第一章 引言

1.1 研究背景及意义

模拟人类智能，让机器具有思想和意识、能独立思考并完成具有人类智力才能完成的工作是人工智能的宏伟目标***[1]***。计算机视觉是这项伟大目标的一个重要研究方向，目标检测与跟踪在计算机视觉研究领域中占有举足轻重的分量，也是人工智能众多方向中极具挑战性的工作。随着社会进入互联网大数据时代，对声音、图像、视频等数据的分析处理能力成为新生时代迫切的需求。人们每天需要处理大量图像和视频信息，在图像和视频中定位识别目标对于人而言是轻而易举之事，但是对于计算机而言，在图像和视频中识别定位目标的能力还达不到两至三岁小孩的水平***[2]***。物体在图像平面投影损失的三维信息，光照变化，目标形状、颜色和尺寸的多样性，以及复杂多变的背景环境等因素都会给目标定位检测与跟踪带来严重的干扰。

近年来人们对“智能”有了新的憧憬，致力于智能设备、智能家居的企业如雨后春笋般迅速成长。目标检测与跟踪受到了诸多学者及企业的关注，美国Google公司的无人驾驶汽车、机械狗等产品让人们看到未来的希望。深圳大疆创新科技有限公司自主研发的无人机在国内也掀起了智能硬件的浪潮，为计算机视觉飞向天空做好铺垫。智能时代美好憧憬是改善人们的生活质量，在现实生活中，目标检测与跟踪应用需求也很广泛，下面列举了一些典型的需求场景。

公众场所：生活水平改善显著，人们对公共安全意识也随之增强，对目标检测与跟踪技术的需求也越来越高。2004年，北京市迎春灯中踩踏事故致37人死亡。11年后悲剧在上海重演，2015年1月1日上海外滩跨年踩踏事件，34人失去了宝贵的生命，数百人受伤，使得一场跨年盛宴成为全国人民的伤痛。目标检测与跟踪技术有助于人们在公众场所做好安全防范，减少不必要的损失，避免类似于过度拥挤造成意外的情况发生。

智能交通：传统交通控制系统已很难再适应高速发达城市的交通环境，智能交通系统主要通过计算机分析录图像序列，跟踪车辆轨迹，统计车流，以保障公共交通秩序。

公安机关：犯罪一直以来是人们痛恨的公敌，公众场所偷盗行窃的行为时有发生，案件调查可能需要在海量视频中追踪犯罪嫌疑人作案记录，人工查询枯燥繁琐，目标智能跟踪能节约大量人力和时间成本，快速定位案发时刻视频片段，提供查案有力线索。

军事应用：世界大战期间珍珠港事件，日本炸沉了四艘战列舰和两艘驱逐舰，炸毁188架飞机，约有2400名美国人丧生，1250人受伤。如果能有效跟踪地方飞机行踪，确定飞机方向，就不会造成如此重大损失。战场环境更为复杂，要求更强的适应性和对快速目标跟踪的稳定可靠跟踪。飞机、导弹等目标的轨迹和姿态的跟踪在战场中是制胜的关键。

目标检测与跟踪的应用场景很多，其算法和模型的评价标准是准确率和速率，我们希望在正确率尽可能高的情况下，算法的响应速度尽可能快已达到实时跟踪的效果，乃至超过人类定位跟踪目标的水平。传感器、摄像头已被广泛应用于目标检测与跟踪场所。近年来，传感器的敏感性，摄像头的像素及精度都有跟大的提高并遵循摩尔定律，将不断提高目标检测与跟踪的速度和精度，并给人们以无限的发展与想象的空间，而且无人机出现将把人类视觉送上天空，在天空将为人类打开一扇心灵的窗户。

1.2 国内外研究现状及难点

智能视频监控需要在繁杂的环境场景中实时观察车辆和行人等目标的行为以及相互之间的关系[0]。智能视频监控技术包括目标检测与跟踪、目标分类、人体运动分析以及行为识别，涉及计算机视觉、模式识别、人工智能等各领域的知识。智能视频监督在美国、欧洲和日本催生了很多大型的研究项目，并近年来几次成为国际会议和研讨会的主题。

早在1997年，美国卡耐基梅隆大学在视频监控和监视（Video Surveillance and Monitoring, VSAM）项目中，采用多传感器协同合作，研究智能视频监控技术，用于监控战场和未来城市等人流与车辆密集环境[1,2]。通常这种危险复杂的环境下，需要实时监控场景状况，单独依靠人的力量很难在速度和规模上达到令人满意的效果。类似典型的场景包括：停车场与机场安保、仓库敏感监控区域、战场敌人行为的自动侦查、无人驾驶飞机监控等。同一时期，意大利热那亚大学生物物理与电子工程系的Carlo等人研究的PASSWORDS（Parallel and Advanced Surveillance System With Operator assistance for Revealing Dangerous Situations）项目致力于开发监控视频的智能分析系统[3]，用于处理特殊场景下的危险情形，远程警报监控，便于安保机构管理控制。该系统分三个阶段：1)初检测阶段旨在提取彩色图像目标移动区域；2)初跟踪阶段在时间序列上对各个目标进行标记跟踪；3)语义理解阶段对图像场景进行分析理解，并对特殊事件进行自动监控，能迅速报警通知相关人员采取必要措施。2005年，IBM华盛顿研究中心Tian等人针对实时智监控视频提出稳健高效的前景分析算法[4]。该算法采用单个固定摄像头获取图像，并通过高斯混合模型进行背景建模，并结合亮度和纹理信息减小阴影带来的干扰，使得该算法适应快速光照环境的变化。

