

申请上海交通大学博士学位论文

基于立体视觉的目标检测与跟踪

论文作者 蔡岭

学 号 0060329029

指导教师 杨新 教授 赵宇明 副教授

专 业 模式识别与智能系统

答辩日期 2011年01月12日

Typeset by L^AT_EX 2_≤ at April 11, 2011

With package **CASthesis** v0.1j of CT_EX.ORG

Submitted in total fulfilment of the requirements for the degree of
Doctor
in Pattern Recognition and Intelligence System

Object detection and tracking based on stereo vision

LING CAI

Supervisor:

Prof. XIN YANG, Associate Prof. YUMING ZHAO

DEPART OF AUTOMATION
SHANGHAI JIAO TONG UNIVERSITY
SHANGHAI, P.R.CHINA

Jan. 12th, 2011

上海交通大学

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：_____

日期：_____年____月____日

上海交通大学

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保 密 ，在_____年解密后适用本授权书。

本学位论文属于

不保密 .

(请在以上方框内打“√”)

学位论文作者签名: _____

指导教师签名: _____

日 期: _____ 年 _____ 月 _____ 日

日 期: _____ 年 _____ 月 _____ 日

基于立体视觉的目标检测与跟踪

摘要

目前，由于传统监控系统的简单记录功能已经无法满足社会公共安全的更高层次要求，因此具有多目标识别与跟踪的新一代智能监控系统逐渐成为了重要的研究课题。智能地检测、跟踪场景中的目标能提高对关键场所的监控能力，及时发现潜在的危险因素，提高人民群众的安全感，防止恐怖主义的侵袭。

本论文着重研究如何采用计算机立体视觉以及模式识别等相关技术构成一套全自动的多目标在线检测与跟踪系统，实现对监控场景中多目标的统计、定位、跟踪以及运动轨迹分析，直至提取出目标的精确轮廓。主要研究工作有摄像机成像原理建模与参数标定分析，双目成像的立体视觉匹配与三维场景快速重建，核函数的全自动聚类分析以及多特征融合的能量泛函与偏微分求解。具体的创新点如下：

1. 提出了基于核函数的离散点聚类算法将世界坐标系下的离散点聚合不同的集合，通过这些集合与监控场景中的目标建立一一对应关系，并以这些聚类的中心位置、方向等参数确定监控场景中的目标数量、位置与运动轨迹。相机坐标系下的特征点投影到地板平面的位置与高度作为已知参数，并构建世界坐标下的空间特征点密度分布函数。以mean-shift方法在不计算空间真实密度值的前提下直接估计其梯度方向，用爬山法搜索密度空间的所有局部最大值。以这些局部最大值为基础划分空间中的所有特征点形成聚类集合，检测监控场景中的目标。同时目标的跟踪问题也相应简化为空间最大值的更新过程。
2. 提出以系统化的分层逐步组合模型融合颜色、纹理、边缘、帧差等多个特征构成统一的特征描述概率测度。选取目标与背景像素组成正负样本集合，并分别提取颜色和多尺度纹理特征构成非参数概率空间。组合多

个特征构成背景与目标特征描述向量估计图像像素的条件概率，贝叶斯分类器以计算的后验概率区分背景像素与目标像素。另一方面，高斯混合模型在帧差特征中提取变化较大的区域，与图像边缘构成目标边缘的测度函数增强目标边缘同时抑制背景边缘。自适应权重函数评价不同特征的可靠性给予恰当的权重值组合，形成完整目标和背景的判断模型。

3. 在活动轮廓模型框架下构建了一种区域测地活动轮廓模型，以区域模型在全局中快速推动活动轮廓大致定位目标轮廓，以测地模型在局部逐步细化结果直至完全与目标精确对齐。梯度下降流求解区域测地模型的偏微分方程，迭代初始化轮廓直至收敛到最小能量曲线。此外，基于目标运动特征的最优初始化轮廓设置可大大减少迭代次数，实现在线的目标轮廓跟踪。
4. 设计一种新的场景特征提方法以图像邻域空间的四种基本梯度基元为依据，分离出纹理丰富的特征点作为匹配对象。两段式的匹配算法将左右视图的扫描线切分成不同片段后逐段分层对应匹配，随后归一化的交叉相关方法计算特征点的精确视差还原其相机坐标系下的三维坐标。

大量的实验证明本文所提的监控系统可在室内外环境中在线检测跟踪多个目标，对环境中的光照变化、阴影覆盖、目标遮挡等复杂条件都表现较好的抗干扰性能。此外，监控系统只需设置监控目标的空间形状参数，可广泛应用于解决行人或车辆等多类型目标的检测跟踪问题。

关键词： 相机标定，极线矫正，视差匹配，立体视觉，核密度估计，聚类，特征融合，贝叶斯模型，能量泛函，曲线演化，水平集

Object detection and tracking based on stereo vision

ABSTRACT

Nowadays, multi-object detection and tracking has become an essential subject in artificial surveillance system field, however, traditional surveillance system can hardly satisfy the higher level demand by the current social security. Therefore, the more advanced artificial surveillance systems are required to replace the existing systems and to help people create a more efficient and safer living style.

This paper focuses on how to develop a complete automatic on-line multi-object detection and tracking system based on 3D computer vision and pattern recognition. This system can realize functions such as counting, locating, tracking, trajectory analysis and contour extraction in surveillance scenes. The main research work includes camera modeling, parameter calibration, stereo vision matching, 3D scene reconstruction, kernel function clustering, multi-feature fusion, energy functional and partial differential equation. Main contributions are as follows:

1. We propose a kernel-based clustering algorithm for discrete points. This algorithm can produce different sets to group discrete points in different world coordinates. By defining the one-by-one correspondence among sets and objects in surveillance scene and determining the cluster center location, orientation, we can obtain the objects numbers, location and trajectory in surveillance scene. We project the feature points in camera coordinates onto the floor plane and regard the height and location as the known prior parameters, then construct the spatial distribution function in world coordinate. Mean-shift method is used to estimate the gradient orientation

without calculation the real spatial density, and search the local maximum in the density space by the points sets and the objects in scene. In this way, the object tracking problem can be simplified as a maximum updating procedure.

2. We propose a systematical hierarchical combination model to fuse the color, texture, edges and frame differences. It is a uniform feature description metric. The positive and negative sample sets of objects and background are selected to extract color and multi-scale texture feature, so that a non-parametric probability space can be constructed. The model combines several features and constructs the background and objects feature descriptor vectors to estimate the conditional probability of image pixels. Bayesian classification can calculate the posterior probability to differentiate background and object pixels. In other aspect, Gaussian Mixture Model extracts the significant changing areas by frame difference feature. It produces a metric function to enhance the object edges and suppress the background edges. An adaptive weight function can judge reliability of different feature and combine them according appropriate weight to form the complete object and background model.
3. We construct a Geodesic region model under the active contour model framework. Such region model can quickly evolves the active model globally and roughly locate the object contour. Later on, it can switches to Geodesic contours model and accurately refine the result until it coincides the object completely. Furthermore, Geodesic contour model is solved by applying gradient descent flow and iterating the initial contour until it converges to its minimum energy curve. Besides, the optimal initial contour according to object motion features can greatly reduce the iteration and realize the online contour tracking.
4. We design a new scene feature extraction method, which based on the four basic gradient elements of image neighborhood space to match objects with

abundant texture. Two-step matching algorithm divides scanline of right and left views to different segments in order to hierarchically match. Later on, NCC is applied to accurately restore the 3D coordinates in camera coordinate.

Numerous of experiments indicate that the surveillance system proposed in this paper can online detect and track multi-object in both indoor and outdoor conditions. The system is also robust in illumination changes, shadow and occlusion situations. Furthermore, we only need to set the spatial shape parameter of objects to be monitored. The proposed system can be broadly used to solve pedestrian and vehicle detection and tracking problems.

KEY WORDS: Camera Calibration, Epipolar Rectification, Disparity Estimation, Stereo Vision, Kernel Density Estimation, Clustering, Feature Fusion, Bayesian Model, Energy Functional, Contour Evolution, Level Set.

目 录

摘要	i
ABSTRACT	iii
目录	vii
插图索引	xii
表格索引	xii
主要符号对照表	xiii
第一章 绪论	1
1.1 研究背景	1
1.2 国内外研究现状	2
1.3 当前存在的主要问题	6
1.4 本文的主要工作及研究内容安排	9
第二章 单相机标定	13
2.1 摄像机原理及数学模型	13
2.2 平面标定方法	16
2.2.1 建立标定约束条件	16
2.2.2 估计相机内外参数	17
2.3 相机标定实验及评价	19
第三章 双目视觉的极线几何与矫正	25
3.1 双目视觉的极线约束	25

3.2 图像矫正	33
3.3 立体匹配及场景重建	36
3.3.1 全局分段视差匹配	38
3.3.2 局部模版细化匹配	40
3.4 双目矫正实验及评价	41
第四章 多目标的检测与跟踪	51
4.1 三维立体坐标系统	52
4.2 多目标的检测	58
4.3 多目标的跟踪	64
4.4 构造位置与方向核函数	65
4.5 多目标系统的实验及评价	68
第五章 多特征融合的轮廓跟踪	79
5.1 区域特征分析及区域能量泛函	81
5.2 边界特征分析及边界能量泛函	86
5.3 曲线演化及最小化能量泛函	90
5.4 运动特征与曲线初始化	92
5.5 单目标系统的实验及评价	93
第六章 总结	101
参考文献	103
致谢	121
攻读学位论文期间发表的学术论文目录	123

表格索引

4.1 系统的量化结果	72
5.1 完整的轮廓跟踪算法	95
5.2 实验的处理时间对比	100

插图索引

1.1 系统的硬件构造流程图	9
1.2 系统的软件处理流程图	10
2.1 摄像机成像原理图	13
2.2 相机标定的标靶图	19
2.3 左摄像机标定图	20
2.4 右摄像机标定图	21
2.5 重投影点偏差分布图	23
3.1 三维重建的不确定问题	26
3.2 双目视觉成像原理图	29
3.3 相机对应的极线图	32
3.4 视察匹配中的遮挡问题	39
3.5 基础矩阵误差	42
3.6 极线矫正的实验结果	45
3.7 特征点的相机坐标	46
3.8 特征点匹配	48
3.9 估计视察在三维重建中的误差	49
4.1 地面标定原理图	54
4.2 监控场景中的特征点	56
4.3 相机坐标系下的特征点	57
4.4 世界坐标系下的特征点	58
4.5 概率分布图	60
4.6 连续帧中的概率空间函数	66

4.7	三维的目标核函数	67
4.8	特征点标定地板平面	69
4.9	单目标跟踪的对比结果	73
4.10	遮挡情况下的多目标跟踪	74
4.11	两队交错列多目标跟踪	75
4.12	室外环境中的多目标跟踪	76
4.13	复杂条件下的多目标跟踪	77
4.14	复杂条件下的多目标跟踪	78
5.1	本文所提的模型架构	81
5.2	选择正负样本集合	83
5.3	对比区域方法与本文所提方法	86
5.4	对比区域方法与本文所提方法	87
5.5	室内外相机的高斯均值变化	88
5.6	室内外相机的高斯方差变化	89
5.7	室内外相机的权重值 λ 变化	90
5.8	灰度视频中低对比度目标的跟踪结果	96
5.9	灰度视频中遮挡目标的跟踪结果	97
5.10	粗糙初始化下的跟踪结果	98
5.11	刚性目标的跟踪结果	99
5.12	移动摄像机的跟踪结果	100

主要符号对照表

f_x	相机X坐标轴放大系数
f_y	相机Y坐标轴放大系数
c_x	相机X坐标轴中心
c_y	相机Y坐标轴中心
u_x	图像坐标轴系X轴坐标
u_y	图像坐标轴系Y轴坐标
x_w	世界坐标系X轴坐标
x_y	世界坐标系Y轴坐标
x_z	世界坐标系Z轴坐标
A	相机内参数矩阵
R	相机外参数旋转矩阵
r	相机外参数旋转向量
T	相机外参数平移矩阵
t	相机外参数平移向量
K	镜头畸变系数矩阵
H	单应性变换矩阵
h	单应性变换向量
e	极点

I_i	第 <i>i</i> 帧图像
E	能量函数
H_x	目标空间核函数
H_θ	目标角度核函数
λ	权重因子
\mathbf{v}	速度向量
RF	区域特征
BF	边界特征
obj	目标集合
bck	背景集合
D	图像帧差特征
g	边界特征测度函数
Φ	水平集函数
N	曲线切向量

第一章 绪论

1.1 研究背景

随着计算机硬件的高速发展，数字摄像机的制造、生产成本大幅降低而其成像质量却不断提高。在最近这十年中，这些成像设备被逐步应用于各种环境中，如：公众环境检测、交通流量估计、安全防护监控等。目前这些系统采用两层结构构成：前端由一个或多个摄像机连续采集成像数据，然后通过数据传输通道交由后台服务器进行处理和存储。当系统操作或管理人员面对系统的大量连续数据时，只能给出宏观上的估计而无法给出细化的量化数据。在一些实时监控的公众场景中，由于目标数量较多、运动速度快、方向多变，这样的问题就显得更加突出。在美国9.11恐怖袭击事件发生后，对公众场景的监控问题被提升到国家安全的高度，并受到许多研究员和学者的关注。同时国内举行的一系列国际大型活动，如2008年北京奥运会、2010年上海世博会也对安全保障提出了较高的要求。在这样的应用背景下，智能的目标检测与跟踪系统能有效的解决公众场景的监控问题，对目标的数量、位置、速度、运动方向等指标进行准确测量。

在现在很多监控场景中，以人体和车辆作为主要观察对象的需求居多。行人在场景中能以多个方向自由运动，并相互干扰、阻挡，一旦单位面积的行人过多不仅会影响行走速度、产生拥挤阻塞，更严重时会发生相互踩踏造成重大安全事故。2004年，北京市密云县举办的迎春灯展中发生的踩踏事故造成37人死亡、15人受伤的悲剧。印度北部喜马偕尔邦一座庙宇的2008年宗教活动中，造成至少145人死亡的严重踩踏事故，其中大部分遇难者是妇女和儿童。不仅如此，人体的检测和跟踪对行人行为判定、高级人机交互也具有广泛的意义。个别个体在公众场所的一些异常行为，往往会造成一些不良后果，如停车场的偷盗行为，路口的闯红灯行为等。在人机交互方面多年来一直只限于鼠标键盘等传统方式，通过人的手势姿态等自然方式与计算机的交流将大大推动人机之间的互动行，提高机器的智能水平。将来，基于视频的多点触摸技术将出现在microsoft的下一代计算机surface computer上，帮助计算机理解一些复杂输入

指令。

另一方面，高速发展的经济使得汽车的大量普及，也造成了道路交通的经常性拥堵。现有的交通系统采用单一固定的车辆调度方式，无法对突发性、异常性的情况作出及时的处理。因此，未来的交通系统应该具有更高的智能来应对可能出现的各种状况，他不仅监控交通流量而且统计不同车型的数量，对不同车道的车辆行为进行跟踪与预测以保障最大的车流量，或是检测一些异常行为防止可能出现的交通事故。此外，利用图像数据提高汽车自身的智能水平也能有助于驾驶员提高驾驶安全，减少事故中对路人的危害。

当然，目标检测与跟踪系统在军事上也具有重大意义，如导弹的末端精确制导、战场环境下的敌我目标分类、卫星侦察等。由于篇幅有限，这里就不作详细介绍。

1.2 国内外研究现状

早在1997年，Carnegie-Mellon大学和麻省理工学院在美国国防部高级研究项目组的资助下组建了视频监控与监视（Video Surveillance and Monitoring）项目组，负责开发监控未来城市和战场的自动视频理解技术^[1, 2]。通常这样的复杂、危险环境下，人的观察能力在速度和规模上很难胜任。一些具体的应用包括：建筑物、停车场安保，仓库、机场中敏感区域监控，城市战场中狙击手目标检测，战场自动侦察等。项目开发的先进平台整合多个传感器数据，自动提取最重要的信息通知用户做出决策。同一时期，法国国家计算机科学与控制研究所的PASSWORDS项目组使用单个彩色摄像机自动理解户内外场景中移动目标的行为^[3]。该系统分为三个层次：初级的图像处理算法在图像序列中提取一些移动区域，随后跟踪模块在时间序列上跟踪这些移动区域。最后，语义模块解释这些区域的行为。一旦有特别的事件发生，系统发出警报通知用户。Pfinder系统^[4, 5]基于高速SGI成像系统实时跟踪人体以提供一个友好的人机交互接口、低码率的图像编码、自动安保系统和虚拟引导服务。

目标识别与跟踪最终要解决的是四个W问题,即那个目标在何时、何地做什么（who, when, where, what）。为了解决这些问题，IBM和Maryland大学联合开发的W4系统^[6]结合形状分析算法检测移动目标，同时跟踪多个行人，并监控他们在户外的各种活动。当目标遗留、交换、移动某些物体时，该系统

都能准确地识别、记录这些可疑活动。Olson和Brill提出的智能相机项目^[7]利用单高斯模型自适应背景方法检测行人目标，结合图论拓扑结构预测、识别一些常见事件和行为。英国Reading大学Computational Vision组的REASON研究项目^[8, 9]（Robust Methods for Monitoring and Understanding People in Public Spaces）建立犯罪预警系统辅助警察及时发现犯罪行为。多个摄像机和计算机形成的处理网络自动分析人体行为、推断犯罪发生的时间，及时向管理人员通报。

除了研究人体的监测与跟踪以外，一些科研机构也将车辆行为的识别与理解作为重要的研究方向，从而开发出更具智能的交通管理系统以提高较通系统的使用效率、安全性，减少能源的消耗和环境的污染。美国俄亥俄州立大学的交通监控实验室^[10]是较早从事这方面工作的研究机构之一，他们早在90年代中期就提出用计算机视觉理论解决车辆跟踪和交通监控问题^[11]。在同一时期，纽约城市学院与中国交通部、国家自然科学基金委合作研究视觉交通监管系统^[12, 13]（Vision System for Automatic Traffic Monitoring）监控公路路面多车道流速、确保车流通畅、预防交通堵塞等问题。Minnesota大学的交通路口监控系统重点分析行人与车辆的分类与识别问题，统计不同时刻路过行人的车辆的数量设计最优化的交通信号灯管理方案^[14, 15]。除了积极预防交通事故的发生以外，对于早期交通事故的监测与报警也有助于重大交通事故的发生，将事故的损失减少到最小程度。台湾交通大学智能型系统控制整合实验室研究了一系列方法^[16, 17]判断道路的壅塞程度、交通流量和车速等信息，监控对象不仅包括一些大型车辆，还包括了一些类似摩托车的小型交通工具。

目标检测与跟踪系统不仅在总体上监控、评估、管理道路车流状况，也能辅助驾驶员驾驶车辆、减少交通事故发生。辅助驾驶系统通过安装在车身上的成像系统观察车外情况，估计前方车辆距自身的距离、两旁行人运动方向等重要信息辅助驾驶员快速作出决策。FORESTI等人^[18]结合目标识别算法和跟踪算法搭建了早期的驾驶系统，并在一系列复杂场景下的试验中获得较好的结果。随后的高速公路辅助驾驶系统^[19]综合考虑颜色、边缘和运动信息检测道路边界、标示牌和相邻车辆，估计场景中的各种信息之间相互关系。国内的清华大学2005年成立汽车电子实验室开发的车载视频安全驾驶辅助系统通过前向摄像机计算车辆对于中央车道线的偏离距离，前方行人、车辆并及时给出报警信

号提醒驾驶者注意安全。常用的成像系统基于普通的摄像机或是低端红外相机，而一些航空和卫星成像则采用更高分辨率相机。由于拍摄环境恶劣、距离较远，造成图像比较模糊，目标区域太小无法分辨。对于这种情况下的目标检测与跟踪，更具有军事应用价值。浙江大学的刘济林老师也随后在2006年开发了一套基于立体视觉的客流统计系统。该系统基于窗口的匹配算法计算视差深度，并以感知聚类为指导同单目图像处理的融合，有效解决了图像中背景干扰及与人体躯干误判等问题、增加深度方向的可靠性、提高了跟踪的准确度^[20]。

多年来山东大学的常发亮老师与柯晶老师一直研究基于双目视觉的运动目标检测与跟踪。常老师2006年时将单目中的camshift算法应用到双目视觉中，以检测场景中的运动目标并识别、跟踪这些运动目标^[21]。2007年时又开发了一种基于分层网络最小割的立体匹配方法对运动目标进行检测，将捕获图像进行两层金字塔分解，对顶层图像对采用最小割全局最优搜指导下层图像的匹配区域，进而得到运动目标所在区域^[22, 23]。之后，柯晶老师从改进的Harris交点算法改善“L”型角点的检测效果，提高人体检测的准确性跟踪场景中的人体运动^[24]。最近，他们又利用之前的Harris检测方法对角点进行立体匹配得到精确的匹配点，然后这些点的约束下对非角点像素进行基于区域相关的立体匹配，得到整体稠密的视差图^[25]。

哈尔滨工业大学机器人研究所用一种改进的人体检测方法，在保证定位精度和可靠性不变的前提下，减小了图像处理过程中的计算量，缩短了整个系统的计算开销，建立了有效地视觉跟踪系统^[26, 27]。除了检测人体以外，立体视觉也能应用到红色番茄的自动、快速识别中。该系统能帮助机器人识别果实、确定果实的空间位置，并最终指导机器人完成采摘过程^[28]。

每年，国际上举行多个会议介绍和讨论目标检测与跟踪方法与系统的最新进展，其中影响力较大的有：IEEE International conference on Computer Vision and Pattern Recognition (CVPR); IEEE International conference on Computer Vision (ICCV); European Conference on Computer Vision (ECCV) ; Asian Conference on Computer Vision (ACCV); IEEE International Conference on Image Processing (ICIP); International Conference on Pattern Recognition (ICPR)等。此外，相关的期刊也积极组织special issue讨论一些新方法和新问题的现状。在国内，中国自动化学会和中国科学院自动化研究所依托中科院自

动化研究所模式识别国家重点实验室多次举办智能视觉监控学术会议，讨论图像序列在动态场景中的目标进行定位、识别和跟踪，以及此基础上的分析和判断目标的行为。一些企业的研究院也积极从事检测与跟踪方面的研究工作，微软亚洲研究院（Microsoft Research Asia）2007年在北京举办了ICCV国际会议，OMRON公司也在西安、杭州等高校中举办传感与控制国际会议。

为了对各种算法进行公平的评价，一些学者和机构建立不同的公共测评库，提供统一的评价平台和评价算法测试算法的性能。IEEE于2000年三月在法国Grenoble举办IEEE International Conference on Automatic Face and Gesture Recognition中推出第一个跟踪与监控的评测平台PETS2000（First IEEE International Workshop on Performance Evaluation of Tracking and Surveillance）^[29]。在随后的2001年CVPR中公布了第2个评价数据库PETS2001^[30]，其中包含了5个在不同光照变化、遮挡、活动条件下的训练集和测试集。截止到2009年，PETS的测试集已经更新到了PETS2009测试集^[31]，其日趋完善的评价数据被广泛用于新算法的性能评价中。另一个常被使用的测试平台由AVSS^[32](IEEE International Conference on Advanced Video and Signal based Surveillance)提供，测试集含有图像数据和语音数据，并区分了同一场景的不同难度测试集。值得一提的是，最新的ASSV2009^[33] 测试中包含了美国标准技术局（National Institute of Standards and Technology）的公开评价数据作为标准测试数据，以提供更权威的评价指标。此外，一些知名大学也提供公共测评库，如俄亥俄州立大学的OTCBVS 数据集合提供一些红外数据测试集；AMI公司多模式的大型视频会议测试数据集；国内的复旦大学在accv2007会议上公布了一套行人检测的评价数据库和相应检测算法^[34]。

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Bhattachary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 当前存在的主要问题

