[第1章 绪论 3](#_Toc503812779)

[1.1 课题的研究背景及意义 3](#_Toc503812780)

[1.2 国内外研究现状 4](#_Toc503812781)

[1.3 论文研究内容及结构安排 5](#_Toc503812782)

[第2章 总体设计 7](#_Toc503812783)

[2.1 主要内容 7](#_Toc503812784)

[2.2 业务流程 7](#_Toc503812785)

[2.3 整体划分 8](#_Toc503812786)

[2.4 具体设计 10](#_Toc503812787)

[2.4.1 双摄像机标定 10](#_Toc503812788)

[2.4.2 场景监控和运动跟踪 11](#_Toc503812789)

[2.4.3 双摄像机协同工作 12](#_Toc503812790)

[2.4.4 视频摘要生成 13](#_Toc503812791)

[2.5 本章小结 13](#_Toc503812792)

[第3章 双摄像机标定 14](#_Toc503812793)

[3.1 双摄像机标定介绍 14](#_Toc503812794)

[3.2 摄像机标定原理 14](#_Toc503812795)

[3.3 双摄像机标定方法介绍 18](#_Toc503812796)

[3.4 双摄像机标定结果评定 21](#_Toc503812797)

[3.5 本章小结 21](#_Toc503812798)

[第4章 运动检测和运动跟踪技术 22](#_Toc503812799)

[4.1 运动目标检测技术 22](#_Toc503812800)

[4.1.1 帧差法 22](#_Toc503812801)

[4.1.1.1 相邻帧间差分法 22](#_Toc503812802)

[4.1.1.2 三帧差分法 23](#_Toc503812803)

[4.1.2 背景减除法 24](#_Toc503812804)

[4.1.2.1 混合高斯背景建模 25](#_Toc503812805)

[4.1.2.2 改进的混合高斯背景建模 26](#_Toc503812806)

[4.2 运动目标跟踪技术 27](#_Toc503812807)

[4.3 运动目标检测和跟踪分析 27](#_Toc503812808)

[4.4 运动目标检测和跟踪结果评定 27](#_Toc503812809)

[4.5 本章小结 27](#_Toc503812810)

[第5章 双摄像机协同工作及摘要系统实现 27](#_Toc503812811)

[5.1 系统概要 27](#_Toc503812812)

[5.2 系统环境和技术 27](#_Toc503812813)

[5.3 系统设计 27](#_Toc503812814)

[5.4 系统模块设计 27](#_Toc503812815)

[5.5 系统性能分析 27](#_Toc503812816)

[5.6 本章小结 27](#_Toc503812817)

[第6章 总结与展望 27](#_Toc503812818)

[6.1 总结 27](#_Toc503812819)

[6.2 展望 27](#_Toc503812820)

# 绪论

## 课题的研究背景及意义

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我们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[1, 2]。

科技进步和多媒体技术的飞速发展，视频数据在极速增长的同时也在增加着人们处理视频数据的难度。因此，对视频数据进行管理和组织的各种技术也应运而生。例如，电子图书馆的视频检索程序、电视台的广告监控、电影的精华摘要。通常计算机对视频的操作还停留在存储、编辑、播放这个层次上，计算机只能处理视频中的图像里的像素信息[3]，而其中包含的人类理解的内容难以用计算机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，视频摘要技术应运而生。视频摘要，就是以半自动或者自动的方式，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意义的信息，并将这些信息以某种方式合并成紧凑的、能充分描述出视频语义内容的视频概要[4]。

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[5]，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，双摄像机和多摄像机逐渐引起各研究机构[6]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[7, 8]。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。

但是，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时效性提出了更高的要求。

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。

## 国内外研究现状

视频摘要（Video Summarization）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。随着时代的发展和科技的进步，尤其是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。视频摘要作为一个飞速发展的研究课题，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研究的Infomedia工程对视频摘要技术进行了初始化研究、哥伦比亚大学的VideoQ[9]、IBM的CueVideo[10]、早稻田大学的生物信号相关方法、布雷西亚大学的隐性马尔科夫（HMMS）的相关方法，许多著名研究项目逐渐完善和改进了视频摘要技术的研究和应用，随后Columiba大学、PihlipS研究院、微软研究院、AT&T实验室、IBM的Almaden研究中心、德国Mannhe如大学和加州大学Berkeley分校等大学或机构都展开了此方面的研究，并开发了多种形式的摘要和各式各样的生成算法。国内的许多顶尖大学如清华大学、浙江大学、复旦大学、中国科学院计算所和国防科技大学等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一定的进展，以及以张宏江为代表的科研人员做了大量的相关研究。