目标检测与跟踪的目的是要解决4个W的问题，即何时、何地、谁、做什么（When, Where, Who, What）。IBM阿尔马登研究中心与马里兰州大学计算机视觉实验室合作开发的W4系统[5]，可以在低端设备上对灰度图像序列室外复杂场景进行目标检测与跟踪。该系统结合目标形状与目标各个部位的特征（头、身子、手、脚等）对目标进行建模，从而减小遮挡带来的干扰，同时跟踪多个行人，并能在户外场景监控行人的行为活动。当目标静止停留、相向移动、相互靠近时，该系统仍能准确分析判断行为活动。英国雷丁大学计算机视觉的REASON项目[6]开发犯罪行为预警系统，利用多个摄像机和电脑组成相机网络，自动理解目标危险行为、推测可能发生犯罪的时间和地点并及时向相关安保人员通报。

除研究人体跟踪以外，车辆的监控跟踪是智能交通系统重要研究领域，也引起众多学者及科研机构的关注。车辆检测系统从功能上来讲主要分为三大类：交通监控系统，驾驶辅助自动导航系统以及交通信息系统[07]。

交通监控系统：20世纪70年代中期，日本研发的智能汽车使用机器视觉系统检测障碍自动导航，该视觉系统使用立体TV相机硬链接逻辑电路实时处理信息，采用的推算函数使得车辆能以10km/s的速度自动避开障碍物，从固定的起点到固定终点快速驾驶[7]。80年代末期，日本想要将智能汽车的思想转移到民用车辆，开始研究PVS(Personal Vehicle System)[8]，对基于视觉的民用汽车进行大量综合性实验，PVS使用两套计算机视觉系统，其一用于识别道路两侧的路标，另一套用于主干道障碍物的检测。实验数据表明，数据表明PVS能以10~30km/h速度自动驾驶。90S初期，自动高速公路车辆系（Automated Highway Vehicle System, AHVS）[9,10]使用单个TV摄像机，利用边缘检测算法检测路标，保持车辆在主干道正常行驶，行驶速度可以达到50km/h。

辅助驾驶系统：目标检测与跟踪除了监控管理道路车流状况外，也能辅助驾驶员更方便地驾驶车辆，提高交通安全系数，减少交通事故地发生。辅助驾驶系统通过车身上安装的成像系统，估计相邻车辆距离自身的距离以及主干道两侧行人运动方向等信息，帮助驾驶员迅速反应，针对不同紧急情况作出相应的决策[11-13]。Foresti与Regazzoni等人研究自动驾驶系统中目标识别与跟踪的协同问题，针对特定的视觉信号部分，识别出相应的象征意义，并跟踪车辆运动轨迹，并在大量复杂场景试验中取得了较好的结果[14]。随后，Betke等人研究的高速公路辅助驾驶系统[15]结合车辆颜色、边缘和运动信息，检测路标、相邻车辆等信息，估计场景中各种不同的交通信息之间的关系，从而辅助驾驶员在高速公路驾驶中应对不同的紧急交通状况。2005年，国内清华大学研发车载视频驾驶辅助系统利用前向摄像机的信息计算车辆偏离轨道中央的距离，以及距离前方行人、车辆的距离，在特殊情况能迅速给出警报信息[16]。

交通信息系统：除自动导航交通监控系统以外，也不断有学者在交通信息系统方向付出努力[17-19]，通过监测预先设定区域光照因素的变化，能通过摄像机或探测器迅速检测车辆的移动。

随着硬件成本降低，技术不断增强，立体视觉在目标检测与跟踪领域的应用越来越多,国内对基于立体视觉的目标检测与跟踪贡献也非常突出。山东大学柯晶与常发亮两位老师多年研究基于立体视觉运动目标的检测与跟踪。2006年，刘冬冬与常发亮老师首次在双目视觉中应用camshift算法，用于检测识别应用场景中的运动目标，并进行跟踪[20]。随后，王哲基于分层网络，将图像进行金字塔分层，并对顶层金字塔图像最小分割全局搜索相似度最大的区域进行匹配，根据匹配结果对运动目标进行检测[21,22]。柯晶老师在特征点改进方面作出了突出贡献，2007年，柯老师的学生黄祖伟通过改进Harris角点提取算法，提高行人检测的精度，从而更好地跟踪场景中的运动目标[23]。2009年，赵聪在之前改进的Harris角点提取算法基础上匹配特征点，并将匹配好的特征点作为约束点，对非Harris角点的像素进行最大区域相关性匹配，最终可以得到稠密视差图[24]。

目标检测与跟踪发展迅速，国际上每年有多个的学术会议涌现出新的研究成果，如CVPR, ECCV, ACCV, ICIP, ICPR等，还有相关的组织机构也会发起一下学术讨论最新的科研进展。为了客观评价学术成果性能，一些机构建立公共数据库，用于评估测试算法性能指标。自2000年起，法国举办IEEE会议推出第一个视频检测与跟踪数据库测试平台PETS(Privacy Enhancing Technologies Symposium)，随后不断新增数据库，现已更新至PETS2015[25,26]，包含不同条件下（如遮挡、光照变化、阴影、危险动作）的场景测试用例，其日趋完善的评测数据库被广泛用作各个新算法的性能评价指标。CDNET[27]是场景变化检测算法评价数据库，该网封装了一个严谨而且全面的测试数据集，包含两个可用的数据集：2012年数据集和2014年数据集。其中2012年数据集测试指标包括动态背景、相机抖动的干扰、物体运动以及阴影干扰，2014年数据集增加天气干扰、夜晚光照条件不好情况的测试数据。Thomas博士发起的VIVID目标跟踪评价网站[28]提供标记真值的目标跟踪数据集，并提供开源代码用于测试评价算法性能。CvPapers的数据集[29]包括人脸检测与识别，猫脸检测，行人检测，显著性检测等测试图集，以及目标跟踪，前景分离算法的测试视频[29]。2007年，复旦大学王黎明公布了行人检测的评价数据库，并提供了相应的检测算法[30]。