近些年来, 更多的相关方法在上节所介绍的理论框架基础上进行逐步拓展, 期望能在各种实验条件下获得更好的性能。总体来说, 目前能在实际环境中使用的监控系统还比较少见, 分析其主要原因是: 监控系统中还存在一些复杂问题需要更好地解决方案。尤其是以下三个问题十分棘手: 监视场景中光照的变化; 目标与其阴影之间的区分; 在线的多目标检测跟踪以及遮挡处理。对于这几个问题的现有解决方法归纳如下:

1. 监视场景中光照变化是实际系统最常遇见的问题。由于目标的颜色特征与光照之间存在一定关联, 所以目标颜色会随着光照的变化而发生改变,

从而造成跟踪算法丢失目标。为此，研究者们一直都在寻找更好的方法来获得比较稳定的颜色特征。在文^[43]中，常见的RGB颜色空间经过归一化操作减少光照对目标颜色的影响、提高跟踪的鲁棒性。Moreno-Noguer^[44]则提出将3维颜色空间投影到2维的Fisher Plane平面上以获得对于光照不变的自适应颜色空间(A Target Dependent Colorspace)。在文^[45]中，一个评价函数在多个候选的颜色空间中在线挑选出最优的颜色空间对目标进行跟踪。

2. 目标的阴影同被跟踪目标有一样的运动特征，常常会被错误地认为是前景的一部分。这将引起检测算法和跟踪算法将认为是目标的一部分而产生错误的处理结果，如将阴影当做目标进行监控，或是将多个目标当成一个目标，甚至是目标的丢失。很多研究者提出不同的方法来减少阴影在监控系统中带来的负面作用。在文^[46]中，开发了一个非参数的统计方法对背景建模，通过多个阈值分离出背景、前期、阴影和高光区域。而文^[47]中介绍了一种检测flat surface阴影的方法以最大有验概率分类为基础，并采用阴影空间上的约束为条件，提取最大概率的阴影区域。考虑到阴影不连续与背景不连续之间的差别，启发式方法在文^[48]中被用于分离前景目标和阴影。这些经典的目标阴影检测方法在关于阴影检测的综述^[49]中进行了全面的对比和详细的评测。最近，文^[50]建立一个物理模型描述阴影在太阳和蓝天的光照下的反射情况，并通过多个子方法逐步精化以得到一个最优检测结果。
3. 目前大多数算法主要处理单一目标的跟踪问题，而处理多目标时，算法的计算量和时间复杂度会相应增加从而导致性能急剧下降。另一方面，实际场景中多目标之间的相互遮挡、合并、分裂以及其它复杂问题也都是跟踪算法需要处理的问题。早期的多目标跟踪^[51, 52]是利用视频之间的帧差判断变化较大的像素，以此来将场景中的所有移动目标作为一个整体在活动轮廓模型的框架下进行统一处理。随后，Maccormick在粒子滤波的框架下提出基于概率互斥原则(Probabilistic exclusion principle)的观察模型以推理多目标遮挡之下原目标的最优表达。另一个多目标的跟踪策略则是基于图论和最优化原则下的，多目标在相邻帧之间的对应关系以实现对群体目标的跟踪。

以上介绍的三个问题普遍存在于监控系统中，而现有方法只能在特定条件下解决其中的一部分问题，因此在实际环境中能大规模应用的系统还很少见。为了能满足社会对智能监控系统日益重视的要求，一些基于新硬件设备的方法也逐步出现。其中，红外成像仪和多摄像机的监控系统受到较大关注，相关的处理算法也经常发表在国内外重要学术期刊上。红外成像利用红外线与温度的关联关系，将场景中的热源以图像的形式表达。成像仪不处理可见光信号，回避了光照、阴影等监控系统中常见问题，但他对背景温度敏感，多目标重叠与遮挡也无法妥善解决。相反的，多摄像机成像系统从多个视角观察监控场景，获得较单摄像头更全面的信息，尤其在处理多目标的遮挡问题时更为有利。同时，光照的变化同时被在多个摄像机记录，其负面影响能被算法互相抵消。考虑到场景中的阴影通常位于地面或墙壁上，多摄像机通过3D测量算法能通过距离和高度消息过滤这些区域而减少阴影的干扰。

随着摄像机的价格越来越低，基于多摄像机的监控系统的成本已能被许多应用项目所接受。一些相关的实验系统和演示系统也经常出现国际学术会议上。在文^[53]中介绍的双摄像机立体成像系统，深度信息能轻松地将目标与背景分割开来，然后结合肤色检测模块与人脸检测模块对不同距离内的人体进行跟踪。类似地，文^[54]提出了一种能对非遮挡、无纹理的平面区域计算深度信息的密集立体估算算法，并能快速估计前景区域、计算目标运动轨迹。Anurag Mittal利用多个相隔较远的摄像头和基于区域的立体视觉算法^[55]寻找属于特定目标的3D点，以检测和跟踪多个行人目标。为了解决移动机器人在复杂背景下人体检测的问题，文^[56]表述一种基于在完整深度图中应用自适应尺度滤波方法在视差图中提取一些人体头部的候选区域，然后参照地板平面过滤一些噪音产生的伪目标，最后确定头部的中心位置。尽管这些基于系统视觉的系统都能实现对多目标的检测与去跟踪，但他们需要借助一些如人脸检测、颜色分析和边缘分割等辅助模块或算法解决检测跟踪的问题。而这些多种算法在实际应用的整合中存的许多冲突影响系统的不稳定，而单一的理论框架更能获得更高的准确性和有效性。

大多数立体视觉的处理算法都基于稠密深度图的基础上，通过将目标远近距离映射为不同的灰度值，以快速的分割算法分离出监控场景中的前景区域和背景区域。事实上实现场景中存在的匀质无纹理区域具有相同的颜色，视差估

计算法很难准确地估计出这些区域的距离信息。另一方面在视差图中分割出的区域对应着空间中的正常目标，一些遮挡目标由于对应区域的面积较小致使监控系统无法准确检测和跟踪。与其费力地尝试计算整个场景的完整深度图^[54]，不如在场景中选取有限个特征点计算其空间坐标既节省计算量也提高了深度信息的准确性。因此一种更准确更鲁棒的检测与跟踪方法应当考虑计算场景特征点深度信息，将监控环境中目标与监控系统的特征点进行对应从而确定目标的数量、定位目标的位置并提取目标的轨迹。

近些年，也有许多学者对立体视觉系统的现状和进展进行全面的综合论述，如：2002年对图像匹配策略的综述^[57]，2004年对立体视觉进展的综述^[58]，摄像机自标定的综述^[59]等。基于知识的视觉测量综述。

1.4 本文的主要工作及研究内容安排

本文主要是针对基于立体视觉多摄像机的监控系统进行研究：构建双目视觉的立体成像硬件系统，提出基于核方法的聚类算法与基于能量泛函的轮廓算法分别用于检测、跟踪空间中多目标运动轨迹和提取单一目标的连续完整轮廓。所研究内容涉及计算机视觉、最优化理论、模式识别、参数和非参数概率方法和微分几何等多个研究方向，最终形成一套能下多种复杂场景下工作的智能监控系统。

研究工作包括了硬件成像系统和软件算法系统两大部分，其中硬件系统由两台摄像机、两张图像采集卡和一台计算机构成，而软件系统由多目标检测与跟踪算法和单目标的轮廓跟踪算法构成。搭建立体成像硬件系统的流程如图1.1所示：

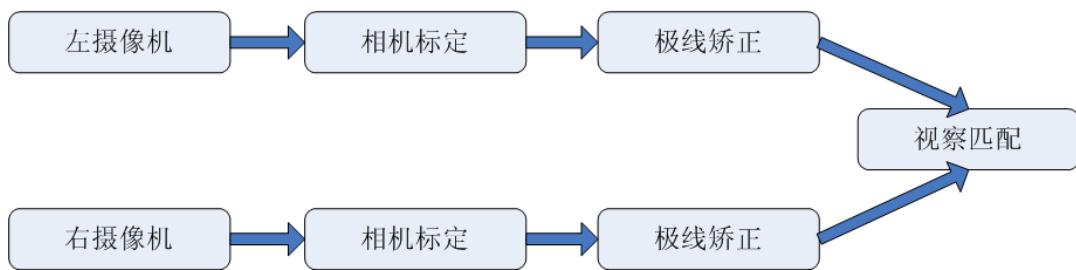


图 1.1 系统的硬件构造流程图

1.1 The construction flowchart of system's hardware

第二章主要介绍相机数学成像模型及内外参数的标定方法。借助于已知标靶的在不同位置和角度所拍摄的图像数据，相机标定算法能估计出最优的相机内外参数矩阵系数和光学镜头的畸变系数。

第三章在相机标定的基础上再次利用标靶的图像数据拟合两个摄像机之间的对应关系和极线约束方程。根据计算得到的基础矩阵，矫正两个摄像机的成像平面构成虚拟的理想双目成像系统，使得相机的成像平面在同一平面上且相机的内部参数相等。提出的双目匹配算法采用多尺度方法由总体到局部的逐步细化匹配算法，既能保证全局快速匹配同时又能获得极高的局部匹配精度。硬件系统的构建方法为软件处理过程提供了较准确的计算数据，为后续的目标检测与跟踪打下良好的硬件基础。

立体视觉成像系统同时拍摄两张场景图像匹配视差后可计算出相机坐标系下的三维坐标，软件系统对硬件系统采集到的数据进行后处理以确定监控场景中目标的数量、目标的位置以及运动轨迹。目标检测模块由核聚类算法将空间特征点进行聚类操作生成特征点的集合与场景中的目标相对应，当聚类超过一定规模时可认定目标存在性。尔后多目标跟踪算法预测目标新位置产生对应聚类集合更新目标运动轨迹，同时单目标跟踪算法针对单一目标精确确定其轮廓和对应区域提供更详细的信息。软件系统的运行流程如图1.2所示：



图 1.2 系统的软件处理流程图

1.2 The processing flowchart of system's software

第四章提出基于核函数的聚类算法将空间离散点聚合成多个集合。由目标的空间几何特征构造的近似核函数结合非参数概率估计方法将特征点形成的离散空间转换成连续空间，mean shift迭代优化方法在不计算具体概率值的前提下根据核函数梯度拟合出空间梯度，并自适应的控制移动步长逐步收敛到空间中的局部最大值。以每个离散点为起始点寻找其对应的空间最大值，而收敛到相同最大值的特征点构成一个聚类集合，并以最大值的位置和方向作为聚类的

位置和方向，此时监控场景中的目标即可由这些聚类所代表。至于多目标的跟踪则是在前一帧目标位置的附近寻找当前帧的局部最大点问题。该算法的实验证明了其在室内、室外不同背景、不同光照、不同姿态条件下的有效性和鲁棒性。

第五章提出了基于能量泛函的轮廓跟踪算法在连续多帧中提取目标对应的精确轮廓和区域。算法根据目标特征的不同特点将颜色、纹理、帧差和边缘等特征分为区域特征和边缘特征两大类。收集目标和背景像素组成正负样本集提取颜色特征和纹理特征，并对这些特征的发布概率构成条件概率密度函数。贝叶斯模型将目标和背景的条件概率转换成像素对目标和背景的后验概率，产生像素的能量函数。积分给定目标轮廓内部和外部区域的像素能量作为曲线的能量泛函。而帧差和边缘融合测量图像域中的移动边界概率，并在测地线模型中整合新的测度函数形成轮廓的边界能量泛函。自适应的权重函数平衡区域能量泛函和边界能量泛函在整体泛函中的权重，梯度下降流演化初始化曲线直至其收敛到目标区域上提取精确轮廓。此外，目标的运动特征在连续两帧中平移目标曲线构成较优的初始化曲线，减少计算的迭代次数和收敛时间。不同的实验也证明了算法在低对比背景、严重遮挡和移动摄像等情况下能在连续帧中提取目标的精确轮廓和所在区域。

第六章对整个工作给出总结指出其中存在的不足，以及未来有待加强以及可进一步提高的研究工作和方向。

第二章 单相机标定

2.1 摄像机原理及数学模型

通常，一台摄像机主要由光学镜头、图像传感器（CCD或CMOS）、数据处理芯片和封装外壳四部分构成。光源发出的光线在物体的表现反射后，光学镜头将入射的光线聚焦到图像传感器上产生不同的模拟电信号，然后数据处理芯片采样后生成数字信号并通过通信接口输出数字图像信号。整个过程涉及到三维现实世界中的物体、二维传感器平面（或称为成像平面）和二维数字图像，即分别对应着三个不同的坐标系统：世界坐标系统、相机坐标系统和图像坐标系统。整个成像过程可用图2.1所示的理想小孔相机模型来说明：

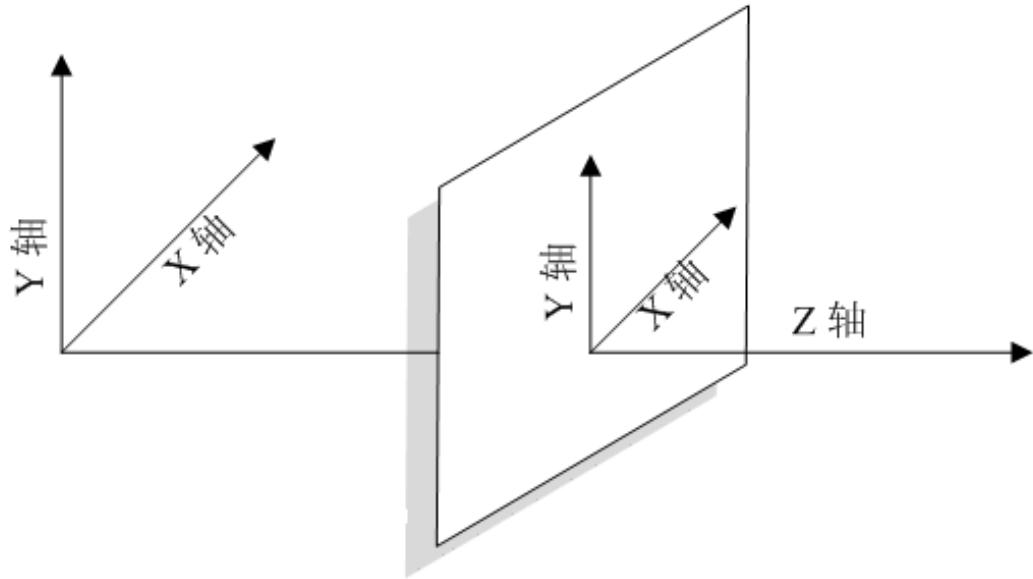


图 2.1 摄像机成像原理图

2.1 the imaging schematic diagram

输入信号来自于世界坐标系统而输出的图像数据是图像坐标系统，将这两个坐标系统和整个成像过程的数学模型用一个简单的线性数学等式关联为：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} (\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \mathbf{T}) \quad (2.1)$$

其中 u, v 为数字图像中像素的坐标。世界坐标系下目标点 x_w, y_w, z_w 通过旋转矩阵 $\mathbf{R} = [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]$ 和平移向量 \mathbf{T} 转变为相机坐标系中，再由相机的内部参数

矩阵 $\mathbf{A} = \begin{bmatrix} f_x & \gamma & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$ 投影到图像坐标系中。内参矩阵表达了照相机内在不变的一些性质，如 c_x 和 c_y 是相机平面的原点， f_x, f_y, r 分别是两个坐标轴的放大系数和关联系数。与内参数矩阵不同，旋转矩阵 \mathbf{R} 和平移向量 \mathbf{T} 与世界坐标的选取有关。选择不同的原点和坐标方向将产生不同的 \mathbf{R} 和 \mathbf{T} 与相机坐标对应，因此他们也常被称为相机的外参数。

实际上这个过程中也整合了成像平面的坐标系，它其实做为一个中间坐标系在世界坐标系和图像坐标系之间架起中间桥梁。世界坐标由转矩阵 \mathbf{R} 和平移向量 \mathbf{T} 转为成像平面坐标系，再经过内参数矩阵 \mathbf{A} 与图像坐标系对应。

尽管式(2.1)为成像过程建立了对应的数学模型，但实际环境中的光学镜头由于成像平面上不同区域对光线的弯曲程度不一致，造成数字图像从画面中心到画面边缘的扭曲程度逐渐增强。这些径向畸变和切向畸变通常不仅使摄像机的成像质量下降，也给成像过程带来较大误差而造成式(2.1)无法成立。为了更准确地表达整个成像过程，当世界坐标系 $[x_w \ y_w \ z_w]$ 旋转平移到相机坐标系之后，在成像平面上的投影 x_c, y_c 被畸变系数 $\mathbf{K} = [k_1 \ k_2 \ k_3]$ 映射到：

$$\begin{aligned} x'_c &= x_c(1 + k_1r^2 + k_2r^4 + k_3r^6) \\ y'_c &= y_c(1 + k_1r^2 + k_2r^4 + k_3r^6) \end{aligned} \quad (2.2)$$

其中 $r^2 = x_c^2 + y_c^2$ 是成像点到成像中心的距离， k_1, k_2, k_3 分别是二阶、三阶、四阶的径向畸变系数。考虑到切向畸变对于畸变的影响较少和计算过程的便捷性和复杂度，在畸变公式中忽略切向畸变只考虑径向畸变。这些畸变系数与系数矩阵 \mathbf{A} 一样，都与世界坐标的选取无关而只与相机自身的特性相关，因

此畸变系数 \mathbf{K} 也被称为内参数。

从成像公式（2.1）也可知：为了从已知的二维数字图像中恢复三维场景，这些相机参数中不可缺少的计算元素必须预先确定。计算机视觉中称这个参数确定的问题为相机标定（camera calibration）^[60]，即估计摄像机几何光学的内部参数以及与世界坐标系关联的外部参数。最早相机标定问题是在成像测量领域^[61, 62]，随后被逐渐引入到计算机视觉中解决一些视觉问题，如自标定机器人、立体成像系统^[63–66]。早期标定方法分为两大类：

- 标靶标定法利用一个已知形状与坐标的物体作为标靶，构造误差函数求解最优参数以使成像误差达到最小。通常标靶物体由两、三个互相垂直面构成，制造复杂而且要求安装精细。
- 自标定法则是保持摄像机拍摄场景不变，精确控制摄像机的移动过程对比前后图像的变换来估计最佳的内部参数。但该方法在标定过程需要估计较多的未知参数，同时标定算法的计算可靠性较差。

Tsai^[67]在1987年所提出的两步标定法被认为是相机标定重要理论工作，被广泛应用于早期的相机标定过程中。该方法基于径向畸变点与真实点之间共线的假设模型采用两步策略分别求解相机的内外参数，第一步首先求解外参数 \mathbf{R} 和平移向量 \mathbf{T} 中的两个分量，然后再求解相机的内部参数 \mathbf{A} 。大部分计算过程都线性方程，因此求解参数的计算复杂度较低求解速度较快，但该算法在标靶平面与成像平面处于平行时求解的误差较大稳定不高^[68, 69]。1999年，张正友^[70, 71]提出的平面标定方法采用简单的平面物体作为标靶，多次拍摄不同位置的标靶图像以非线性优化方法来估计摄像机参数。平面标靶的标定方法结合了传统标靶标定法与自标定法的思想，通过简单的二维平面标靶的简单手动操作就能获得较高的标定精度。国内的中科院自动化所的胡占义在标定方面有诸多的研究成果。早期提出了线性摄像机自标定方法^[72]，二次曲线的纯旋转摄像机自标定方法^[73]，无穷远平面的单应矩阵摄像机自标定^[74]，条件数的摄像机自标定^[75]。在此基础上，他们又采用比网格更简单的圆形^[76, 77]作为标靶，并且不需要测量棋盘网格的物理长度和建立特征对应关系，进一步的降低了相机标定的复杂度实现了完全自动化的标定过程。而吴毅红老师将单一标靶扩展为两

个平行圆的多标靶模式，利用摄像机模型对于准仿射不变性特点计算曲线的交点得到圆环点的图像，无需任何匹配估计相机内外参数。

2.2 平面标定方法

尽管现在新的标定方法不断出现，寻找更简单、更稳定、更准确的标定方法，但仍没有对比试验表明他们比张氏标定法更精确和更鲁棒。在张正友标定方法中的使用平面标靶，可使用普通打印机打印网状棋盘格并固定在平板平面构成。通过手动随意的移动标靶拍摄对应的图像，通过Levenberg-Marquardt (LM) 优化算法经过十多次即可收敛获得相机的内部参数和外部参数。相比于已有的标靶标定方法，该方法中的标靶加工简单，使用方便；而对于自标定方法，它又具有极高的稳定性和鲁棒性。因此，本文所述的目标检测与跟踪系统也采用该方法标定相机参数，并在此基础上搭建立体视觉系统。为了能更好的介绍本文所述系统，现将张氏标定方法做一个简短的介绍。

2.2.1 建立标定约束条件

将图像坐标和世界坐标表达为齐次坐标，式 (2.1) 重写为：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{T} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2.3)$$

在不影响计算结果的前提下，可以人为的选取标靶的平面为世界坐标系平面，则标靶上所有特征点的Z轴坐标为0即 $z_w = 0$ 。如此可减少一个相机外参数变量和降低计算过程的复杂度，将式 (2.3) 简写为：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \quad (2.4)$$

由上式可知，从世界坐标系平面到成像平面的映射事实上是单应性变换 (homography)。换言之，图像坐标系的点与标靶平面的点可用单应矩阵关

联 \mathbf{H} :

$$\mathbf{H} = s^{-1} \mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} \quad (2.5)$$

另记 $\tilde{u} = \frac{1}{\bar{w}} \mathbf{h}_1^T \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$, $\tilde{v} = \frac{1}{\bar{w}} \mathbf{h}_2^T \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$, $\tilde{w} = \mathbf{h}_3^T \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix}$ 。世界坐标系

点 $\begin{bmatrix} x_w & y_w \end{bmatrix}$ 在 \mathbf{H} 矩阵的映射下得到估计的像素点 \tilde{u}, \tilde{v} , 以估计像素与真实像素 $\begin{bmatrix} u & v \end{bmatrix}$ 之间的距离构造评价函数

$$f(\mathbf{H}) = \sum_i (u_i - \tilde{u}_i)^2 + (v_i - \tilde{v}_i)^2 \quad (2.6)$$

为了最小化距离函数并求解出最优的矩阵 \mathbf{H} , 常用的最小二乘拟合法不能适用于这个非线性问题。因此, 只能采用一些非线性的迭代优化方法如的Levenberg-Marquardt算法^[70, 78]和Newton Raphson(NR)算法^[61, 79]从初始解开始逐步逼近最优解。

由矩阵 \mathbf{H} 的定义可知, 它是相机内参数与外参数的组合, 再根据旋转矩阵的正交性质将外参数 $\mathbf{r}_1, \mathbf{r}_2$ 用 $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ 表达为:

$$\begin{aligned} \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 &= 0 \\ \mathbf{h}_1^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_1 &= \mathbf{h}_2^T \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{h}_2 \end{aligned} \quad (2.7)$$

由以上这六个约束式, 再将 $\mathbf{A}^{-T} \mathbf{A}^{-1}$ 看作为一个整体变量, 用其中的六个未知数构造成向量 $\mathbf{b} = [B_{12} \ B_{12} \ B_{22} \ B_{13} \ B_{23} \ B_{33}]$, 则约束式 (2.7) 能重写为线性表达式:

$$\mathbf{v} \mathbf{b} = 0 \quad (2.8)$$

其中 \mathbf{v} 由 $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ 中的元素根据约束式 (2.7) 组合而成

2.2.2 估计相机内外参数

根据矩阵 \mathbf{A} 的定义以及它与矩阵 \mathbf{B} 之间的关系 ($\mathbf{B} = \mathbf{A}^{-T} \mathbf{A}^{-1}$) 可知:

$$\begin{aligned}
B_{11} &= \frac{1}{f_x^2}, \\
B_{12} &= -\frac{\gamma}{f_x^2 f_y}, \\
B_{13} &= -\frac{c_y \gamma - c_x f_y}{f_x^2 f_y} \\
B_{22} &= -\frac{\gamma^2}{f_x^2 f_y^2} + \frac{1}{f_y^2} \\
B_{23} &= -\frac{\gamma(c_y \gamma - c_x f_y)}{f_x^2 f_y^2} - \frac{c_y}{f_y^2} \\
B_{33} &= \frac{(c_y \gamma - c_x f_y)^2}{f_x^2 f_y^2} + \frac{c_y^2}{f_y^2} + 1
\end{aligned} \tag{2.9}$$