卡耐基梅隆大学最早开始视频摘要技术的研究，其开发出的Infomedia工程[11]通过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，清华大学提出了一种基于人物关系分析的视频自动摘要算法。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。

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模型的协同跟踪、基于区域的双摄像机协同、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【】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【】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。

多目标跟踪和多摄像机协同定位也是目前多摄像机协同跟踪中存在的难点，本文对于多摄像机协同工作的重心也将放在解决两个摄像机同步定位的问题。

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【】，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，,其中Kran和Shah【】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。

## 论文研究内容及结构安排

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，其中三个主要的三个难点和重点是在于双摄像机标定方法的研究与实现，双摄像机协同工作的研究和实现，视频摘要技术和双摄像机的结合。基于以上研究内容，我将本论文研究内容分为以下几个部分：

查询和研究运动目标检测和跟踪技术；

查询双摄像机标定相关文献，研究双摄像机相关理论，实现双摄像机标定通过标定建立适用于双摄像机协同工作的坐标系；

研究双摄像机协同工作的方法，并实现标定后双摄像机能够正常协同工作；

研究视频摘要技术，根据研究的双摄像机标定方法和双摄像机协同工作方法生成带有目标详细信息的视频摘要。

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，介绍了当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总体设计。

第三章双摄像机标定的研究，本章介绍了双摄像机标定的原理和概况，实现两个摄像机标定出彼此的相对位置。

第四章双摄像机协同工作的研究，本章介绍了根据标定出的坐标系对双摄像机控制方法，实现双摄像机协同工作。

第五章系统框架与实现，本章介绍了系统框架和实现方法，系统从标定到运动目标检测和跟踪再到根据跟踪到的目标如何进行协同工作，最后生成带有目标细节信息的视频摘要。

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。

# 总体设计

本章主要介绍课题研究的双摄像机实时视频摘要系统的总体设计，描述出系统设计的整体思路和各模块的设计思想。

## 主要内容

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信息要兼顾到监控区域的范围，因此必然会导致目标的清晰度的不足。大场景和清晰度是一对相互制衡的两个指标，要想进行大场景的监控大部分情况下会以清晰度的损失来作为更大监控范围的代价。

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双摄像机协同工作的优势来弥补在监控大场景的同时获取高清晰度的目标信息方面的不足。本文将针对这一目的设计出能够利用双摄像机（一个场景摄像机，一个云台摄像机）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来实现云台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。系统主要内容如下：

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统一坐标系对云台摄像机球面极坐标系的映射，使得标定出的统一坐标系的坐标信息与云台摄像机的水平角度和俯仰角度，以便云台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，比较各种算法间的优劣，找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采集的图像对场景中的运动目标进行检测和跟踪，并将图像中的坐标信息转换成标定的统一坐标系中的坐标信息。
3. 设计云台控制的接口，使系统可以根据两个摄像机标定的坐标系中的位置信息，将云台摄像机运动到坐标系中目标坐标相对应的俯仰坐标对应的位置。
4. 设计数据融合方法，使得两个摄像机采集的数据形成一个统一的能相互关联的整体，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。

## 业务流程

整个系统的业务流程如XXX所示，本系统首先要进行系统参数进行配置；然后使用双摄像机进行图像采集；对双摄像机进行标定；在场景监控摄像机中对运动目标检测和跟踪；云台摄像机协同场景摄像机对跟踪的目标进行跟踪抓拍；对两个摄像机采集的信息进行数据融合生成视频摘要。

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX流程图

## 整体划分

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，使各模块之间相互配合，又尽量少的互相影响。双摄像机实时视频摘要系统划分为四个模块，双摄像机标定模块、场景监控和运动跟踪模块、双摄像机协同工作、视频摘要生成模块。各模块功能如下：