一个完整的智能系统含有两个子课题，即：目标的检测（object detection）与目标跟踪（object tracking）。这两个研究课题涵盖的知识面广泛，大多数研究者都是将他们分开进行研究，并提出相应解决方法，同时大多数情况下国际会议和期刊也都将他们作为独立子课题研究、讨论。通常，检测算法负责在图像、视频等信号中提取特定的目标，标示其所在区域；而跟踪算法以目标区域作为初始化条件，在后续的信号中持续定位目标直至其在信号源中消失为止。下面分别讨论这两个子课题的相关算法的研究现状：

• 在目标检测方面，为了能利用背景减除方法快速检测出场景中运动的目标，Gaussian mixed模型 [35] 和Bayesian模型 [36] 分别被用来对背景的变化规律建立动态模型。Heikkila和Pietikainen则设计的LBP（local binary pattern）纹理描述子 [37] 建立背景模型以减少光照带来的系统不稳定性。这类方法着重检测的运动目标，但并不能进一步区分目标的类别。如：背景方法在交通监控系统时往往将行人或者光照变化做为目标，产生错误的统计结果。而面对一类特定目标的样本进行检测的问题，机器学习的方法则比较常用的解决方案。这类方法利用一些分析工具对目标提取特征如小波特征 [38] 、运动特征 [39] 、梯度方向直方图HOG特征 [40] 等，产生目标特征描述子，然后利用一些监督分类器如：神经网络、支持向量机或是Adaboost学习正负样本特征，最后并对检测样本进行分类以识别特定的特定类型目标。

• 早期时候的Isard和Blake将跟踪问题看作为非线性和非高斯情况下的动态系统状态估计问题，用粒子滤波（Particle Filter）方法 [41] 成功实现了在复杂环境下对单一目标的轮廓跟踪。而随后，Comaniciu和Meer用Bhattach-ary系数度量跟踪目标与候选目标相似程度，而mean shift迭代爬山过程寻找与目标最匹配的位置以实现连续的跟踪过程。另一方面，借助于active contour model在图像分割方法优越的性能，一些学者将跟踪问题转变成目标轮廓为参数的能量泛函最小化问题 [42] ，并用活动轮廓做连续的演化形变在连续帧中提取目标轮廓。

1.3 本文主要工作

1.4 章节安排

**第二章 单目摄像机标定**

**2.1 摄像机数学模型**

计算机视觉源于大千世界中光源发出的射线光线，通过介质照射到物体表面，大部分光线被物体吸收，只有少部分被反射光线会进入人眼，被视网接收，形成不同颜色，为我们所察觉。通常摄像机成像的原理也是如此，胶片如同人的视网膜，在相机曝光时捕捉进入相机的光线，并形成图像。

**2.1.1 针孔模型**

通常，使用一个简洁的数学模型来描述相机成像原理——针孔模型[01-03]。针孔模型假设远处物体反射的光线中只有一条光线通过针孔平面的光心，并最终投影到成像平面上。成像平面与光心之间的距离为假想摄像机的焦距。如图2.1(a)所示，相机焦距为，物体长度，物体到相机焦点的距离，利用相似三角形关系可知成像平面物体长度为。为了计算方便，该模型可等价于图2.1(b)所示模型，将成像平面移至焦平面与物体之间，则计算直接免去负数带来的麻烦。

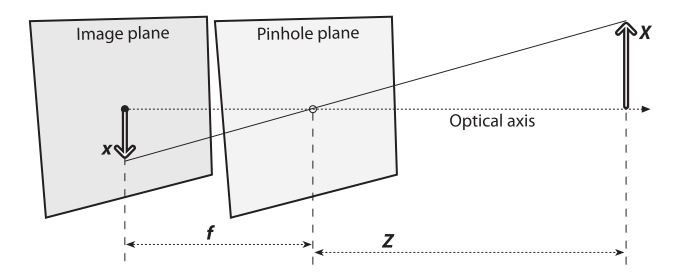
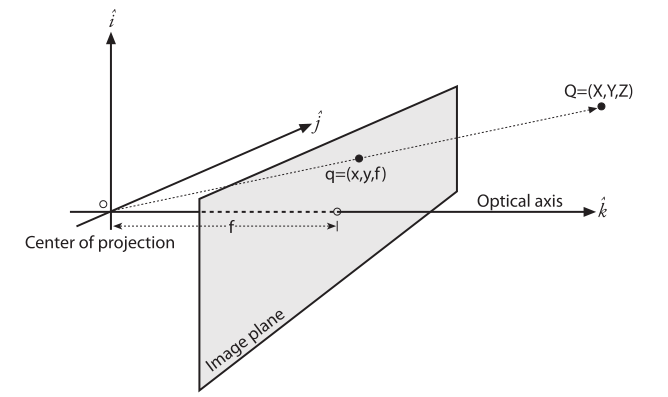
 

图2.1 摄像机成像原理

**2.1.2 摄像机矩阵**

针孔模型非常简单，然而现实生活中摄像机如果使用一个针孔却不使用针孔，因为针孔不便于快速曝光收集反射光线，成像速度太慢，通常采用透镜方便迅速收集更多的光线。一般情况下，透镜中心不在光轴上，引入两个参数，分别表示光心距离透镜中心相对位置的横、纵坐标。假设一个点世界坐标为，投影到成像平面坐标为，则有

（2.1）

其中分别为透镜两个不同方向的像素焦距，该投影过程被称为投影变换。由于成像平面所有点的交比不变，用表示该平面上的投影点，则上述模型可以转换为齐次坐标形式：

，其中，为相机坐标参数[04]，

展开可以发现，用于恢复投影像素坐标。

**2.2 畸变模型**

理论上来讲，可以不引入畸变模型，但实际摄像机成像过程中，没有任何一个透镜是毫无瑕疵的。况且即使透镜完美无瑕，安装过程也不一定能保证透镜平面与成像平面完全平行。透镜加工和人工安装上可能存在一定的误差，因此需要了解畸变原理，通过数学方法校正图像。