利用这六个关于**B**的等式，可进一步求解出矩阵**A**中的六个未知数，即

$$\begin{aligned}
c_y &= (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \\
s &= B_{11}/[B_{11}B_{33} - B_{13}^2 - f_y(B_{12}B_{13} - B_{11}B_{23})] \\
f_x &= 1/\sqrt{sB_{11}} \\
f_y &= \sqrt{B_{11}/s(B_{11}B_{22} - B_{12}^2)} \\
\gamma &= -sB_{12}f_x^2 f_y \\
c_x &= \gamma c_y / f - sB_{13}c_y^2
\end{aligned} \tag{2.10}$$

在估计出内参数矩阵**A**并依照单应矩阵的构造等式 $\mathbf{H} = s^{-1}\mathbf{A} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix}$ ，可估计出外参数矩阵 $\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} = s\mathbf{A}^{-1}\mathbf{H}$ 。

理想情况下，相机的参数可通过以上标定过程获得，但实际过程中光学镜头存在的畸变问题将造成以上标定出的参数仍存在较大的误差，不能准确地描述成像中的三维场景与二维图像之间存在的紧密联系。因此，光学镜头的畸变系数的估计也是不能回避的问题。以上估计的内参数矩阵**A**可以将图像系下的点 $[u \ v]$ 映射到成像平面点 $[x'_c \ y'_c]$ 上，但估计出来的外部参数矩阵 $\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix}$ 将世界坐标系下的点 $[x_w \ y_w \ z_w]$ 投影到成像平面为 $[x_c \ y_c]$ 。这两者之间的误差就是由相机畸变系数所造成，其过程可根据畸变模型（参见式（2.2））表达为：

$$\begin{bmatrix} x_{c,0}r^2 & x_{c,0}r^4 & x_{c,0}r^6 \\ y_{c,0}r^2 & y_{c,0}r^4 & y_{c,0}r^6 \\ \dots & \dots & \dots \\ x_{c,n}r^2 & x_{c,n}r^4 & x_{c,n}r^6 \\ y_{c,n}r^2 & y_{c,n}r^4 & y_{c,n}r^6 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} = \begin{bmatrix} x'_{c,0} - x_{c,0} \\ y'_{c,0} - y_{c,0} \\ \dots \\ x'_{c,n} - x_{c,n} \\ y'_{c,n} - y_{c,n} \end{bmatrix} \quad (2.11)$$

同求解矩阵 \mathbf{H} 方法类似，畸变系数的拟合过程也采用Levenberg-Marquardt算法使在成像平面上的两个映射点之间的总体误差最小。

2.3 相机标定实验及评价

本章所述系统中的标定标靶均为如图2.2所示的9*7棋盘，且每个方格打印后宽度为28mm。

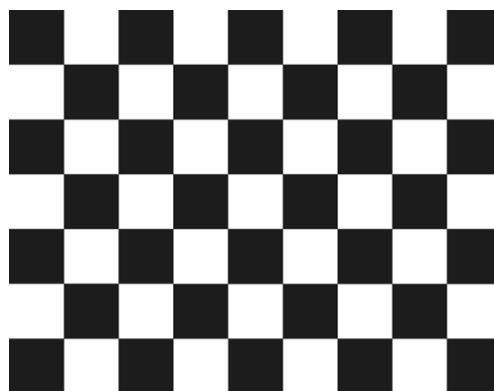


图 2.2 相机标定的标靶图

Calibration board 2.2 Calibration board for camera calibration

将打印的图像平整粘贴到平板面上作为标定用的标靶，使用两个摄像机同时拍摄手动控制的标靶在不同位置和角度上的多个图像作为标定图像（如图2.3和图2.4所示）。

通过上节所述的标定法尽管能先后标定出内外相机参数，但求解过程中两次用到可LM算法求解非线性优化方程。在第一步求解矩阵 \mathbf{H} 时，不恰当的初始值会造成LM算法算法收敛到局部最优解，从而造成随后估计的参数与真实



图 2.3 左摄像机拍摄的标定图

2.3 Calibration images from the left camera

值有很大偏差。为了验证标定的结果是否正确，在标定完成后需要进行一次重投影操作评价标定的质量高低、判断标定的相机参数是否有效。

重投影操作即是将标定时标靶的世界坐标系点用成像公式（2.1）和畸变公式（2.2）重新投影到图像坐标系中，计算这些投影点与真实图像像素点之间的距离。如果所有的重投影点都与真实点距离较小时，可认定标定成功；反之，则说明标定失败并需要重新拍摄标定图像进行标定。重投影操作的计算流程为：

步骤-I 用相机外参数将标靶上的世界坐标系点坐标转换到相机坐标下，

即：

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ 1 \end{bmatrix} \quad (2.12)$$



图 2.4 右摄像机拍摄的标定图
2.4 Calibration images from the right camera

步骤-II 归一化坐标计算投影平面中特征点坐标:

$$\begin{aligned} x_c &= x/w \\ y_c &= y/w \end{aligned} \quad (2.13)$$

步骤-III 由畸变模型 (2.2) 估计真实的成像平面坐标。

步骤-IV 内参数矩阵 \mathbf{A} 将成像平面坐标映射到图像坐标, 获得重投影的点

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \\ 1 \end{bmatrix} = \mathbf{A} \begin{bmatrix} x_c \\ y_c \\ 1 \end{bmatrix} \quad (2.14)$$

步骤-V 计算重投影的点与真实点之间的平均值距离和偏差值 (见式 (2.6)), 评估标定参数的质量。当均值或偏差超过规定阈值时, 可

认为标定失败；反正，则证明标定成功。

图2.3和图2.4中的标定图计算得到的相机参数重投影后的结果如图2.5 所示，该图以真实点为坐标原点展示投影点与真实点的偏差分布情况（同一张标定图的投影结果用同样的颜色显示）。

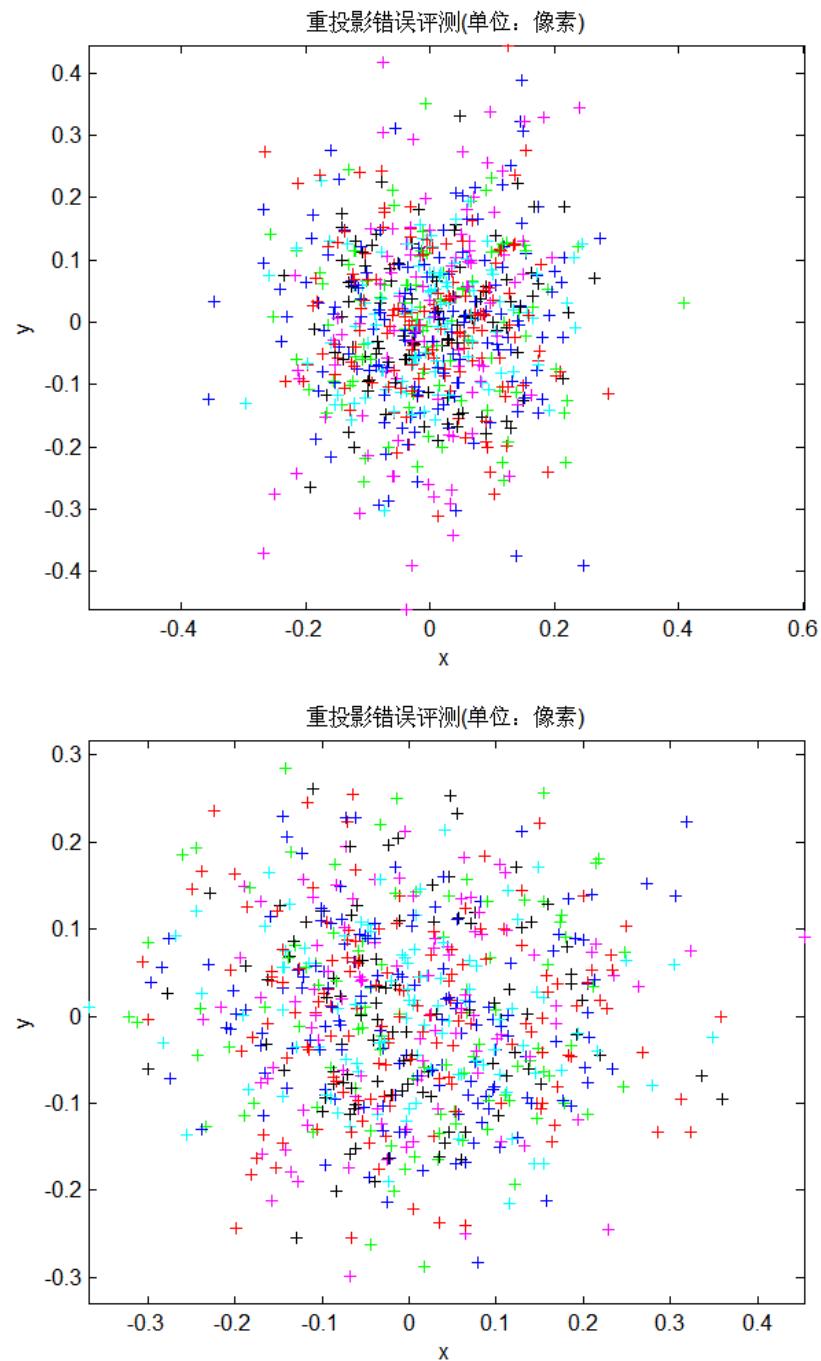


图 2.5 重投影点偏差分布图
2.5 The error distribution of re-projection points

第三章 双目视觉的极线几何与矫正

第二章介绍的摄像机标定算法确定了相机的内在性质以及外部世界坐标系与相机坐标系之间存在的紧密联系，于是一些研究方法尝试着从二维图像中重建场景或物体的三维结构。但目前这些方法都必须利用一些场景的先验知识，在给定的约束条件下开始三维重建过程。从二维到三维中存在的不确定性难题也可从数学模型上给出理论上的解释，将相机的内部参数和外部参数整合在一起表达为投影矩阵 \mathbf{P} ，相机的成像公式（2.1）可简单地重新表达为：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (3.1)$$

从式（3.1）可知，必然存在多个世界坐标点 $\begin{bmatrix} x_w & y_w & z_w \end{bmatrix}$ 以不同的尺度系数 s 与同一个图像像素点 $\begin{bmatrix} u & v \end{bmatrix}$ 相对应。

图3.1对该问题给出了比较直观的解释：虽然空间中的A点和B点位置不同，只要这两点与相机光心都位于同一条直线上经过投影后必将对应着同一个图像像素点。在由二维图像进行三维重建时，通过该投影点是无法确定它究竟与世界坐标系下的A点对应还是与B点相对应。因此，三维重建在没有场景知识或条件约束的情况下是不能完成的任务^[80, 81]。

3.1 双目视觉的极线约束

通过以上对摄像机成像原理分析可知：解决三维空间的重建问题需要二维图像以外的其他数据作为附加条件，在多个世界坐标系对一个图像像素点对应的多对一问题中为像素点确定空间中唯一相对应的点。在计算机视觉领域有很多重建方法，其中比较常用的方法包括：从运动中重建^[82-85]、从阴影中重建^[86-88]、从多视角中重建^[80, 89-91]。运动重建要求重建物体与摄像机之间存在相对运动，而空间中不同点移动后在图像位置上的变化存在差异能帮助完成重

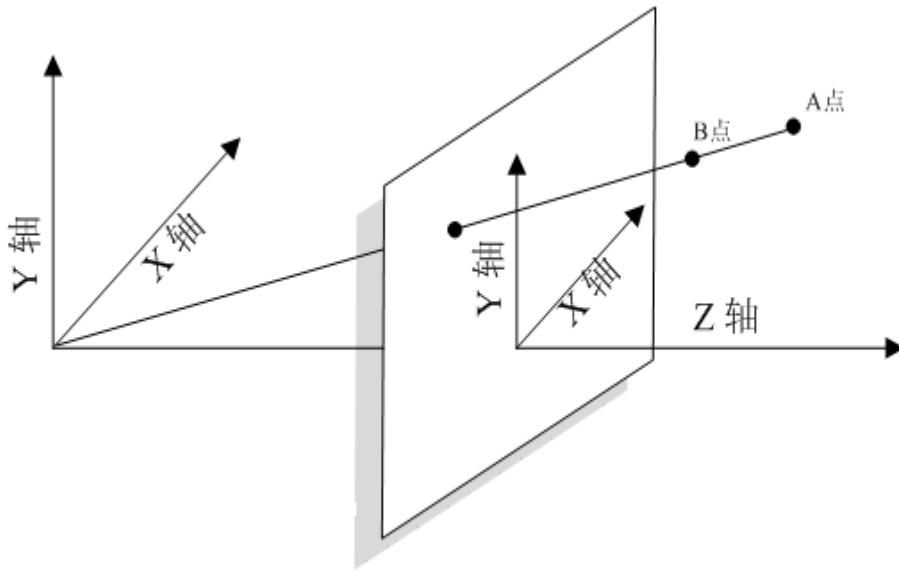


图 3.1 三维重建的不确定问题

3.1 The uncertainty problem in three-dimensional reconstruction

建过程。阴影中重建则是通过物体上点的在光照下明暗变化关系估计其空间形状。相对于前两种方法，多视角重建方法以增加成像设备为代价换取更快的重建速度、更广泛的应用环境、更鲁棒的处理结果。该方法与生物视觉成像原理最相似，能实时获得场景中的三维信息、直接估计目标的空间位置。基于多视角方法的这些优点，它一直被用于解决移动机器人的视觉问题。

双目视觉是最简单、最基本的多视角系统，它从两个不同方向观察同一个场景获取三维信息，这与人类的视觉模式最为相似。因此采用双目视觉的方法能有效地模拟人类对监控场景的处理过程，以检测监控目标的出现、跟踪其运动轨迹。假设两个摄像机同时拍摄同一个标靶获取标定图像，并由第二章所介绍的标定方法分别估计两个相机的内外参数，且标定图像经过矫正消除了光学镜头畸变引起的变形，则双目系统的成像过程可表达为：

$$c_1 \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = \mathbf{A}_1(\mathbf{R}_1 \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \mathbf{T}_1) \quad (3.2)$$

$$c_2 \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = \mathbf{A}_2(\mathbf{R}_2 \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \mathbf{T}_2) \quad (3.3)$$

联立这两式，同时消除世界坐标系参数 $\begin{bmatrix} x_w & y_w & z_w \end{bmatrix}$ 可得到两个摄像机图像之间的关联关系为：

$$\begin{bmatrix} u_1 & v_1 & 1 \end{bmatrix} F \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = 0 \quad (3.4)$$

其中 3×3 矩阵 F 关联两个摄像机中对应同一目标点的两个像素点，在计算机视觉理论中常被称为**基础矩阵**（Fundamental matrix）。世界坐标系下的目标点在两个摄像机成像平面中对应着像素点 $\begin{bmatrix} u_1 & v_1 & 1 \end{bmatrix}$ 和 $\begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix}$ ，必须满足基础矩阵 F 的极线约束^[92-94]。假设左摄像机中给出一点 $\begin{bmatrix} u_1 & v_1 \end{bmatrix}$ ，则根据极线约束式（3.4）可得：

$$\begin{bmatrix} u_1 & v_1 & 1 \end{bmatrix} F \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = l_1 \begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix}^T = 0 \quad (3.5)$$

右摄像机成像图像中的对应点 $\begin{bmatrix} u_2 & v_2 \end{bmatrix}$ 必在直线 $l_1 = \begin{bmatrix} u_1 & v_1 & 1 \end{bmatrix} F$ 之上。同理，如果给定右摄像机中的点 $\begin{bmatrix} u_2 & v_2 \end{bmatrix}$ ，则左摄像机成像图像的对应点 $\begin{bmatrix} u_1 & v_1 \end{bmatrix}$ 也必在直线 $l_2 = F \begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix}^T$ 上。

在这两个摄像机中也存在十分特殊的点在计算机视觉称它是极点(epipoles)

\mathbf{e}_1 和 \mathbf{e}_2 。由于无法为它在另一图像中找到对应点，因此极点与基础矩阵相乘的结果为0，也可表达为：

$$\mathbf{e}_1 F = 0 \quad (3.6)$$

$$F\mathbf{e}_2^T = 0 \quad (3.7)$$

此外，左右摄像机的极点还有另一个重要的性质是：两个摄像机中所有极线都相交于极点。极点在成像原理中事实上是两个摄像机与空间点的点所构成平面与两个成像平面的交点，或者是在左摄像机的光心 \mathbf{C}_1 在右摄像机上的投影点 \mathbf{e}_2 ；反之右摄像机的光心 \mathbf{C}_2 在左摄像机的投影点为 \mathbf{e}_1 。在如图3.2所示的双目视觉结构图上，两个摄像机坐标系的原点（即光心）和空间中的A点构成了一个平面，而该平面与两个相机成像平面的交线就是基础矩阵控制的极线。两个光心 \mathbf{C}_1 和 \mathbf{C}_2 与成像平面分别相交于极点 \mathbf{e}_1 和 \mathbf{e}_2 ，同时极线1和极线2也都经过两个极点。

基础矩阵反映了两个相机之间的关联关系，也是相机内参数和相机距离方位等众多因素的综合结果。只要两个摄像机的特性与距离方位不发生改变，基础矩阵保证也就相应地保持不变。计算基础矩阵的过程与相机标定十分类似，也同样需要利用第二章中提到的标定图像。3*3的矩阵 F 含有9个未知数，考虑到其乘以任意的尺度仍然满足式(3.4)中的约束条件，因此必须将矩阵归一化以固定最后一个参数为1，其余的8个参数中仍有7个为自由参数。为了求解基础矩阵中的这7个未知数最少需要7个求解方程，即需要在左摄像机和右摄像机中需要7个对应点作为条件构成求解方程。现在假定已知7个对应点的条件，首先给出简单计算公式求解线性方程组求解基础矩阵，将公式(3.4)改写成齐次线性等式的形式：

$$\mathbf{u}_i^T \mathbf{f} = 0 \quad (3.8)$$

其中：

$$\mathbf{u}_i = [u_{i1}u_{i2} \ v_{i1}u_{i2} \ u_{i2} \ u_{i1}v_{i2} \ v_{i1}v_{i2} \ v_{i2} \ u_{i1} \ v_{i1} \ 1]^T \quad (3.9)$$

$$\mathbf{f} = [F_{11} \ F_{12} \ F_{13} \ F_{21} \ F_{22} \ F_{23} \ F_{31} \ F_{32} \ F_{33}]^T \quad (3.10)$$

F_{ij} 为基础矩阵中第*i*行第*j*列的元素。再将7个点对组合在一起构成完整等式：

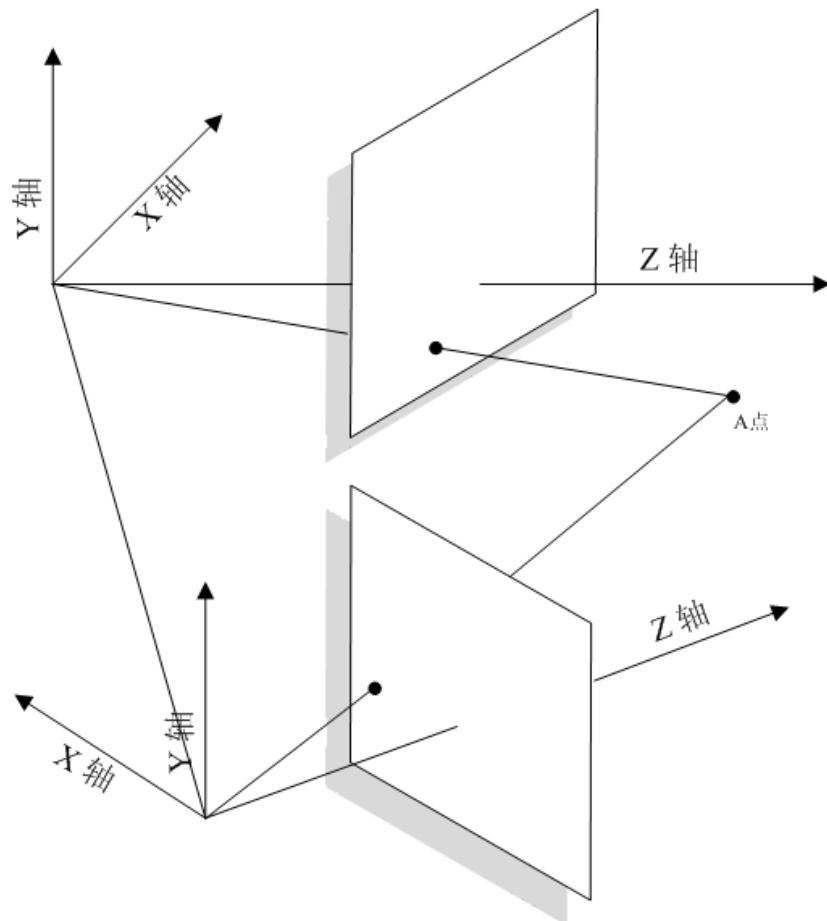


图 3.2 双目视觉成像原理图
3.2 the schematic diagram of binocular vision

$$\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{u}_5 & \mathbf{u}_6 & \mathbf{u}_7 \end{bmatrix} \mathbf{f} = 0 \quad (3.11)$$

将参数向量 $\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{u}_5 & \mathbf{u}_6 & \mathbf{u}_7 \end{bmatrix}$ 做奇异值分解可得到2个张成零空间的分向量 \mathbf{f}_1 和 \mathbf{f}_2 。整个零空间可由 \mathbf{f}_1 和 \mathbf{f}_2 对应的矩阵 F_1 和 F_2 线性组合成空间的任意元素：

$$F = aF_1 + (1 - a)F_2 \quad (3.12)$$

考虑到矩阵 F 的秩为2，则得到求解组合参数 a 的等式为：

$$\det [aF_1 + (1 - a)F_2] = 0 \quad (3.13)$$

最后由求解出组合参数 a 后，公式 (3.12) 能给出对应的基础矩阵解。

以上提到的7点方法虽然能求解出基础矩阵，但要求给出的计算数据 $\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 & \mathbf{u}_4 & \mathbf{u}_5 & \mathbf{u}_6 & \mathbf{u}_7 \end{bmatrix}$ 必须完全准确，不允许计算数据中含有任何对应错误和误差。这样苛刻的要求在特征点匹配过程中难以实现，因此也造成该方法在实际的求解过程中结果十分不稳定。在第二章所述单相机的标定过程中，每幅标定图包含了 $9*7$ 个特征点能构成63个方程，大大超过求解 F 所需要的方程数。通常比较简单的求解方法是使用最小二乘拟合来求解最佳系数矩阵，但考虑到已有的对应关系不一定都准确，其中可能出现一些错误的对应关系也存在一些对应关系的噪音，因此比较稳定的解决方法是采用随机取样调查 (random sample consensus) [95] 或是最小中值二乘法 (Least Median of Squares) [96–98] 将错误的匹配对计算的影响减少到最小程度，从而保持计算数据的准确性和有效性。最小中值二乘法当错误的匹配数据少于正确的匹配数据时能保证计算精度，而随机取样调查甚至能在匹配数据多于正确的匹配数据时仍保持较高的计算准确性。但考虑到本章所述的标定中使用的棋盘式标靶在匹配中绝大多数特征点匹配能够保证正确性，只可能少数点的匹配结果精度不高，再由于最小中值二乘法计算速度较快且当匹配错误低于50%时计算可以保持较高的稳定性，所以该方法对解决棋盘格标靶的基础矩阵是十分有效的。

假设第 i 个匹配点在约束式计算的误差结果为 $merr_i = \begin{bmatrix} u_{1i} & v_{1i} & 1 \end{bmatrix} F \begin{bmatrix} u_{2i} \\ v_{2i} \\ 1 \end{bmatrix}$ ，