* 双摄像机标定

双摄像机标定模块主要负责构建系统所需要的参考坐标系，分别先标定出场景监控摄像机和云台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将分别将两个摄像机图像坐标和摄像机坐标建立联系，将摄像机坐标系和世界坐标系建立映射关系，这样就建立了图像坐标系和世界坐标系之间的映射关系。有了映射关系就可以根据场景摄像机中跟踪到的运动目标映射出目标在世界坐标系的位置，来操控云台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的缺陷，一个是标定方法一般要求条件比较苛刻，比较难易操作，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。这些问题都是这个模块面对的主要难题和需要解决的问题。

* 场景监控和运动跟踪

场景监控和运动跟踪模块主要功能是对整个大场景进行监控和运动目标的检测和运动目标的跟踪。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，以供云台摄像机对确定的目标进行实时跟踪获取高清的图像信息。

* 双摄像机协同工作

双摄像机协同工作模块作为整个系统的控制核心，它综合了场景监控和运动跟踪模块采集的运动目标的坐标信息和移动信息，来控制云台摄像机对目标进行实时跟踪，获取到目标的高清图像信息。本模块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，对需要云台摄像机进行跟踪的先后进行排序，然后根据排序对云台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。本模块还需要对跟踪抓拍的成功率进行评估，尽可能的使系统维持在一个相对较高的抓拍成功率。

* 视频摘要生成

视频摘要生成是综合各模块采集的信息生成合格的视频摘要。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和抽象，记录成文本或者图片或者以短视频形式进行存储。对记录的文本、图片或者短视频，根据时间对运动目标进行跟踪，对跟踪成功的目标进行抓拍，对抓拍成功的目标进行事件记录，这样在时间轴上就形成了相对有序的，有效的目标信息检索。这样就可以使我们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可以快速的检索出运动目标在场景中存在的关键的、有效的、清晰的数据信息，图像信息。

## 具体设计

### 双摄像机标定

双摄像机标定模块是本课题的重点也是双摄像机实时视频摘要的核心模块之一，双摄像机标定模块的标定精度和决定了双摄像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广泛的学习和深入的研究，最后基于标定所需要条件的难易程度和标定精度考虑，决定以张正友教授对标定的研究为基础实验对本课题有效的双摄像机标定方法。

本文的双摄像机标定模块采用两步标定法对双摄像机进行标定，双摄像机标定分为如下几个部分如图xxx所示：



Xxxxxxxxxxxxx图

首先，在标定的时候先不考虑相机存在畸变，对两个摄像机的内参数中的五个线性参数进行标定，获得粗略的线性参数。

然后，使用标定出的粗略的线性内参数对相机的畸变系数等线性参数进行标定。由于标定线性参数的时候没有考虑畸变系数，所以线性参数是不准确的，所以再次对线性内参数进行标定，这次将畸变系数考虑进去对前面标定出的线性参数进行修正，再利用修正后的线性参数对畸变系数进行修正，反复迭代，直到两个摄像机的参数达到标定精度要求的阀值，完成最终对摄像机内参数和外参数的标定。

在完成对两个摄像机内参数标定后，对两个摄像机外参数进行进行标定，确定两个摄像机相对的水平位置和旋转位置。

最后使用标定出的内参数和外参数来对两个摄像机的坐标系映射到世界坐标系中，实现两个摄像机图像中目标位置的映射。

### 场景监控和运动跟踪

场景监控和运动跟踪模块是云台摄像机抓拍的基础之一，要想调用云台摄像机进行运动抓拍首先要先获取目标的位置信息，这就需要对场景中的运动目标进行捕捉和跟踪，由于这部分需要使用场景监控摄像机进行场景监控，场景监控摄像机作为云台摄像机获取运动目标高清信息的一个补充，就可以使用清晰度一般但是稳定性不错的摄像机作为场景监控摄像机，只需要快速的在场景中找到运动目标的大致位置并进行持续的跟踪，实时的反馈运动目标在场景监控摄像机坐标系中的坐标系转换成根据经过标定后的坐标关系和极坐标关系转换成的云台摄像机对应的坐标系中的位置信息。