**2.2.1 径向畸变**

实际应用中摄像机通常在图像边缘畸变明显，主要是由透镜“鱼眼”带来的影响，如图2.2所示。对于一般透镜而言，偏离透镜中心越远，畸变越严重，这种情况在一些低端的透镜中表现更为严重。但在高端的透镜中不是特别明显，因为系统为畸变做了很多校正径向畸变的工作。

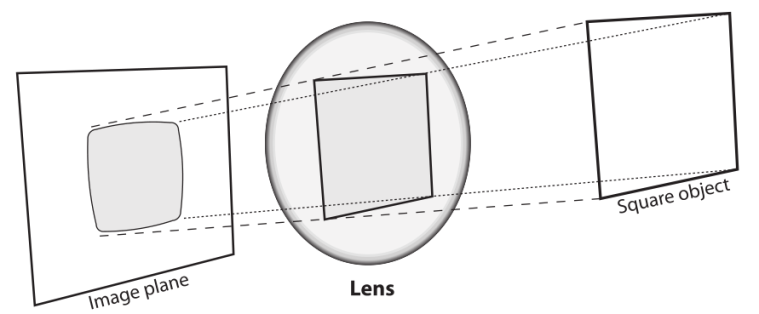


图2.2 径向畸变

对于径向畸变，光学中心畸变为0，越往边缘，畸变越严重。现实透镜中，这种畸变较小，以光学中心向边缘展开，利用Taylor级数展开对径向畸变建立模型[05,06]。

.（2.2）

函数关于径向对称，且，近似保留7项Taylor展开式，故畸变模型可简化为

（2.3）

对于高端相机通常保留两个参数即可，一般便宜的相机（如鱼眼相机）通常保留三个系数。成像平面任意一点的像素位置坐标可调节为

（2.4）

（2.5）

其中为畸变点在成像平面的像素坐标，为校正后的该点的新像素坐标。图2.3显示径向畸变分布图。

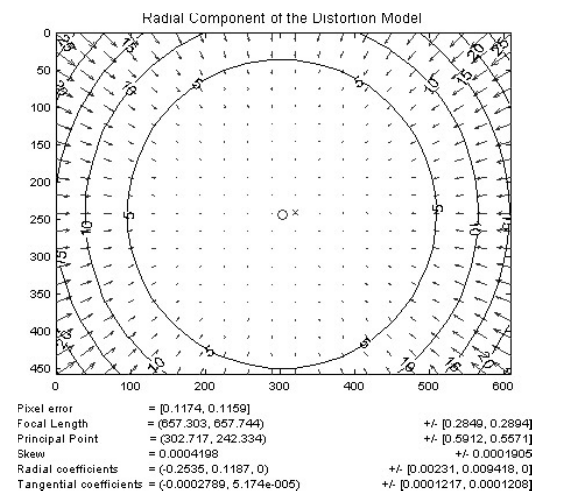


图2.3 径向畸变图

**2.2.2 切向畸变**

除径向畸变引起图像失真之外，人工安装透镜时，与成像平面不平行也会导致图像畸变，如图2.4所示：

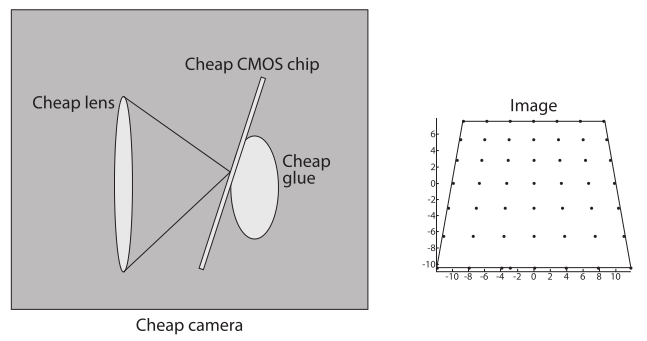


图2.4 透镜不平行于成像平面产生切向畸变

Brown引入两个额外参数来描述切向畸变[07]，如下：

（2.6）

（2.7）

其中为畸变点在成像平面的像素坐标，为校正后像素坐标。图2.5所示为切向畸变分布图。

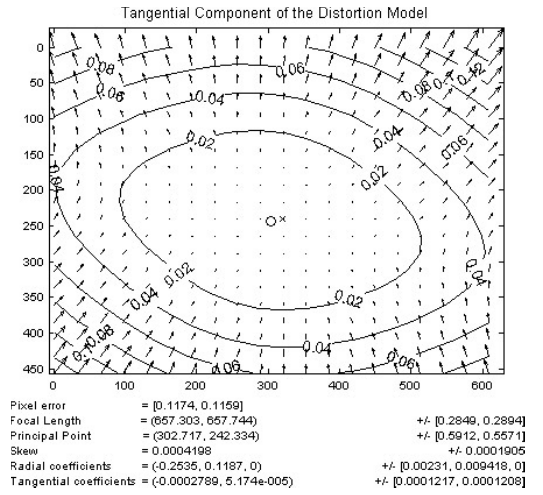


图2.5 切向畸变分布图

摄像机矩阵参数和畸变参数共九个参数统称为摄像机内参，如果是要更精确校正，可以定义更多摄像机内参数。

**2.3 摄像机标定**

**2.3.1 单应变换**

对摄像机模型进行简单的改造，可以用一个单应变化表示为

（2.8）

其中，为单应矩阵，可分为两部分：一部分表示定位观察的平面物体的物理变换，另一部分为摄像机投影变换用表示；为物体世界坐标的齐次形式，为成像平面点像素坐标的齐次形式。实际应用中，不需要关注所有的空间坐标，只需定义平面坐标，可以令进行简化：

（2.9）

通过上述单应变换公式，将实际的点与图像平面上的点联系起来，可以避免计算摄像机内参数而直接计算单应矩阵。

**2.3.2 棋盘格标定**

为从已知的投影二维图像恢复物体的真实场景信息，必须确定摄像机的内参数，该过程称为相机标定[08]。很多视觉应用中，目标检测与跟踪的结果与相机标定过程息息相关。根据标定方式不同，摄像机参数标定方法主要分两大类：靶面标定法和自标定法。