则最小中值的目标函数为：

$$\min_i \text{median } merr_i^2 \quad (3.14)$$

最优的基础矩阵应当使整体计算误差的中值达到最小。理论上已经证明统计中值对错误的匹配结果和噪音是十分有效的计算方法，但该方法不能给出直接的目标函数表达式而且必须搜索所有可能的估计值。在计算中所用的计算数据有 $63*15=945$ 对，所有可能的组合包含有 C_{945}^7 种可能。直接如此庞大的搜索空间是一件不太可能的任务，只有随机选取从全部的匹配样本中选取合适的子集计算误差中值作为一种替代方案。

最小中值二乘法算法求解基础矩阵的计算流程如下：

步骤-I 使用蒙地卡罗随机方法在所有对应点对中随机抽取m个7对点作为构成计算的基础矩阵的最小构成集合。

步骤-II 每个集合都使用7个点对构成向量u并采用以上所提7点法计算出对应的基础矩阵解F。

步骤-III 计算 $\begin{bmatrix} u_{1i} & v_{1i} & 1 \end{bmatrix}$ 与 $F \begin{bmatrix} u_{2i} & v_{2i} & 1 \end{bmatrix}^T$ 的欧式距离， $\begin{bmatrix} u_{21i} & v_{2i} & 1 \end{bmatrix}$ 与 $\begin{bmatrix} u_{1i} & v_{1i} & 1 \end{bmatrix} F$ 的欧式距离之和作为F的评价误差。

步骤-IV 对所有的矩阵F的误差排序，选取误差中值所对应的解作为最优解。

借助基础矩阵对两个摄像机的约束关系，摄像机的任意一个像素将在另一摄像机中找到对应的极线，如果一个摄像机中的一条经过点 $\begin{bmatrix} u_1 & v_1 & 1 \end{bmatrix}$ 的任意一条直线l为：

$$l : \begin{bmatrix} u_1 + at & v_1 + bt & 1 \end{bmatrix} \quad (3.15)$$

将该直线公式代入到基础矩阵约束（3.4）中可得：

$$\begin{aligned} & \left[\begin{array}{ccc} u_1 + at & v_1 + bt & 1 \end{array} \right] F \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = 0 \\ & l_1 \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} + \left[\begin{array}{ccc} at & bt & 0 \end{array} \right] F \begin{bmatrix} u_2 \\ v_2 \\ 1 \end{bmatrix} = 0 \end{aligned} \quad (3.16)$$

在直线 l_1 中任取2个不同的点，代入到上式中可得到

$$\left[\begin{array}{ccc} at & bt & 0 \end{array} \right] F \begin{bmatrix} u_{2i} \\ v_{2i} \\ 1 \end{bmatrix} \quad (i=1,2) \quad (3.17)$$

由此可计算确定在另一摄像机中直线系数 a 和 b 。不仅是两个摄像机中的两个点有对应关系，两条特定的直线也可建立起对应关系。

在第二章标定图的标靶棋盘格中提取特征点可比较容易的建立两个摄像机之间的对应关系，随后最小中值二乘法估计最优的基础矩阵，再将基础矩阵控制的极线绘制到标定图像显示如下：

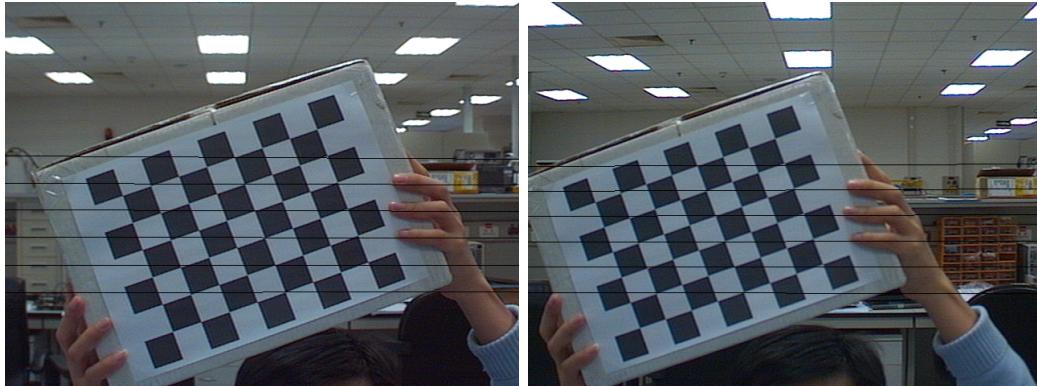


图 3.3 相机对应的极线图

3.3 the epipolar lines in calibration images

两个摄像机在拍摄时位于同一水平面上，造成图3.3中所示的极线看似平行但其实在水平位置上有2-3个像素落差。另外，左右摄像机对应的水平线也

不在同一个水平线上。极线的这样两个问题给后续的匹配过程带来了极大的困难，不利于快速的计算三维世界坐标，为此左右图像能借助下节介绍的图像矫正方法矫正极线。

3.2 图像矫正

基础矩阵给出的极线约束将对应的点的搜索范围从二维平面缩减到一维的直线之上，大大简化了搜索对应点的空间复杂度。但这些极线基本上不与图像坐标系中的x轴或y轴具有共同的方向，因此搜索时需要在直线上估计离散的x, y轴坐标再进行匹配。如果图像中所有的极线都能与x, y轴平行时坐标系统将极大简化，极线坐标与图像坐标统一在同一个坐标系下。矫正后另一个更重要的优点是两个摄像机的对应只需要在图像的x轴或y轴上匹配相似度，降低了匹配算法的复杂度而且匹配结果更准确^[99]。

给定两个摄像机所摄图像，通过一些变换操作在原有图像的基础上生成新的图像使得其中的极线与x轴或y轴平行（一般为x轴平行），在计算机视觉中这样的变换被称为**图像矫正或极线矫正**。该问题是立体视觉中比较常见的问题，但一直以来比较实用的算法较少。早期的Ayache和Lustman^[100]提出了一些人为认定的约束条件并直接求解出变化矩阵矫正图像，但这些条件中提到的紧支性和必要性没有详细证明，因此也无法给出具体的评价和验证方法。另一些法则严格要求摄像机的几何位置，如两个摄像机成像平面的y轴平行^[101]。

目前比较常用的矫正算法可分为两类：

- 基于标定的矫正算法。图像在矫正变化时假设相机的内参数，外参数已知，并根据这些相机参数计算矫正的图像的变换矩阵。基于标定参数的矫正算法计算过程比较简单，对相机间的距离、角度不敏感，大多数情况下都能矫正成功，但计算过程必须依靠预先标定的相机内外参数。
- 非标定的矫正算法。该类方法类似于张正有的相机标定方法，通过一系列在两个摄像机中对应的特征点对来估计矫正模型。由于省略了两个相机的标定过程，节省了一些系统的处理过程，比较适合一些只需计算稠密视差图而不需要知道准确世界坐标和相机坐标的系统。在缺少相机参数及相机精确成像模型的帮助下，矫正算法对两个摄像机之间的位置、角

度以及焦距都比较敏感。一旦两个摄像机的参数相差较大，矫正算法获得结果将变得不可接受。

相机参数能提供相机平面的坐标系关系，矫正图像中新坐标系比较容易确定。相反的，图像矫正在缺少相机参数的情况下存在更多的自由度，矫正矩阵的搜索空间将成倍增加。正是鉴于非标定矫正算法存在的巨大挑战，众多的研究者提出不同的解决方法试图来解决该问题。多摄像机成像奠基者之一的Hartley最早在1993年提出非标定矫正问题^[102]并给出一些初期的解决方法和计算模型。随后他在1999年详细介绍了非标定极线矫正理论模型^[103]和比较实用的矫正算法并成为了公认的经典矫正算法之一。而最近的Wu和Yu不需要计算基础矩阵并考虑镜头畸变系数带来的负面影响以最小化视差原则矫正极线^[104]。相对于非标定矫正算法，标定矫正算法的理论基础和数学模型都相对简单，并且计算的稳定性较高。经典的标定矫正算法是由Fusiello在2000年提出的紧支标定算法^[105]。一些提高的标定矫正算法在此基础之上相继出现取得更好的结果，如考虑到矫正后能最大程度的保留原有图像数据的计算模型。

尽管许多学者提出不同的非标定矫正算法，但即使最新的非标定矫正算法也无法取得和标定矫正算法一样的稳定性和准确性^[106]。考虑到本文所介绍的监控系统已经完成了标定工作，且对立体视觉中数据的稳定性和准确性都有较高要求，因此标定的矫正算法更适合于立体视觉的监控系统。下面将详细介绍标定矫正的算法理论框架。

由第二章所介绍的摄像机模型可知，其焦平面是一个与成像平面平行且经过光心的平面，而光心 \mathbf{c} 的位置可用旋转矩阵 \mathbf{R} 和平移向量 \mathbf{T} 来表示为：

$$\mathbf{c} = -\mathbf{R}^{-1}\mathbf{T} \quad (3.18)$$

同样，平移向量 \mathbf{T} 也可用光心表达为 $\mathbf{T} = -\mathbf{R}\mathbf{c}$ ，将 \mathbf{T} 的新表达式代入到成像公式（2.1）中用光心位置消去平移向量可得：

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{A}(\mathbf{R} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} - \mathbf{R}\mathbf{c}) \quad (3.19)$$

实际安装的两个摄像机很难保证成像平面共面,这意味着两个相机的外部参数 \mathbf{R}_1 和 \mathbf{R}_2 不同,但矫正的目的就是为了让 \mathbf{R}_1 和 \mathbf{R}_2 一致以对齐两个成像坐标的坐标轴。为了达到这样得目的,成像的过程使用新的外部参数取代已有外部参数而内部参数可任意给定,通过扭曲图像来保持与原来的世界坐标系一致,该过程可表达为:

$$\begin{aligned} \mathbf{R}^{-1}\mathbf{A}^{-1}(s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} + \mathbf{A}\mathbf{R}\mathbf{c}) &= s\mathbf{R}^{-1}\mathbf{A}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} + \mathbf{c} = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \\ \mathbf{R}'^{-1}\mathbf{A}'^{-1}(s \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} + \mathbf{A}'\mathbf{R}'\mathbf{c}) &= s\mathbf{R}'^{-1}\mathbf{A}'^{-1} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} + \mathbf{c} = \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \end{aligned} \quad (3.20)$$

上式中的 \mathbf{R}' 和 \mathbf{A}' 是矫正后得到的相机内部参数和外部参数,而 (u', v') 是矫正后的图像坐标。在整个矫正过程中只有世界坐标系 (x_w, y_w, z_w) 和摄像机光心 \mathbf{c} 保持不变,其它的参数都必须要重写计算。矫正后的新相机坐标系的三个坐标轴的要求为:

1. 两个摄像机在位置上只有水平位移而不存在y,z轴上的偏差,所以x轴的方向应当于两个摄像机光心的连线方向重合,或是一个摄像机的光心在另一个摄像机的x轴上。
2. y轴必须与x轴垂直还需要制定一个其他得方向联合确定。除了x轴以外的另一个方向虽然可以任意指定,但与原来的方向相差很大时会造成原来的图像矫正后可利用的数据较少。例如,新的y轴如果与x轴和原有的y轴垂直则新的成像平面几乎与旧的成像平面垂直,也就是原来图像的数据基本不能用。通常比较合理的方向可选择已有的z轴和新的x轴确定y轴。
3. 一旦x轴和y轴确定后, z轴方向可由它们张成的平面唯一确定。第二章中设世界坐标系的x-y平面设定为标靶平面,这也使z轴计算在矫正中的关系不大。

将要求1.和要求2.转成数学等式可表达为:

$$\mathbf{r}'_1 = \frac{\mathbf{c}_1 - \mathbf{c}_2}{\|\mathbf{c}_1 - \mathbf{c}_2\|} \quad (3.21)$$

联立等式 (3.20) 中的两式, 消去世界坐标系点 $\begin{bmatrix} x_w & y_w & z_w \end{bmatrix}$ 得到矫正前图像和矫正后图像之间的关系为:

$$\begin{aligned} s\mathbf{R}^{-1}\mathbf{A}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} + \mathbf{c} &= s\mathbf{R}'^{-1}\mathbf{A}'^{-1} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} + \mathbf{c} \\ \mathbf{R}^{-1}\mathbf{A}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \mathbf{R}'^{-1}\mathbf{A}'^{-1} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} \\ \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} &= \mathbf{A}'\mathbf{R}'\mathbf{R}^{-1}\mathbf{A}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \end{aligned} \quad (3.22)$$

3.3 立体匹配及场景重建

当左右相机经过极线矫正后, 寻找对应匹配点问题的复杂度大大减低。视差图由原来的极限匹配问题转换为在矫正图同一条水平线的匹配问题, 即在两张矫正图的同一水平线上搜索最相似的点对。计算机视觉理论称这样的匹配过程为左右视角的视差估计问题, 而解决该问题的方法可分为两大类:

- 对应匹配 (correspondence) 法。将左右视角的图像作为匹配对象, 寻找两者之间的最优映射对应关系使得目标函数的能量最小。两副图像的匹配过程中需要克服遮挡、边缘噪音的干扰问题, 尽可能地对有效数据建立一一映射关系。这样的匹配通常都属于非线性优化问题, 已有的数学方法都很难求解最优的匹配结果。因此, 大多数方法都只能是采用近似目标函数, 以局部优化的方法来给出次优解。早期的匹配算法将该问题以概率问题建模^[107], 假定左右矫正图都是关于某一真实值的正态分布并优化最大概率目标函数来估计匹配关系。随后的对应算法基本以能量函数来建立模型, 并给出优化方法求解最优解。匹配法从全局估计映射关

系、计算视差、有效地解决了局部无纹理区域的匹配问题，但该类方法只能得到像素级视差精度，而不能计算亚像素级的匹配结果。

- 模板相关（correlation）法。以某一像素为中心选取其邻域，同另一张图像中某像素的邻域计算相关系数。在相关系数最大的一对像素之间计算位置偏差作为像素视差。此类方法需要预先选取匹配的中心点、相关操作中窗口大小，以及匹配中搜索的范围和方向。SSD（Sum of Squared Differences）是最常使用的邻域相关方法，并衍生出一些改进版本。在文^[105]中提出的自适应、多窗口的SSD匹配方法，能根据邻域内成员的特性自动的设置计算窗口大小，既保证了匹配过程中的精度也减少了算法的计算复杂度。但实际过程由于摄像机曝光参数或传感器敏感度的差异，匹配的图像亮度存在一定的差值，这将大大影响SSD算法的准确性，而NCC（normalized cross correlation）方法通过归一化操作减少了亮度所带来的干扰因素。对应匹配法相反的是，模板相关法注重于局部区域的匹配而不能适用于无纹理区域，但逐像素的计算过程能插值出亚像素的匹配精度。

对应匹配法将两副图像作为匹配图像构建一个目标函数或能量函数，使得最佳的匹配结果能使得该目标函数最小化。文^[108]对早期的一维扫描线法^[109]，模拟退火法^[110]，中值场演化法^[111]等作出了详细的对比试验，评价每个算法的计算结果与真实值之间的误差对比，并最后指出模拟退火的非线性优化方法整体上优于其他算法。而随后的graph cut方法^[112]采用离散优化模型最小化Gibbs能量函数以计算稠密的全局视差图。2009年香港科技大学的Jia Jiaya研究组在这些方法的基础上将视差估计与图像matting问题统一在最大后验概率框架中^[113]，将图像分解为背景层、目标层和遮挡层以提供最精细的视察计算结果。

模板相关法只关注局部的匹配关系因而比于对应匹配法数学模型更简单、计算复杂度更低。一直以来新的方法也比较少，鲁棒的M-estimator^[114, 115]估计方法是比NCC更优的局部相似性度量方法，它能计算多个相似性度量所组成的等式系统的根以获得更稳定的计算结果。模版相关匹配方法的另一个分支是局部描述子方法，通过在局部选择尺度映射变化不变特征点生成特征向量进行

对比获得更复杂环境下的匹配关系，如SIFT特征描述子^[116, 117]，SURF特征描述子^[118, 119]。关于这些局部描述子方法的比较与综述可参见文^[120]。

这两类估计视差的方法具有不同的特点，适用于不同的视差计算要求。对应匹配法能建立全局的匹配结果但精度不高，计算复杂度高、处理时间较长；而模板相关法针对局部有明显特征的少数特征点像素匹配更准确、精度更高的结果。事实上，这两种方法并不是互相排斥的可以通过分步匹配进行融合，形成从粗糙到细致、分段匹配的计算过程。

考虑到所构建立体视觉系统是用于实现目标检测与跟踪系统，因此视差计算过程必须具有极高的计算效率和具有使用意义的匹配结果。为了满足这样的要求，本节中提出构建两段式的匹配算法有效地估计视差：首先，一个高效率的全局匹配法获得基本的匹配结果。基于该结果，模板方法对一些关键特征点逐个匹配得到更精确的结果。全局建立的初步匹配不仅能帮助相关法建立最佳的初始匹配窗口大小、水平搜索区域，而且能防止匹配时出现的跨区域偏差。同时，相关方法得到的匹配结果又能进一步向周围邻域传播、细化计算结果。

3.3.1 全局分段视差匹配

在左右视角下所拍摄图像经过极线矫正，将世界坐标下同一目标点所对应的像素置于同一水平线上。这意味着匹配过程只需在一个方向上按序搜索，但由于其中存在如图所示的一些遮挡区域（在一个视角存在，而在另一个视角不存在）会使得匹配算法失效造成后续的匹配出现错误。

整体匹配的策略计算效率比较高，同时获得所有点的匹配结果，但匹配过程中对图像的光照、对比度变化等较敏感，无法匹配不同曝光参数或不同光照条件的图像。经典的Birchfield视察估计算法^[121–123] 中提出目标函数 $\gamma(M)$ 由三个子项构成：遮挡惩罚项、常数奖励项和相似性度量。随后的Graph Cut^[112] 算法中使用两个子项的目标函数来组合目标函数，即灰度相似性度量和视察平滑性约束。这些方法都基于灰度相等假设：要求两副图像中对应像素的灰度相等并以此来计算目标函数的多个子项约束。但这个比较理想化的假设在实际工作中的多摄像机系统中难以保证，这也使得最后的计算结果含有较多估计错误。

考虑到光照等因素只是对图像的均值发生改变，而对灰度的变化规律不会产生较大影响。基于光照特性的这个观察，图像中的高频信息比低频信息对图

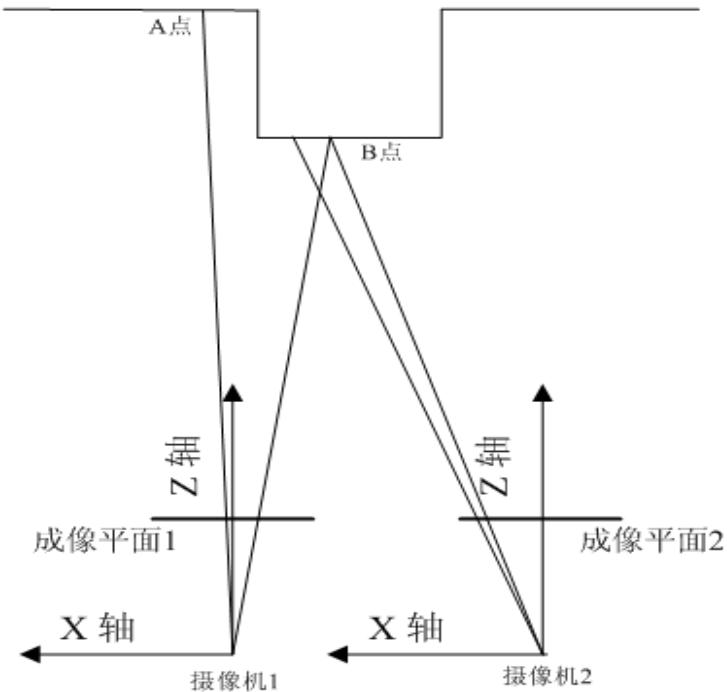


图 3.4 视察匹配中的遮挡问题

3.4 Some occlusion phenomena in the disparity estimation

像的匹配更为重要。常用的频率分析方法是傅里叶分析和小波分析，他们都能有效地提取连续或离散信号中频率成分而被用于在频谱计算中。这些方法所需的计算量限制了它们在实时视差估计中的应用，但借助其（尤其是小波分析）从粗糙对比到逐步细化思想可产生多尺度的片段对比方法。

将定存在任一片段 $f_i(p)$ ，其分解表达方式可表达为片段的中间均值加上每个像素的偏差值，即：

$$f_i(p) = avg_i + var_i(p) \quad (3.23)$$

其中均值 avg 可被认为是低频分量，而偏差 var_i 则是高频分量。在此基础之上，可进一步构造片段的近似表达结构为：

$$f_i(p) \approx \tilde{f}_i(p) = \begin{cases} avg_i + 1 & var_i(p) > 0 \\ avg_i - 1 & var_i(p) < 0 \end{cases} \quad (3.24)$$

在近似表达形式 $\tilde{f}_i(p)$ 中寻找最大的平坦区域 $C = \max(C_i)$ 且 $C_i = \{|p - q|, \sum_p^q \tilde{f}_i(p+1) - \tilde{f}_i(p) = 0\}$ 。以该区域作为确定的对应区域并生成新的片段，当该片段存在的最大片段都小于设定的阈值时将不予分裂。重复该过程直至所有的片段都不能分裂时收敛并结束计算。这时产生的片段按顺序对应产生区段之间的映射。在左右两图中分别分解片段时会产生不一致的分解结果（其中一个分解为2段而另一个分解为3段），这样影响了后续的对应关系。此时，将3段分解中的一段合并成2段与另一段分解保持一致。

3.3.2 局部模版细化匹配

通过对一维灰度信号的高频分解，建立匹配关系获得初步的视差估计结果。更细化的结果则需要进一步计算局部的模版相关性，寻找最大相关位置作为最优的匹配结果。最简单的SSD相关性计算为两个局部模版的绝对差平方之和：

$$d(u, v) = \sum_{x,y} (I_1(x, y) - I_2(x - u, y - v))^2 \quad (3.25)$$

从上式可知：如果两个模版越相似则计算的平方差之和越小；反之模版差距越大则计算值越大。将该式重新展开后可得：

$$\begin{aligned} d(u, v) &= \sum_{x,y} I_1(x, y)^2 - \sum_{x,y} I_1(x, y)I_2(x - u, y - v) \\ &\quad + \sum_{x,y} I_2(x - u, y - v)^2 \end{aligned} \quad (3.26)$$

其中 $\sum_{x,y} I_1(x, y)^2$ 和 $\sum_{x,y} I_2(x - u, y - v)^2$ 是左右图像中邻域灰度的平方和。由于这两个变量在计算过程中变化相对较小（尤其是 $\sum_{x,y} I_1(x, y)^2$ 在计算中为定值），因此这两项在相似性判断中贡献作用不大可将其忽略不计。换言之，相关计算过程中最重要的部分是交叉相关项

$$corr(u, v) = \sum_{x,y} I_1(x, y)I_2(x - u, y - v) \quad (3.27)$$

但相关项在计算过程中存在的缺陷表现为 $\sum_{x,y} I_2(x-u, y-v)^2$ 在计算中对光照的变换不能保持稳定、匹配度波动较大，从而造成错误的匹配结果。例如，两个正确匹配点的匹配相关值却小于其与一个亮点的相关值。

因为相关计算对图像的亮度变化比较敏感，因此同样两个模版在不同光照下得到的匹配结果差别较大。但实际工作中两个摄像机在不同视角下光照和曝光参数不同造成图像亮度不同是立体视觉中比较常见的问题。基于以上这些问题，在相关计算中将图像邻域归一化为单位长度向量后再计算交叉相关性，即：

$$Ncorr(u, v) = \frac{\sum_{x,y} [I_1(x, y) - \bar{I}_1] \sum_{x,y} [I_2(x-u, y-v) - \bar{I}_2]}{\left(\sum_{x,y} [I_1(x, y) - \bar{I}_1]\right)^2 \sum_{x,y} [I_2(x-u, y-v) - \bar{I}_2]^2)^{0.5}} \quad (3.28)$$