基于以上目的，场景监控和运动跟踪模块需要满足实时、快速和稳定三个要求才能满足以上目的，这就需要实时、快速和稳定的运动目标检测和跟踪算法相结合才能实现。整个模块也要根据这三个目的去设计，实现方法如xxxx图所示：

Xxxxxx图

首先，先建立场景的背景的模型，并做到尽可能少的受到光照和抖动的影响，使能够得到一个尽可能稳定的背景模型，然后对新获取的图像和背景模型做帧差，使用帧差法进行空间约束，运动目标检测，然后进行滤波获取相对稳定的运动目标轮廓，提取轮廓区域特征进行运动目标跟踪，实时获取运动目标的位置信息。

### 双摄像机协同工作

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。从摄像机的分工来说，场景监控摄像机主要是检测和跟踪运动目标并实时反馈运动目标的位置信息，云台摄像机主要根据场景监控摄像机采集的数据进行运动目标跟踪及抓拍，但是如果有多目标需要跟踪的话就会产生需要跟踪的目标无法在短时间内全部调用云台摄像机进行跟踪，导致会有一定量的目标会丢失掉。为了更好的协调两个摄像机进行数据的采集，我们设计了一种双摄像机协同工作的方法。模块设计如xxx图所示：

Xxxxxxxxx图

首先根据场景监控摄像机实时反馈的运动目标的位置信息来对运动目标的轨迹进行拟合，对运动方向进行预测，对运动速度等运动参数进行计算，最终得出每个运动目标速度、方位、在监控区域的时间，并对出监控区域的时间进行预测，根据这些所有目标的运动数据进行合理的管理、预测、排序，再合理的调用云台摄像机进行跟踪抓拍，对跟踪效率进行提升。

### 视频摘要生成

视频摘要生成模块主要是生成整个视频监控的信息摘要集合，包含了对图像信息进行采集，将采集的信息作为监控显示播放，根据场景监控摄像机和云台摄像机采集的图像信息进行图像分析和分类，生成包含图像和短视频等多种信息形式相结合的摘要信息。

首先，通过对场景监控摄像机和云台摄像机的ip网络接口采集两个摄像机的图像信息，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，利用双摄像机协同工作模块协同场景监控摄像机对云台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信息结合起来，生成带有高清图像信息的多形式的摘要信息。

## 本章小结

# 双摄像机标定

## 双摄像机标定介绍

图像信息通过两个摄像机进行采集，采集回来的图像信息使用标定后的信息使得两个摄像机采集的图像信息产生彼此摄像机之间的映射关系，然后再根据云台摄像机的摄像机坐标系和云台摄像机的求坐标的映射关系建立场景监控摄像机坐标系到云台摄像机求坐标之间的映射关系。这是场景监控摄像机和云台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摄像机自身的内参数和相对的外参数的过程。摄像机自身的内参数是指摄像机内部的几何光学信息，外参数是两个摄像机在空间中相对的平移和三维偏转信息。本章重点介绍双摄像机标定的原理和各种方法的优劣，以及本文选用的标定方法介绍和标定结果分析。

## 摄像机标定原理

摄像机标定是计算机视觉研究的基础方法之一，标定在三维重建和目标跟踪等应用中有重要的作用。要对一个摄像机或者两个摄像机进行标定，首先要建立几个坐标系的概念：图像像素坐标系、图像物理坐标系、相机坐标系、世界坐标系。

图像像素坐标系：图像像素坐标系是以图像的左上角为原点，u轴平行于图像平面指向为右，v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，图像及图像像素坐标系可表示如xxx图所示。

Xxx图

图像物理坐标系：图像像素坐标系只能表示目标在图像中的像素坐标信息，并不能表示图像中目标位置的物理信息。因此，我们在图像像素坐标系的基础上建立一个具有物理意义的图像物理坐标系，图像物理坐标系的原点在图像的中心点，x轴和y轴分别平行与图像像素坐标系的u轴和v轴，坐标使用(x,y)来表示，图像和图像物理坐标系可表示如xxx图所示。