靶面标定方法借助于观察已知三维信息几何形状的物体在成像平面的投影来确定相机参数。通常情况下，使用两至三个相互垂直的平面用于标定，也可能会构造误差最小函数来求解最优相机参数[09,010]。

自标定法不使用任何标定物，而是在静止的场景通过移动相机获取不同的图像并进行对比。场景不变性会对单一图像相机参数会产生两个限制，因此对于同一相机不同时刻拍摄的图像可以计算相机参数[011-013]。

尽管新的相机标定方法不断出现，但是张正友在1999年提出的棋盘格标定方法依然被广泛应用到计算机视觉项目中[014,015]。通过已知边长且黑白相间棋盘格平面，在不同视场投影的图像，利用棋盘格的角点像素坐标与真实世界坐标之间的关系求解相机内参。首先假设摄像机没有畸变，将单应矩阵写成列形式，即有：

（2.10）

旋转矩阵各列向量是相互正交的，且长度相等，则，，因此：

（2.11）

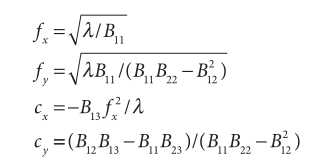
（2.12）

为方便计算，令，事实上有通用形式的封闭解：

（2.13）

利用改写约束条件为，其中是通过棋盘格标定得到相机旋转矩阵参数组成的。如前所述，棋盘格视场大于2个，则有解。

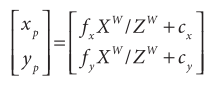
摄像机内参数可以从矩阵的封闭解中直接求解：



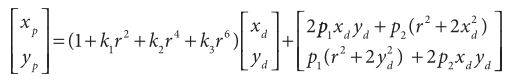
其中：



除了标定相机矩阵参数，光学畸变也是不可能避免的问题，因此我们还需标定摄像机畸变参数。理想针孔模型中，令为点的位置，为畸变位置，那么有：



通过以下变化可得到消除畸变的标定结果：



**2.4 相机标定实验及评价**

本章采用使用棋盘格对摄像机进行标定，棋盘格大小为7\*6，每个格子大小为30mm，将棋盘格平面放置在相机前合理位置，使得相机可以拍摄棋盘格不同角度的多个图像，作为相机的标定图像，如图2.7所示：



图2.7 摄像机拍摄的标定图

通过计算矩阵，反推摄像机矩阵参数合畸变参数，并利用标定好的内参数对图像进行校正，如图2.8为校正结果，从结果明显可以看出校正前图像畸变明显，校正后图像更接近真实场景图。

图 2.8 相机校正前后对比图

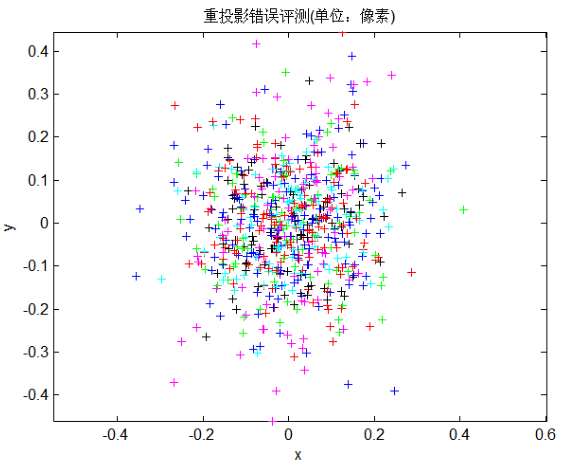
为了判断标定后结果是否正确，评价标定结果好坏，需要利用公式（2.1）计算物体世界坐标在成像平面的投影，并与校正后的重投影点进行对比，计算这些投影点与重投影点之间的距离，如果距离较小，则认为标定正确。反之，则认为标定失败，需要重新标定。计算重投影的过程公式（2.9），然后归一化坐标：



并由畸变模型估计真实成像平面坐标后利用摄像机内参数矩阵转换至成像平面的像素坐标，获得重投影点坐标：



计算重投影点与真实投影点之间平均距离值与偏差值，如果均值或偏差值超过一定阈值，则认为标定失败。如图2.9所示为重投影点偏差分布图（同一张标定图的点用同一种颜色表示）。



**第三章 立体相机标定**

  在上一章节中，本文讲述了摄像机成像原理及单个摄像机标定和校正的过程。从本章开始，我们将进入立体视觉的话题。人的双眼从不同角度观察同一场景，通过不同视场图像信息的相关性可以恢复场景的三维信息，使得我们感知的世界是三维立体的。立体相机就是利用这一原理，采用多个相机同时拍摄同一场景图像，并恢复三维信息。在本章中，将介绍多个相机成像的三维几何原理，同时介绍立体相机标定过程，对不同视场的图像进行校正对准。

**3.1 立体成像原理**

假设有两个非常完美的相机，没有任何畸变，而且相互平行，两个图像平面共面，光轴相互严格平行（光轴是从投影中心向主点发出的一条射线称为主光线），如图3.1所示：

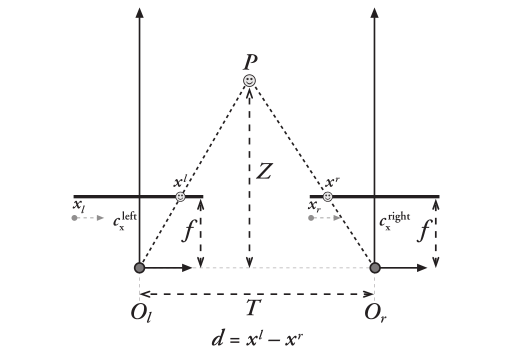


图3.1 双目成像原理

假设左右摄像机主点都与图像中心重合，焦距相等，即，并且两个相机都经过校正对准，即同一点在两个相机成像平面的纵坐标相同，这种理想状态称为前向平行对准状态。空间任意一点在左相机图像平面投影为，在右相机图像平面投影为（和坐标表示参考原点分别为两图像中心），根据三角形相似定理，则有：