在左摄像机图像 I_1 中选取像素点 $\begin{bmatrix} x & y \end{bmatrix}$ ，在右摄像机图像 I_2 中依次顺序选取 u, v 计算相关值 $Ncorr(u, v)$ 产生一个系列相关系数 $Ncorr(u, v); (u = 1, 2 \dots N, v = y)$ ，并构成离散函数 $f(u)$ 。在一些对精度要求更高的应用中，可以在离散匹配值的基础上构造连续拟合函数估计亚像素的最优匹配结果，产生更高精度的视察估计值。

3.4 双目矫正实验及评价

按照第一节所介绍的方法计算基础矩阵，对应基础矩阵的评价是统计基础矩阵约束公式（3.4）的平均误差 Fe ，其计算公式为：

$$Fe = \sum_{i=1,2 \dots NP} \left(\begin{bmatrix} u_{1i} & v_{1i} & 1 \end{bmatrix} F \begin{bmatrix} u_{2i} \\ v_{2i} \\ 1 \end{bmatrix} \right)^2 / NP$$

其中 NP 为总的匹配特征点数量。在实际中得到基础矩阵的误差如图3.5所示

第二节介绍的图像矫正能随意为矫正后的相机成像模型指定相机内参数，但考虑到随后世界坐标系的计算便捷性特令矫正后的左摄像机内参数等于矫正后的右摄像机内参数，即 $\mathbf{A}'_1 = \mathbf{A}'_2 = \mathbf{A}'$ 。以新的相机参数代入到矫正公式（3.22）对原图像进行矫正操作以保证任意的对应点在新的图像处于同一水平

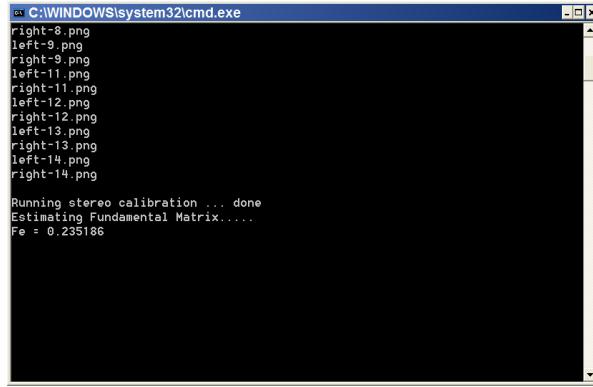


图 3.5 基础矩阵误差

3.5 Fundamental matrix errors

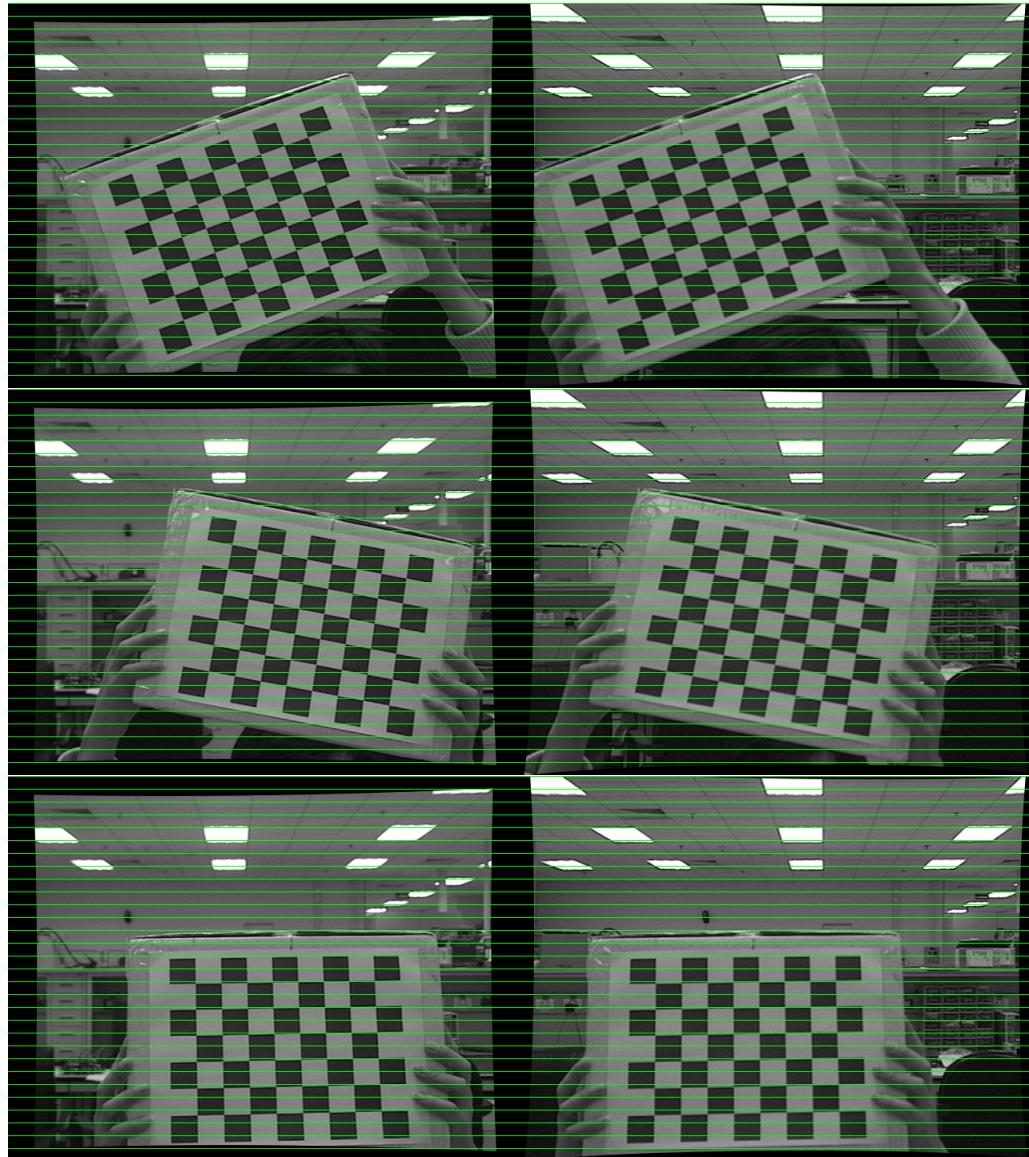
线上，即矫正后的图像的左极线与右极线共线。将第二章得到的左摄像机标定图2.3和右摄像机标定图2.4进行矫正得到新图像显示在图3.6中。在得到的矫正结果图中可看到：两个相机对应的极线在同一条水平线且对应的世界坐标系下的点也基本准确。

基于矫正后的新相机内参数 \mathbf{A}' ，图像上的像素点也可用相机模型计算其在新相机坐标系下的空间位置：

$$\begin{bmatrix} x_{c1}/z_{c1} \\ y_{c1}/z_{c1} \\ 1 \end{bmatrix} = (\mathbf{A}')^{-1} \begin{bmatrix} u'_1 \\ v'_1 \\ 1 \end{bmatrix} \quad (3.29)$$

$$\begin{bmatrix} x_{c2}/z_{c2} \\ y_{c2}/z_{c2} \\ 1 \end{bmatrix} = (\mathbf{A}')^{-1} \begin{bmatrix} u'_2 \\ v'_2 \\ 1 \end{bmatrix} \quad (3.30)$$

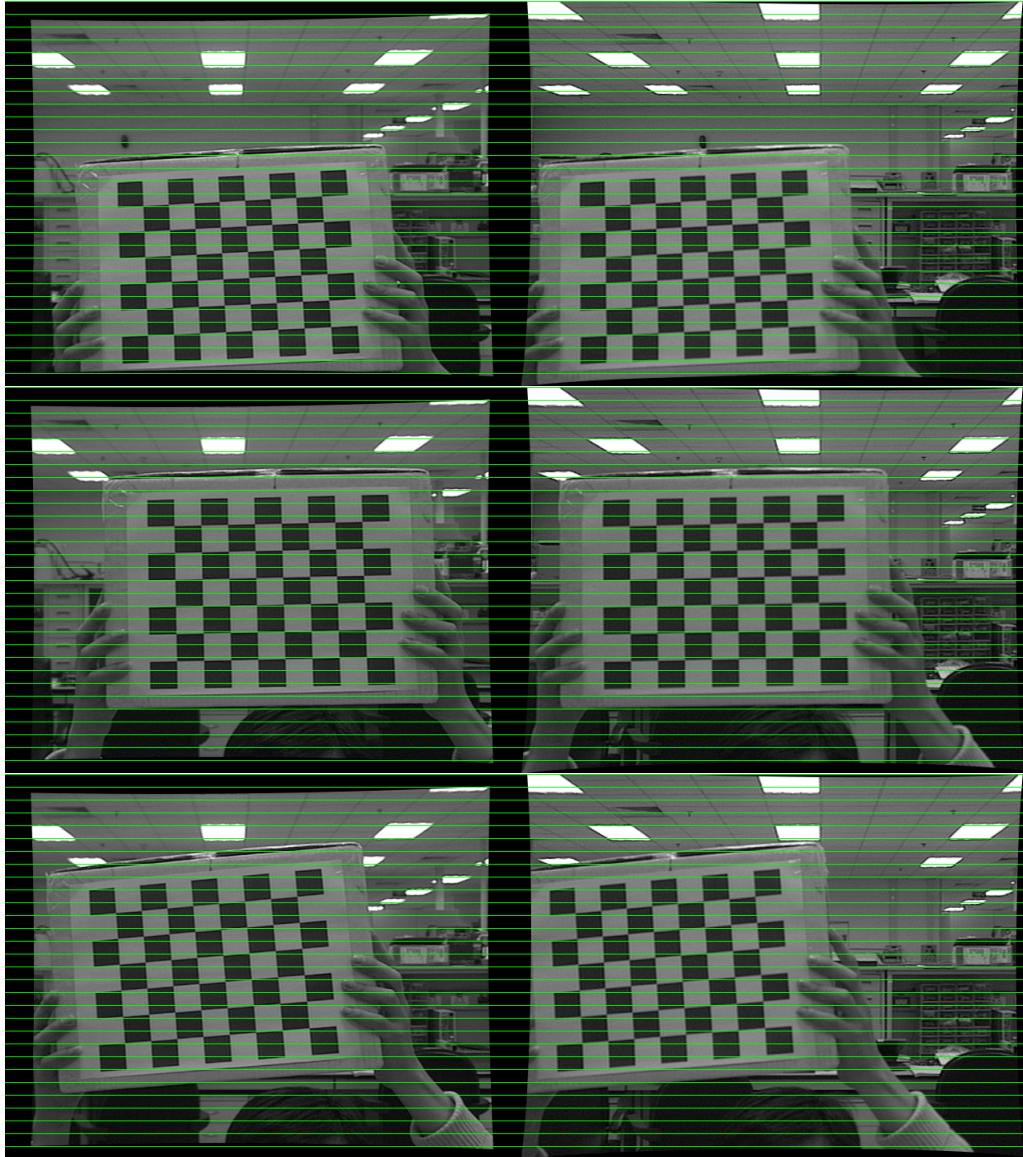
以上的式 (3.29) 和 (3.30) 分别是矫正后左摄像机和右摄像机图像上的像素点与世界坐标系之间的映射投影关系。同时，由矫正时提出的三个条件也可知两个摄像机的相机坐标系也存在关联关系，即左相机的相机坐标系与右相机的相机坐标系只是在X方向上有平移运动。换言之，两个相机坐标系在矫正后Y和Z轴位置完全相同，只是X值之间的差距是两个摄像机之间的光心距离，因此两个相机的相机坐标系之间的关系为：



$$\begin{bmatrix} x_{c1} \\ y_{c1} \\ z_{c1} \end{bmatrix} = \begin{bmatrix} x_{c2} + b \\ y_{c2} \\ z_{c2} \end{bmatrix} \quad (3.31)$$

其中 $b = \mathbf{c}_1 - \mathbf{c}_2$ 为两个摄像机之间的光心距离, 或在计算机视觉中被称为基线 (baseline) 距离。

如果左右相机中的像素点 $\begin{bmatrix} u'_1 & v'_1 \end{bmatrix}$ 和 $\begin{bmatrix} u'_2 & v'_2 \end{bmatrix}$ 对应同一个世界坐标系下



的点，则可联合公式（3.29）、（3.30）和（3.31）求解出空间摄像机坐标系下的三维坐标 $\begin{bmatrix} x_{c1} & y_{c1} & z_{c1} \end{bmatrix}$ 或 $\begin{bmatrix} x_{c2} & y_{c2} & z_{c2} \end{bmatrix}$ 完成对三维空间的重建。首先展开两个相机坐标系下X轴的计算等式并相减可得：

$$\begin{aligned} f_x \frac{x_{c1}}{z_{c1}} + C_x - f_x \frac{x_{c2}}{z_{c2}} + C_x &= u'_1 - u'_2 \\ f_x \frac{b}{z_{c1}} = u'_1 - u'_2 & \\ \Rightarrow z_{c1} = z_{c2} = f_x \frac{b}{u'_1 - u'_2} & \end{aligned} \quad (3.32)$$

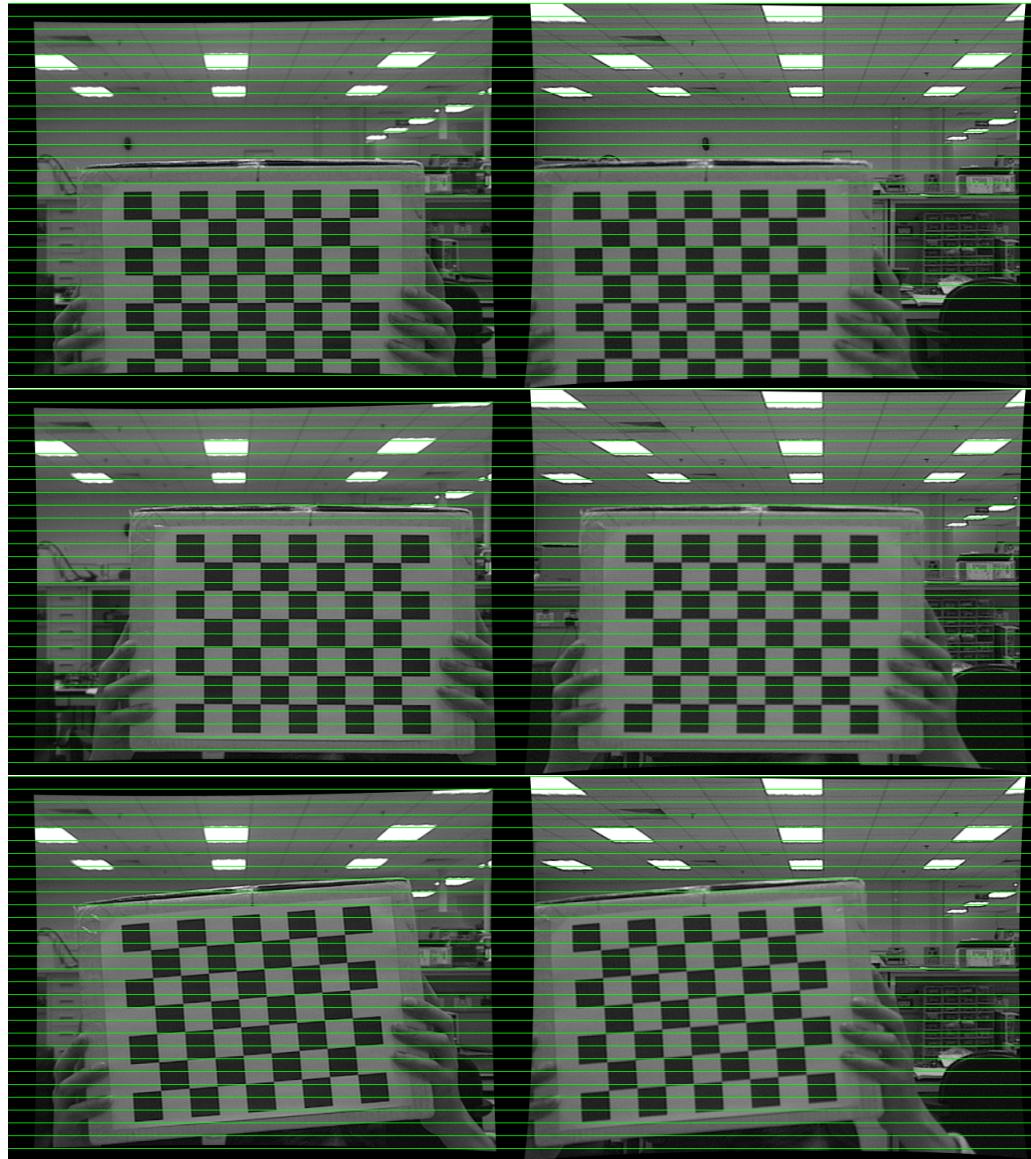


图 3.6 极线矫正的实验结果

3.6 Experimental results of epipolar rectification

一旦计算出Z轴的坐标 z_{c1} 和 z_{c2} 以后代入到式 (3.29) 和 (3.30)，便可计算出X轴和Y轴的坐标 x_{c1} , x_{c2} 和 y_{c1} , y_{c2} :

$$x_{c1} = x_{c2} + d = (u'_1 - C_x) \frac{z_{c1}}{f_x} \quad (3.33)$$

$$y_{c1} = y_{c2} = (v'_1 - C_y) \frac{z_{c1}}{f_y} \quad (3.34)$$

分别将左右相机的第一副标定图中的所有特征点用矫正后的相机内参数，计算矫正后的相机外参数 \mathbf{R} 和 \mathbf{T} 得到特征点在相机坐标系下的位置。图3.7显示了所有特征点在相机坐标位置，其中左相机的特征点用红色显示，右相机的特征点用绿色显示。图中可看到对应点特征点在空间只是在X轴上有差距，但Y轴和Z轴几乎相等。这也证明矫正后的两个相机的XYZ轴都符合图像矫正所提的三个假设。

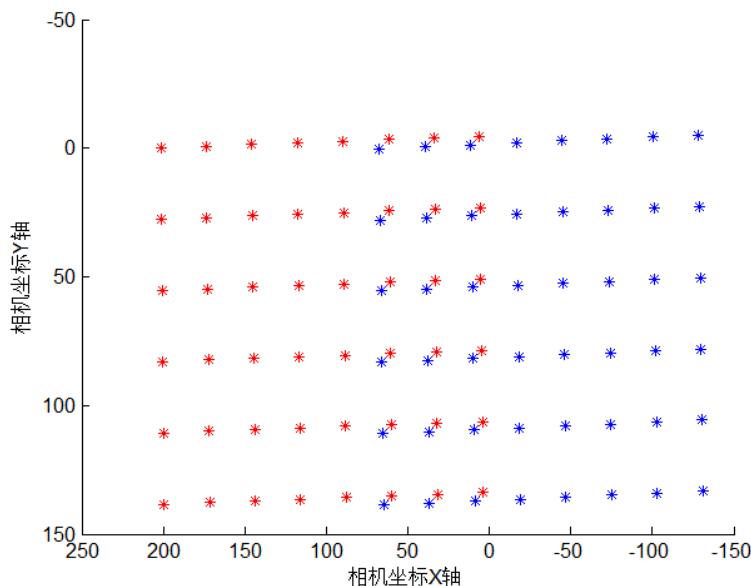
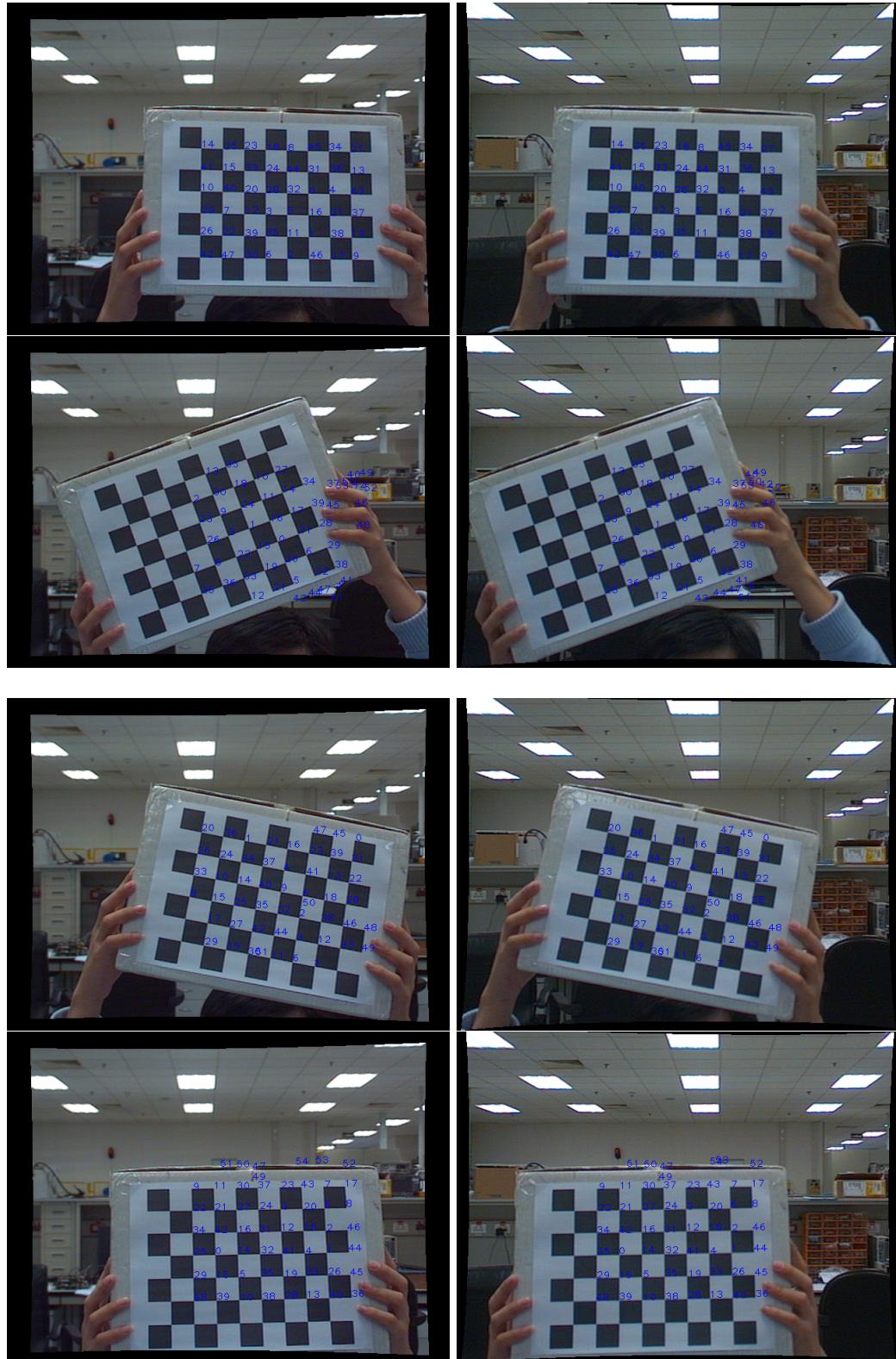


图 3.7 特征点的相机坐标 (单位: mm)

3.7 The camera coordinate of feature points

最后测试第三节提出的视察计算方法是否能准确的匹配左右相机中的同一目标点。同样以标定图为测试图，在左相机的图像中提取特征点，然后通过提出的视差匹配方法在右相机的图像中寻找最佳匹配点。同一个特征点标定在左右两副图像用标记同样的数字编号。匹配前六副标定图中特征点的结果显示如下：