Xxx图

相机坐标系：相机坐标系是以相机的光心为坐标系原点，X轴和Y轴分别平行于图像坐标系的X、Y轴，Z轴为相机的光轴。相机坐标系是图像和相机产生联系的基础也是世界坐标系和相机坐标系产生联系的基础，它的存在使得图像坐标系和世界坐标系产生了可以联系的基础。相机坐标系可表示如图xxx所示。

Xxx图

世界坐标系：世界坐标系也称为绝对坐标系，它可以表示场景中的绝对坐标，可以通过刚体变换(旋转和平移)将世界坐标系中的坐标转换为相机坐标系中的坐标。有了这种联系就可以将世界坐标系中的坐标信息转换成相机坐标系，再通过相机坐标系映射到图像坐标系中，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金丝线性模型，因为它只包含了透视投影变换和刚体变换，并不包含相机的畸变因素。针孔模型是通过透视投影成像的，针孔模型的相机参数分别用内、外参数矩阵描述。针孔成像模型如xxx图所示。

Xxx图

针孔模型是各种相机模型中最简单的一种，针孔模型是通过透视投影成像的，针孔模型的相机参数分别用内参数和外参数矩阵描述。针孔模型中图像像素坐标系和图像物理坐标系之间的关系如图XXX所示：

图xxx

图 1 参考坐标系

图xxx中，O0为图像坐标系的原点，(u,v)为图像像素坐标系中的P点的坐标。光心的图像坐标为(u0,v0),每一个像素在u轴和v轴方向上的物理尺寸为dx和dy，在不考虑畸变的情况下，图像中任意一个像素在图像像素坐标系和图像物理坐标系之间的变换关系为

用矩阵的形式可以表示为

图xxx也描述了相机坐标系和世界坐标系之间的关系。为了可以清楚的描述三位环境中的相机和物体的位置以及姿态的关系，我们选择一个基准坐标系作为世界坐标系，该坐标系由Xw轴、Yw轴和Zw轴组成。相机坐标系和世界坐标系之间的关系可以使用正交单位旋转矩阵R与三维平移向量T来描述。因此，如果已知空间某点P在世界坐标系和相机坐标系下的其次坐标分别为(Xw， Yw,，Zw)T和(Xc， Yc,，Zc)T，则有

式中，R为正交单位旋转矩阵，T为三维平移向量。

对于空间中的任意一点P，在平面成像后的图像中的物理坐标可以写成

由式xxxxx可得

式中，P为3\*4矩阵，被称为投影矩阵；s为扭转因子；；；K由ku，kv，s，u0，v0决定，因为K只与相机的内部结构有关，所以K也被称为相机的内参数矩阵； [R T]由相机相对于世界坐标系的方位决定的，称为相机的外参数矩阵。

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

## 双摄像机标定方法介绍

双摄像机标定是计算机视觉研究的基础，相机标定在三维重建和目标的跟踪及定位方面具有重要的意义。相机标定根据标定的方式分为传统的标定方法、自标定法和主动视觉标定法。传统标定法又被称为强标定，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，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。

1. 线性标定法

直接线性变换法（DLT）是由Abdel-Aziz和Karsra于1971年提出的标定方法。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。Hall，Ito，ODFaugeras和G Toscani对线性标定法进行更为深入的研究，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，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。

1. 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，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。

1. 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的平面模版两步法脱离了必须在传统的高精度标定台上进行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。

本文综上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。

求取内参矩阵M时需要先求取一个名叫单应性矩阵的参数，单应性也是我们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xxx所示：

|  |
| --- |
| 图 2 单应性示意 |

我们将平面标定板上的特征点的坐标记作(xwi,ywi,zwi)，成像平面上的坐标记作(xci,yci,zci)，图像坐标记作(ui,vi)根据公式有如下关系：

式中，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

式中，Ii=[ui vi 1]T为点Pi的图像坐标；Pi=[xwi ywi 1]T为点Pi的标定板中特征点的坐标系坐标；H=M[n o p]=[h1 h2 h3]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n和o是正交的，于是有了下面的约束：

由式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，我们可以使得

由于B是一个对称矩阵，可以定义一个六维向量：

式中，vij=[hi1hj1 hi1hj2+hi2hj1 hi2hj2 hi3hj1+hi1hj3 hi3hj2+hi3hj2 hi3hj3]。这样，相机的内参矩阵约束条件可以重写为