（3.1）

其中称为视差，为空间点距离两相机中点的距离，即深度信息。上式很明显可以看出深度信息与视差成反比，在趋于0的时候，视差改变很小都意味着深度信息改变很大。反之，当视差比较大时，微小的变化表示深度信息基本没有变化。若以右相机为原点，计算摄像机坐标与图像坐标之间关系：

（3.2）

**3.2 立体标定**

**3.2.1 对极几何**

在3.1节中，我们讨论了双目相机成像原理，但这基于一系列完美假设的情况下建立模型的。实际应用中并非都这样理想，摄像机不可能像图3.1那样严格前向平行对准。在接下来的章节中将介绍在实际立体成像中如何映射到严格前向平行对准的状态。为了得到较好的结果，必须同步两个相机，如果相机不同步，那么只能在静态场景中进行实验，一旦场景发生变化，就会导致严重的问题。为完成两个相机的数学对准，我们需要了解更多的立体视觉几何知识，一旦深入立体相机成像原理，便能完成平行对准。

在立体视觉成像系统中，对极几何定理[016]描述空间中同一点在两个相机中成像的位置关系。如图3.2所示，对极几何由两个摄像机针孔模型组成，图中所示与分别表示两摄像机针孔，即投影中心。两投影中心连线与成像平面的交点称为极点。

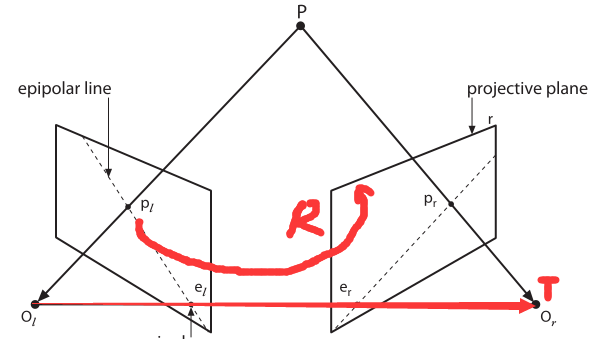


图3.2 对极几何原理图

由于图像平面体现是二维信息，图像中任一像素点对应空间很多位置，以右摄像机而言，图像上像素坐标对应的空间点可能在射线上任意一个点，很难确定实际位置。同样，左摄像机图像平面也会有一个投影点，该点可以是射线上任意一点在左摄像机成像平面的投影，因此该投影可能是左图像平面直线上任意一点，该直线称为点的极线。同样，如果以左摄像机投影点为参考，则对应右图的投影点可能在直线上，该直线为点的极线。空间点与两摄像机连线所拓展的平面称为极面。综上所述，对极定理的性质可以总结如下：

（1）摄像机视图内任意一个三维空间点都包含在极面内，并且极面与每个图像平面相交的直线为极线。

（2）对于给定一幅图像上的特征点，其对应点在必定另一幅图像对应的极线上，该特性被称为“极线约束”。

（3）极线约束意味着在二维平面上搜索对应点时可以简化为在一维直线上匹配对应点，我们只需沿着极线搜索匹配对应的特征点，这将为特征点匹配大大地节省时间，并且排除了很多虚假匹配的点。

（4）任意两点在左右两摄像机平面出现的次序保持一致。如果两个点和点在两个成像平面上都可见，并且按顺序水平投影在其中一个成像平面上，那么在另一个成像平面的投影也按次序水平出现。

**3.2.2 本征矩阵与基础矩阵**

上一小节介绍过对极几何定理，实际应用中通过二维图像恢复三维信息也需要利用对极几何定理。为了描述极线约束的特性，需要建立两幅图像对应特征点之间的几何关系，给出数学上的联系。

给定空间一个三维信息的点，要推导两个点在两个成像平面上的投影位置和之间的联系。从和之间的相互关系开始，和分别是两个摄像机坐标系中投影平面的物理坐标，假设以相机为中心进行推导。如图3.2所示，左摄像机坐标系经过旋转角度，转移向量与右相机坐标系重合，则另一台摄像机坐标系原点为。选取左相机一系列观测点，则在右相机对应点观测坐标，满足关系。如我们所知，上一小节介绍的平面可以联系所有的相关事务，为描述这个平面，采用法向量和平面上的所有点表示：

（3.3）

回顾向量和在极面上，如果有一个向量与和都垂直，那么两向量点乘可以表示求取平面公式中的法向量。这样，通过点的所有可能点以及包含这两个向量的关系可表示为：

（3.4）

我们的目的是选取的系列点和的联系来获取成像平面点和之间的联系。通过等式将点绘制到图像上，可方便地重写等式为，并带入到，可得：

（3.5）

数学中，经常将叉积形式改写为矩阵相乘的形式，因此可定义矩阵为：

（3.6）

将推导出的矩阵带入到叉积公式中，则有：

（3.7）

上式中乘积即为本征矩阵，我们定义，则：

（3.8）

通过投影方程和将上式简化，并利用进行分解得到最终成像平面上特征点的关系：

（3.9）

本征矩阵描述的是两台摄像机所有的几何信息，但是并没有涉及任何摄像机本身的信息，事实上是秩亏矩阵，公式（3.9）无解。实际应用中，我们最终要描述的是像素坐标之间的几何关系，从相机成像平面几何坐标投影至像素坐标需要用到摄像机的内参。为使用像素坐标代替成像平面坐标，二者通过摄像机内参建立联系，已知，则，因此关于的等式变成：

（3.10）

令，则有：

（3.11）

其中，为基础矩阵，不仅描述了不同摄像机之间的几何关系，同时也包含了摄像机内参数信息。假设如果相机中成像点像素坐标为，则右相机成像仪上像素坐标对应的点必然在直线上。在两图像中的极点由于无法在对应的令一幅图像中找到相应的点，因此极点与基矩阵相乘的结果为0，表示为，。

