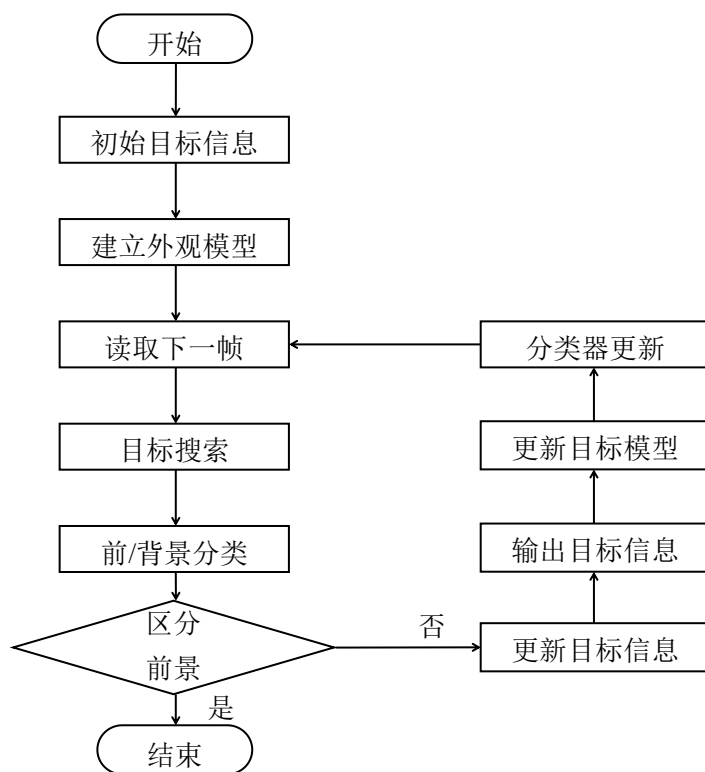


图 2-5 基于判别模型的目标跟踪流程

### 2.3.4 双摄像机协同工作和视频摘要生成

双摄像机协同工作是指场景监控摄像机和云台摄像机协同起来共同采集信息。从相机的分工来说，场景监控摄像机主要是检测和跟踪运动目标并实时反馈目标的位置信息和速度信息，目标抓拍相机主要根据场景监控摄像机采集的目标实时位置进行运动目标协同跟踪及抓拍。但是，面对多个目标需要跟踪时就会产生需要跟踪的目标无法在短时间内全部调用目标抓拍进行跟踪抓拍，导致会有一定数量的目标会丢失掉。为了更好的协调两个摄像机进行数据的采集，本文初步规划了双摄像机协同工作的视频摘要生成方法。该方法如图 2-6 所示：

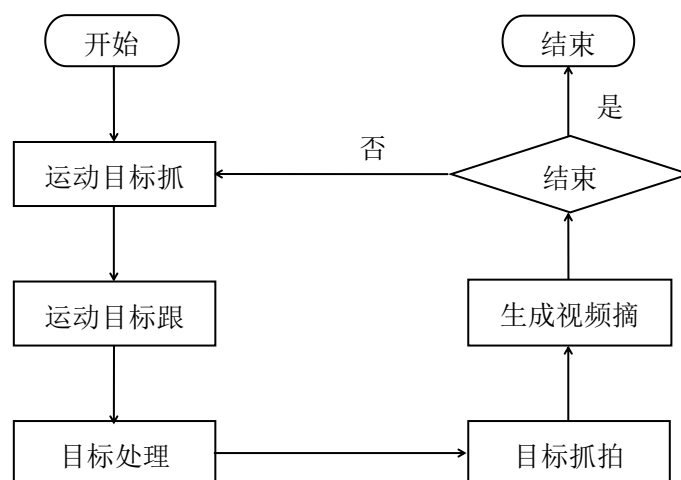


图 2-6 视频摘要生成流程图

视频摘要目的是生成整个视频监控的信息集合，包含了对场景的图像信息进行采集，将采集的信息作为监控显示播放，根据场景监控摄像机和目标抓拍摄像机采集的高清图像信息对图像的细节信息进行手机，生成包含场景和目标以及目标高清信息等多种信息形式相结合的摘要信息。

生成视频摘要是首先通过对场景监控相机和目标抓拍相机的 ip 网络接口采集两台设备的实时图像信息，将两个摄像机根据网络协议编码的数字码流转换成可以使用 opencv 处理的图像信息流，以用作视频摘要分析时使用。然后对场景监控摄像机采集的图像信息进行运动目标检测、运动目标跟踪，对运动目标轨迹信息进行记录，估算运动目标速度，然后将以上操作的数据作为场景监控摄像机产生的视频摘要的一部分。

最后利用双摄像机的协同工作协同场景监控摄像机对目标抓拍摄像机进行抓拍控制，抓取运动目标的高清信息，并将高清信息结合场景监控摄像机采集的目标信息结合起来，生成带有高清图像信息等多种形式的摘要信息。

## 2.4 本章小结

本章从整体的设计目标开始，描述了一个基于双摄像机的实时视频摘要系统的需求和内容。将系统做模块化划分，划分成四个功能和需求不同的部分，有双摄像机标定、运动目标检测、运动目标跟踪、双摄像机协同工作生成视频摘要，每个部分都有自己的特点和作用。针对每个部分的作用和各部分之间的相互间的

内在关系阐述了每个部分的设计思路和设计特色。其中有的部分设计的时候标准相对较高，存在一定的创新难度，预计各部分均存在相对的宽容空间，可以容许在可接受范围内的误差出现。

## 第三章 双摄像机标定

### 3.1 双摄像机标定介绍

本文所介绍的双摄像机实时视频摘要系统通过两个摄像机进行图像信息采集，基于采集到的图像完成对两个相机的标定后建立两个摄像机图像的映射关系。这是场景监控摄像机和目标抓拍摄像机协同对目标进行跟踪抓拍的基础，而这一基础是否牢靠完全依赖于标定是否准确。双摄像机标定的过程就是求解两个相机内在光学参数和两者之间相对彼此的位姿和角度的过程。本章重点介绍两个摄像机标定的基本原理和各种方法的不同，根据这些方法在标定中的是否满足本课题的最低要求，以及本文选用的标定方法和改进介绍。

### 3.2 摄像机标定原理

本文设计的双摄像机实时视频摘要系统需要对两个摄像机进行标定，要对一个摄像机标定或者对两个相机进行联合标定，首先要建立如下几个坐标系的概念：图像的像素坐标系、图像的物理坐标系、相机本身坐标系、公共的世界坐标系。

图像的像素坐标系：图像的像素坐标系是以图像的左上角为原点， $u$  轴为水平方向坐标轴，平行于图像平面指向为右， $v$  轴为竖直方向的坐标轴，垂直于  $u$  轴向下。图像像素坐标系中的坐标可以使用  $(u,v)$  表示，图像的像素坐标系可表示为如图 3-1 的形式。

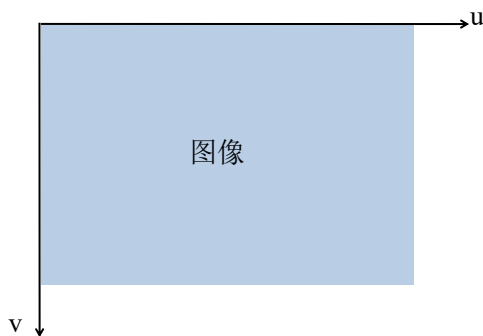


图 3-1 图像像素坐标系示意图

图像的物理坐标系：图像的像素只能表示目标的像素位置信息，并不能代表图像中目标的现实位置信息。因此，在其基础上建立一个具有物理意义的图像的物理坐标系，将图像的物理坐标系的原点设为图像的中心点， $x$  轴朝水平方向平行于图像的像素坐标系中的  $u$  轴， $y$  轴在竖直方向平行于图像像素坐标系的  $v$  轴，任意位置坐标可以用  $(x,y)$  来表示，图像物理坐标系可表示如图 3-2 所示。

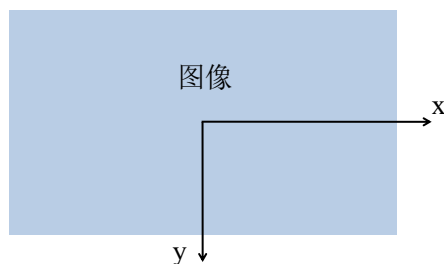


图 3-2 图像物理坐标系示意图

相机坐标系：相机坐标系是以其镜头的中心点作为其坐标系原点， $X$  轴平行于图像坐标系的  $X$  轴， $Z$  轴为相机的光轴， $Y$  轴垂直于  $Z$  轴和  $X$  轴组成的平面。相机坐标系是图像和相机产生映射的基础也是世界坐标和相机坐标相互转换的媒介，它的存在使得图像中的位置信息可以和世界坐标系中的位置进行转化。相机坐标系可表示如图 3-3 所示。

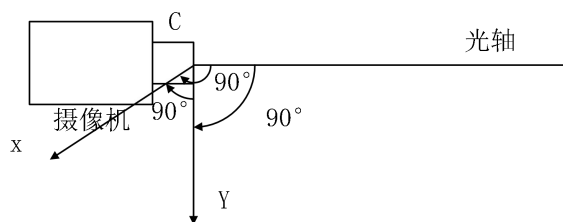


图 3-3 相机坐标系示意图

世界坐标系：世界坐标系也称为绝对坐标系，它可以代表任意场景中的目标的绝对位置。通过刚体变换(旋转和平移)将世界坐标系中的位置信息转换为相机坐标系中的位置信息。有了这种关联就可以将绝对坐标系中的位置信息转换成相机相对的位置信息，再通过相机坐标系映射到图像中目标的位置，使得图像中的目标信息和世界坐标系产生映射关系。



本文所使用的标定方法使用的均是相机模型中的针孔模型成像，针孔模型成像是相机的一个典型模型，因为它只包含了场景到图像的透视投影以及旋转平移的变化，并不包含对失真的描述。针孔成像模型如图 3-4 所示。

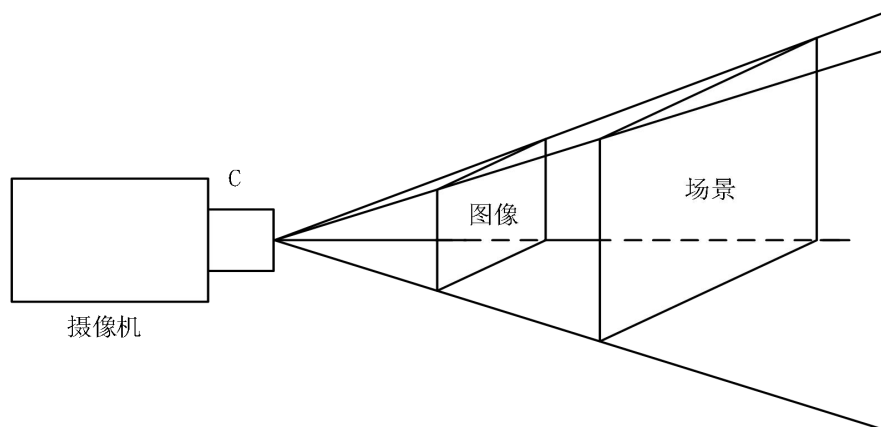


图 3-4 针孔模型成像示意图

由图 3-4 可以看到针孔模型是各种相机模型中最简单的一种，其通过透视投影成像，针孔模型的相机参数分别用内参数和外参数矩阵描述。该模型中图像的像素坐标和图像的物理坐标之间存在的对应关系如图 3-5 所示：

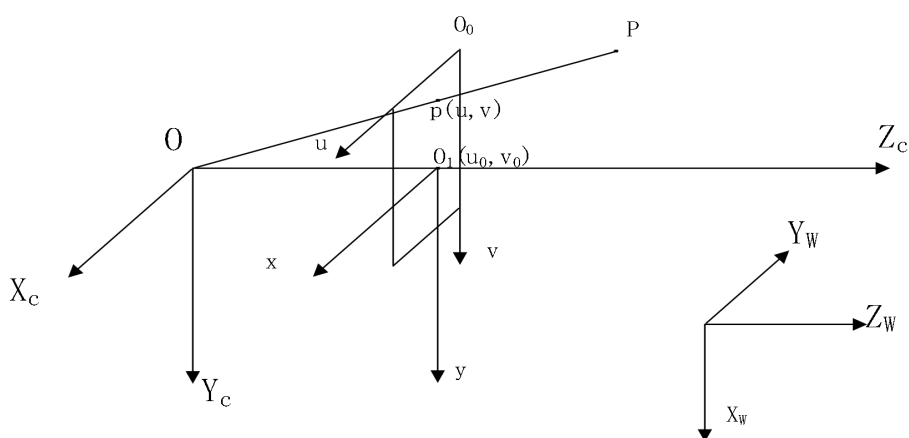


图 3-5 参考坐标系之间的关系示意图

在图 3-5 中,  $O_0$  为图像坐标系的原点,  $(u,v)$  为图像中的  $P$  点的坐标。光心在图像中的位置可以表示为  $(u_0, v_0)$ , 图像中每一个像素代表的物理尺寸在对应的坐标轴上可以表示为  $dx$  和  $dy$ , 在不存在畸变的时候, 图像中任意一个像素在图像的像素坐标和图像的物理坐标之间存在的变换关系如式(3-1)所示:

$$\begin{aligned} u &= \frac{x}{dx} + u_0 \\ v &= \frac{y}{dy} + v_0 \end{aligned} \quad (3-1)$$

用矩阵的形式可以表示如式(3-2):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3-2)$$

图 3-5 也描述了相机坐标如何通过统一的坐标体系转换到世界坐标的。为了可以清楚的表示出三维环境中的相机和目标的相对位置以及姿态, 就需要选择一个独立的基准坐标系作为参考, 该坐标系由  $X_w$  轴、 $Y_w$  轴和  $Z_w$  轴组成。相机坐标和世界坐标还可以通过空间运动关系进行转换即旋转和平移, 使用正交的单位旋转矩阵  $R$  与三维平移向量  $T$  将对应的相机坐标转换到世界坐标系中去。因此, 如果已知空间某点  $P$  的世界坐标为  $(X_w, Y_w, Z_w)^T$  它的相机坐标表示为  $(X_c, Y_c, Z_c)^T$ , 则有如式(3-3)的关系:

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

式(3-3)中,  $R$  为正交单位旋转矩阵,  $T$  为空间中  $x$ 、 $y$ 、 $z$  三个方向上的平移向量。

对于空间中的任意一点  $P$ , 在它对应的平面图像中的物理坐标可以写成式(3-4):

$$\begin{aligned} x &= f \frac{X_c}{Z_c} \\ y &= f \frac{Y_c}{Z_c} \end{aligned} \quad (3-4)$$

由(3-3)式可得:

$$\begin{aligned}
Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} f/dx & s & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \\
&= K [R \quad T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = P \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}
\end{aligned} \tag{3-5}$$