如果有标定n幅图像，可以得到n组上式方程，将其写成矩阵的形式为

式中，V是2n\*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：

式中的。

在标定出相机的内参数后，根据单应性矩阵，可得外参数矩阵为

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资定位。

根据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：

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。

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。

式中，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：

式中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，为右边相机的内参矩阵。

## 双摄像机标定结果评定

## 本章小结

# 运动检测和运动跟踪技术

本章将介绍课题中使用到运动目标检测和跟踪技术，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。

## 运动目标检测技术

运动目标检测技术是实现视频进行分析和运动目标跟踪的基础技术。运动目标检测技术经过长时间的发展、实践和应用，已经相当成熟的应用在各个领域。针对不同的场景和环境，各种运动目标检测算法的效果也不尽相同，不仅要针对场景中的光照、环境变换，还要对摄像机抖动、光影扰动有一定的抵抗能力。带着这个目的我们选取几种运动目标检测的算法进行筛选和改进使之适合我们的双摄像机实时视频摘要系统。

### 帧差法

帧差法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，研究者又将帧差法分为两帧法、三帧差法和多帧差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，这些像素灰度发生变化的区域我们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。

#### 相邻帧间差分法

帧差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。

|  |
| --- |
| 图 3 相邻帧差算法示意  第i-1帧图像pi-1  第i帧图像pi  帧差  阈值判决  差分图像Mi |

在i时刻，图像pi的每个像素值减去pi-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Mi，这张Mi图像可以根据像素的值来区分前景和背景。

其中公式中的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。

#### 三帧差分法

相邻帧间差分法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帧差分法的改进版本三帧差分法。

图 4 三帧差分法示意

第i-1帧图像

第i帧图像

第i+1帧图像

M1图像

M2图像

差分图像Mi

差分

差分

与

三帧差分法改进于相邻帧差法，首先将第i帧图像和第i-1帧图像进行差分运算得到M1图像，将第i+1帧和第i帧图像做差分运算得到M2图像，再将求得的M1和M2进行与运算得到最终的Mi图像，由于在第一步做差分之后产生的M1图像和M2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。

### 背景减除法

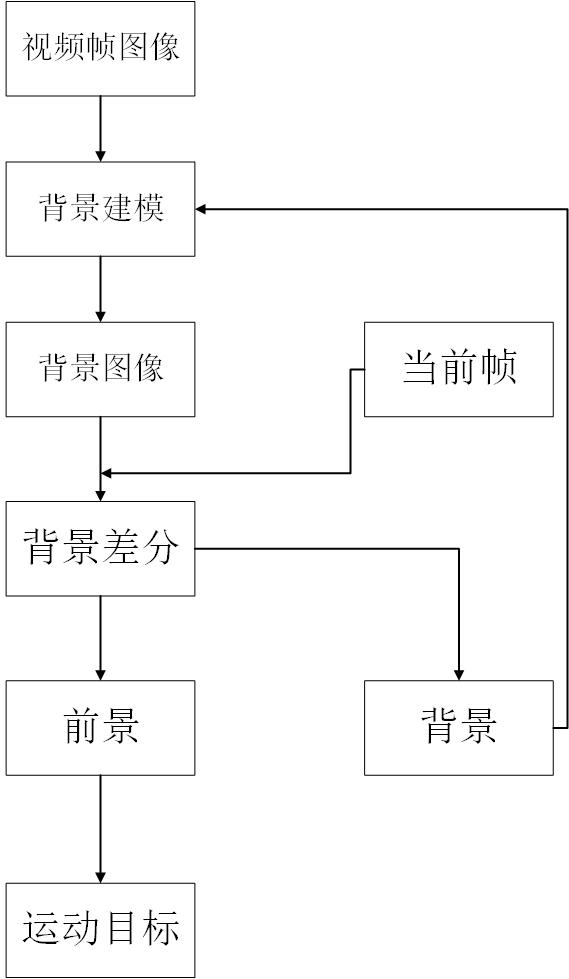
背景减除法是利用建立的背景参数模型来作为一个场景的代表，使用当前帧与背景模型做差分来实现运动目标的检测。同样是做差分运算，但是背景减除法与帧差法不同的是做帧差的对象是当前帧和背景模型。背景模型大多是根据统计学原理依据场景中像素的统计信息进行背景建模，因此，大多数的背景减除法求运动目标比较容易受光照和抖动等干扰的影响，但是，由于背景减除法建立背景模型的时候大都采用统计学的方法对背景进行建模，分割运动目标的时候会得到更清晰的轮廓信息。