**3.2.3 立体相机标定**

立体相机标定是描述相机参数及相机状态的过程，通过立体标定，可以清楚的知道两个相机之间的几何位置关系，并对图像进行校正，使得两相机图像相对应的特征点前向平行对准，两相机光轴相互平行。如图3.2所示，立体相机之间的几何关系可以用旋转矩阵和转移向量来描述。此处的转移向量和平移向量与第二章中单目摄像机中不同，此处描述的是两个相机之间的位置关系。

立体相机标定过程如下：空间物体坐标系中任何一个三维特征点在两相机平面投影的空间坐标分别为和。相应的，通过两相机坐标系位置关系可以表示，联合三个等式可分别求解旋转和平移的简单关系：

（3.12）

（3.13）

其中和为做相机的外参，和表示右相机外餐，对于不同视场的图像，这些参数会改变。通过第二章中标定相机参数那一章节，可以求出相机外餐，从而代入式（3.12）与式（3.13）则可返回立体标定的结果。

**3.3 立体相机校正**

理想情况下，两相机光轴平行并且行完全对准时，计算立体视差是非常简单的。然而，实际情况总是不能达到理想状态，完美的相机共面平行时几乎不存在的，因此需要对立体相机进行前向平行对准校正，以保证对准之后计算的可靠性与方便性。所谓的立体校正是要将两台摄像机的图像平面进行重投影，使得两个平面精准的落在同一平面上，并且行完全对准到前向平行的状态。完成校正的结果是使得极线位于无穷远处，即两个光轴平行。由于可以选择的平行平面数量的限制，使得我们需要加入更多的条件约束，如畸变最小化和重叠最大化等。

有很多学者提供校正立体相机的方法，比较有典型代表意义的有两种算法：非标定立体校正法[017]和标定立体校正法[018]。

**非标定立体校正法**：该算法旨在计算单应矩阵，使得计算两幅图像视差时可以将对极点映射至无穷远。优点是可避免标定两个相机内参数，通过不同视场图像对应点的信息来完成。缺点是校正之后场景图像的比例仍然不知，不能恢复图像唯一的三维信息，如图3.3所示。

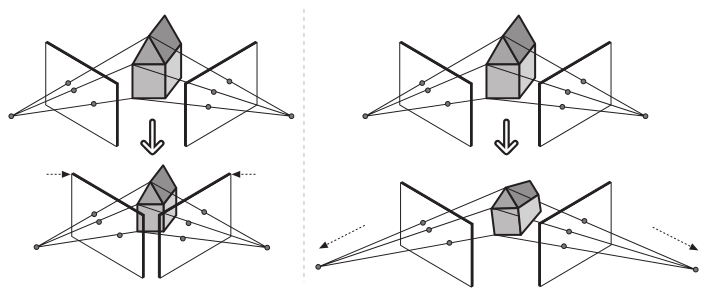


图3.3 立体重建非唯一

（可以加东西OpenCV P469）

**标定立体校正法：**所谓标定立体校正法是指借助标定好相机参数的结果，对图像进行校正。给定立体图像之间的关系及旋转矩阵和转移向量，使得两幅图像分别投影次数和重投影畸变最小化，同时最大化观测面积。（OpenCV P469）

**3.4 立体校正实验与结果分析**

本文采用第二种方法，对图像进行校正，校正之后左右摄像机内参数相等，并且两幅图像上相应的特征点满足前向平行对准的状态，即校正之后左右图像极线在共线，如图3.4所示。

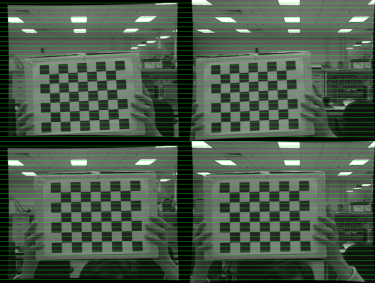


图3.4 双目相机校正结果

假设校正之后新相机内参数为，重新利用相机投影模型计算新相机坐标系下图像的像素坐标：

（3.14）

（3.15）

公式（3.14）与公式（3.15）分别是校正后左右摄像机世界坐标系与图像上像素之间的映射关系。同时，校正之后两摄像机像素坐标严格平行对准，换言之，两个相机坐标系之间除轴不同之外，其他两个轴轴与轴完全相同，所以两个相机坐标系之间的关系可描述为：

（3.16）

其中为两个摄像机之间的光心距离，也是在计算机视觉中所谓的基线距离。假设两相机中像素点与为匹配点，则可结合公式（3.14）、公式（3.15）和公式（3.16）可求解出三维相机坐标：

（3.17）

计算出相机坐标轴坐标之后，代入公式（3.14）、公式（3.15）可以计算出相机坐标的轴坐标与轴坐标：

（3.18）

（3.19）

通过立体相机标定，使用校正后得到的相机内参信息计算相机外参数和，并计算相机坐标。图3.5显示了标定板上角点的相机坐标位置，为方便展示，只显示相机坐标轴和相机坐标轴坐标。其中左摄像机图像特征点用红色显示，右摄像机图像特征点用绿色显示。由图可以看出相对应的特征点只有相机坐标轴不同，相差120mm（等于基线长度），而相机坐标轴与相机坐标轴几乎相等。