式(3-5)中, P 为 3\*4 矩阵, 被称为投影矩阵; s 为扭转因子;  $k_u=f/dx$ ;  $k_v=f/dy$ ; K 由  $k_u$ ,  $k_v$ , s,  $u_0$ ,  $v_0$  的值来确定, 因为 K 只与相机的机械和光学参数有关, 所以 K 也被称为相机的内参数; [R T] 由相机参考与世界坐标系相对的位置来确定, 也被叫做外参数矩阵。

通过以上方法可以计算出相机的内参矩阵, 也就是相机固定的内部参数, 它是双摄像机标定和定位的基础。

### 3.3 双摄像机标定方法介绍

根据标定的方式不同可以将标定分为传统的标定方法、自标定法和主动视觉标定法三种<sup>[19]</sup>。传统标定又被称为强标定<sup>[20]</sup>, 强标定具有计算复杂、需要标定块、标定难度大, 但是适用于任何相机模型等特点; 自标定法又被称为弱标定<sup>[21, 22]</sup>, 具有精度不高、非线性标定、鲁棒性不强, 但是仅需建立图像之间的对应关系、灵活方便等特点; 主动视觉标定法<sup>[23]</sup>不能应用于相机无运动数据和无法控制相机运动数据的场合, 但是通常可以线性求解, 鲁棒性较好。

目前, 相机的标定提出了很多具体的方法, 根据相机模型将标定方法具体为三种类型: 线性标定法、非线性标定法和两步标定法<sup>[24, 25]</sup>。

#### (1) 线性标定法

Abdel-Aziz 和 Karsra 于 1971 年提出基于直接线性变换 (DLT) 的标定方法<sup>[26]</sup>。因为直接线性变换法相对比较简单, 所以直接线性变换法在线性标定法中算是应用较多的一种方法。Dainis 和 Juberts 对线性标定法进行更为深入的研究<sup>[27]</sup>, 并取得了不错的效果。线性标定法是根据解线性方程来获取转换参数<sup>[28]</sup>, 运算量小, 速度快, 但是由于缺少对摄像机畸变问题的考虑和对噪声的抗干扰不明显, 这种方法对长焦距的小畸变镜头相对适合。

#### (2) 非线性标定法

非线性标定法，主要基于非线性模型，非线性模型越准确，计算量也越大。虽然这种方法可以补偿镜头产生的畸变<sup>[29]</sup>，可以采用更为复杂的映像模型，但是也更依赖更为良好的初始估计，如果迭代过程设计的不恰当，优化的过程可能就不稳定。

### (3) 两步标定法

两步标定法将标定分为两步，第一步先不考虑镜头畸变，求解摄像机内参数，再使用非精确的相机内参数去估计相机的外参数，通过迭代的方法将两个参数进行逼近，最后得到在一定阈值内的内外参数数据。其中 Z Zhang 的平面模板两步法<sup>[30, 31]</sup>不再使用之前常用的高精度标定台作为标定图像采集的方法，只需通过拍摄多种位置和姿态的标定板来进行相机内部参数的标定计算。

本文综述几种标定方法的特点和现实标定环境，决定选取 Z Zhang 的张氏标定法也被称为张正友标定法来进行摄像机内部参数的计算。

张氏标定法是张正友教授在 1998 年提出并逐渐完善的一种线性标定方法，作为一种简单且精度较高的标定方法这种方法被大家广泛认同和应用。张氏标定法采用二维平面面板作为标靶，通过在不同视点采集的图像上标靶的信息，实现了相机的标定。这里的标定是指对相机内外参数的标定，其中内参数使用内参矩阵  $M$  表示。

$$M = \begin{bmatrix} k_x & k_s & u_0 \\ 0 & k_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3-6)$$

求取内参矩阵  $M$  时需要先求取两个平面的单应性矩阵，其中单应性也是本文要用到的一个重要的理论。单应性是一种针孔成像的一种特性，它是只考虑两个平面之间的映射关系而不考虑深度对映射的影响，单应性如图 3-6 所示：

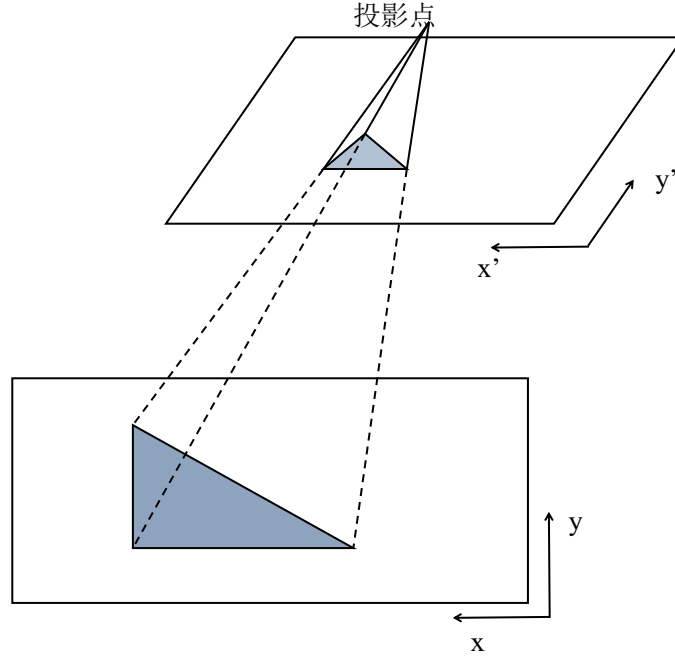


图 3-6 单应性示意图

两个空间中的平面通过同一个投影点作针孔映射，那么必然存在两个平面上的唯一映射关系，由于存在这样的关系本文将这种针孔投影存在的特性称为单应性。通过投影点到两个平面的投影必然存在两个维度 $(x,y)$ 到 $(x',y')$ 的映射关系，本文将这个映射关系求出就叫做单应性矩阵。

用标定板的方式进行求解时，一般将平面标定板上的特征点的坐标记作 $(x_{wi}, y_{wi}, z_{wi})$ ，成像平面上的坐标记作 $(x_{ci}, y_{ci}, z_{ci})$ ，图像坐标记作 $(u_i, v_i)$ 根据公式有如下关系：

$$s \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = M \begin{bmatrix} n & o & a & p \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ 0 \\ 1 \end{bmatrix} = M \begin{bmatrix} n & o & p \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ 1 \end{bmatrix} \quad (3-7)$$

式(3-7)中， $s$  为深度系数； $M$  为相机的内部参数也就是内参矩阵。

可得公式：

$$sI_i = HP_i \quad (3-8)$$

式(3-8)中， $I_i = [u_i \ v_i \ 1]^T$  为点  $P_i$  的图像坐标； $P_i = [x_{wi} \ y_{wi} \ 1]^T$  为点  $P_i$  的标定板中特征点的坐标系坐标； $H = M[n \ o \ p] = [h_1 \ h_2 \ h_3]$  为世界坐标系到图像空间的单应性矩阵。

求出投影和真实世界中坐标关系后就可以得到单应性矩阵，单应性矩阵中又

包含了相机的内部参数矩阵，由于单位向量  $\mathbf{n}$  和  $\mathbf{o}$  是正交的，于是有了下面的约束：

$$\begin{cases} h_1^T M^{-T} M^{-1} h_2 = 0 \\ h_1^T M^{-T} M^{-1} h_1 = h_2^T M^{-T} M^{-1} h_2 \end{cases} \quad (3-9)$$

由式(3-9)中的约束可以求得相机的内参数矩阵，在求解内参数矩阵时，可以使得

$$B = M^{-T} M^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \quad (3-10)$$

由于  $B$  是一个对称矩阵，可以定义一个六维向量：

$$\begin{aligned} b &= [B_{11} \quad B_{12} \quad B_{13} \quad B_{14} \quad B_{15} \quad B_{16}]^T \\ h_i^T B h_j &= v_{ij}^T b \end{aligned} \quad (3-11)$$

式(3-11)中， $v_{ij} = [h_{i1}h_{j1} \quad h_{i1}h_{j2}+h_{i2}h_{j1} \quad h_{i2}h_{j2} \quad h_{i3}h_{j1}+h_{i1}h_{j3} \quad h_{i3}h_{j2}+h_{i2}h_{j3} \quad h_{i3}h_{j3}]$ 。这样，相机的内参矩阵约束条件可以重写为

$$\begin{bmatrix} v_{12}^T \\ (v_{11} + v_{12})^T \end{bmatrix} b = 0 \quad (3-12)$$

如果有标定  $n$  幅图像，可以得到  $n$  组式(3-12)方程，将其写成矩阵的形式为

$$Vb = 0 \quad (3-13)$$

式(3-13)中， $V$  是  $2n \times 6$  的矩阵。

当采集的标定图像大于 3 的时候， $V$  的最小特征值对应的特征向量即为  $b$ ，然后根据  $b$  和  $B$  的定义可以求出相机的内参数为：

$$\begin{cases} v_0 = (B_{12}B_{13} - B_{11}B_{23}) / (B_{11}B_{22} - B_{12}^2) \\ k_x = \sqrt{c / B_{11}} \\ k_y = \sqrt{cB_{11} / (B_{11}B_{22} - B_{12}^2)} \\ k_s = -B_{12}k_x^2k_y / c \\ u_0 = \frac{k_s v_0}{k_y} - B_{13}k_x^2 / c \end{cases} \quad (3-14)$$

式(3-14)中的  $c = B_{33} - [B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})] / B_{11}^2$ 。

在标定出相机的内参数后，根据单应性矩阵  $H = \gamma M[n \quad o \quad p]$ ，可得外参数矩

阵为

$$\begin{cases} \gamma = \frac{1}{\|M^{-1}h_1\|} = \frac{1}{\|M^{-1}h_2\|} \\ n = \gamma M^{-1}h_1 \\ o = \gamma M^{-1}h_2 \\ a = n \times o \\ p = \gamma M^{-1}h_3 \end{cases} \quad (3-15)$$

求得了相机的内参数和外参数就可以求得图像坐标系中目标到世界坐标系中目标的坐标重投影。

根据张正友教授的标定方法能够得到两个摄像机的内参数和相对于标定板的世界坐标系的内参数，这样还不算两个摄像机的标定，因为还没有建立两个摄像机之间的联系，无法得到两个摄像机在坐标和位姿的关系，也就无法根据场景摄像机采集的运动目标的位置信息使用云台摄像机进行目标位置定位。针对单目标标定中两个摄像机无关联的问题，本课题实现了对两个摄像机采集的图像进行区域定位，可以使得两个摄像机采集的图像中像素的坐标实现相对于摄像机坐标系的位姿定位。

由数学模型中的坐标映射关系可以求得一个图像坐标系与世界坐标系的转换矩阵：

$$s \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = M \begin{bmatrix} n & o & a & p \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ 0 \\ 1 \end{bmatrix} = M \begin{bmatrix} n & o & p \end{bmatrix} \begin{bmatrix} x_{wi} \\ y_{wi} \\ 1 \end{bmatrix} = H \begin{bmatrix} x_{wi} \\ y_{wi} \\ 1 \end{bmatrix} \quad (3-16)$$

其中的  $H$  就是世界坐标系到图像坐标系的转换矩阵，那么也就存在一个逆矩阵使得图像到世界坐标系存在一个转换矩阵  $H^{-1}$  可以使图像坐标系能够通过转换矩阵  $H^{-1}$  将图像中的坐标转换到世界坐标系中。

$$sH^{-1} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} x_{wi} \\ y_{wi} \\ 1 \end{bmatrix} \quad (3-17)$$

但是，转换矩阵中不仅包含了相机的内参数矩阵，还包含了针对标定时所用标定板所代表的标定板世界坐标系与相机之间的外参数也就是旋转平移参数，并

不具有通用性，所以并不容易得到一个通用的两个摄像机之间的转换矩阵。

针对这样的问题本文给出的解决方案是先根据内参矩阵的作用，即使用内参矩阵对图像坐标系中的坐标转换成相机坐标系中的坐标，转换后分别得到两个相机采集的图像中的坐标相对于相机坐标系中的坐标，再根据标定时标定板的位姿得到两个相机的旋转和平移关系就可得出一个新的通用的转换矩阵使得两个相机坐标系可以进行坐标转换。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = M \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3-18)$$

式(3-18)中， $M$  为内参矩阵，该公式描述了如何使用内参矩阵将图像中的图像坐标转换成相机坐标系中的相机坐标。

那么两幅图像中坐标的关系可以表示为：

$$\begin{bmatrix} x_l \\ y_l \\ 1 \end{bmatrix} = M_l^{-1} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = M_l^{-1} R T \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = M_l^{-1} R T M_r \begin{bmatrix} x_r \\ y_r \\ 1 \end{bmatrix} \quad (3-19)$$

式(3-19)中  $M_l^{-1}$  为左边相机的内参矩阵的逆矩阵， $R$ 、 $T$  分别为两个相机的旋转矩阵和平移矩阵， $M_r$  为右边相机的内参矩阵。

### 3.4 双摄像机标定的实验及分析

本节将在两个摄像机上对提出的双摄像机标定算法进行验证性实验和性能测试。实验基于两个不同的摄像机，使用 12\*18 的黑白方格标定板进行两个摄像机同一场景下对标定板信息采集，然后对采集的标定板信息进行单目标定，标定后进行重投影误差计算，将误差保持在可接受的范围，对两个摄像机采集的图片进行双目标定，计算两个相机之间的映射矩阵。





图 3-7 模拟标定用双摄像机

本文采用两个焦距不同的相机用于场景模拟和双摄像机标定实验，如图 3-7 所示，两个摄像机焦距和内参数虽然不同，但是采集的图像分辨率都为  $640 \times 480$  像素，标定时将两个相机放置在相邻位置同一高度，照射同一片区域。

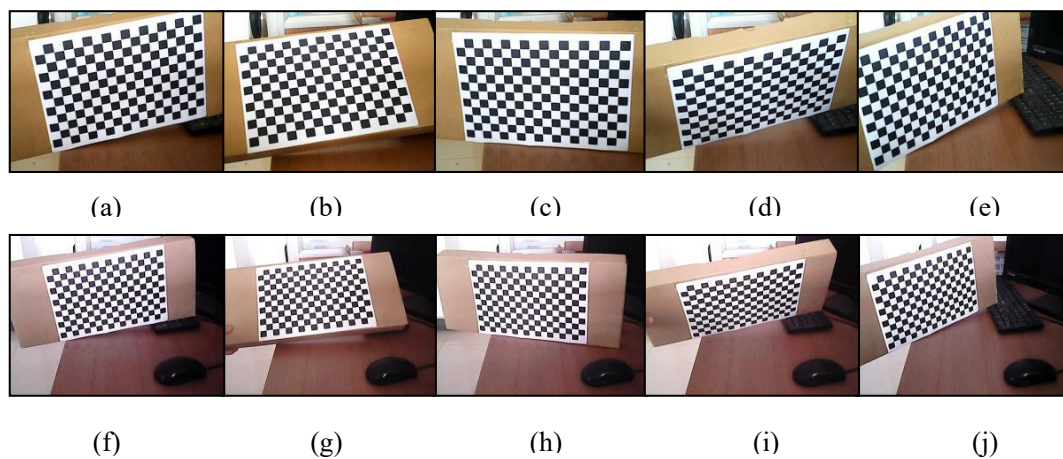


图 3-8 两个摄像机采集的图像：(a)-(e)为模拟场景监控相机拍摄图片；(f)-(j)为模拟目标抓拍相机拍摄图片。

如图 3-8 所示，(a)-(e)为模拟场景监控相机拍摄的一组图片，(f)-(j)为模拟抓拍相机拍摄的一组图片，每个图片均上下对应，为同一时刻两个相机拍摄的同以物体不同角度所拍摄的图片。在完成了两个摄像机图像的采集工作后继续对采集的图像进行处理。