最后的总测试通过匹配到的视察和矫正后新的相机内外参数从左右图像中还原出三维信息。图3.7得到特征点在相机空间中的位置，因此在测试中可以利



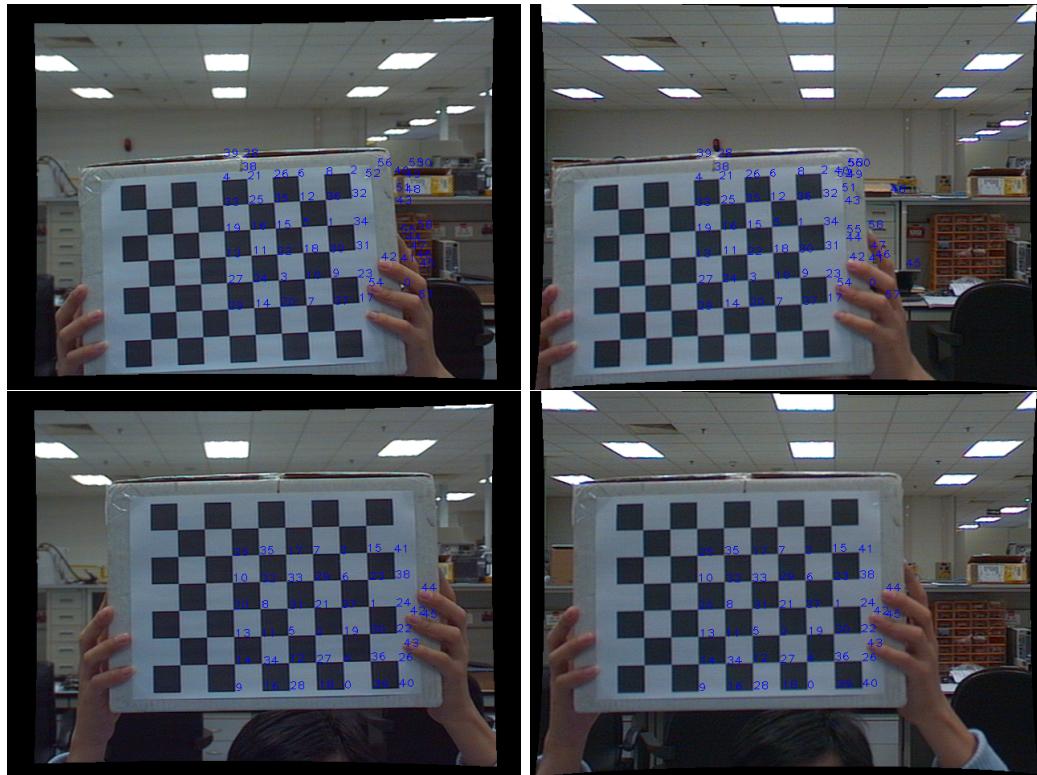


图 3.8 特征点匹配计算视差

3.8 Estimating Disparity by matching feature points

用这些点的位置作为真实值。在左相机中提取特征点，然后通过提出的视差匹配方法在右相机寻找最佳匹配点、计算特征点视差，最后用计算这些点的Z轴坐标并减去真实特征点的坐标，其比较结果如图3.9所示。从误差分布可看到在目标Z轴真实值大约为850mm时，视察估计在估计Z轴时带来的平均误差小于10mm，最大误差大约15mm。事实上可通过插值亚像素的匹配精度而进一步减少误差，即使视差估计中没有考虑亚像素精度但在监控环境中这样的误差是在合理方法范围之内。

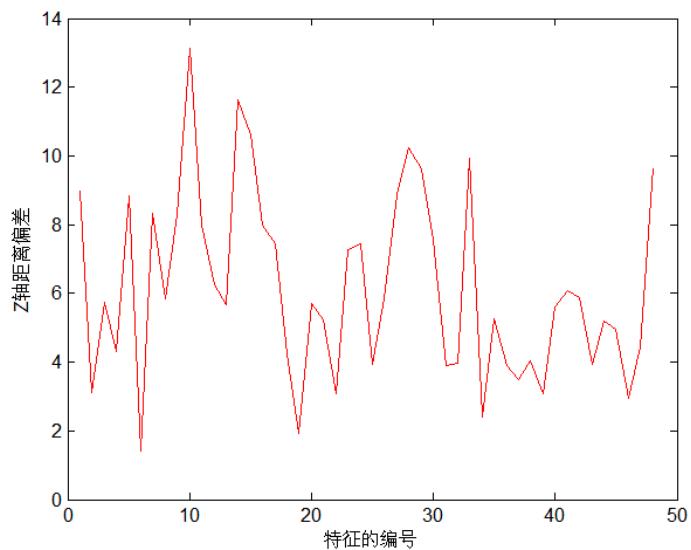


图 3.9 估计视察在三维重建中的误差 (单位: mm)

3.9 Errors of three-dimensional reconstruction caused by disparity estimation

第四章 多目标的检测与跟踪

多目标的监控是计算机视觉研究中十分前沿的研究课题，每年都有大量的最新研究成果发表在各种重要的学术期刊和会议上。监控系统涉及到众多的子课题，目标检测和目标跟踪是其中两个最重要的研究内容。目标检测用于定位监控目标而跟踪方法则是后续信号中持续地捕捉目标的最新位置。这两个方法的研究内容相关关联，跟踪方法需要检测模块提供初始的目标位置，而检测方法则可以独立完成跟踪任务。

目前的监控系统都存在三个难以解决的问题，即1) 复杂背景下的光照变化；2) 阴影干扰；3) 严重遮挡下的多目标跟踪。这也造成目前已有的系统无法在实际的应用环境下稳定的工作，从而缺少真正意义上的智能监控系统。

随着对公众安全、智能交通、自动化流失线等方面越来越高的要求，一些新的硬件设备也逐步被应用到智能监控系统之中。这其中红外成像技术和多摄像机系统是两种比较有前景的研究方向。红外技术对热源成像，减少了对可见光的依赖对光照变化比较鲁棒而多摄像机能从多个角度观察监控场景能产生更全面的信息有利于解决多目标之间的遮挡问题^[124, 125]。红外成像只能处理热源物体且对环境问题较敏感，但是多摄像机系统能针对检测任意目标并对其跟踪。这也意味着基于多摄像的监控系统有更广泛的应用场景更鲁棒的工作性能。

双目立体系统由两个摄像机构成可看作是最简单的多摄像机系统，依照两个摄像机之间的基线距离又可划分为：**长基线**（wide baseline）和**短基线**（short baseline）系统。通常，长基线系统直接估计一些稀疏点集在两个摄像机中的对应关系，而短基线系统则倾向于构建稠密深度视图对应相机视角。但前者即使采样大量计算时间也很得到准确的对应关系。相反地，短基线系统由于相机视角变化较小所拍摄照片相似度较高因此更容易获得准确的视差，但对于某些不具有明显纹理的区域也往往无法估计准确视差而生成稀疏视差图。尽管一些方法能解决该问题，但所需要的大量计算量和处理时间限制了他们在实时监控系统中的应用^[126]。

总的来说，基于特征点建立对应的方法比基于视差图的方法不仅计算效率更高，而且提供了更精确的空间位置定位。此外，在一些如低对比度或遮挡等极端条件下，基于特征点的方法的性能更稳定。基于这些想法，本章提出一个短基线立体视觉的多目标检测与跟踪系统。所提的方法不需要估计稠密的深度视图提取目标区域，而是提取一些特征点计算其深度后投影到二维的地板平面。投影点的高度和位置可产生一些聚类，并由这些聚类确定目标的数量、位置和方向。与已有的监控系统相比，新系统能有效地解决以上提到的三个问题：

1. 同一个场景中的光照变化对于两个监控摄像机的影响是相同的，所以在匹配时光照变化的影响能相互抵消而不会对匹配视差时产生干扰。
2. 阴影在场景中通常都在地上或墙壁上，即阴影的特征点在平地上的投影高度给零。立体视觉能计算特征点的在空间中的高度，只需在聚类操作之间过滤掉高度较低的投影点则不会将阴影误认为是监控目标。
3. 因为从俯视角度下这些聚类不存在交集，所以特征点投影到地板平面后目标不存在相互遮挡问题。通过对聚类位置的不断更新来实现多目标跟踪，更新策略所需较少的计算开销使得新监控系统能在线检测、跟踪多个目标。

本章所介绍的立体成像平台基于第二章所述的相机标定和第三章所述的极线约束与校正匹配的理论基础之上，由两个固定的摄像机搭建而成并通过PC机上的双采集卡同时采集图像数据进行的目标检测与识别过程。在系统开始工作之前，需要做一些简单的初始化工作建立监控场景的数学模型。本章第一节将详细介绍这个建模过程。

4.1 三维立体坐标系统

监控所使用的两个估计摄像机之间的基线距离较短，摄像机之间的视角差别较小从而在矫正后的图像中匹配也比较准确。但是对于无纹理的区域中的像素点依然不能建立对应关系，即恢复整个场景的三维结构是十分困难的事情仅能得到场景的稀疏深度图。一些新方法试图用大计算量方法^[127]解决稀疏深度图的问题同时也限制了这些方法很难应用于在线的监控系统之中。

另一方面，在图像中提取如角点和边缘点的特征点能比较容易地确定其三维坐标，这些特征点所需的较小计算开销也有利于在线的多目标监控系统。考虑到特征点在视察匹配、计算开销、稳定性等方面的优点，本章所提的目标检测与跟踪方法都以特征点为处理对象。

监控场景通常是一些如广场、车站、公路等公众区域，它们具有一个共同的特点是：需要监控的区域都是地板平面的某一个特定兴趣区域。因此在所提的监控系统有个重要的假设：监控场景中的场地是平面地面，则该平面在立体视觉系统中能用简单的平面方程表达其空间位置。为了确定地板平面在相机坐标下的平面方程，需要类似于相机标定一样的平面标定过程。所监控的平面一般面积比较大，对应标定平面的标靶也需要面积较大的模式结构。虽然这样的固定式标靶标定比较准确，但较大的标靶也会造成许多使用中的不便问题。在地面标定时不采用固定标靶模式，而是通过人为布置的一些特征点在地面上，然后在标定图中手工选取这些特征点构成标定的特征点集。图4.1的原理图解释了地板平面标定的过程，图中的红色点为人工设置的多个特征点，它们都平置在地板平面之上高度可忽略不计。

常见的提取特征算法有harris^[128]算子和susan^[129, 130]算子或最新的SIFT算子^[131, 132]和SUFR算子^[133, 134]。将图像看做一个二维的信号，则该信号的自相关函数可写为：

$$c(x, y) = \sum_W [I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \quad (4.1)$$

其中 x, y 是图像 I 中像素，而 x_i, y_i 是以 x, y 为中心 w 为宽度的局部窗口。相关的偏移量 $I(x_i + \Delta x, y_i + \Delta y)$ 是图像中对于 x, y 的偏移 $\Delta x, \Delta y$ 的图像像素值，利用Taylor展开式可用图像在 x_i, y_i 的偏导数 $I_x(x_i, y_i), I_y(x_i, y_i)$ 将它展开为：

$$I(x_i + \Delta x, y_i + \Delta y) \approx I(x_i, y_i) + \begin{bmatrix} I_x(x_i, y_i) & I_y(x_i, y_i) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (4.2)$$

将等式(4.2)代入到式(4.1)中，可得到：

$$c(x, y) = \sum_W [I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \quad (4.3)$$

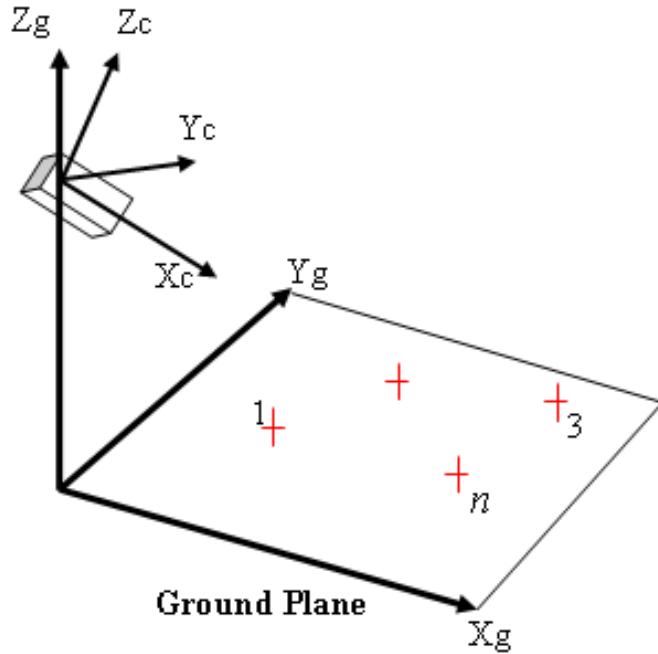


图 4.1 地面标定原理图

4.1 the schematic diagram of plane calibration

$$c(x, y) = \sum_W [I(x_i, y_i) - I(x_i, y_i) - \begin{bmatrix} I_x(x_i, y_i) & I_y(x_i, y_i) \end{bmatrix}] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}]^2 \quad (4.4)$$

$$c(x, y) = \sum_W [\begin{bmatrix} I_x(x_i, y_i) & I_y(x_i, y_i) \end{bmatrix}] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}]^2 \quad (4.5)$$

$$c(x, y) = \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \begin{bmatrix} \sum_W (I_x(x_i, y_i))^2 & \sum_W I_x(x_i, y_i) I_y(x_i, y_i) \\ \sum_W I_x(x_i, y_i) I_y(x_i, y_i) & \sum_W (I_y(x_i, y_i))^2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (4.6)$$

$$c(x, y) = \begin{bmatrix} \Delta x & \Delta y \end{bmatrix} \mathbf{C}(x, y) \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (4.7)$$

其中矩阵 $\mathbf{C}(x, y)$ 近似表达了像素局部邻域的亮度结构。Harris 用这个矩阵的两个特征值来判断像素是否属于角点。有时场景中比较单一提取的角点数量太

少，计算出的三维特征点不能完全地表达整个场景的空间和目标之间的关系。事实上由于图像是离散信号因此在局部的邻域中只有四种变化结构，即水平边缘、垂直边缘和两个对角边缘，因此可统计窗口内四种变化结构的均值作为特征点的判断依据，而不只是考虑矩阵 $\mathbf{C}(x, y)$ 的特征值，定义新的邻域矩阵 $\tilde{\mathbf{C}}(x, y)$ 为：

$$\tilde{\mathbf{C}}(x, y) = \frac{1}{W^2} \begin{bmatrix} \sum_w [I(x_i + 1, y_i) - I(x_i, y_i)] & \sum_w [I(x_i + 1, y_i + 1) - I(x_i, y_i)] \\ \sum_w [I(x_i + 1, y_i) - I(x_i, y_i + 1)] & \sum_w [I(x_i, y_i + 1) - I(x_i, y_i)] \end{bmatrix} \quad (4.8)$$

该矩阵中的4个元素分别测量了窗口 w 中四种边缘的强度均值，如果其中最小的均值都大于设定的阈值则可以认为是特征点。反之，则这样的像素不能认为是特征点。

通过以上提到的特征提取算法能比较迅速地提取那些人为制造的地板特征点，并通过手工标定出其所对应的图像坐标为 $(u_1, v_1)(u_2, v_2) \cdots (u_n, v_n)$ 。利用第二章标定出的相机内参数和相机成像模型和第三章中立体视觉理论，将特征点的图像坐标转换成相机坐标 $(X_{c1}, Y_{c1}, Z_{c1})(X_{c2}, Y_{c2}, Z_{c2}) \cdots (X_{cn}, Y_{cn}, Z_{cn})$ 。

已知这些点都位于地板平面上，同样在相机坐标下它们也必在同一平面 $\mu X_c + \nu Y_c + \omega Z_c = 1$ 之上。通过这些标定出的特征点和他们对应的相机坐标，构造平面的拟合方程为：

$$\text{Min} \quad \sum_{i=1}^N (\mu X_c + \nu Y_c + \omega Z_c - 1)^2 \quad (4.9)$$

而最优系数的求解表达式为：

$$\begin{aligned} \mu \bullet \tau = & \sum X_{ci} [\sum Y_{ci}^2 \sum Z_{ci}^2 - (\sum Y_{ci} Z_{ci})^2] \\ & + \sum Y_{ci} [\sum Y_{ci} Z_{ci} \sum X_{ci} Z_{ci} - \sum Z_{ci}^2 \sum X_{ci} Y_{ci}] \\ & + \sum Z_{ci} [\sum X_{ci} Y_{ci} \sum Y_{ci} Z_{ci} - \sum Y_{ci}^2 \sum X_{ci} Z_{ci}] \end{aligned} \quad (4.10)$$

$$\begin{aligned} \nu \bullet \tau = & \sum X_{ci} [\sum X_{ci} Z_{ci} \sum Y_{ci} Z_{ci} - \sum Z_{ci}^2 \sum X_{ci} Y_{ci}] \\ & + \sum Y_{ci} [\sum X_{ci}^2 \sum Z_{ci}^2 - (\sum X_{ci} Z_{ci})^2] \\ & + \sum Z_{ci} [\sum X_{ci} Y_{ci} \sum X_{ci} Z_{ci} - \sum X_{ci}^2 \sum Y_{ci} Z_{ci}] \end{aligned} \quad (4.11)$$

$$\begin{aligned}
 w \bullet \tau = & \sum X_{ci} [\sum X_{ci} Y_{ci} \sum Y_{ci} Z_{ci} - \sum Y_{ci}^2 \sum X_{ci} Z_{ci}] \\
 & + \sum Y_{ci} [\sum X_{ci} Y_{ci} \sum X_{ci} Z_{ci} - \sum X_{ci}^2 \sum Y_{ci} Z_{ci}] \\
 & + \sum Z_{ci} [\sum X_{ci}^2 \sum Y_{ci}^2 - (\sum X_{ci} Y_{ci})^2]
 \end{aligned} \tag{4.12}$$

其中

$$\begin{aligned}
 \tau = & \sum X_{ci}^2 \sum Y_{ci}^2 \sum Z_{ci}^2 + 2 \sum X_{ci} Y_{ci} \sum X_{ci} Z_{ci} \sum Y_{ci} Z_{ci} \\
 & - \sum Y_{ci}^2 (\sum X_{ci} Z_{ci})^2 - \sum X_{ci}^2 (\sum Y_{ci} Z_{ci})^2 - \sum Z_{ci}^2 (\sum X_{ci} Y_{ci})^2
 \end{aligned} \tag{4.13}$$

相机坐标系统以相机中心为原点， X, Y 轴张成了相机的成像平面 Z 轴则垂直成像平面由相机内部指向外部。在如图4.2所示的场景中，目标和背景中的大门、墙壁等物体上都分布着大量的特征点。在立体视觉中通过匹配出的视差和相机参数都只能还原出特征点的相机坐标下位置，而图**则说明了在相机坐标系下的整个场景中特征点以及地板平面的空间位置关系，同时该图也说明了不同物体的特征点在相机系统下都混叠在一起很难分类特征点是属于目标还是属于背景。



图 4.2 监控场景中的特征点

4.2 Feature points in the surveillance scene

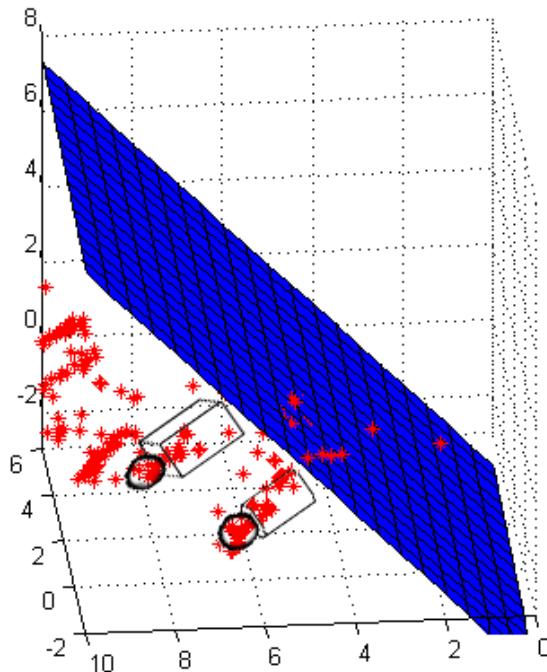


图 4.3 相机坐标系下的特征点

4.3 Feature points expressed in the camera coordinate system

利用前面标定的地板平面方程，这些特征点可以投影地板平面得到投影点坐标 $(x_{p1}, y_{p1})(x_{p2}, y_{p2}) \cdots (x_{pn}, y_{pn})$ 和特征点相对地面平面的高度 $z_{p1}, z_{p2} \cdots z_{pn}$ 。以这三个值构成特征点的新坐标可间接得到特征点的世界坐标系下的空间位置，而该坐标系以摄像机光心的投影点为坐标原点，地板平面为XY平面垂直高度为Z轴所构成。图4.4 中显示了空间中的特征点转化到世界坐标系后的空间位置。

空间中的投影点一部分来自监控目标，另一部分来自于如门窗和墙壁等背景物体，但同一物体的特征点在新的坐标系下之间不会存在交集和重叠问题。为了将背景带来的特征点去除但又不影响系统目标检测的准确性，预先指定如图4.4中的蓝色区域为真正监控范围。所有位于监控区域之外的投影点将被过滤，不再做后续的处理。背景物体的特征点以及阴影带来的特征点都将投影到设定的监控区域之外因此它们将不会影响系统检测与跟踪的准确性。

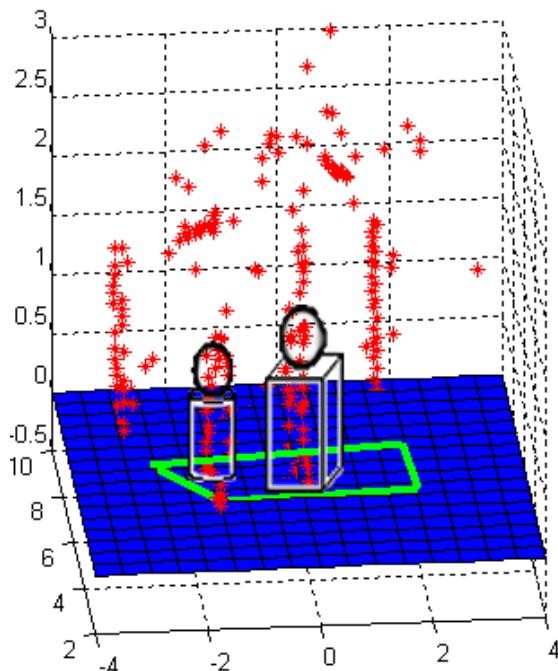


图 4.4 世界坐标系下的特征点

4.4 Feature points expressed in the world coordinate system

4.2 多目标的检测

在世界坐标系中的目标特征点被投影到地板平面上以后，对应同一目标的投影点彼此相邻。基于特征点空间上相邻的性质，这些投影点能被聚合成不同的集合以代表监控场景中目标的位置和方向。由于监控场景中目标数量并不是固定不变而是随着目标的出现与离开随时发生变化，所以聚类产生的集合数量也是不能预先确定的。而经典的聚类算法如k-mean, isodata等经典算法都是需要集合数目作为初始化条件，因此他们并不能用来解决这里的聚类问题。

另一方面，由立体视觉得到的投影点是与世界坐标相对应，他们之间的距离与特征点在世界空间的俯视角度下的距离相同，于是目标的空间形状能作为聚类中重要的约束条件。例如，不论是老人或是年轻人、女性或是男性，俯视视角下他们都能用一个椭圆形状近似表达，其长轴的方向则表示人体在空间中的不同站立方向。

目标检测中的聚类问题经过以上分析可抽象为：在不确定集合数量但已知集合尺寸的情况下将投影点聚合成不同目标集合。

核密度估计（Kernel density estimate）是一种常用于估计离散样本点概率密度的非参数方法。该方法与参数密度估计方法不同的是，它不需要为概率空间建立假设模型，而且能对某个局部区域进行概率估计，因此常被用于在特征分析中。采用指定窗口内样本点的密度近似窗口中心概率值。因为窗口中的每个样本点由核函数赋予其对估计概率的权重值，所以密集区域在多个样本贡献作用下的概率值大于稀疏区域在少量样本贡献作用下的概率值。概率空间中的局部极大值则是聚类集合的中心和方向，所有投影点能由爬山法向局部最大值逼近并与其关联，而逼近同一个局部最大值的投影点聚合成一个集合。为了有效地在核密度空间中搜索局部最大值，Fukunaga和Hostetler^[135]提出了mean shift方法能从概率空间中任意点出发，通过估计局部密度梯度向量并自适应调整逼近步长、逐步收敛到局部最大值。该方法已成功用于解决数据分析^[136]、图像分割^[137]、目标跟踪^[138, 139]中的关键问题。