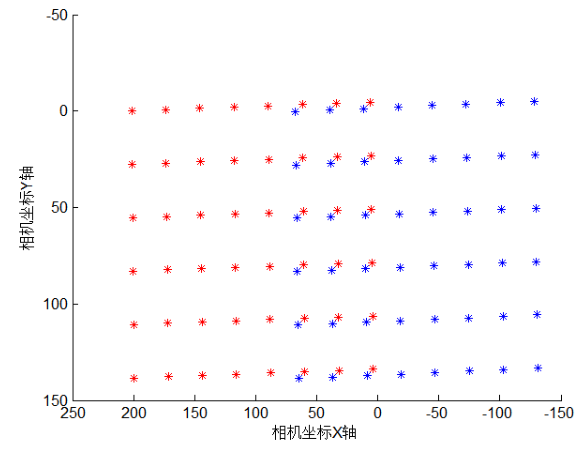


图3.5 标定板特征点相机坐标（单位：mm）

最后首先对左右相机图像进行标定校正，并通过手动在左图像标定板中选择特征点，利用视差自动在右图中搜索对应特征点，相同的点位置做同一数字标记。图3.6显示匹配的结果。









图3.6 特征点匹配计算视差

接下来，通过匹配到的结果利用公式（3.18）计算视差，并还原特征点相机坐标系下三维信息。因此在真实测试环境中，可以利用真实的三维坐标与计算的三维坐标进行对比以评价校正结果的优劣程度。真实环境下标定板聚类相机位置大约1000mm时，通过视差估计计算的相机坐标平均误差大约在10mm左右，最大误差大约15mm，如图3.7所示。事实上还可以通过亚像素插值的方法来提高立体视差的精度从而进一步减小误差。

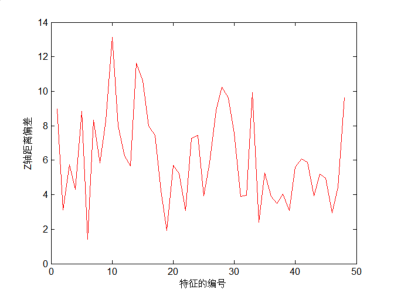


图3.7 三维重建中的误差（单位：mm）

第四章 目标检测与跟踪

3.1 场景特征

3.2 特征匹配

3.2 非监督聚类

3.2.1 核密度估计

3.2.2 核函数构造

3.2. 3 meanshift爬山

3.3目标跟踪

3.3.1 单目标跟踪

3.3.2 多目标跟踪

第五章 地面标定

4.1 相机坐标系统

4.2 世界坐标系统

4.2.1坐标转换

4.2 LDA自适应标定

4.2.1 LDA算法

4.2.2 LDA参数估计

第六章 实验结果与分析

5.1 世界坐标系统校正

5.2 多目标检测与跟踪

5.3 分析

第七章 总结与展望

[0] Introduction to the Special Section on Video Surveillance

[1] Advances in Co op erative Multi-Sensor Video Surveillance.

[2] A System for Video Surveillance and Monitoring.

[3] The PASSWORDS project

[4] Robust and Efficient Foreground Analysis for Real-time Video Surveillance

[5] W4: Real-time surveillance of people and their activities

[6] Fusion of Multiple Tracking Algorithms for Ro-bust People Tracking

[07] A Real-Time Vision System for Nighttime Vehicle Detection and Traffic Surveillance

[7] Studies on the intelligent vehicle

[8] Driving control system for an autonomous vehicle using multiple observed point information

[9] An automated highway vehicle system using computer vision-Recognition of white guidelines

[10] An automated highway vehicle system using computer vision-A vehicle control method using a lane line detection system

[11] On-line vehicle and pedestrian detections based on sign pattern

[12] Automatic Vehicle Guidance: The Experience of the ARGO Autonomous Vehicle

[13] An evolutionary approach to visual sensing for vehicle navigation

[14] Foresti G. L. M.V. and Regazzoni C., “Vehicle recognition and tracking from road image sequences”, IEEE Transactions on Vehicular Technology, 1999, 48(1), 301–318.

[15] Betke M. H.E. and Davis L.S., “Real-time multiple vehicle detection and tracking from a moving vehicle”, Machine Vision and Applications, 2000, 12(2), 69–83.

[16] 赵虹, “立体视觉技术在客流统计系统中的应用”, 浙江杭州：浙江大学.

[17] Vehicle detection video through image processing: The auto scope system

[18] A window-based image processing technique for quantitative and qualitative analysis of road traffic parameters

[19] Vehicle-type identification through automated virtual loop assignment and block-based direction-biased motion estimation

[20] 刘冬冬, “基于双目视觉和CamShift算法的目标检测与跟踪”, 山东济南：山东大学, 2006.

[21] 王哲, “立体视觉匹配及基于立体视觉的运动目标检测与跟踪方法研究”,山东济南：山东大学, 2007.

[22] 王哲., “一种基于立体视觉的运动目标检测算法”, 计算机应用, 2006,26(11), 2724–2726.

[23] 黄祖伟, “基于双目立体视觉的目标跟踪算法研究”, 山东济南：山东大学, 2007.

[24] 赵聪, “基于双目立体视觉的运动目标检测与跟踪”, 山东济南：山东大学, 2009.

[25] <http://www.cvg.reading.ac.uk/PETS2015/a.html>

[26] <http://www.filewatcher.com/b/ftp/ftp.cs.rdg.ac.uk/pub-0.html>

[27] <http://changedetection.net/>

[28] <http://vision.cse.psu.edu/data/vividEval/main.html>

[29] <http://www.cvpapers.com/datasets.html>

[30] Object Detection Combining Recognition and Segmentation

[01]针孔模型

[02]针孔模型

[03]针孔模型

[04] A four-step camera calibration procedure with implicit image correction

[05] Lens distortion for close-range photogrammetry

[06] Close-range camera calibration

[07] Decentering Distortion of Lenses

[08] Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry

[09] Three-Dimensional Computer Vision: a Geometric Viewpoint.

[010] A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the shelf tv cameras and lenses.

[011] Self-calibration of a moving camera from point correspondences and fundamental matrices

[012] A theory of self-calibration of a moving camera

[013] An algorithm for self calibration from several views

[014] A Flexible New Technique for Camera Calibration

[015] Flexible Camera Calibration By Viewing a Plane From Unknown Orientations

[016] 对极几何

[017] Theory and practice of projective rectification

[018] A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses