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。

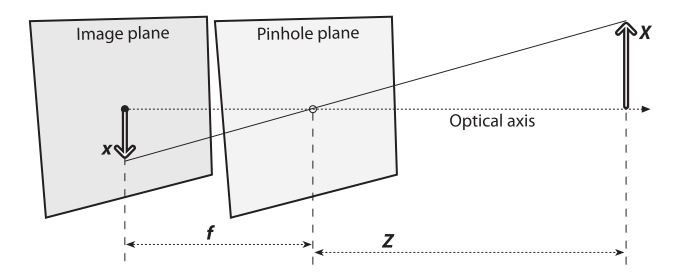
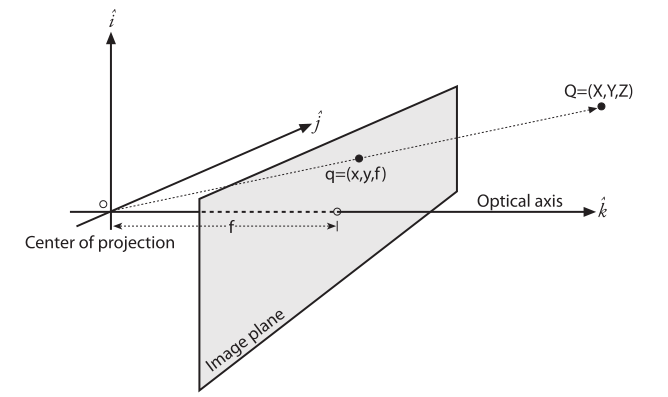
 

图2.1 摄像机成像原理

**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



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。

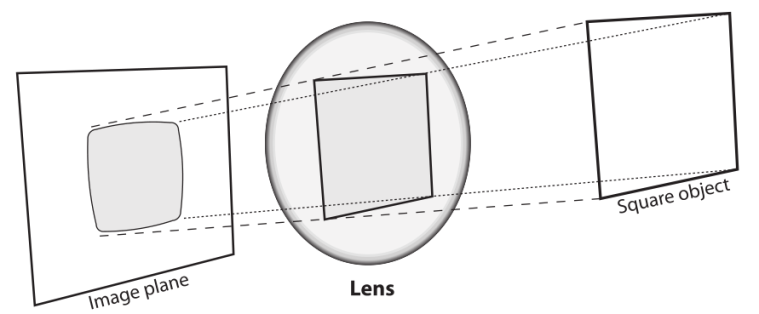


图2.2 径向畸变

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]。

.

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



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





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。

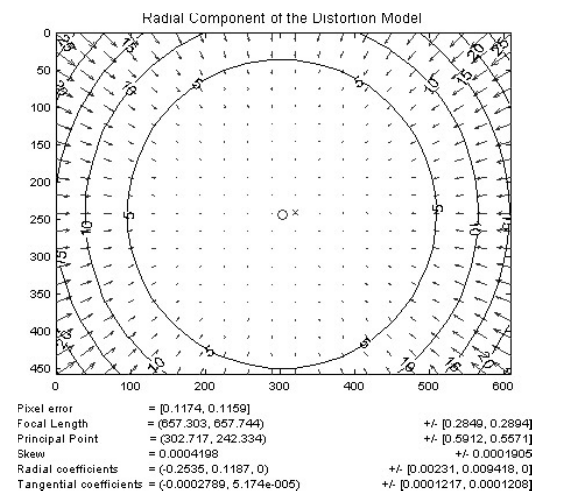


图2.3 径向畸变图

**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：

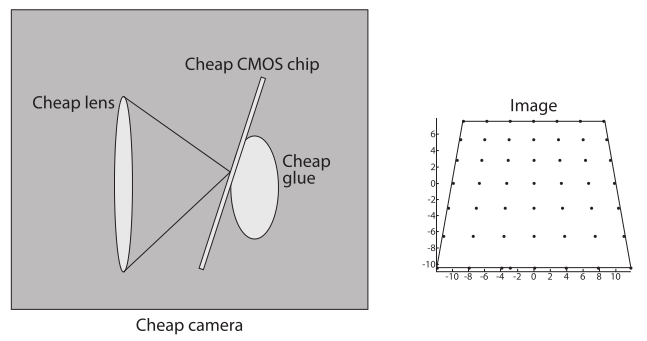


图2.4 透镜不平行于成像平面产生切向畸变

Brown引入两个额外参数来描述切向畸变[07]，如下：





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。

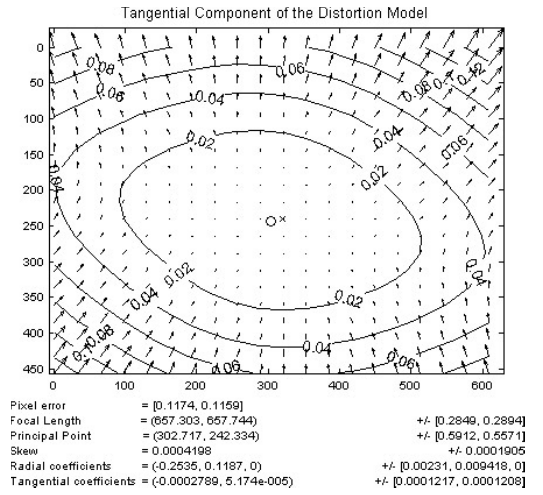


图2.5 切向畸变分布图

**2.3 摄像机标定**

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



其中，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，可令进行简化。

2.4 图像校正与误差分析

第三章 立体相机标定

第三章 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

第四章 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

第五章 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

第六章 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 photogram-metry

[06] Close-range camera calibration

[07] Decentering Distortion of Lenses