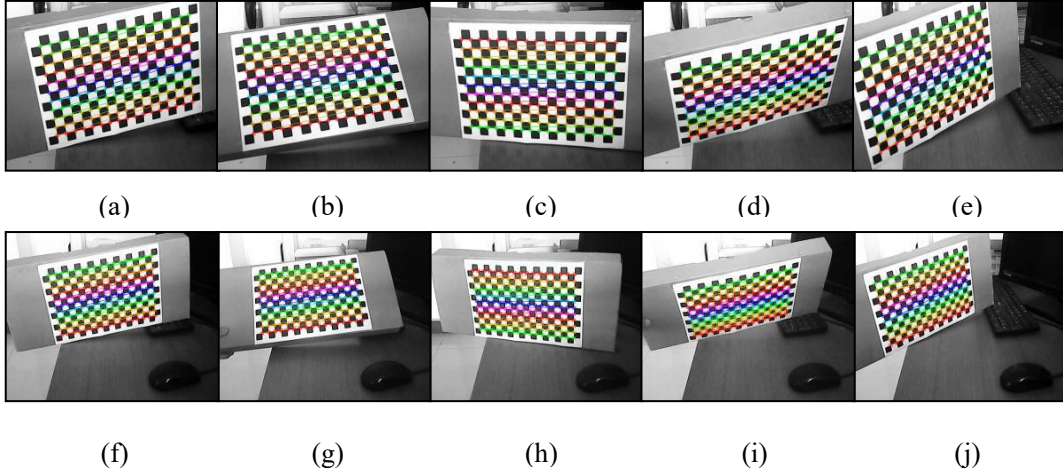


图 3-9 两个相机查找对应特征点；(a)-(e)为场景监控相机采集图片处理结果；(f)-(j)为目标抓拍相机采集图片处理结果。

如图 3-9 所示，对两组图片中对应的标定板上的特征点进行检测，定位统一标定板同一时刻在两个相机图像中的标定板特征位置。

得到这些特征点的数据后就可以根据这些特征点数据分别对两个相机单独标定，先得到相机的内部参数。其中场景监控相机内参矩阵如，目标抓拍相机内参矩阵如所示：

$$M_l = \begin{bmatrix} 1239.96 & 0 & 329.17 \\ 0 & 1231.93 & 181.48 \\ 0 & 0 & 1 \end{bmatrix} \quad (3-20)$$

$$M_r = \begin{bmatrix} 831.89 & 0 & 295.39 \\ 0 & 826.30 & 162.02 \\ 0 & 0 & 1 \end{bmatrix} \quad (3-21)$$

通过两个摄像机的内参数对两个相机实现场景重建如图 3-10 所示，重建出两个摄像机的相对位置和标定板相对于两个摄像机的相对位置：

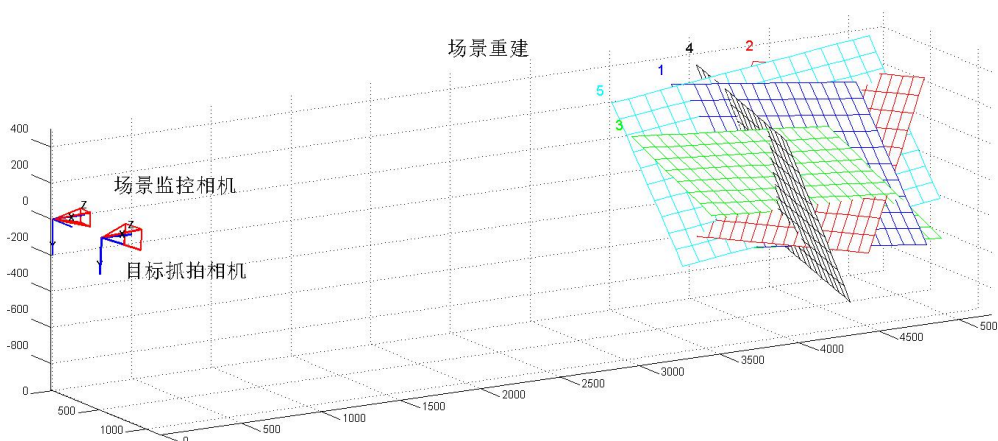


图 3-10 场景重建示意图

两个相机的内参数求解出后，再求解两个相机彼此间的映射矩阵如式(3-22)：

$$H = \begin{bmatrix} 0.66 & 0.22 & 56.90 \\ 0.002 & 0.67 & 35.37 \\ -0.00007 & 0.00002 & 1 \end{bmatrix} \quad (3-22)$$

根据两个相机坐标的映射矩阵求得两个相机重映射的图像如图 3-11：

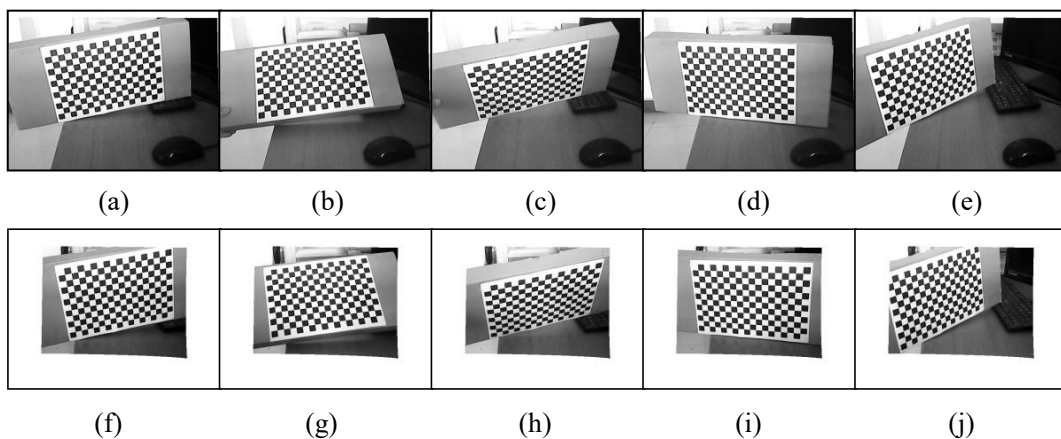


图 3-11 重映射图像；(a)-(e)为目标抓拍相机采集图片处理结果；(f)-(j)为场景监控相机采集图片处理结果。

使用两个相机的内参数矩阵和重建的结果来确定两个图像的偏移和旋转，以场景监控相机的图像为基础对场景监控相机进行空间变换使之能够与目标抓拍相机采集的图像重建到同一图像坐标系中。

重投影误差分析：

针对标定后场景监控相机的重投影图像中标定板的位置重新进行角点位置定位和对应模拟目标抓拍相机采集的图像中角点位置进行对比，求解重投影后模拟

场景监控相机图像的映射平均误差。

表 3-1 标定重投影误差

图片序号	平均投影误差 (像素)
f	5.90
g	2.88
h	8.19
i	8.14
j	6.06
f~j 均值	6.32

由表 3-1 可见,采用的五张标定用的图片在两个摄像机中图像经过转换矩阵转换后,两个模拟摄像机对应的每张图像中标定板上每个对应点在图像中的像素位置计算相对的位置后,每幅图像中特征点相对位置的平均误差如表 3-1 所示,最大误差不超过 10 个像素,最小误差为 2.88 像素,图像从 f~j 中的 187\*5 个特征点的平均误差为 6.32 个像素。基本上能够满足本系统对双摄像机标定的精度需求。

3.5 本章小结

本章介绍了摄像机标定的含义以及标定的基础,标定主要是处理相机、图像、和世界三个坐标系之间的关系,使各坐标系之间能够建立练习,拥有统一的参考系,使得场景监控摄像机采集的目标位置能够有一个与之对应的世界坐标系并且映射到目标抓拍相机的坐标系中去。根据对张正友标定法的利用,基本上实现了双摄像机标定的目的,两个摄像机中的目标存在一个位置上的映射关系。虽然标定的精准性还存在一定的误差,但是面对这种误差可以使用双摄像机协同工作中的一些处理误差的策略对误差进行控制和避免。这样就可以以简单的方式标定,标定后取得预期的标定目标。

## 第四章 运动目标检测和跟踪

本章将介绍课题中使用到运动目标检测和跟踪技术<sup>[32]</sup>，因为在完成两个摄像机的标定之后，基本上确立了两个摄像机的相对位置关系和采集的图像之间的映射关系，也就可以对两个摄像机进行后续的协同工作研究，而两个摄像机协同对运动目标进行跟踪和抓拍的另一重要基础就是运动目标的检测和跟踪。只有首先在场景监控摄像机中准确地检测到运动目标并稳定地进行运动目标跟踪，实时稳定地获取目标的位置信息，再根据标定数据同步运动目标在两个摄像机中的相对位置，才能根据两个摄像机协同工作策略对检测到的目标进行跟踪和抓拍工作。

### 4.1 运动目标检测技术

运动目标检测技术一般认为有这三类：光流法、背景差分法、帧间差分法<sup>[33, 34]</sup>。这三种算法分别从不同的角度实现了对场景中运动的目标进行检测，但是产生的效果各不相同。面对不同的场景，这几种算法都表现出各自的优劣，因此要选取适合本课题场景的运动目标算法还需要根据场景特点和算法特点进行综合考虑。

#### 4.1.1 帧差分法

帧差分法<sup>[35, 36]</sup>是运动目标检测技术中较早提出的运动目标检测算法，也叫做帧间差分法，经过长时间的研究和应用后，研究者又将帧差分法分为两帧法、三帧差分法和多帧差分法来满足各种不同场景和应用的需求。

帧间差分法是将视频中的图像序列中的连续两帧或者多帧进行差分运算，然后对差分后得到的数据进行阈值分割处理，得到两帧之间像素灰度值发生一定变化的像素的区域，这些像素灰度值发生变化的区域一般认为它就是相对于背景而言的前景，也就是场景中的运动目标。帧间差分法是一种相对简单，运算量相对较小的运动目标检测算法，具有很好的实时性，但是需要根据不同的速度和时机选择合适的帧间差分的帧差。

##### (1) 相邻帧间差分法

帧差分法中最简单方法是相邻帧间差分法，该算法非常简单且容易理解，计算量小而且实时性很好，由于在背景比较稳定的场景中不需要经常更新场景信息，

所以背景稳定的场景比较常用这种算法。相邻帧间差分法如图 4-1 所示：

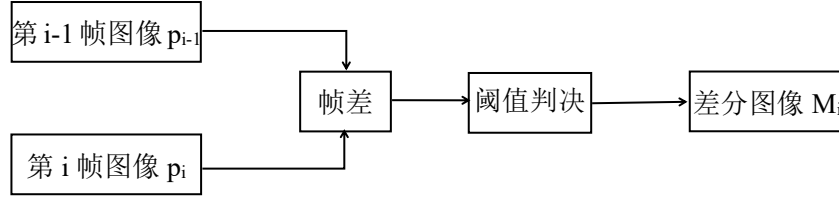


图 4-1 相邻帧间差分法模型原理图

在  $i$  时刻，图像  $p_i$  的每个像素值减去  $p_{i-1}$  相应位置的像素值，得到一张新的图像根据阈值判决条件对两幅图像中的像素差值进行阈值判断，得到一张差分后的图像  $M_i$ ，这张  $M_i$  图像可以根据像素的值来区分前景和背景。

$$M_i = \begin{cases} 255, & |p_i - p_{i-1}| > T; \\ 0, & \text{other.} \end{cases} \quad (4-1)$$

式(4-1)中的  $T$  为判决阈值，当像素的差值大于阈值判决条件  $T$  时像素值变为 255，当像素差值小于阈值判决条件  $T$  时将像素值调整为 0，这样就得到一张差分后的二值图，将前景和背景区分出来。

## (2) 三帧差分法

相邻帧间差分法<sup>[37]</sup>是运动目标检测中的基础方法，拥有快速和简单的特点，但是也存在两帧间间隔过大或者运动目标过快的速度，会产生“空洞”和“双影”的情况，这会导致错检运动目标或者检测的目标不完整。针对以上问题，研究者们推出了一种三帧差分法。

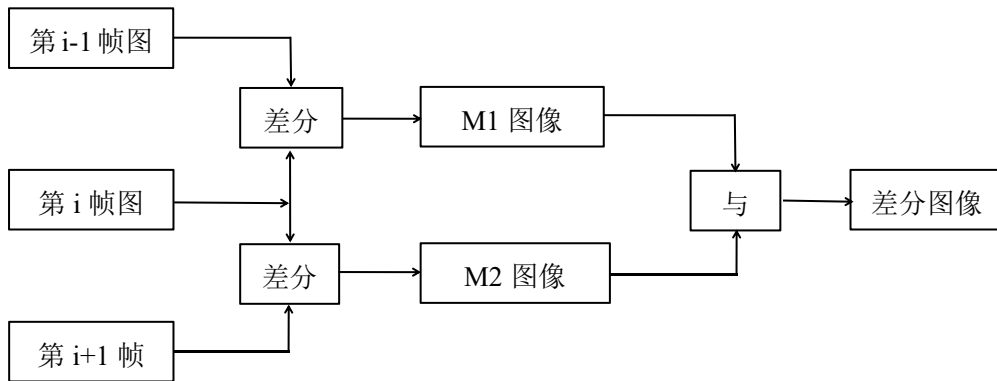


图 4-2 三帧差分法模型原理图

三帧差分法改进于相邻帧差分法<sup>[38]</sup>，如图 4-2 所示，首先将第  $i$  帧图像和第  $i-1$  帧图像进行差分运算得到  $M_1$  图像，将第  $i+1$  帧和第  $i$  帧图像做差分运算得到  $M_2$



图像，再将求得的  $M_1$  和  $M_2$  进行与运算得到最终的  $M_i$  图像，由于在第一步做差分之后产生的  $M_1$  图像和  $M_2$  图像大部分区域都是背景，只有部分存在不同，在做了“与”操作后，能够有效的抑制相邻帧差法导致的拉长影和空洞情况。这种方法效果不仅明显优于相邻帧差法还兼顾了相邻帧差法的运算量小，速度快的特点。

4.1.2 背景减除法

背景减除法是利用建立的背景参数模型来作为一个场景的代表<sup>[39]</sup>，使用当前帧与背景模型做差分来实现运动目标的检测。同样是做差分运算，但是背景减除法与帧差法不同的是做帧差的对象是当前帧和背景模型。背景模型大多是根据统计学原理依据场景中像素的统计信息进行背景建模，因此，大多数的背景减除法求运动目标比较容易受光照和抖动等干扰的影响，但是，由于背景减除法建立背景模型的时候大都采用统计学的方法对背景进行建模，分割运动目标的时候会得到更清晰的轮廓信息。

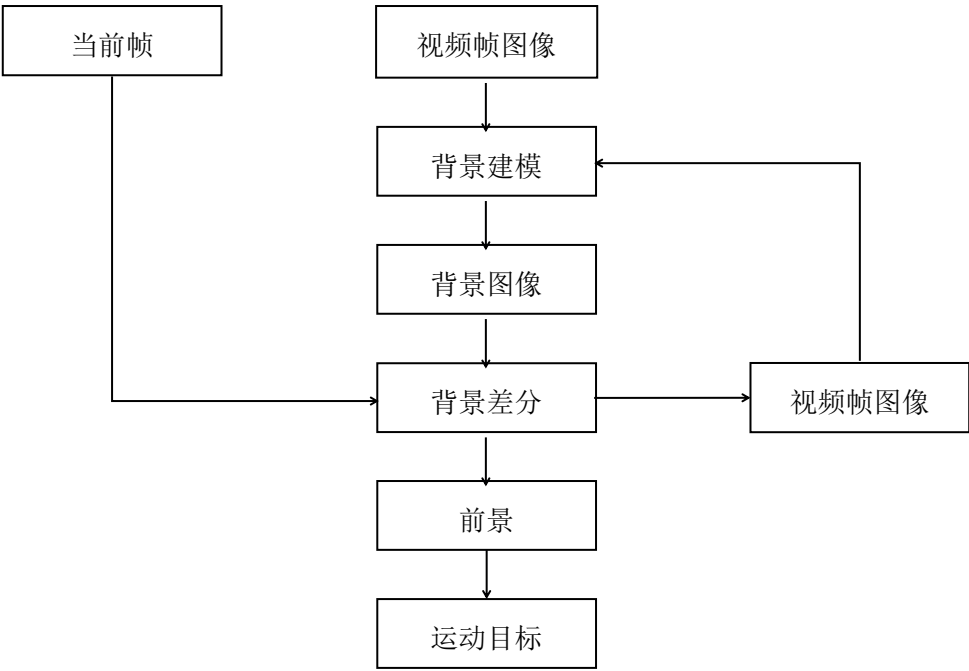


图 4-3 背景减除法示意图

背景减除法进行运动目标检测如图 4-3 所示，首先要选择合适的模型进行背景建模，然后根据背景建模模型从视频帧序列中得到背景模型，再将当前需要被检测的帧和背景模型帧进行差分，经过多次的迭代区分过程，可以得到前景也就是运动目标和背景，背景也将加入背景建模的过程实现对背景模型的维护，实现动

态的稳定模型输出，这样动态环境中也可以比较稳定的区分出运动目标比较清晰的轮廓信息。下面将介绍典型的背景差分方法及其改进方法。

### (1) 混合高斯背景建模

混合高斯背景建模是由 Stauffer 等<sup>[40]</sup>在高斯背景建模的基础上发展而来的一种优秀的背景建模方法。作为最常用背景减除时求取运动目标的背景建模方法，有着出色的环境适应性和通用性，它弥补了单高斯模型中对抖动的应对表现不足的情况，能更清晰的收敛出运动目标的轮廓信息。

一般混合高斯背景建模算法为图像序列中的每帧图像的每一个像素点定义了  $K$  个状态， $K$  的取值范围通常 3~5。假设每个像素位置上的像素值使用  $X_t$  表示，那么这个像素值的概率密度函数可以使用  $K$  个高斯函数来表示：

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \times \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (4-2)$$

式(4-2)中  $\omega_{i,t}$  为模型在  $t$  时第  $i$  个高斯分布的权值， $\mu_{i,t}$  为模型在  $t$  时分第  $i$  个高斯分布的均值矩阵， $\Sigma_{i,t}$  为其在  $t$  时第  $i$  个高斯分布的协方差估计矩阵， $\eta(\omega_{i,t}, \mu_{i,t}, \Sigma_{i,t})$  是时间为  $t$  的时候序号为  $i$  的高斯分布用来计算概率密度的方法，所以：

$$\begin{cases} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{i,t}|^{\frac{1}{2}}} * \exp\left[-\frac{1}{2}(X_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t - \mu_{i,t})\right] \\ \Sigma_{i,t} = \sigma_{i,t}^2 I \end{cases} \quad (4-3)$$

式(4-3)中  $n$  代表图中的颜色通道数值，当  $n > 1$  时当作每个通道之间不受彼此影响的同时拥有有一样的方差  $\sigma_{i,t}^2$ 。

如果采集的  $X_t$  匹配  $|X_t^c - \mu_{i,t-1}^c| < 2.5 * \sigma_{i,t-1}$ ，那么  $X_t$  被称为与满足第  $i$  个高斯分布，然后对  $X_t$  进行迭代，迭代结果如式(4-4)：

$$\begin{cases} \omega_t = (1 - \alpha)\omega_t + \alpha \\ \mu_t = (1 - \alpha)\mu_{t-1} + \rho X_t \\ \sigma_{i,t}^2 = (1 - \rho)\sigma_{i,t}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \end{cases} \quad (4-4)$$

式(4-5)中  $\rho$  是模型的学习速率，也被称作收敛速率、学习因子，可以表示为：

$$\rho = \alpha * \frac{1}{(2\pi)^{\frac{n}{2}}} * e^{\left[-\frac{1}{2}(X_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t - \mu_{i,t})\right]} \quad (4-5)$$

如果  $X_t$  在场景中没有实现收敛，那么这个模型和所有队列中的模型进行对比，替代掉相似性较小的模型或则直接插入。实现以上操作后，以  $\beta$  为依据将高斯分



布按照降幂来重新插入，按照式(4-6)将前  $B$  个模型作为满足条件的  $X_t$  作为背景。

$$B = \arg_b \min \left\{ \sum_{k=1}^b \omega_{k,t} > T \right\} \quad (4-6)$$

式(4-6)中  $T$  为划分背景模型的相似程度的权重值。

混合高斯背景建模虽然有比较好的背景描述能力，但是由于要对每一帧的每个像素进行混合高斯建模，运算量相当之大，尤其是在图片的彩色模式下。因此，这样明显的缺陷也导致了混合高斯背景建模的实时性相当低，如果需要在监控中实时的进行运动目标检测，那么就必须在背景建模的过程中提升背景建模的效率，减少背景建模中的运算量。

## (2) 改进的混合高斯背景建模

基于混合高斯背景建模的优秀的背景区分能力，很多学者对混合高斯建模算法进行了改进，但大多数学者的改进都是以提升建模后背景的稳定性和算法精度上。针对本课题需要选用算法精度达到一定程度并且运算速率较高，满足本系统的实时检测运动目标的实时性要求。

目前，针对减少建模运算量的改进也出现了很多很不错的算法，如黄玉等<sup>[42]</sup>将 YCbCr 空间用于混合高斯背景建模方法，白向峰等人<sup>[41]</sup>提出了参考偏差均值作为混合高斯背景建模中前景参数收敛的方法，这两种方法均在一定程度上减少了混合高斯背景建模的运算量，但是建模的过程中仍需对每一帧中的转换到更小的域中的每个像素点进行混合高斯建模，运算量仍然比较大。在查阅文献和比较算法的优劣后，研究时发现董俊宁<sup>[43]</sup>提出的基于空间约束的混合高斯背景建模方法在建模的过程中减少参与运算的帧数的过程能够极大的减少建模中的运算量，比较适合本课题的实时检测要求。

基于空间约束的混合背景建模采用了两种背景模型辅助建立混合高斯背景模型的方法，将首先得到的一块可疑区域作为一种空间上的约束，使得每一帧在进行第二重混合高斯背景建模的时候进行的混合高斯建模只在第一重建模的空间约束的基础下进行建模，这样会减少每帧中混合高斯建模的像素的个数，极大的减少运算量，提升了背景收敛的速率和对运动目标发现的敏感性。

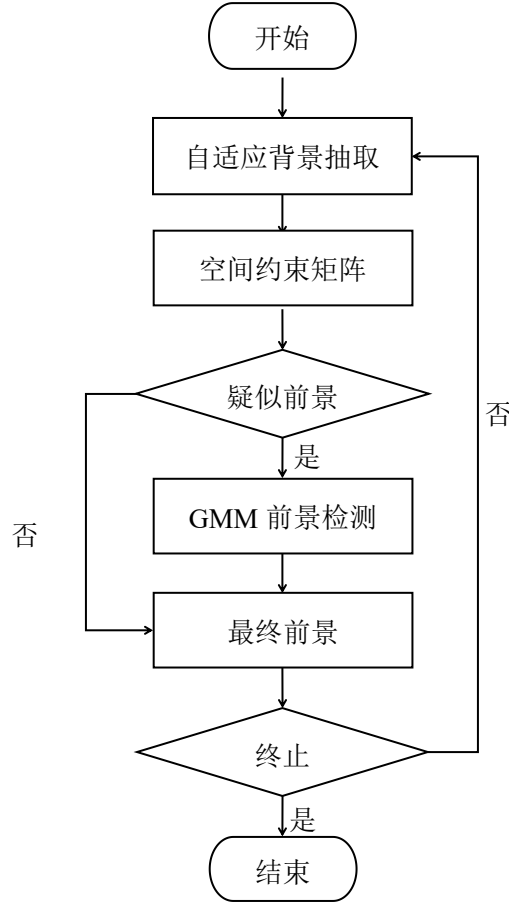


图 4-4 基于空间约束的混合高斯背景建模流程图

运动目标检测示意如图 4-4 所示，模型通过自适应背景抽取和背景减除的方式对背景第一次建模，得到前景的空间约束矩阵  $M$ ，然后模型只对空间矩阵中疑似前景的部分进行混合高斯背景建模更好的描述出前景也就是运动目标的信息。

基于空间约束的混合高斯背景建模增加收敛速度的关键在于通过提前运算得到了第一重背景，在第一重背景的基础上通过运算得到空间约束矩阵，根据空间约束矩阵约束做混合高斯运算的范围才大大的减小了运算量。那么空间约束矩阵的求取就是整个算法的关键。

图像  $M$  中任意点  $(x, y)$  可以使用式(4-7)中的数学模型表示：

$$M_{x,y} = \begin{cases} 0 & , |F_{x,y} - Z_{x,y}| < TH_{x,y} \\ 1 & , else \end{cases} \quad (4-7)$$

如果  $M_{x,y}=1$  那么点  $(x, y)$  代表这一时刻存在物体可能在背景上方； $M_{x,y}=0$  时，点  $(x, y)$  代表了这个位置大概率是不存在目标的。这种方法是基于传统的背景减法进行检测，但是在其基础上对阈值进行了自动化的更新，这种改进的好处是能够

更好的适应背景的变化，运用场景更灵活。

假设最终得到的运动目标检测结果帧是  $G$ ，那么对  $G$  帧图像中的任意像素点  $(x,y)$ ，其第  $t$  帧对应的运算可以描述为式(4-8)：

$$G_{t,x,y} = \begin{cases} 0, M_{t,x,y} = 0 \cup [\exists i \in [1, B] | |X_{t,x,y}^C - \mu_{i,t-1,x,y}^C| < 2.5\sigma_{i,t-1}] \\ 2^d - 1, \text{其他} \end{cases} \quad (4-8)$$

模型由于先进行简单的帧差运算，简单的固定背景减除法等方法先把疑似前景进行空间约束，这些运算相对混合高斯模型运算相当的小，随后进行的完整混合高斯建模运算也是相对完整的混合高斯运算只有少数空间约束矩阵中的部分像素。因此，基于空间约束的混合高斯建模算法在速度上较混合高斯背景建模算法提升了四倍以上的速度，精度也较混合高斯背景建模略有提高，在满足准确检测目标的同时大大的提升了运动目标检测的实时性。

## 4.2 运动目标检测应用及分析

本节将介绍基于空间约束的混合高斯背景建模方法在本系统中的应用以及效果分析。为了满足本课题的实时性要求既要快速又要稳定的完成背景建模任务，使用改进的背景建模算法针对本课题大场景对场景建立稳定的背景模型。根据选定的背景建模方法，对录制的场景中的录像信息进行背景建模的实验和效果分析，属于验证性实验不与其他算法进行横向比较。

本课题的场景选在本校主教学楼楼下开阔地带，三条路口的交汇点，由于场景比较开阔，地形比较复杂。场景中如果需要监控整个区域的目标的话就会丢失对目标监控清晰度，如果需要对目标拍摄高清信息，那么就需要加大摄像机的高额投入或者丢失部分监控区域的监控信息。



图 4-5 监控场景示意图

如图 4-5 所示监控场景区域，其中黑色框里的是运动目标，本文中对运动目标检测和跟踪的相关实验和分析都是基于图中的场景来进行的。

根据选取的基于空间约束的混合高斯背景建模方法来对场景进行建模，其基础是根据混合高斯的模型来对场景进行建模，但是在建模前先进行疑似的前景检测。疑似前景检测的方法是先对相邻的两帧进行差分来求出一个约束矩阵，这个约束矩阵包含了图像中疑似前景的目标和前景目标。

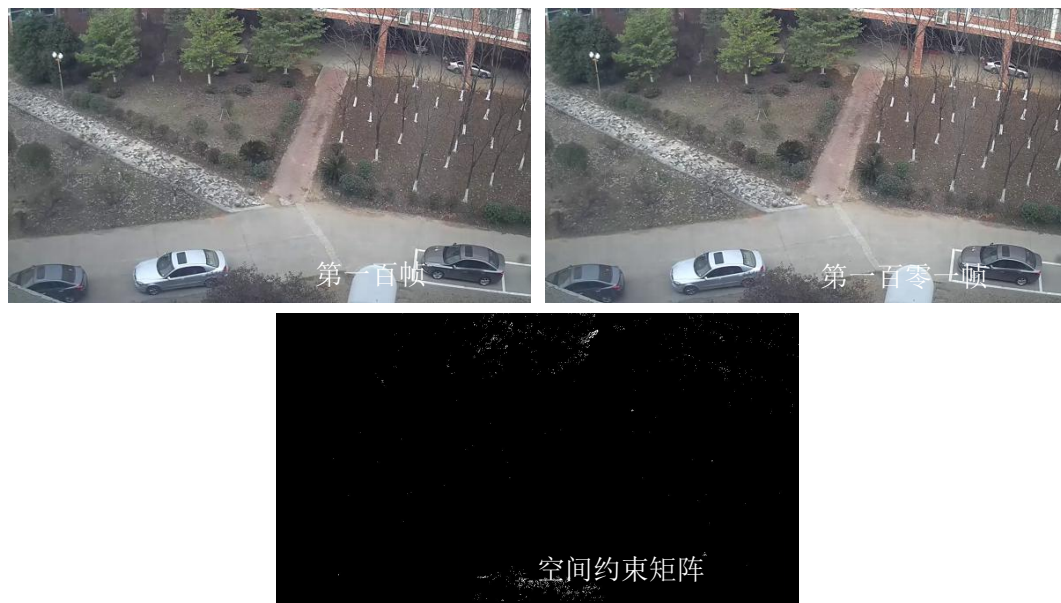


图 4-6 相邻帧间空间约束矩阵示意图

如图 4-6 所示，对采集的录像中的第一百帧和一百零一帧进行自适应阈值的差分算法，求得疑似前景的空间约束矩阵，从图像中可以看出，由于风的作用树叶存在的抖动，这些动作也被识别为疑似前景区域。



图 4-7 带有运动目标的空间约束矩阵示意图

如图 4-7 所示,在视频中的第三百帧和第三百零一帧得到图像的空间约束矩阵中可以明显的发现有运动目标的位置有明显的连通区域。将连通区域连带树叶抖动产生的干扰一同作为疑似前景的区域生成一个只包含白色目标区域的空间约束矩阵,作为场景进行混合高斯建模的一个空间约束,用来约束参与混合高斯建模时需要建模的像素点的数目。

在得到当前帧的空间约束矩阵后,针对每一帧的空间约束矩阵中标注的疑似前景的位置首先进行形态学处理滤除一些抖动和光照的影响,再对空间约束过的场景进行混合高斯背景建模。

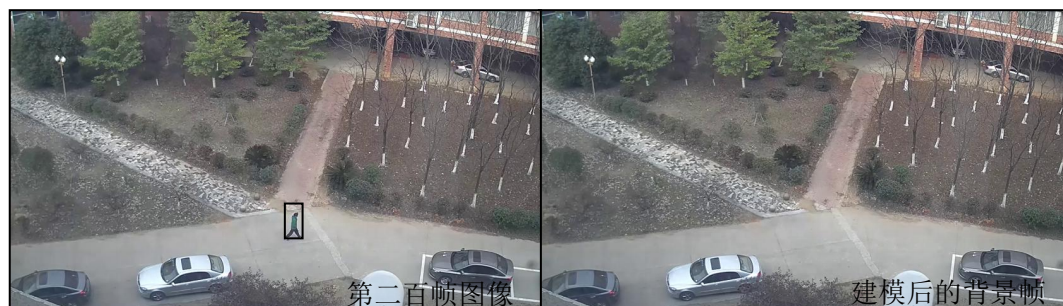


图 4-8 背景建模后得到的背景帧

如图 4-8 所示,根据对视频录像中的第二百帧及之前的图像进行连续的背景建模,得到稳定的能够描述场景中背景信息的背景帧,该图像除了少部分区域由于受到光照和抖动的影响存在一些背景偏差之外,大部分区域得到的背景是稳定的。



根据得到的背景帧对场景中存在的运动目标进行检测并将检测的信息更新到背景模型中。

根据背景建模分离出的背景帧作为背景减除法检测前景目标的背景模型，对需要进行运动目标检测的图像做背景差分法来获取运动目标的轮廓信息，获取的轮廓信息进行形态学处理后即得到前景目标区域，通过在待检测的图像中的前景区域进行提取，即可分离出前景目标。

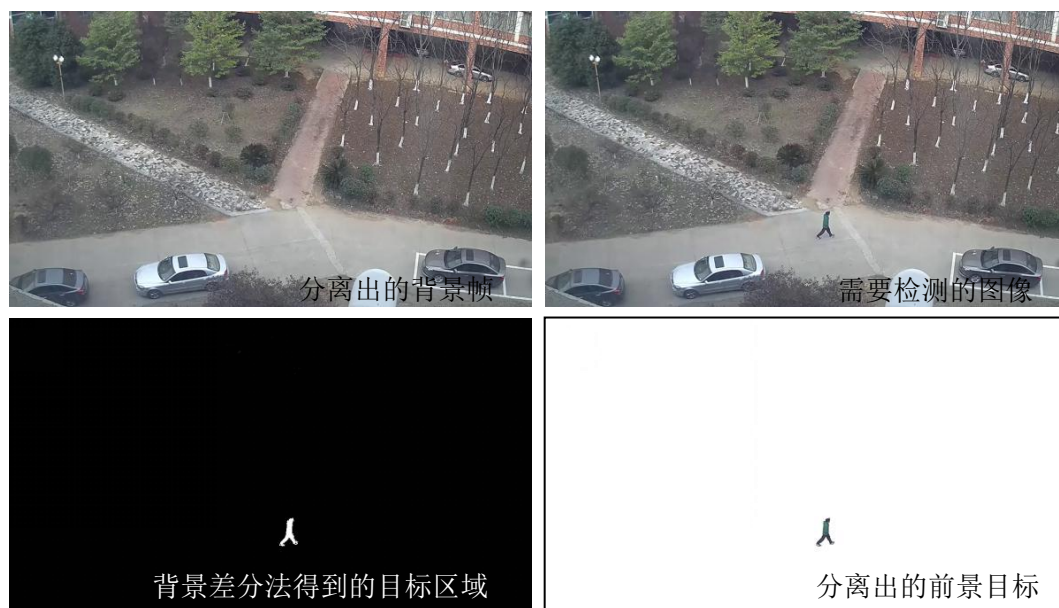


图 4-9 背景差分法分离前景和背景

如图 4-9 所示基于空间约束的混合高斯背景建模的方法对背景进行建模得到分离出的背景帧，除了树叶部分存在少许模糊，基本还算稳定。然后，对需要检测的图像进行背景差分法检测运动目标，途中显示的目标区域是进行差分法后得到的目标区域进行形态学处理后的图像，过滤了树叶的抖动和部分小的干扰。再根据得到的运动目标空间信息对需要检测的图像进行提取得到前景目标信息，也就是要检测的运动目标。

### 4.3 基于 Camshift 的运动目标跟踪

本节将介绍本课题用到的运动目标跟踪技术，运动目标跟踪效果的好坏直接影响到后续对视频的语义识别等基于运动目标跟踪的处理。由于跟踪目标和环境的多样性，各种目标跟踪技术也层出不穷，基于不同的场景和要跟随的目标不同的目标跟踪技术所展示出的效果也大相径庭。经过研究和参考各种类型跟踪算法的应用场景后，选择 Camshift 算法运动目标跟踪方法来实现本课题在目标跟踪上

的需求。

Camshift 算法的提出是 Bradski 针对 Meanshift 算法缺乏模型更新以及固定的和函数窗宽导致的对尺度变化目标跟踪定位失败的问题进行的改进。该算法在 Meanshift 算法的基础上提出了可自动适应的跟踪窗口大小，有效的解决了移动目标尺度变化带来的准确率的问题。

### 4.3.1 Meanshift 算法原理分析

Meanshift 这一概念起初由 Fukunaga 等提出的<sup>[44]</sup>，其只是单纯的对一个偏移的均值向量进行描述。随后，有人将该算法的核心理论运用迭代的方式动态的对偏移向量进行迭代，在不断改进的过程中 1995 年 Cheng Yizong<sup>[45]</sup>引入了核函数的概念设置偏移量的权值系数来达到使该方法具有更广泛的适用的目的，在经过诸多研究者的改进和扩充，Comaniciu 等人将该算法成功的运用于运动目标跟踪。因 Meanshift 算法在运动跟踪时无需参数，匹配目标实时性高，在匹配跟踪非刚性目标时候效果显著的特点，所以被广泛应用于目标跟踪领域。

Meanshift 算法的基础是概率密度偏移模型，其基本模型也被称为概率密度的偏移向量。

设  $x_i(i=1, \dots, n)$  是属于  $d$  维空间的  $R^d$  中的样本点，有  $n$  个这样的样本点，那么  $x$  使用 meanshift 模型描述的基本表达式为：

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (4-9)$$

在公式(4-9)中， $k$  代表球形区域内包含的采样点的数目， $k \leq n$ 。式中的  $(x_i - x)$  代表样本点中  $x_i$  相对于基准点  $x$  的偏移向量。均值偏移向量  $t$  实际上是包含在范围内靠近中心的  $k$  个点相对于点  $x$  的偏移先求和再求平均的结果。其结果证明概率梯度和方向上点的个数会相对更加集中，即概率密度便宜方向和概率梯度方向存在一致性。

根据以上的算法模型描述，如果使用该模型进行运动目标跟踪的话可以进行如下几步：

- (1) 目标模型建立，即在目标区域使用 Meanshift 模型对该区域所有像素的颜色直方图统计，进行核密度估计，采用不同的核函数得到的核密度也不尽相同。
- (2) 候选目标表示，在候选目标区域中也使用 Meanshift 模型对候选区域所有像素进行颜色直方图的统计，估算核密度。
- (3) 相似性度量，相似性度量的方法有很多，一般的方法是在获取了目标模型

的核密度和候选目标的核密度计算两个核密度的距离，其中将核密度也称为偏移向量，计算两个向量之间夹角的余弦，其值在 $[0,1]$ 之间，越接近 1 说明候选区域和目标模型越相似。

(4) 目标的定位，目标的定位首先是依据目标的 **Meanshift** 模型在目标的质心的邻域进行同等区域的核密度估计，得到以邻域为中心点的所有区域的核密度值和目标核密度进行相似性度量，使得新的候选区域朝核密度相似度更高的区域偏移，最终得到的新的目标区域与目标模型达到一定的相似性，终止在图像中的搜索。

### 4.3.2 Camshift 算法搜索过程描述

**Camshift** 算法也被称为连续自适应的均值偏移算法，虽然该算法以 **Meanshift** 算法为核心，但是与其不一样的是在该算法中可以针对动态的概率分布图，自适应地根据目标的大小变化调节跟踪窗口的范围。在目标尺度存在变化的场景中，**Camshift** 算法的精度是明显优于 **Meanshift** 算法的。连续自适应均值偏移算法的流程如描述：



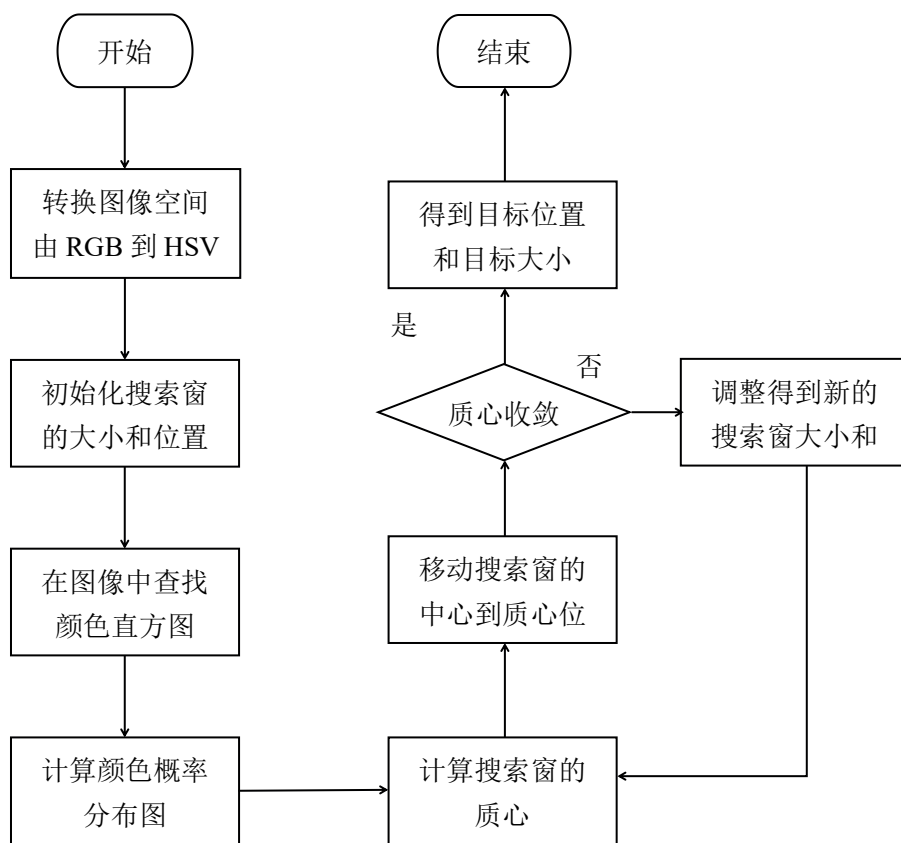


图 4-10 Camshift 算法流程图

如图 4-10 所示，第一步把图像从 RGB 空间映射到 HSV 空间，提取出要跟踪的目标所在区域 H 分量的直方图。再在后续帧中 HSV 图像中搜索与提取出的目标的 H 分量直方图相似的区域，并计算该区域的质心，如果质心位置和中心位置距离过大不满足迭代终止条件或者超过了迭代次数则重新调整搜索窗的大小并搜索新的位置，知道满足迭代要求。

这样的方法虽然能够很准确的对目标进行持续跟踪，但是也存在一些问题需要解决才能满足实际中的应用。比如说 Camshift 在跟踪的过程中如果像素过高、目标过小或者目标速度达到一个比较高的情况那么根据 Camshift 算法将存在一定的延迟，对应用中实时性有一定要求的场景将存在一定的制约。

### 4.3.3 改进的 Camshift 算法描述

根据 Camshift 算法应用于目标跟踪场景中存在的一些优缺点进行总结，选取了改进后的 Camshift 算法<sup>[46]</sup>以满足本课题在实时性方面的要求。运用的改进后的方法在保持跟踪效率的同时优化 Camshift 算法的实时性，原理是使用了空间约束