图 5 背景减除法示意

背景减除法进行运动目标检测，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。

#### 混合高斯背景建模

混合高斯背景建模是由Stauffer等【Adaptive background mixture models for real-time tracking】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。

经典的混合高斯背景建模算法为视频中每一帧图像中的每个像素点定义K个状态，这个K值一般取3~5。假设每个像素位置上的像素值使用Xt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：



式中,分别为t时刻第i个高斯分布的权值、均值矩阵、协方差估计矩阵, 为t时刻的第i个高斯分布的概率密度函数，因此：



式中n为颜色通道数，当n>1时模型假设各颜色通道相互独立且具有相同的方差。

若当前像素值Xt满足(c表示各颜色通道)，则认为该像素值与第i个高斯分布匹配，并对首个匹配的高斯分布进行如下参数更新：

式中为学习率,是参数更新的学习因子，即：

对其他高斯匹配则只进行权值衰减：，同一像素的各点高斯分布的权值应该满足。

若与其他所有的高斯分布都不匹配，则令这个与其他分布不匹配的参数加入模型，或者取代优先级最低的高斯分布。完成参数更新后，按照优先级对高斯分布进行降幂排序，根据下式取前B个高斯分布作为背景模型。

上式中T表示权重阈值，一般取为0.7。

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，如果我们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。

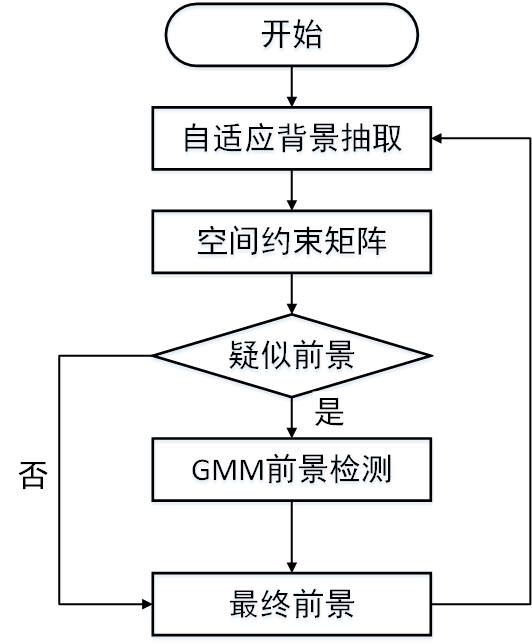
#### 改进的混合高斯背景建模

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。针对本课题我们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，如白向峰【】等人提出了以偏差均值作为混合高斯背景建模中前景匹配阈值参数的方法，黄玉等【】提出了基于YCbCr颜色空间的混合高斯背景建模方法，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，我们发现董俊宁【】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建立双重背景模型的方式，在第一种背景模型的运算中将疑似前景的区域进行空间上的约束，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 6 基于空间约束的混合高斯背景建模示意



基于空间约束的背景建模对运动目标检测示意如图所示，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。

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：

式中，n为图像深度，的时候为黑色背景，时为白色前景即运动目标。

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。

## 运动目标跟踪技术

## 运动目标检测和跟踪分析

## 运动目标检测和跟踪结果评定

## 本章小结

# 双摄像机协同工作及摘要系统实现

## 系统概要

## 系统环境和技术

## 系统设计

## 系统模块设计

## 系统性能分析

## 本章小结

# 总结与展望

## 总结

## 展望

[1] Z K, K M, J M. Tracking-Learning-Detection. IEEE Transactions on Pattern Analysis & Machine Intelligence,2012, 34 (7): 1409-22

[2] Bai T, Li Y F. Robust visual tracking with structured sparse representation appearance model. Pattern Recognition,2012, 45 (6): 2390-2404

[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] Kettnaker 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] 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

[10] 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

[11] 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