的方法减少了 Camshift 算法在匹配目标的时候产生的大量运算。

具体的跟踪流程如图 4-11 所示：

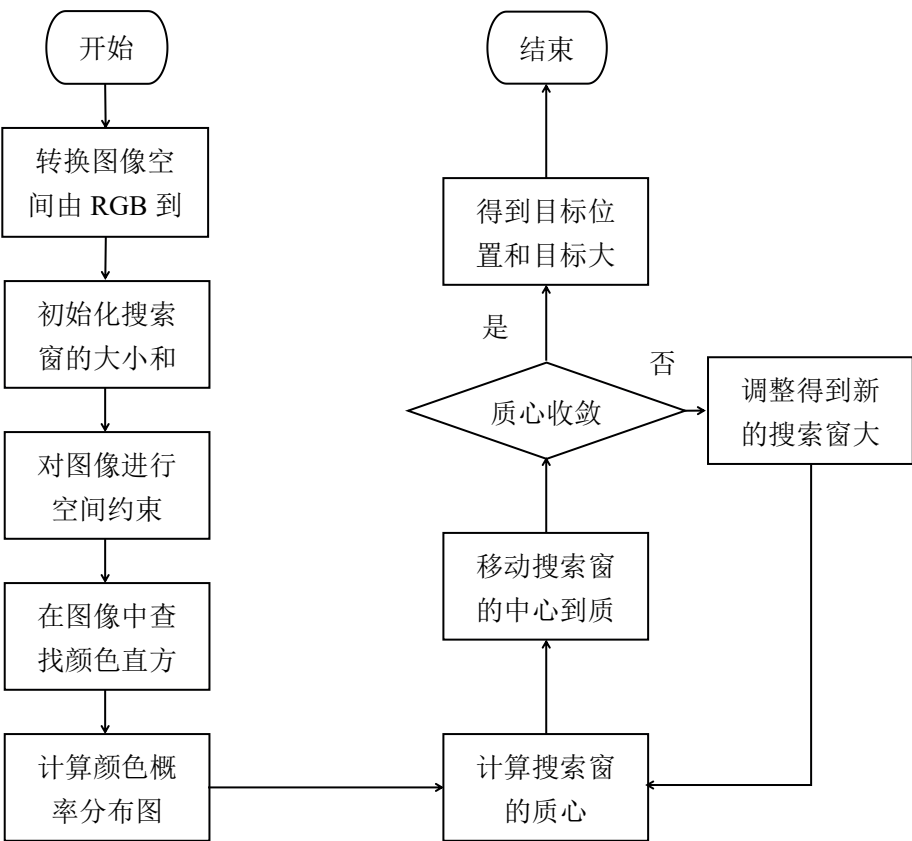


图 4-11 改进的 camshift 算法流程图

如图 4-11 所示，在 Camshift 算法进行目标检索之前对搜索区域进行一定的空间约束，使搜索区域有一定的方向性和约束性，以此来减少搜索中的计算量，达到更小的延迟和更高的实时效果。

4.4 运动目标跟踪应用及分析

本文的运动目标跟踪算法选择的是改进的 Camshift 算法，针对运动目标检测的应用和分析是基于场景中的监控录像以及第四章提及的运动目标检测方法提取的运动目标。本节对跟踪算法的应用和分析的主要目标是验证 Camshift 算法是否能够适应课题在准确性和实时性上的需求。

本节的应用验证环节的视频内容仍然是选在本校的主教学楼下方的开阔地带，由于存在三条路交汇，这块区域是两栋教学楼之间的主要通过区域，也是教

学楼两侧的通过区域，地域比较开阔，干扰较少，适合用于本章对于 Camshift 算法进行识别和跟踪的验证和分析。

在本节的实验和应用的过程中，首先利用稳定的背景区域提取发现运动目标，并通过第四章的方法将目标的运动区域求解出来，然后将分离出的运动目标作为跟踪算法的匹配目标进行运动跟踪。



图 4-12 提取场景中检测到的运动目标

如图 4-12 将场景中检测到的运动目标提取出来，用作目标跟踪算法进行跟踪，并且求解出目标轮廓的质心，方便通过质心标注出目标在场景中运动的轨迹。

如图 4-13 所示，首先将带有运动目标的图像从 RGB 空间转换到 HSV 空间，再计算 RGB 空间中的颜色概率分布，并以概率分布作为匹配目标的一个标准在目标区域的一个邻域内进行概率相似性匹配，直至在搜索区域内找到目标。根据概率密度求解出目标的质心，质心代表了目标在颜色概率分布中的形态信息，根据其质心描述的目标的运动轨迹，并将带有目标轨迹的运动信息作为视频摘要的一部分内容。

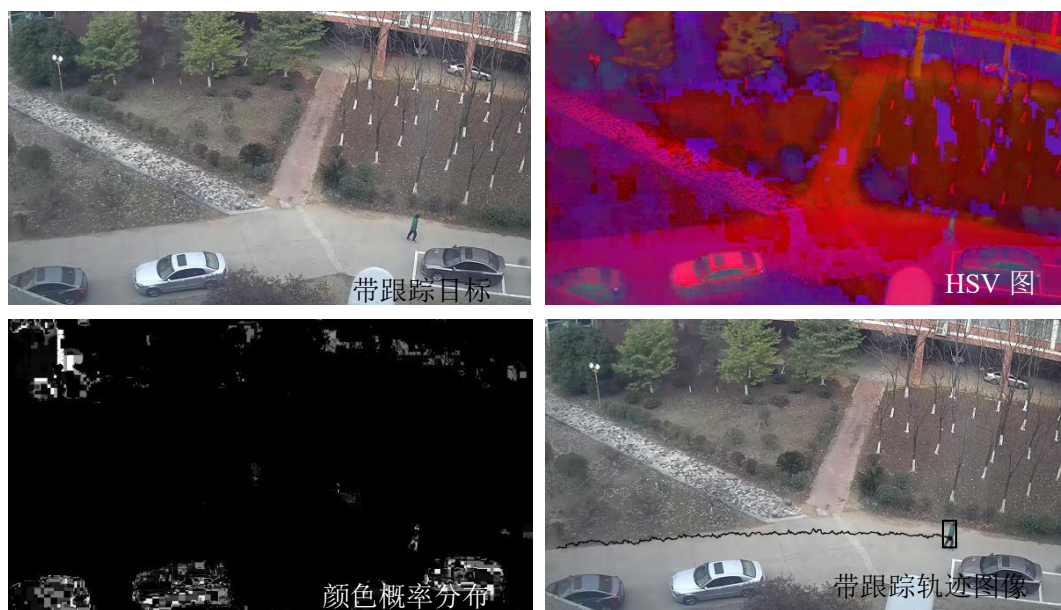


图 4-13 运动目标跟踪及轨迹生成

从图 4-13 的实验结果可以看到，使用改进的 Camshift 算法可以有效的稳定的对检测到的运动目标进行跟踪，基本上满足了课题在设计运动跟踪部分时提出的技术要求。它能够比较快速的对给定的目标进行识别，也能对目标进行比较稳定的进行跟踪，也可以对目标在场景中的运动轨迹进行比较精确的勾勒，在实现了以上几点以后，运动跟踪部分就能够胜任完整准确的记录运动目标在场景中的运动信息任务。

## 4.5 本章小结

本章介绍了是如何利用现有的算法进行改进来适应本课题对运动目标检测和跟踪的需求。本文使用的运动目标检测算法主要是依据帧差法作为主要的方法来提取目标的轮廓信息和定位疑似目标的空间信息，其次依托混合高斯背景建模的方法对场景进行合理的背景建模，依据基于空间约束的混合高斯背景建模的方法提高背景建模的效率，完成了满足课题需求的运动检测模块。参考目标检测应用时的情况来看，空间约束时的疑似前景检测时对光照和抖动来说相对比较敏感，如果抖动过大对形态学处理的要求比较高，那么实时性也难以得到保证，在相对稳定的场景中运用这种改进的目标检测方法还是相当有效果的。根据第二章中设计目标本文选取了现在应用比较广泛的 Camshift 运动跟踪算法进行运动目标跟踪，实践应用中该算法表现基本满足系统对目标跟踪监控的需求，能够在可接受的间

内锁定选定的目标，对目标的质心计算比较可靠，能够稳定的对目标进行跟踪监控，并且能够准确的在场景中完成对目标运动轨迹的勾勒。完成对目标个跟踪之后将目标在场景中的活动信息，运动轨迹等关键信息保存起来作为视频摘要中重要的一部分。

## 第五章 双摄像机协同工作及摘要系统实现

本章在其他章节的基础上进行整体的系统集成和双摄像机协同工作介绍，各模块的设计和实践证明，根据设定的各种技术指标各模块都能很好的满足要求。因此，只需要在各模块的基础上进行集成和完善，就能实现稳定的准确的带有目标高清信息的视频摘要系统。

### 5.1 双摄像机协同工作

本节将从摄像机的分工开始介绍双摄像机的分工原理，两个摄像机分工的基础以及分工的目的。两个摄像机分工后如何依据各自的优势来进行协同工作，使得两个摄像机取长补短，完成一个摄像机无法完成的任务。完成两个摄像机的协同工作进行系统的整合，使之能够形成能够稳定采集带有高清图像信息的视频摘要系统。

#### 5.1.1 双摄像机分工

本课题根据两个摄像机的功能和作用的不同将其分为场景监控摄像机和目标抓拍摄像机。其中场景监控摄像机的作用是对固定的场景进行整体监控，检测到场景中的运动目标并且对检测的目标进行运动跟踪，根据目标在场景中的活动情况生成目标的位置信息，轨迹信息等作为视频摘要的一部分；目标抓拍相机，根据场景监控摄像机对检测到的目标进行抓拍传回的的目标的位置信息，通过控制云台的方式进行高清图像信息抓拍，将抓拍的信息作为视频摘要的一部分。

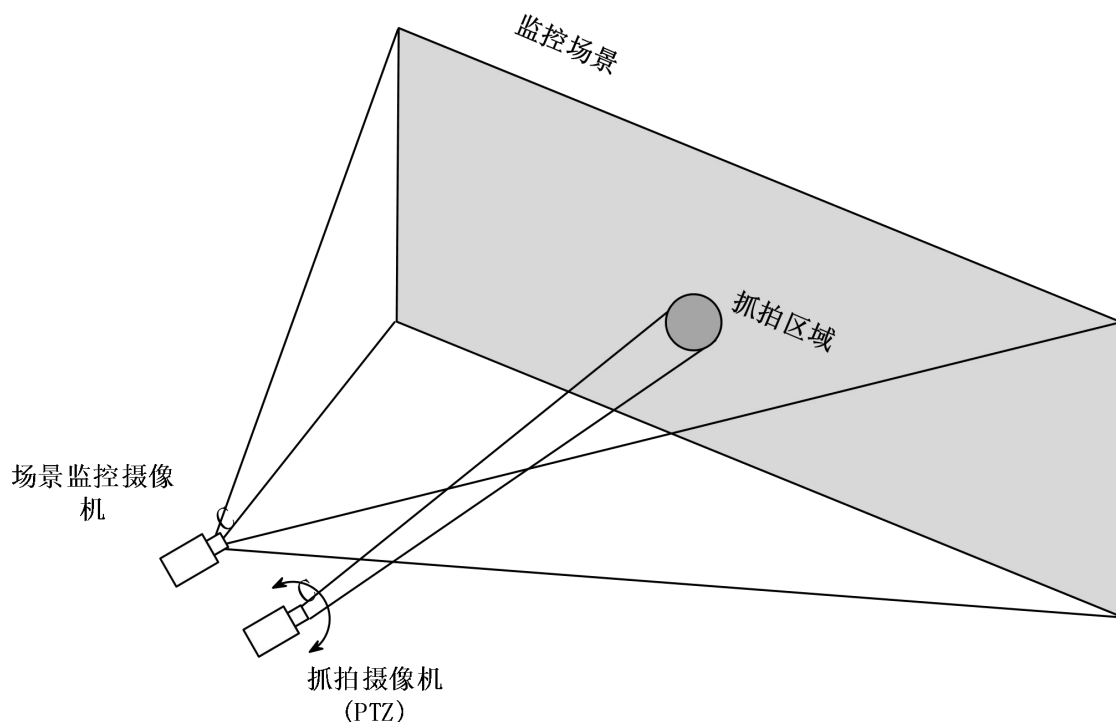


图 5-1 双摄像机分工示意图

如图 5-1 所示，两个摄像机针对监控区域进行协同监控，但是在监控区域上有着明显的区别。场景监控摄像机长期对监控区域进行监控，实时地采集监控区域中的各种宏观信息，当发现有价值目标的时候，参考双摄像机协同工作策略控制目标抓拍相机对目标进行调焦抓拍，其监控区域基本只有监控区域的一部分。

### 5.1.2 双摄像机协同工作策略

根据两个摄像机的作用不同、分工不同，那么两个摄像机的工作方式也就不同。作为一个完整的视频摘要系统，两个摄像机如何有机的结合在一起，进行明确分工并且统一有效的协作运行作为本节要介绍的内容。

双摄像机协同工作需要制定相应的协同工作策略，即如何将场景监控摄像机得到的目标的位置共享到目标抓拍摄像机的坐标系中，如何对两个摄像机共有的坐标系中的目标进行跟踪抓拍。

两个摄像机进行共享坐标通过双摄像机标定的方式实现两个摄像机对同一片区域拥有统一的坐标系。这样就可以只需要通过场景监控摄像机对目标进行检测和跟踪就可以实现目标抓拍摄像机也能实时获得目标的坐标信息，再控制目标抓拍摄像机进行高清目标信息抓拍。



由于目标抓拍相机的机械特性所限，要想控制其进行目标高清信息拍摄时，必须控制云台进行旋转和调焦等操作，这些操作是需要消耗一部分时间的，那么就会导致目标抓拍时并不一定能够照顾到所有目标。因此，本文使用了一种制定优先级的策略来控制云台相机尽可能有效的对场景中的目标进行拍摄。本文制定优先级策略依据场景中目标的两个参数来制定，一个是目标的位置，一个是目标的速度：

(1) 对于目标的位置相对来说比较敏感，如果目标在图像的中心位置或者朝着中心位置移动时，那么就认为抓拍该目标的窗口时间会比较长，所以目标越靠近边缘，同时又没有被抓拍过就设其优先级越高，优先级与距离中心点的距离有关；

(2) 对于目标的速度来说，相同的位置，目标的速度越快那么留给系统进行抓拍的窗口期就越短，认为速度快的目标抓拍的优先级也就越高。

以上这两种优先级并非完全的竞争关系，将两种优先级综合起来比较每一个目标两种优先级加权后的总值进行比较，来确定目标综合优先级。

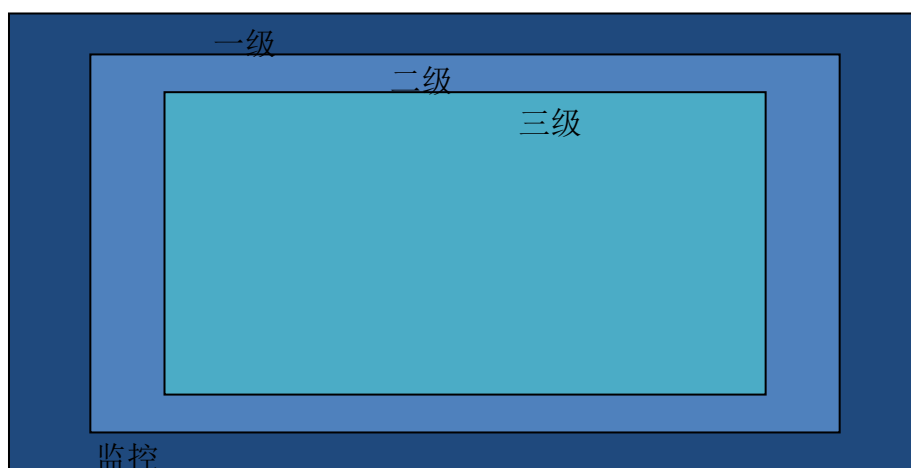


图 5-2 基于空间的优先级示意图

如图 5-2 所示，将监控场景分为三个部分，越靠外围优先级越高。基于速度的优先级设置就基于运动目标跟踪来测算目标的大约速度，将多个目标的速度进行比较进行分级，速度在 0~10 个像素/秒时设置其为三级速度优先级、10~50 像素/秒时设置其为二级速度优先级、速度在 50~100 个像素/秒时设置其为一级速度优先级。

根据抓拍的策略对目标的综合优先级进行排序，生成一个优先级由高到低的排序生成跟踪队列，根据 FIFO 的原则进行目标的抓拍。

最终的总优先级为场景优先级加上速度优先级进行比较，优先级相等的进行



单项比较，以速度优先级为主。

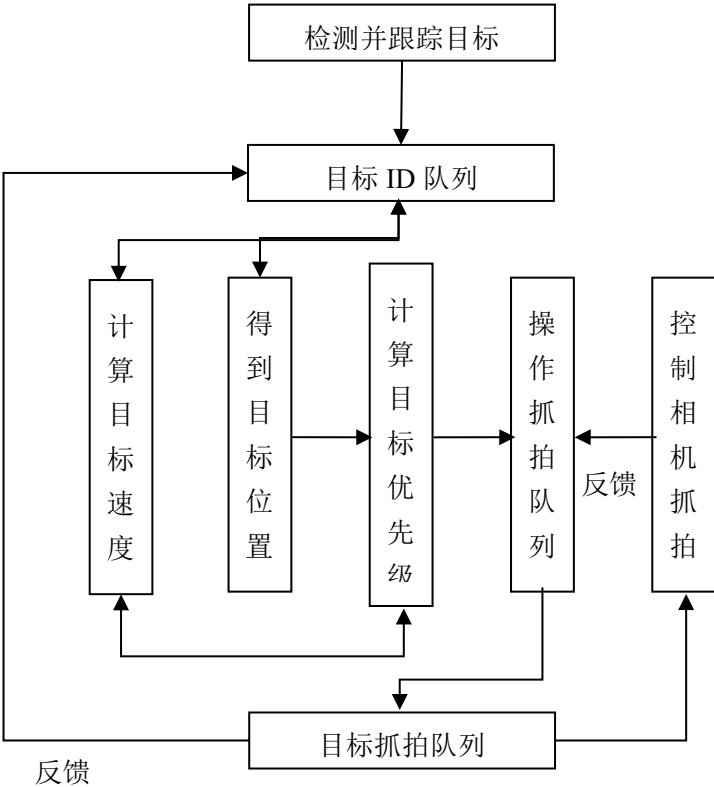


图 5-3 双摄像协同框图

如图 5-3 所示，从检测并跟踪到运动目标后就对目标设置唯一标识目标 ID，将目标的速度和位置信息以目标 ID 为标识将这两种属性实时进行同步。根据目标的速度和位置两个属性对目标 ID 进行优先级排序，排序后根据目标拍摄队列对控制目标抓拍相机进行抓拍目标，并将抓拍结果反馈后操作目标抓拍队列和目标 ID 序列，使抓拍成功的目标在 ID 上进行标记，未成功的继续留在抓拍队列，直到抓拍完成。

5.2 系统硬件及开发环境

5.2.1 系统硬件

系统硬件部分主要包含了三个部分：场景监控摄像机、目标抓拍摄像机、和 PC 计算机。本系统主要程序运行在 PC 计算机上，实现通过两个摄像机采集监控两个摄像机的图像，使用制定的两个摄像机协同工作策略通过摄像机网络接口实现对两个摄像机分别不同的控制和操作，实现双摄像机视频摘要系统。

表 5-1 主要硬件			
设备名称	型号说明	数量	用途
场景监控摄像机	海康卫视监控摄像机分辨率 1280*720	1	场景监控
目标抓拍摄像机	海康威视高清球形云台机分 辨率 1280*720	1	目标抓拍
PC 计算机	Windows10 操作系统 i7 处理 器 4GB 内存 PC	1	运行双摄像机视频摘 要主程序

如表 5-1 所示，场景监控摄像机和目标抓拍摄像机采集的图像分辨率都是 1280\*720 像素的，但是目标抓拍摄像机可以进行高倍调教并且通过云台控制可以实现水平 360 度旋转和垂直 90 度旋转，拍摄远处目标清晰，但是成本较高，场景监控摄像只能固定使用而且画质一般，但是价格低廉。这样高低搭配的两个摄像机既有目标抓拍相机的各种优势也补齐了成本上的短板。

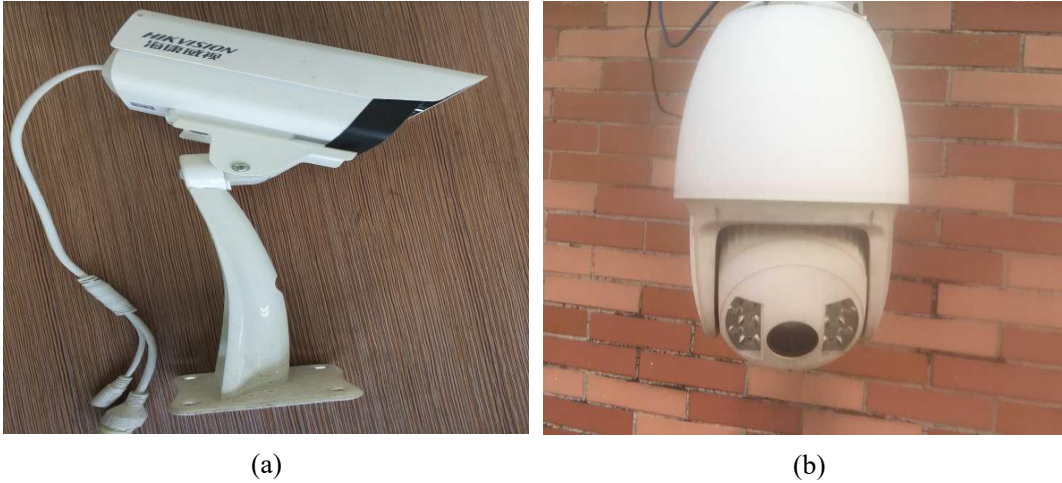


图 5-4 两台摄像机：

(a)为低分辨率的场景监控摄像机；(b)为高清云台目标抓拍摄像机

如图 5-4 所示，两台摄像机一个为固定的监控固定场景用的低分辨率摄像机，另一个为可以控制云台调整焦距的高清云台摄像机，由于两台摄像机功能差距非常大，所以两台摄像机的价格差距也比较大。

5.2.2 开发环境

本系统的是在微软的 windows10 操作系统中对软件进行设计开发的，所使用的软件版本清单如表 5-1 所示：

表 5-2 开发软件清单

软件名称	软件版本
Visual studio	Visual studio2010
MFC(Microsoft Foundation Classes)	MFC version10.0
Opencv	Opencv2.4.8
HCNETSDK(海康威视 SDK)	SDKv5.2.3.3

如表 5-2 所示，本课题使用的是 C++编程语言在 windows10 操作系统上使用微软的 C++ 编程平台 visual studio2010 开发平台进行程序的编写；用到了 visualstudio2010 中的 MFC 接口类进行软件界面的开发；开发的过程中用到了处理图像的一些方法源自于开源图像处理库 opencv 进行图像的操作；对摄像机的操作用到了海康威视提供的开放编程接口实现对两个网络摄像机的控制和操作。

5.3 系统运行效果展示

5.3.1 系统界面

本课题采用 MFC 的框架进行系统可视化控制界面的开发，具体形式如图 5-5：

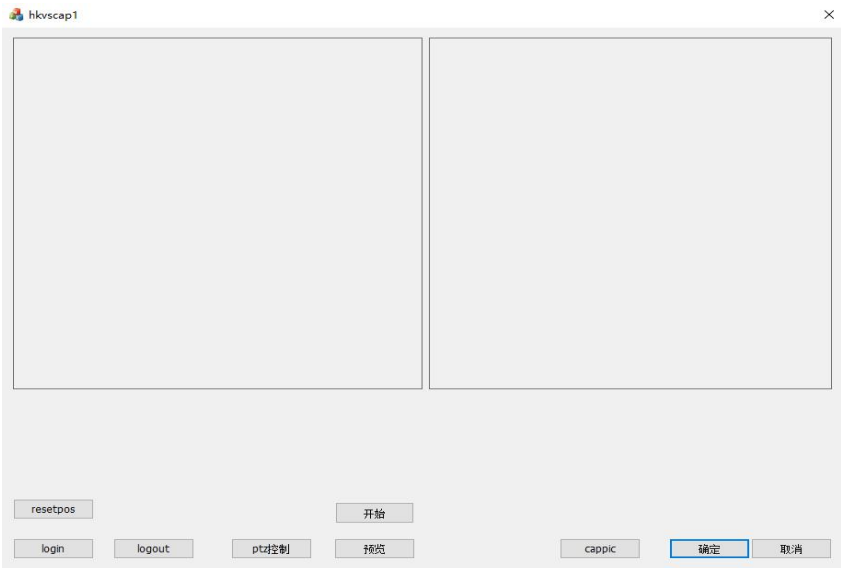


图 5-5 系统界面展示

如图 5-5 所示，本系统的界面包括了两个图像显示框用来预览两个相机采集的实时图像；设置了 login 和 logout 按钮进行两个网络摄像机的连接和断开连接的操作；预设了 PTZ 控制按钮来进行目标抓拍相机的云台控制测试；resetpos 按钮用来复位目标抓拍相机的位置；cappic 用来手工抓拍感兴趣的图片；开始按钮可以开始

视频摘要系统进行两个相机的协同工作生成具有高清信息的视频摘要。

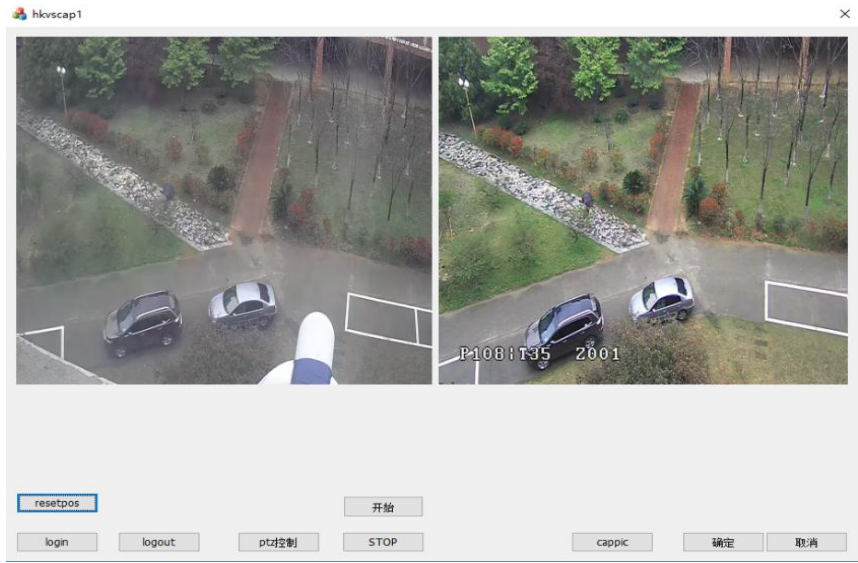


图 5-6 预览复位展示

如图 5-6 所示，在系统的预览界面可以在两个图像控件中预览两个相机同时采集的图像信息。点击 `resetpos` 按钮控制目标抓拍摄像机的云台运动到初始位置。

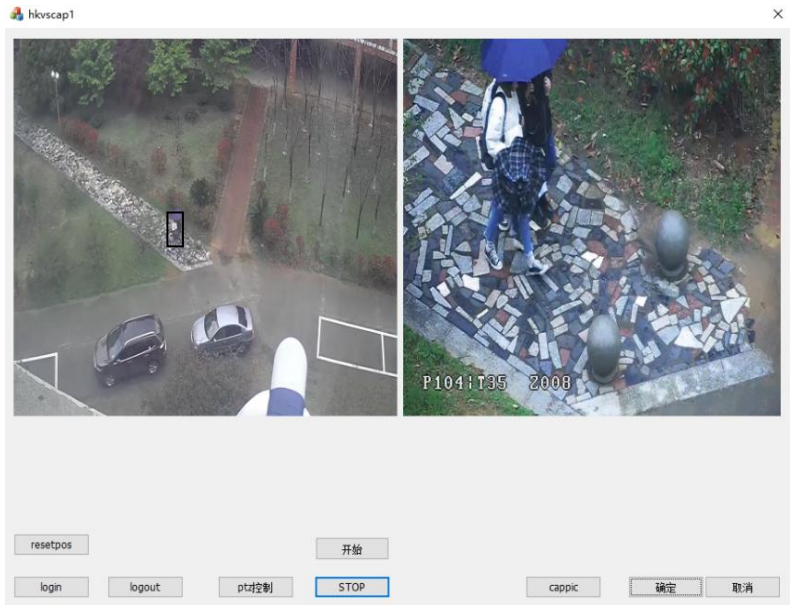


图 5-7 抓拍示意

如图 5-7 所示，系统在对场景进行监控的时候自动控制目标抓拍摄像机对场景中发现的运动目标进行高清抓拍。

5.3.2 视频摘要生成

根据现实场景中出现的运动目标进行检测，对检测到的目标使用第四章描述



的方法进行运动目标跟踪，并描绘运动目标的运动轨迹，测算目标的大约速度。



图 5-8 场景中多目标检测跟踪

如图 5-8 所示，场景监控相机在场景中检测到多个目标并对多个目标设置目标 id 根据双摄像机抓拍策略对多个目标的优先级进行计算，根据得到的不同的 id 对应的最终的优先级排序，从优先队列中选取首要抓拍目标进行抓拍高清信息。



图 5-9 抓拍效果；

(a)目标 id 为 001 的抓拍；(b)目标 id 为 002 的抓拍

如图 5-9 所示，对场景中跟踪到的目标进行抓拍的结果，(a)为目标 id 为 001 的目标抓拍，(b)为目标 id 为 002 的目标高清抓拍信息，和图 5-8 中场景监控相机中采集到的目标相比明显清晰度更好，基本可以辨认目标面部的细节，可以确认目标肤色、上衣、裤子和鞋子的颜色，得到目标的高清记录。

最终根据场景监控相机采集的目标在场景中的运动信息和记录，以及目标抓拍相机抓拍到的目标高清信息作为高清记录。将这些信息融合起来生成具有不同粒度的视频摘要。



图 5-10 内容丰富的视频摘要；

(a)目标场景信息；(b)存储视频信息；(c)-(d)目标高清信息

如图 5-10 所示，最终的视频摘要包括了场景中目标轨迹和运动信息、采集的场景中的视频信息以及目标的高清图像信息，形了可以查阅的多种形式的视频摘要。内容相对一般的视频摘要系统更为丰富，具有目标的更为高清的信息，细节更为清晰，基本上实现了本课题设计的目标。

## 5.4 本章小结

本章主要对双摄像机协同工作做了详细的介绍，对整个系统运行生成视频摘要，两个摄像机如何协同工作发挥出双摄像的优势。两个相机针对工作特点制了多个目标的抓拍策略，以及多个目标优先级的设置和根据优先级来确定最终的抓拍顺序，提高抓拍的效率。介绍整个系统的开发环境和和基础硬件信息，根据开发环境和硬件的特点对本文设计的系统进行修正，用合适的方法合理的完成双摄像机实时视频摘要系统的实现。本章也展示了根据设计的系统的各部分进行集成，设计了核心的控制部分双摄像机协同工作部分来控制两个相机抓拍场景中的目标，实现对场景检测到的目标进行跟踪采集和高清抓拍，将这些信息融合成有效的视频摘要。

## 第六章 总结与展望

### 6.1 总结

本课题研究的背景是基于目前大场景监控中存在的一些不足来解决其中存在的问题。本文选取了两个不同规格的摄像机来针对大场景监控时监控的范围过大导致监控的精度过低，查询的时候低分率的视频或者摘要信息并不能很好的反映出场景中监控的信息。针对以上问题和对应的考虑本文设计了基于双摄像机的实时视频摘要系统，各项工作如下：

(1) 对国内外视频摘要技术和双摄像机技术进行研究和分析，基于张正友教授的标定法对双摄像机标定进行研究和改进。设计和实现两个不同规格的摄像机进行有效的标定，确立出两个相机统一的参考坐标系来实现两个相机的位置共享，为两个相机协同工作打下统一协作的基础。

(2) 研究当下的目标检测算法并了解目标检测算法的特点，根据不同的算法的特点针对本课题的需求进行选择和改进。本文利用了空间约束的混合高斯背景建模的方法对场景进行背景建模，能够快速的稳定的对场景中运动的目标进行检测和发现。

(3) 对现有成熟的运动目标跟踪算法进行研究和学习，根据双摄像机实时视频摘要系统的需求，选取和改进满足本系统要求的运动目标跟踪算法。本文主要针对运动目标的跟踪算法进行实时性和准确性方面的优化，实现实时的对检测到的目标进行实时跟踪，这样实时的对跟踪到的运动目标进行抓拍打下了良好的基础。

(4) 根据两个摄像机的特点对两个摄像机分工和协同工作的特点，制定了合适的双摄像机协同工作方法，以及两个摄像机对目标的高清信息进行抓拍的策略，每个目标抓拍时优先级的制定。基本上能够合理的控制两个摄像机有效的对多个目标进行高清信息抓拍，并根据抓拍的信息和场景监控采集的信息，生成有效的视频摘要。

### 6.2 展望

本文设计的双摄像机实时视频摘要系统，虽然可以解决单摄像机在大场景监

控中摘要的清晰度的问题，但是也出现了双摄像机的一些缺陷，由于研究时间和本人能力所限，本课题还存在一些可以继续改进的地方：

本课题使用了双摄像机的形式进行场景监控目标抓拍，其有单摄像机不具有的灵活性和组合优势，但是其也存在控制上的难题。本系统不仅要场景监控还要对场景中出现的目标进行有效的抓拍，那么控制逻辑就显得十分重要，但是由于本人在控制方面只是的不足，导致控制目标进行抓拍时存在时效性不佳或者目标变多时，抓拍会存在漏抓的情况，这方面还有很大的改进空间。以及由于要用到目标抓拍相机对跟踪的目标进行抓拍来生成视频摘要，那么就要求必须在跟踪的同时进行抓拍，这就需要系统具有良好的实时性，但是云台摄像机的机械性能的限制也阻碍了系统对抓拍时的实时性，所以以后针对机械的实时性和控制逻辑的高效性都还有很大的改进空间。



## 参考文献

- [1] Bai T, Li Y F. Robust visual tracking with structured sparse representation appearance model. *Pattern Recognition*, 2012, 45 (6): 2390-2404
- [2] Z K, K M, J M. Tracking-Learning-Detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34 (7): 1409-22
- [3] Ponceleon D, Srinivasan S, Amir A, et al. Key to effective video retrieval: effective cataloging and browsing. in: *ACM International Conference on Multimedia '98*, Bristol, England, September. 1998: 99-107
- [4] Ajmal M, Ashraf M H, Shakir M, et al. Video Summarization: Techniques and Classification. in: *International Conference on Computer Vision and Graphics*. 2012: 1-13
- [5] Kettner V, Zabih R. Bayesian Multi-Camera Surveillance. in: *Computer Vision and Pattern Recognition*, 1999. *IEEE Computer Society Conference on*. 1999: 259 Vol. 2
- [6] Bano S, Cavallaro A. Discovery and organization of multi-camera user-generated videos of the same event. *Information Sciences*, 2015, 302 (C): 108-121
- [7] Ristani E, Solera F, Zou R, et al. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking. 2016: 17-35
- [8] Zhang K, Zhang L, Yang M H. Real-Time Compressive Tracking. in: *European Conference on Computer Vision*. 2012: 864-877
- [9] Kanade M A S T. Video Skimming and Characterization through the Combination of Image and Language Understanding. in: *IEEE International Workshop on Content-Based Access of Image and Video Database*, 1998. *Proceedings*. 1997: 61-70
- [10] Chang S F, Chen W, Meng H J, et al. VideoQ: an automated content based video search system using visual cues. in: *ACM International Conference on Multimedia*. 1997: 313-324
- [11] Ponceleon D, Amir A, Srinivasan S, et al. CueVideo: automated multimedia indexing and retrieval. in: *ACM International Conference on Multimedia '99*, Orlando, FL, USA, October 30 - November. 1999: 199
- [12] Rav-Acha A, Pritch Y, Peleg S. Making a Long Video Short: Dynamic Video Synopsis. in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2006: 435-441
- [13] Liu A, Yang Z. An interactive method for dynamic video synopsis generation. 2010, 1: V1-656 - V1-661

- [14] Zhao T, Aggarwal M, Kumar R, et al. Real-time wide area multi-camera stereo tracking. in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. 2005: 976-983 vol. 1
- [15] Tan M H, Ranganath S. Multi-camera people tracking using Bayesian networks. in: Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. 2003: 1335-1340 vol.3
- [16] Mittal A, Davis L S. M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene Using Region-Based Stereo. in: European Conference on Computer Vision. 2002: 18-36
- [17] Tyagi A, Keck M, Davis J W, et al. Kernel-Based 3D Tracking. in: Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on. 2007: 1-8
- [18] Khan S, Shah M. Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View. 2003, 25 (10): 1355-1360
- [19] 李鹏, 王军宁. 摄像机标定方法简述. 传感器世界, 2007, 13 (3): 21-24
- [20] Abdel-Aziz Y I. Direct Linear Transformation From Comparator Coordinates in Close-range Photogrammetry. in: ASP Symposium on Close-Range Photogrammetry. 1971:
- [21] Tsai R Y, Lenz R K. A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. IEEE Transactions on Robotics & Automation, 1989, 5 (3): 345-358
- [22] Maybank S J, Faugeras O D. A theory of self-calibration of a moving camera: Kluwer Academic Publishers, 1992. 123-151
- [23] Faugeras O D. What can be seen in three dimensions with an uncalibrated stereo rig? in: Proc. European Conference on Computer Vision, Santa Margherita, Italy. 1992: 563-578
- [24] Li M. Camera calibration of a head-eye system for active vision. in: European Conference on Computer Vision. 1994: 543-554
- [25] Basu A. Active calibration: alternative strategy and analysis. in: Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on. 1993: 495-500
- [26] 张晓明, 沈连娟, 赵高飞, 等. 基于平面模板的两步法标定摄像机的改进. 现代制造工程, 2008, 2008 (10): 9-12
- [27] Ma S D. A self-calibration technique for active vision systems. IEEE Trans Robotics Automat, 1996, 12 (1): 114-120
- [28] Wiersema K, Higgins A B, Covino S, et al. Calibration of EFOSC2 broadband linear imaging polarimetry. 2018:

- [29] Haneishi H, Miyake Y. Distortion compensation of electronic endoscope image. in: Nuclear Science Symposium and Medical Imaging Conference. 1994: 1717-1721 vol.3
- [30] Zhang Z. A Flexible New Technique for Camera Calibration. *Tpami*, 2000, 22 (11): 1330-1334
- [31] Zhang Z. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations. in: The Proceedings of the Seventh IEEE International Conference on Computer Vision. 2002: 666-673 vol.1
- [32] Cheng F H, Chen Y L. Real time multiple objects tracking and identification based on discrete wavelet transform. *Pattern Recognition*, 2006, 39 (6): 1126-1139
- [33] Rui Y, Huang T S, Ortega M, et al. Relevance feedback: a power tool for interactive content-based image retrieval. *Circuits & Systems for Video Technology IEEE Transactions on*, 1998, 8 (5): 644-655
- [34] Chien S Y, Ma S Y, Chen L G. Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits & Systems for Video Technology*, 2002, 12 (7): 577-586
- [35] Cheng Y H, Wang J. A Motion Image Detection Method Based on the Inter-Frame Difference Method. *Applied Mechanics & Materials*, 2014, 490-491: 1283-1286
- [36] Hsu Y Z, Nagel H H, Rekers G. New likelihood test methods for change detection in image sequences. *Computer Vision Graphics & Image Processing*, 1984, 26 (1): 73-106
- [37] Guo C. An Improved Algorithm Based on the Five Frame Difference Method for Dynamic Target Detection. *Journal of Chongqing University of Science & Technology*, 2015:
- [38] Qiu-Lin L I, Jia-Feng H E. Vehicles Detection Based on Three-frame-difference Method and Cross-entropy Threshold Method. *Computer Engineering*, 2011, 37 (4): 172-174
- [39] Chalidabhongse T H, Kim K, Harwood D, et al. A perturbation method for evaluating background subtraction algorithms. in: Joint IEEE International Workshop on Visual Surveillance and PERFORMANCE Evaluation of Tracking and Surveillance. 2003:
- [40] Stauffer C, Grimson W E L. Adaptive Background Mixture Models for Real-Time Tracking. *Proc Cvpr*, 1999, 2: 2246
- [41] 白向峰, 李艾华, 李喜来,等. 新型背景混合高斯模型. *中国图象图形学报*, 2011, 16 (6): 983-988
- [42] 黄玉, 殷荃茗, 周书仁. 基于 YCbCr 的自适应混合高斯模型背景建模. *计算机工程与科学*, 2015, 37 (1): 152-156
- [43] 董俊宁, 杨词慧. 空间约束混合高斯运动目标检测. *中国图象图形学报*, 2016, 21 (5): 588-594

- 
- [44] Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. in: Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. 2002: 2142
- [45] Cheng Y. Mean Shift, Mode Seeking, and Clustering. IEEE Trans.pattern.anal.mach.intell, 1995, 17 (8): 790-799
- [46] Yue Y, Gao Y, Zhang X. An Improved Camshift Algorithm Based on Dynamic Background. in: First IEEE International Conference on Information Science and Engineering. 2009: 1141-1144

## 攻读硕士期间发表论文及参加科研情况

### 攻读硕士期间发表论文及专利情况：

1. 合作者,作者. 基于卡尔曼滤波与感知哈希技术的模板匹配跟踪算法[J].计算机与现代化（已见刊）
2. 导师,合作者,作者. 一种非接触式呼吸检测装置:中国,201510884295.8[发明专利].2015.12.4
3. 导师,合作者,作者.一种非接触式呼吸检测装置:中国,201520998495.1[使用新型专利].2015.12.4.

### 攻读硕士期间参加科研项目情况：

1. 参与导师国家青年自然科学基金项目“基于多质点运动建模的肿瘤形态与位置变化实时跟踪关键技术研究”，项目编号：61402218，2015.1-2017.12。
2. 参与南昌航空大学研究生创新基金“一种改进的非局部均值去噪算法研究”，项目编号: YC2016043.

## 致 谢

花开花落，四季轮回，三年的研究生时光匆匆而过，在这即将毕业之际，内心也是百感交集。回望这三年的研究生学习生活，从刚开始的懵懂无知到如今渐明事理、学习了知识、开阔了眼界，这几年也算是收获匪浅。在科研之余，身边的老师和同学也都让我感受到了科研的乐趣、生活的温暖。在此，由衷的感谢那些耐心帮助我的老师和同学。

要感谢我的研究生导师 老师，在研究生期间孜孜不倦地对我进行学习和生活上的指导，教会了我很多人生的道理、工作的正确态度以及。杨老师以严谨的学术态度，丰富的阅历，雄厚的理论知识，在我论文开题、学术研究和编程的时候都给予了我巨大的帮助。杨老师就像我学习路上的指明灯，不至于摸黑前行。

同时，还要感谢课题组的 老师，张老师对整个课题组的同学都视如己出，悉心的帮助大家论文开题，帮助大家处理论文中的各种难题，让大家都以一个良好的心态面对学习生活中的各种难题。

实验室的师兄师姐也像家里的兄弟姐妹一样对我们这些师弟师妹们，在学习上给我们树立了好的榜样，在生活上给予我们关爱，正是他们的存在才能让我们在刚来学校时快速的找到前进的方向。

感谢参评本论文的指导老师 and 领导，感谢你们抽出宝贵的时间对本论文进行指导和评价。

最后，还要特别感谢我的父母和家人，是他们给我提供了良好的家庭环境，让我快乐的成长和学习。

## 南昌航空大学硕士学位论文原创性声明

本人郑重声明：所呈交的硕士学位论文，是我个人在导师指导下，在南昌航空大学攻读硕士学位期间独立进行研究工作所取得的成果。尽我所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中作了明确地说明并表示了谢意。本声明的法律结果将完全由本人承担。

签名：\_\_\_\_\_日期：\_\_\_\_\_

## 南昌航空大学硕士学位论文使用授权书

本论文的研究成果归南昌航空大学所有，本论文的研究内容不得以其它单位的名义发表。本人完全了解南昌航空大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权南昌航空大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。同时授权中国科学技术信息研究所将本学位论文收录到《中国学位论文全文数据库》，并通过网络向社会公众提供信息服务。

（保密的学位论文在解密后适用本授权书）

签名：\_\_\_\_\_导师签名：\_\_\_\_\_日期：\_\_\_\_\_