为了能快速寻找概率空间中的局部最大值并产生相应的聚类集合，本节中提出一种基于核函数的聚类算法来解决已知集合的尺寸，但未知其数量的前提下对离散数据点聚类的算法。产生的聚类集合不仅确定监控场景中的目标数量，而且能确定其位置和方向。

假设目标的位置和方向是概率空间中的两个独立变量，关于投影点的概率密度函数用方向核函数 H_θ 和位置核函数 H_x 表达为：

$$E(\mathbf{x}, \theta) = \sum_{i=1}^{n_i} H_\theta(d_i(\theta)) \sum_{j=1}^{n_j} w_j H_x(d_j(\mathbf{x}), \theta_i) \quad (4.14)$$

其中 \mathbf{x} 和 θ 是位置和方向变量， $d_i(\theta)$ ， $d_j(\mathbf{x})$ 是方向与位置的归一化距离测度函数，即 $d_i(\theta) = \|A_\theta(\theta - \theta_i)\|^2$ ， $d_j(\mathbf{x}) = \|\mathbf{A}_x(\mathbf{x} - \mathbf{x}_j)\|^2$ 。测度函数中的变量 \mathbf{x}_j 表示第j个投影点的坐标 $\begin{bmatrix} x_{pj} & y_{pj} \end{bmatrix}$ ，变量 θ_i 是在 \mathbf{x}_j 表示第j个投影点的坐标 $\begin{bmatrix} 0 & 2\pi \end{bmatrix}$ 中的第i个方向值。 w_j 则是第j个投影点的权重值。关于核函数 $H_\theta(d_i(\theta))$ 和 $H_x(d_j(\mathbf{x}), \theta_i)$ 的构造将在第四节中给予详细的介绍。

值得一提的是，上式中用来构造概率空间的投影点满足两个条件：1) 投影后的位置必须指定监控区域之内；2) 投影后相对地板平面的高度值大于所设阈值。借助这两个限定条件，那些墙壁和地板上的阴影产生的投影点将会在计算之前被滤去，并不对聚类结果产生任何影响。图4.5 (a) 显示的监控区域

之内的投影点，(b) 则是用KDE方法还原出的连续概率空间分布。

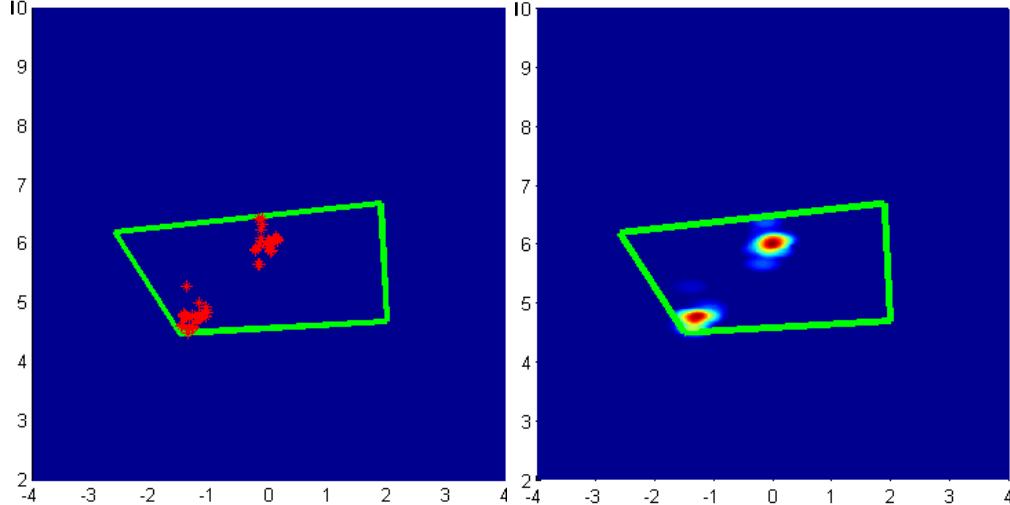


图 4.5 概率分布图
4.5 the probability distribution

在概率空间 $E(\mathbf{x}, \theta)$ 中的局部最大点对应着离散空间中点最稠密的位置。在空间中搜索出所有的局部最大点，从一个起始点开始不断在其邻域范围内寻找更大的点作为新的位置，直至不能找到更大的点时则达到局部最大点的位置。在邻域的不同方向搜索时，其空间梯度的反方向是最速上升方向，即是向局部最大值最快逼近的方向。因此计算概率分布函数对变量 \mathbf{x} 和 θ 的偏导数，即梯度方法：

$$\frac{\partial E}{\partial \mathbf{x}} = \sum_{i=1}^{n_i} H_\theta(d_i(\theta)) \frac{\partial (\sum_{j=1}^{n_j} w_j H_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i))}{\partial \mathbf{x}} \quad (4.15)$$

$$\frac{\partial E}{\partial \theta} = \frac{\partial (\sum_{i=1}^{n_i} H_\theta(d_i(\theta)))}{\partial \theta} \sum_{j=1}^{n_j} w_j H_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i) \quad (4.16)$$

概率密度函数的梯度值在 mean shift 方法被近似为核函数导数之和，也就是上式能重新表达为：

$$\frac{\partial E}{\partial \mathbf{x}} = 2\mathbf{A}_{\mathbf{x}}^2 \left[\sum_{i=1}^{n_i} H_{\theta}(d_i(\theta)) \sum_{j=1}^{n_j} w_j K_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i) \right] \Delta \mathbf{x} \quad (4.17)$$

$$\frac{\partial E}{\partial \theta} = 2A_{\theta}^2 \left[\sum_{i=1}^{n_i} K_{\theta}(d_i(\theta)) \sum_{j=1}^{n_j} w_j H_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i) \right] \Delta \theta \quad (4.18)$$

偏导数中的位置偏移量 $\Delta \mathbf{x}$ 和角度偏移量 $\Delta \theta$ 分别为：

$$\Delta \mathbf{x} = \frac{\sum_{i=1}^{n_i} H_{\theta}(d_i(\theta)) \sum_{j=1}^{n_j} w_j \mathbf{x}_j K_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i)}{\sum_{i=1}^{n_i} H_{\theta}(d_i(\theta)) \sum_{j=1}^{n_j} w_j K_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i)} - \mathbf{x}$$

$$\Delta \theta = \frac{\sum_{i=1}^{n_i} \theta_i K_{\theta}(d_i(\theta)) \sum_{j=1}^{n_j} w_j H_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i)}{\sum_{i=1}^{n_i} K_{\theta}(d_i(\theta)) \sum_{j=1}^{n_j} w_j H_{\mathbf{x}}(d_j(\mathbf{x}), \theta_i)} - \theta$$

其中 K_{θ} 为核函数 H_{θ} 的负导数函数，而 $K_{\mathbf{x}}$ 为核函数 $H_{\mathbf{x}}$ 的负导数函数，即 $K_{\theta} = -H'_{\theta}$, $K_{\mathbf{x}} = -H'_{\mathbf{x}}$ 。概率空间 E 的局部最小值应满足 $\frac{\partial E(\mathbf{x}, \theta)}{\partial \mathbf{x}} = 0$ 和 $\frac{\partial E(\mathbf{x}, \theta)}{\partial \theta} = 0$ 条件，为此需要用一个偏移量 $\Delta \mathbf{x}$ 和 $\Delta \theta$ 来对 \mathbf{x} 和 θ 进行修正得到一个更接近最大值的点 $(\hat{\mathbf{x}}, \hat{\theta})$ 。

$$\hat{\mathbf{x}} = \mathbf{x} + \Delta \mathbf{x} \quad (4.19)$$

$$\hat{\theta} = \theta + \Delta \theta \quad (4.20)$$

然后以新参数 $(\hat{\mathbf{x}}, \hat{\theta})$ 为自变量代入再次代入偏移公式 (4.17) 和 (4.18) 中，寻找更接近峰值的点。当 \mathbf{x} 和 θ 最终到达最大值时偏移量将为0，修正过程也相应结束。整个迭代收敛的过程通常被称为mean shift过程。

最初计算时以投影点 $\begin{bmatrix} x_{pj} & y_{pj} \end{bmatrix}$ 和随机角度 θ 作为变量初始化(\mathbf{x} 和 θ)，经过迭代爬山过程收敛到局部最大值($\hat{\mathbf{x}}, \hat{\theta}$)并建立两者的关联关系。每个投影点以这样的方式划分给一个局部最大值，而属于同一个局部最大值的投影点构成一个聚类集合。同时局部最大值的位置和方向被认为是对应聚类集合的位置和方向。

在场景中的目标彼此相隔很远时，特征点和投影点很容易被划分成不同的聚类集合；但多个目标非常接近时，mean shift过程可能将多个目标的特征点聚合在一起形成一个包含较大的集合。这样的聚类包含多个目标，在尺寸上在并不满足预先设定的要求，它需要被进一步被强行切分成多个小的集合。

大聚类 C_k 中的方向 $\hat{\theta}$ 是核函数覆盖最多投影点的角度，在该方向上任意两点的距离可以定义为：

$$\hat{d}(\mathbf{x}_j, \hat{\mathbf{x}}_k) = \frac{(x_j - \hat{x}_k) + \tan(\hat{\theta}_k)(y_j - \hat{y}_k)}{\sqrt{1 + \tan^2(\hat{\theta}_k)}} \quad (4.21)$$

较小聚类中投影点与聚类中心之间的距离都较小，而大聚类中则对应距离较大。为了评测聚类是否属于大聚类并被进一步切分，基于方向距离定义聚类内部投影点的方差为：

$$var(C_k) = \frac{\sum_{\mathbf{x}_j \in C_k} w_j |\hat{d}(\mathbf{x}_j, \hat{\mathbf{x}}_k)|}{\sum_{\mathbf{x}_j \in C_k} w_j} \quad (4.22)$$

对于较小的聚类，具有大权重的投影点都靠近集合的中心点，即该聚类的方差 $var(C_k)$ 较小。而大的聚类包含多个目标的投影点，造成许多大权重的点都远离集合中心点导致统计方差值比较大。因此，当某个聚类的方差大于设定阈值时被标记为大聚类并将被切分成多个小的聚类。

含有多个目标的大聚类在切分时是为了选择两个新的中心并用它们划分集合中的所有投影点。这两个新的中心位置必须彼此远离，但又需要尽量靠近大权重的投影点。因此定义一个新的目标函数搜索两个新的中心点：

$$F(\hat{\mathbf{x}}_{s1}, \hat{\mathbf{x}}_{s2}) = \sum_{\mathbf{x}_j \in H(\hat{\mathbf{x}}_{s1})} w_j d(\mathbf{x}_j, \hat{\mathbf{x}}_{s1})^2 + \sum_{\mathbf{x}_j \in H(\hat{\mathbf{x}}_{s2})} w_j d(\mathbf{x}_j, \hat{\mathbf{x}}_{s2})^2 - \lambda \log(d(\hat{\mathbf{x}}_{s1}, \hat{\mathbf{x}}_{s2})^2) \quad (4.23)$$

其中 $H(\hat{\mathbf{x}}_{s1})$ 和 $H(\hat{\mathbf{x}}_{s2})$ 是等式 (4.23) 中的以 $\hat{\mathbf{x}}_{s1}$ 和 $\hat{\mathbf{x}}_{s2}$ 为位置核函数，而 λ 是惩罚因子尽量使两个新聚类的中心相隔较远。实际的应用中可以简单定义 λ 为以聚类方差为变量的函数：

$$\lambda = f(var(C_k)) \quad (4.24)$$

当聚类的方差越大时，相应的惩罚因子也应给予更大的惩罚力度。因此惩罚因子的函数 f 可使用其他的单调递增函数。

寻找新的聚类中心也即是确定两个新中心使目标函数 (4.23) 最小化。采用梯度迭代方法计算目标函数的尺度负梯度值调整新中心的位置，其目标函数的梯度表达为：

$$\frac{\partial F}{\partial \hat{\mathbf{x}}_{s1}} = 2[\lambda \frac{1}{d(\hat{\mathbf{x}}_{s1}, \hat{\mathbf{x}}_{s2})} \frac{\partial d(\hat{\mathbf{x}}_{s1}, \hat{\mathbf{x}}_{s2})}{\partial \hat{\mathbf{x}}_{s1}} - \sum_{\mathbf{x}_j \in H(\hat{\mathbf{x}}_{s1})} w_j d(\mathbf{x}_j, \hat{\mathbf{x}}_{s1}) \frac{\partial d(\mathbf{x}_j, \hat{\mathbf{x}}_{s1})}{\partial \hat{\mathbf{x}}_{s1}}] \quad (4.25)$$

$$\frac{\partial F}{\partial \hat{\mathbf{x}}_{s2}} = 2[\lambda \frac{1}{d(\hat{\mathbf{x}}_{s1}, \hat{\mathbf{x}}_{s2})} \frac{\partial d(\hat{\mathbf{x}}_{s1}, \hat{\mathbf{x}}_{s2})}{\partial \hat{\mathbf{x}}_{s2}} - \sum_{\mathbf{x}_j \in H(\hat{\mathbf{x}}_{s2})} w_j d(\mathbf{x}_j, \hat{\mathbf{x}}_{s2}) \frac{\partial d(\mathbf{x}_j, \hat{\mathbf{x}}_{s2})}{\partial \hat{\mathbf{x}}_{s2}}] \quad (4.26)$$

使用目标函数梯度值多次迭代调整新的聚类中心，并最终达到稳定状态后确定新聚类的两个中心点。以这两个位置点和大聚类的方向作为新产生聚类的中心和方向，原来聚类中的所有投影点依照与新聚类中心的位置来划分。如果投影点 \mathbf{x}_j 到聚类 $\hat{\mathbf{x}}_{s1}$ 的距离较近（即 $\hat{d}(\mathbf{x}_j, \hat{\mathbf{x}}_{s1}) < \hat{d}(\mathbf{x}_j, \hat{\mathbf{x}}_{s2})$ ）时，它将划分到该聚类中；反之，它应当划分到聚类 $\hat{\mathbf{x}}_{s2}$ 中。以上的过程不断切分大聚类，直至所有新产生的聚类不再满足设定的分裂条件为止。

4.3 多目标的跟踪

一旦目标在监控场景中被检测、标定后，在随后的连续帧中跟踪算法确定它在时间序列的关系、提取每帧中的目标区域、分析它在空间的移动轨迹。目标在每帧中的检测与对应既能联合工作也能独立工作，即跟踪算法可归结为两大类：

1. 对应跟踪方法。检测算法在每帧中提取可能的目标区域，跟踪算法则为两帧中的同一目标建立对应关系。这类方法中以单个点来代表目标位置，对应前后帧中相同的点产生目标运动轨迹。由于忽略目标所占区域大小，因而在遮挡、进入、退出等情况下很难确定其对应关系。常用的对应方法是基于目标运动光滑性假设^[140] 或测度不确定性建立约束条件寻找最优的对应关系^[141]。
2. 更新跟踪方法。通过迭代更新目标位置和目标区域信息，将目标检测与跟踪对应结为一体共同解决。该方法以目标前一帧的位置为起始点，根据其运动规律在当前帧中搜索一些可能的区域确定目标的最优位置。例如Comaniciu 和Meer^[138, 139] 提出用目标颜色组成的权重直方图作为特征，迭代搜索邻域寻找与目标颜色直方图最相近的区域作为目标在当前帧的位置。

在跟踪过程中目标在前一帧的位置是已知信息，而在运动模型可估计其在当前帧的大概位置。假设在 M 帧中跟踪带有 N 个特征的目标，对应跟踪方法需要做 NM 个寻找最大值的迭代计算；而更新跟踪方法则只需要 $N + M - 1$ 次迭代计算过程。由此可知，更新方法能减少大量的计算开销、实现对多目标的在线跟踪。

目标在 $t-1$ 和 $t-2$ 的位置分别为 $\hat{\mathbf{x}}_{t-1}$ 和 $\hat{\mathbf{x}}_{t-2}$ ，它的运动向量可表示为 $\mathbf{v}_{t-1} = \hat{\mathbf{x}}_{t-1} - \hat{\mathbf{x}}_{t-2}$ 。在线跟踪时采样的周期应该少于150ms，低速目标在这样的采样频率下可以近似为匀速运动，其运动方程可写为：

$$\mathbf{v}_t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{v}_{t-1} \quad (4.27)$$

将 \mathbf{v} 作为目标系统的状态量和观察量，则在更新方法中对目标位置的预测与修正均可由经典的卡尔曼滤波器来完成。具体过程如下：卡尔曼滤波器根据目标过去的运动向量预测新预测新的运动向量，在上帧位置的基础上得到当前帧的预测位置。Mean shift迭代过程以该位置开始寻找局部最大值作为对当前帧的观察位置和观察运动向量。最后再由滤波器修正观察量得到真实的目标位置，并更新目标的位置实现跟踪过程。图4.6 显示了连续四帧的由投影点构成的概率空间，图中的黑色圆点是滤波器预测的目标位置。事实上这些预测位置与局部最大值十分接近，在寻找最大值的过程中只需要几次迭代即可达到最大值位置。

一旦目标在当前帧的位置被确定，所有在此位置附近的投影点不需要计算被直接归为一个集合表示监控场景中的目标。跟踪算法只需要经过一次寻找最大值的计算过程，这不同于目标检测中多次计算过程。以这样的方式，跟踪算法所需要的计算量较小，能满足多目标同时跟踪的计算需求。

4.4 构造位置与方向核函数

在公式（4.14）中的位置核函数 H_x 与方向核函数 H_θ 表示目标特征点的空间性质，权重函数 w_j 则反映了投影点在确定目标位置（聚类集合的中心）时的相对重要性。通常，不同类型的目标有不同空间属性，也应该对应不同的核函数和权重函数。目前，监控系统中对于行人的检测与跟踪是比较热门的研究领域，但由于不能妥善地解决第一章所介绍的几个问题，造成目前能真正实用的系统还较少。本节中以人体作为例子说明如何构建检测人体的核函数和权重函数，以实现一个能在复杂环境下工作的人体检测与跟踪系统。

人体在俯视角度下能近似为一个长轴为 α 、短轴为 β 的椭圆形状，即椭圆的中心对应着人体的头顶，长轴对应着左右肩膀两端。参数 α 和 β 可设置为人体的平均宽度和厚度（需要观察目标在空间中大小的特性）。如上所述，人体从俯视角度看可由一个长轴为0.7m，短轴为0.3的椭圆形状来近似表达，而椭圆的中心对应的是人的头部，长轴的两端则对应的是肩膀。由于人的头部要高于肩膀，因此椭圆的中心是最高点，长、短轴的两端可设置为0，其具体的数学表达式可定义为

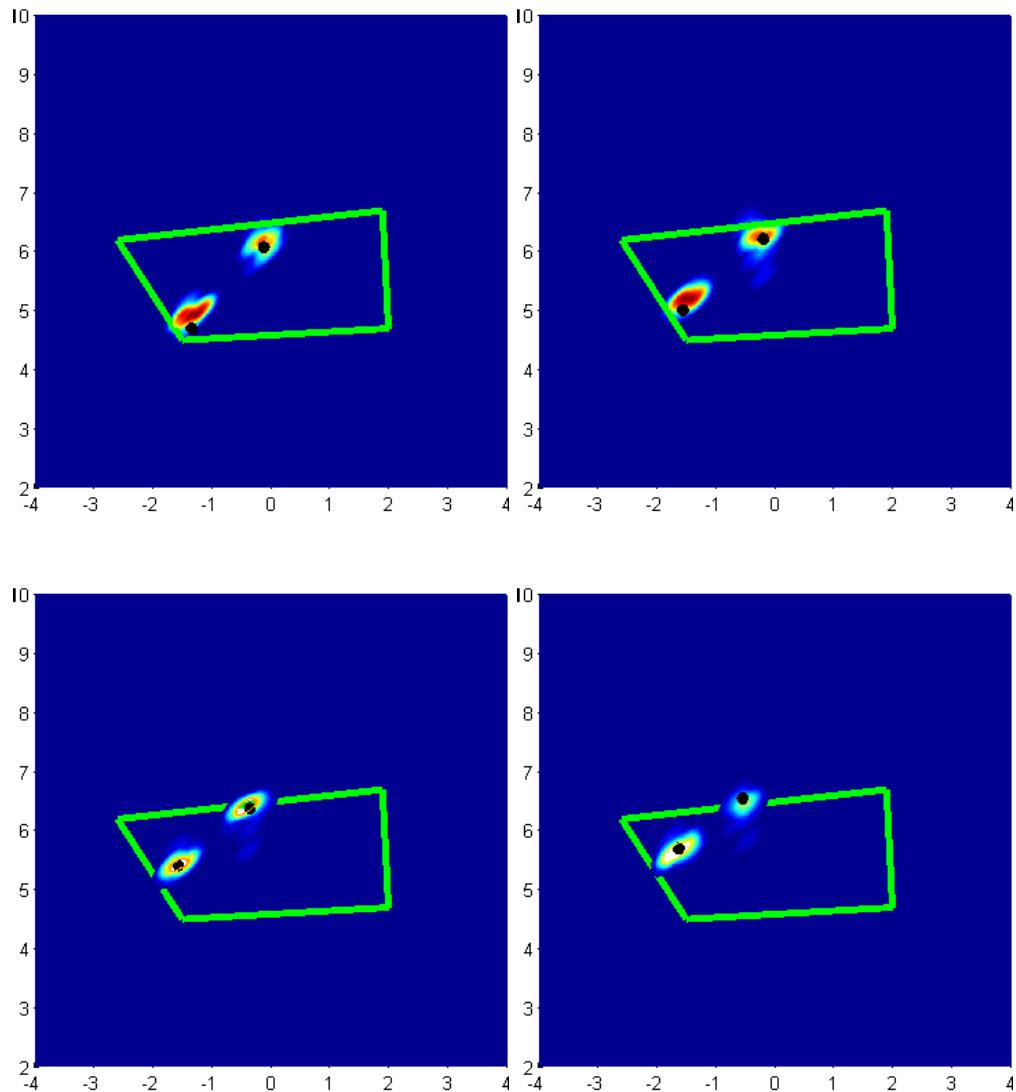


图 4.6 连续帧中的概率空间函数

4.6 The constructed probability function in sequential frames

$$\mathbf{A}_x = 2 \begin{bmatrix} 1/\alpha & 0 \\ 0 & 1/\beta \end{bmatrix} \quad (4.28)$$

再考虑到人体的头部高度高于肩膀两段的高度，因此椭圆的中心也应当高于长轴的两端，即就是高度的核函数 H 应当定义为：

$$H_x(x) = \begin{cases} 1 - x & 0 \leq x \leq 1 \\ 0 & otherwise \end{cases} \quad (4.29)$$

整合以上形状核函数 (4.28) 与高度核函数 (4.29) 可得到完整的目标核函数 H_x ，该函数在三维空间中的分布如图4.7所示。

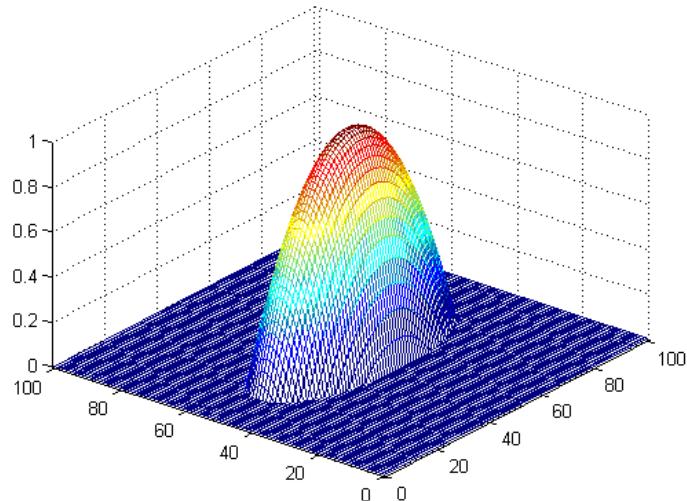


图 4.7 三维的目标核函数

4.7 The three-dimensional kernel function for the object

位置核函数 $H_x(d_j(\mathbf{x}), \theta_i)$ 在方向参数 θ_i 的作用下以椭圆中心将 $H_x(d_j(\mathbf{x}))$ 逆时针旋转 θ_i 角度得到的核函数。事实上核函数的设置在目标检测与跟踪中的关系十分重要，较大的核函数可能将多个相邻的聚类融合成一个大聚类；而较小的聚类也可能错误地将正常大小的聚类切分成多个偏小的聚类。因此，合适的核函数应当在计算开销和检测性能之间取得良好的平衡。实际的系统中的核函

数应该考虑目标的平均形状作为重要的参考依据，以它作为设置的参照标准在检测时取得较好地性能表现。

与形状和高度核函数不同，方向核函数表达了形状核函数对于中心的旋转角度。从 $[0, 2\pi]$ 区域里选取均匀选择 n_i 个角度样本值，定义 H_θ 为：

$$H_\theta(x) = e^{-x} \quad (4.30)$$

与位置核函数的宽度系数不同，方向核函数中的宽度系数不需要特别严格的设置，例如在人体检测与跟踪中可设置为 $n_i/2\pi$ 或是 $n_i/4\pi$ 。具体的系数只与跟踪角度有关，对目标的角度精度要求较高时选择较小的宽度、较多的角度样本；而角度的精度要求较低时则只需选择较大的宽度、较少的角度样本。

权重系数 w_j 是每个投影点在确定目标中心时的相对重要性。对于人体来说，头顶既是人体的中心也是离地板平面最高的点，也就是说目标的中心对应着距离地板平面的最高点，所以聚类的中心应当尽可能的靠近距离地面较高的点。权重函数应当给予越高的点越大的权重以尽可能地吸引聚类中心，其具体的定义可写为：

$$w_j = z_j \quad (4.31)$$

本章所述的检测与跟踪算法没有使用常见的图像特征如肤色和人脸等，而采用目标的空间属性作为目标检测与跟踪的依据。人体可能由于个体的差异在颜色、纹理等图像特征方面的差异较大，而空间特征则相对比较稳定并且个体差异较小。以特征聚类作为检测方法不针对某一类目标，可用于任何目标的检测与跟踪上，例如将以上的核函数设置为盒函数，权重设置为常数则可以用于车辆的检测与跟踪，因此所介绍的新系统比其他已有的系统具有更广泛的应用环境。

4.5 多目标系统的实验及评价

第二章所述的相机标定建立了摄像机与世界之间的关系，随后的第三章所述的矫正与匹配构成了立体的视觉系统还原出整个三维空间场景。基于立体视觉系统采集的三维数据，目标检测与跟踪系统负责将空间的多目标定位并在时

间序列上持续跟踪他们的运动轨迹。本节介绍一系列的实验测试所提检测与跟踪系统的性能和准确性。

尽管系统对跟踪的目标类型没有严格的要求，但在实验中出于对实验条件和复杂度的考虑仍然选择人体作为唯一的监控对象（以车辆作为监控对象时不仅需要更大的实验场地，而且场景中可用的对象数量较少）。采样第二节所提的特征点提取算法在摄像机所摄图像中提取局部纹理丰富的像素作为场景中的特征点，实验的背景既有单一的室内场景也有复杂的户外场景。以椭圆核函数作为跟踪行人的位置核函数，且长轴参数 α 选为0.7短轴参数 β 为0.3。

在开始实验之前将一些与地板平面颜色相差较大的小纸片均匀地撒在地板平面上，人为地制造一些地板平面上的特征点并拍摄监控场景图像。实验中的场景标定如图4.8所示，通过人为手工选取这些地板上的特征点作为标定的数据。然后用第一节介绍的方法标定地板平面，为后续特征点的投影操作提供参考平面。

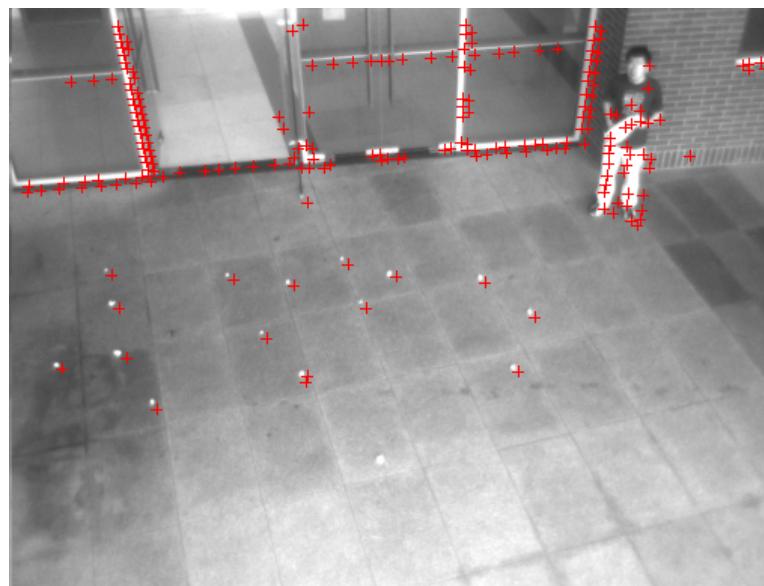


图 4.8 人为制造的特征点标定地板平面

4.8 The ground plane equation with man-made feature points

目前，公共的短基线立体视觉系统测试数据库比较少见，因此实验都采用自主采集的数据用来测评系统的性能。已有的立体视觉系统对摄像机的数量、位置、基线长度和一些辅助模块有不同的要求，因此很难采用一个公平的方式

去评价这些不同系统和算法。为了能公平地比较算法的性能，实验一中用本章所提的算法与一些传统的跟踪算法如mean shift和camshift做单一目标的跟踪对比实验。在图4.9的第一列和第二列中，mean shift算法^[138]和camshift算法^[142]被同样的初始化后在后续的图像序列中跟踪单一目标。从实验的结果可知：当目标与背景相差较大时((b) 和 (e))，他们能被mean shift(跟踪结果用矩形表达)和camshift(跟踪结果用椭圆表达)准确地跟踪；当目标与背景相似时((a)、(c) 和 (d))单目的跟踪算法很快就丢失目标，但双目系统计算通过特征点后投影到地板平面产生聚类集合能跟踪所有的目标。因为所提算法不采用颜色特征而是目标的空间属性作为跟踪特征，所以即使目标与背景相似时依然能被检测并跟踪(如图4.9所示)。此外，所提的监控系统能自动的检测目标不需要手动初始化目标位置，并多帧中跟踪目标所在的位置计算其真实的运动轨迹。

多目标在场景中运动经常出现相互的遮挡情况，这是多目标监控系统中比较难以解决的问题之一。当目标被严重遮挡时，仅有少量的特征能被检测到以至于它很难被检测和跟踪。通常基于颜色特征、人脸特征或是深度特征的系统对于遮挡目标的检测都存在许多问题。图4.10是遮挡情形中的一种，即多对人相向交叉行进。虽然这种目标之间的遮挡往往是短暂现象，但是当两对行人相遇时他们的运动变得十分难以预测。目标可能交叉相遇后各自返回相反方向或者继续向原来方向运动。由于被遮挡的目标在稠密深度图和投影图中仅对应着较小区域，因此造成^[54]中所述方法无法区分两个遮挡目标。尽管遮挡只持续较短的时间，但它仍会引起目标轨迹产生剧烈波动。跟踪算法将附近投影误认为是目标位置，从而造成了一些错误的跟踪结果(见图4.10中第二列)。遮挡只是减少目标特征点的数量，但我们的系统并不依靠特征点的数量检测、跟踪目标，因此已有的这些特征点依然能产生与正常目标一样的独立聚类(见图4.10中第三列)。

除了上述交叉队列的遮挡情况外，另一种遮挡是队伍向摄像机的方向移动，队列前部的目标遮挡后部目标。在图4.11中，一队行人向摄像机的方向移动，而队列中排在第三个的女孩几乎被前面两个男孩全部遮挡，只留下小部分头部区域能被观察到。这种情况下，方法^[54]丢失了目标并给出了错误的目标数量和运动轨迹。遮挡在我们系统中也同样产生了较少的特征点，但是这些少

数特征点与地面相距较大而被赋予极大的权重值，因此系统产生了正确的聚类集合并成功地定位目标在空间中的位置。实验结果如图4.11中第三列所示。

方法[54]为了正确的跟踪遮挡目标，需要在多个不同视角上安装多个立体视觉系统产生完整的俯视图，需要比我们的系统更多硬件设备。此外，系统标定和俯视图的配准也产生了更高的计算量，这造成系统[54]无法在线工作。

以上实验均在室内环境中，场景的背景简单且光照条件比较稳定，而在室外环境中摄像机没有天花板的限制能安置在较高的位置，从而目标之间的遮挡在这样的视角情况下发生的概率更小更有利解决目标的遮挡问题。另一方面，室外的环境比室内环境又复杂许多，如快速地光照变化和复杂的背景会对基于颜色特征的监控系统产生严重的干扰，同样也会增加立体视觉中视差计算的噪音。图4.12中所提系统监控某一建筑的出口、跟踪不同行人的行走方向，而场景中的光照从出口向远处不断增强。系统[54]不能区分相邻的目标，而基于对应策略的跟踪算法又无法跟踪快速移动，因此在图4.12第二列中产生了错误的跟踪结果。在我们系统中指定的监控区域能过滤掉不在其范围内的背景特征点，而特征点视差比稠密视差含有较少的计算错误。视察中的噪音点投影到地板平面后并不会聚集在某个小区域内，它们通常分散在整个区域产生较小的概率密度。图4.12中第三列的实验结果显示我们系统的检测与跟踪能在复杂背景、光照变化的室外条件下正常工作。

大多数检测与跟踪算法都会对目标之间的间距比较敏感。如果目标彼此距离较近，它们可能被聚合成一个较大的集合从而导致系统产生错误的检测与跟踪结果。图4.13是对两个相邻紧密目标的监控结果，文[54]所述系统的监控结果显示在图4.13中的第二列，其中：第一对目标之间的距离大于聚类算法的最小分辨极限因而她们被很容易地分开并独立地进行跟踪；而第二对和第三对目标之间的距离太小造成投影形成一个集合而无法区分。这种情况下，本章所述系统也同样会将实验中的目标聚合成一个聚类集合，但随后分裂算法可自动鉴别出这个大聚类集合并进一步将其切分成多个较小的聚类。这些小的聚类虽然具有不同的中心位置但其方向仍然和原来的聚类一致（见图4.13第三列所示）。目标特征点在最后一帧中被正确的聚类成两个集合，因此在没有进行分裂操作的情况下所产生的聚类具有独立的方向。

实验中进一步模拟了雨天环境下的行人检测与跟踪情况，以测试系统在光

照剧烈变化、大量阴影以及严重遮挡情况下的检测结果。图4.14中逐步变化的光照变化在俯视图中产生了大量的噪音区域。同时行人在雨天时会撑起雨伞遮挡住身体的大部分区域，从而造成在俯视视角下雨伞投影区域比系统所构造的形状核函数区域超出很多。这些的极端条件对监控系统都是带来了极大的挑战。在这种复杂情况下，文[54]所述系统和我们的系统常常将单个雨伞错误地检测为多个目标并进行跟踪。但虚假目标跟踪过程中会被噪音区域或是附近目标所吸引从而逐步远离真实目标的位置。但我们系统中的特征点视差计算比稠密视差计算产生更少的噪音，同时噪音区域产生较小的聚类会在跟踪之前被设定的阈值所过滤去除。随着跟踪的不断持续和视角的改变，虚假目标会逐渐消失在真实目标之中，并会产生太多的错误的结果（见图4.14第三列所示）。

在以上所有的实验中，系统检测与跟踪结果按不同的场景进行量化评估，包括：两方向和多方向的目标队列；撑伞的目标队列。具体的评测结果如表4.13所示，其中准确率是正确被跟踪的目标数量与所有出现的目标数量之比。

条件	准确率
两方向(3ft)	95.08%
两方向(1.5ft)	91.67%
多方向	89.29%
雨伞	80.95%

* 3ft和1.5ft是相邻目标之间的距离

表 4.1 系统的量化结果
4.1 The system quantitative result

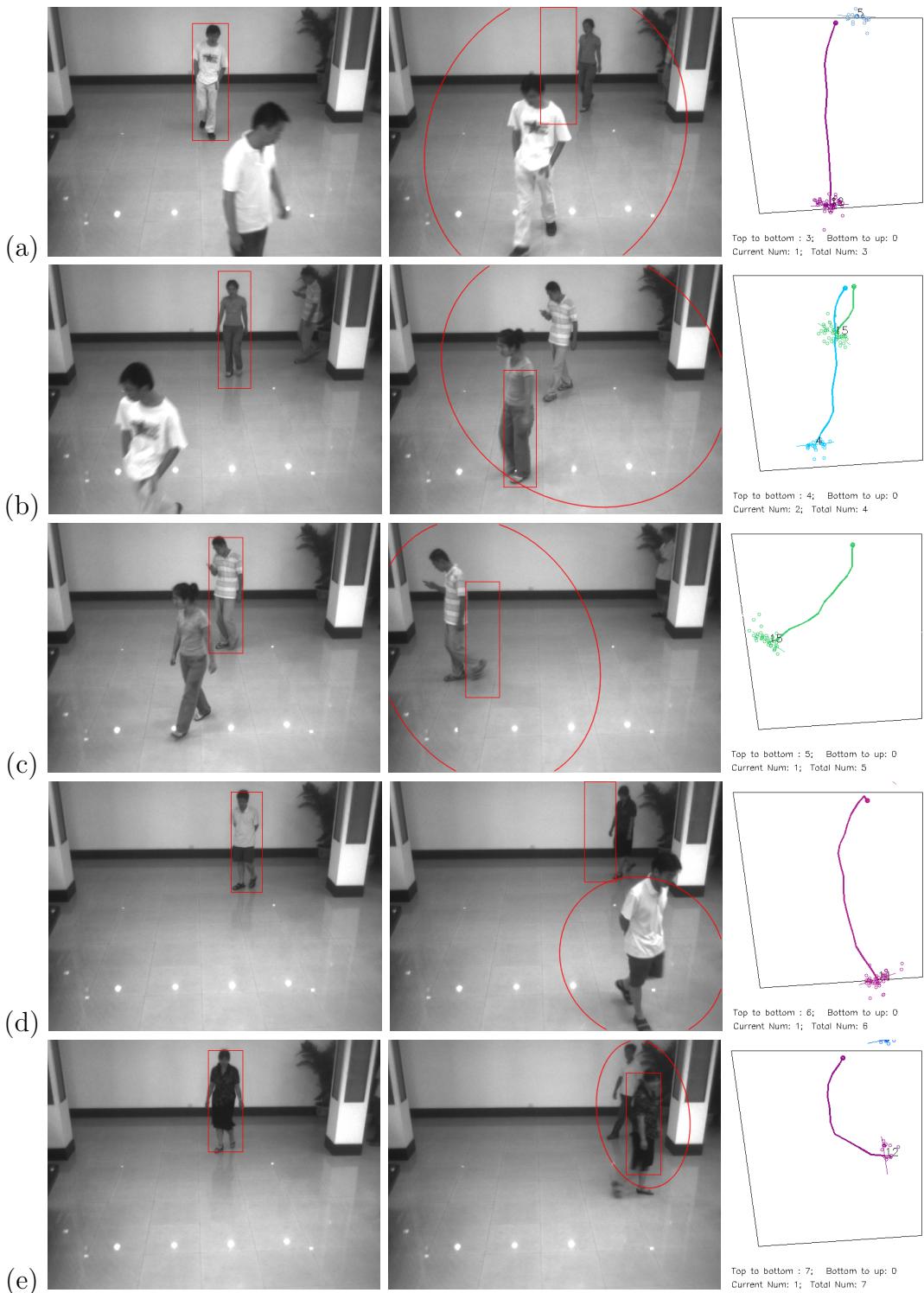


图 4.9 单目标跟踪的对比结果

4.9 The compared result for Single object tracking

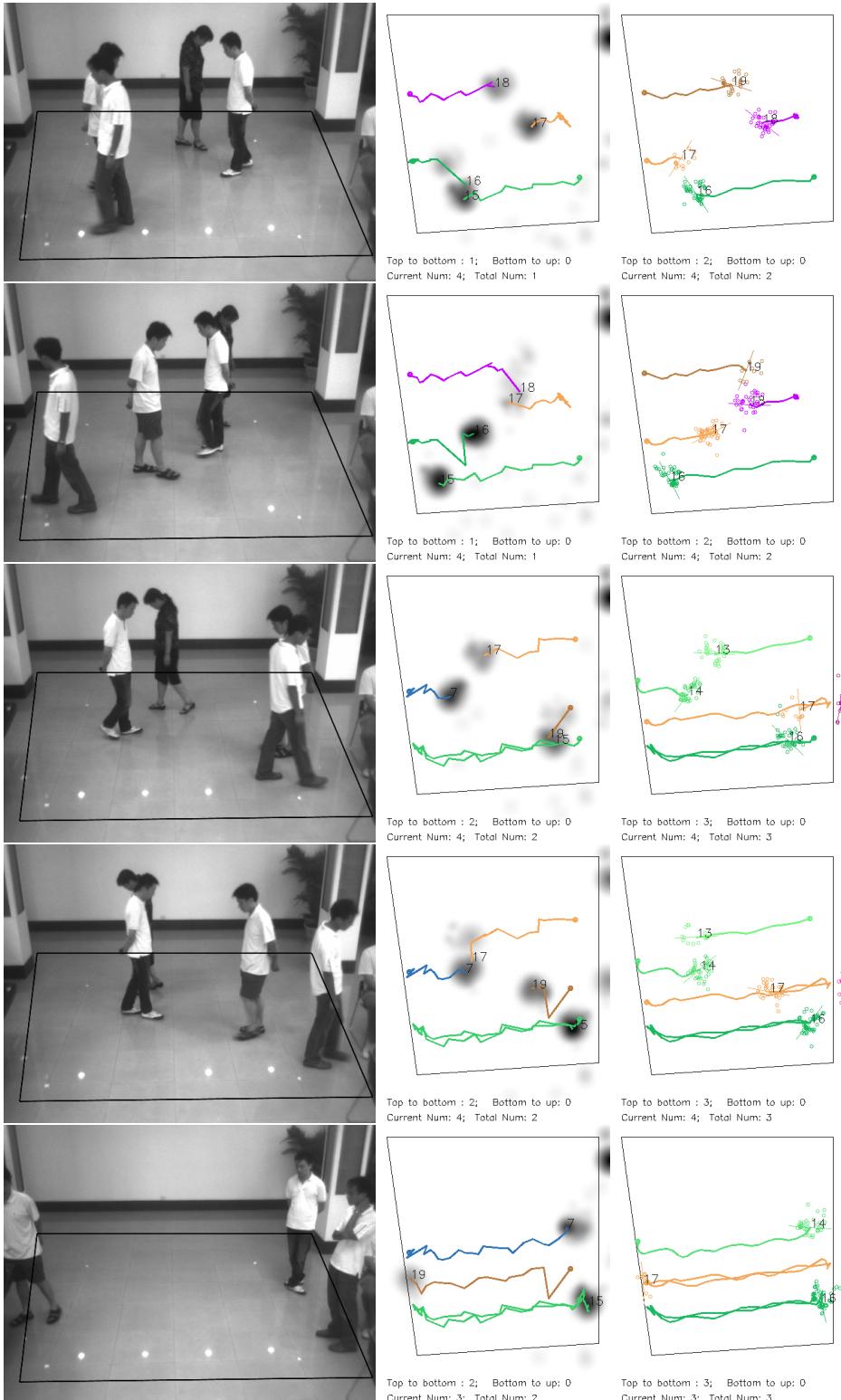


图 4.10 遮挡情况下的多目标跟踪

4.10 Multi-object tracking under occlusion

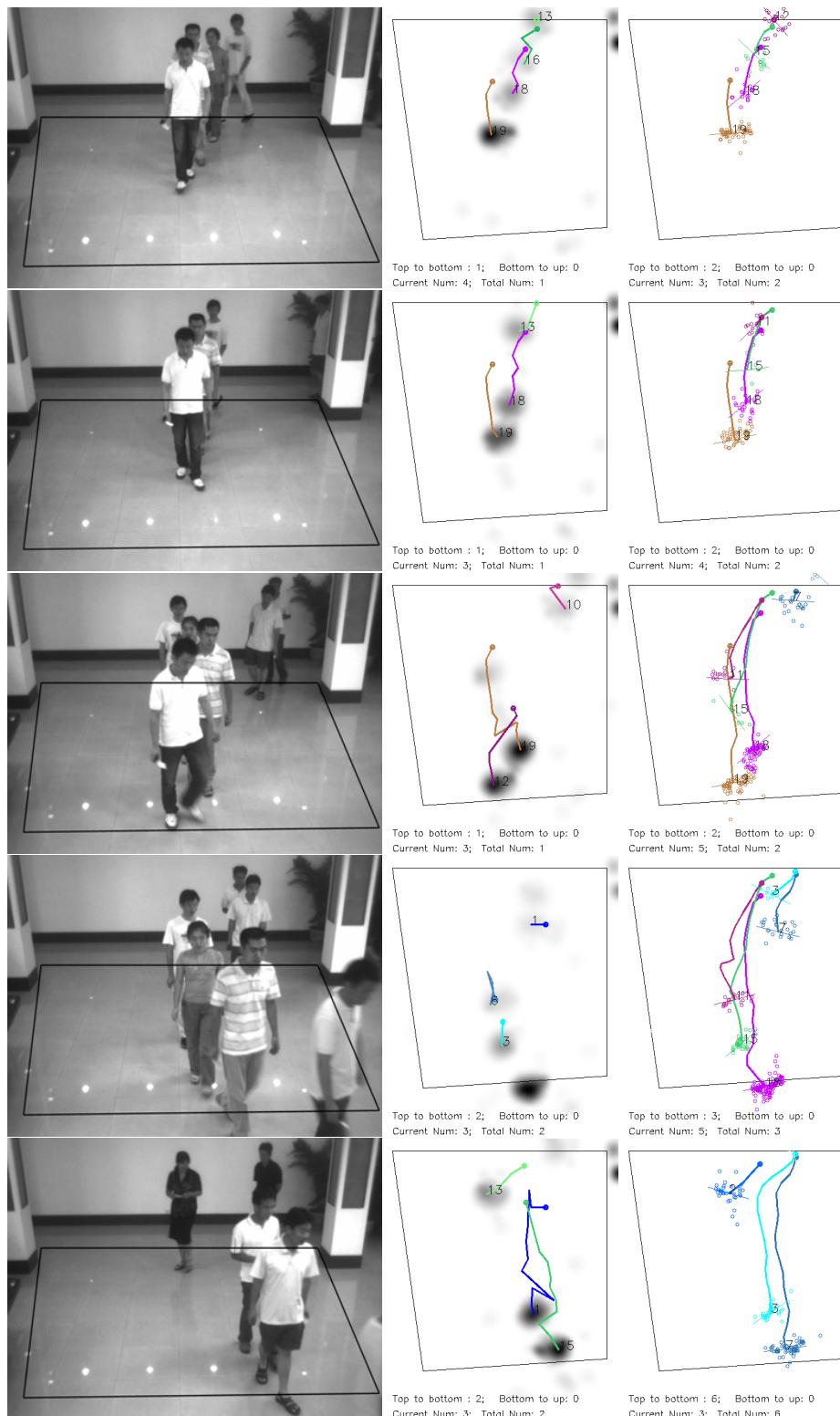


图 4.11 两队交错列多目标跟踪

4.11 Multi-object tracking with two queues moving toward each other

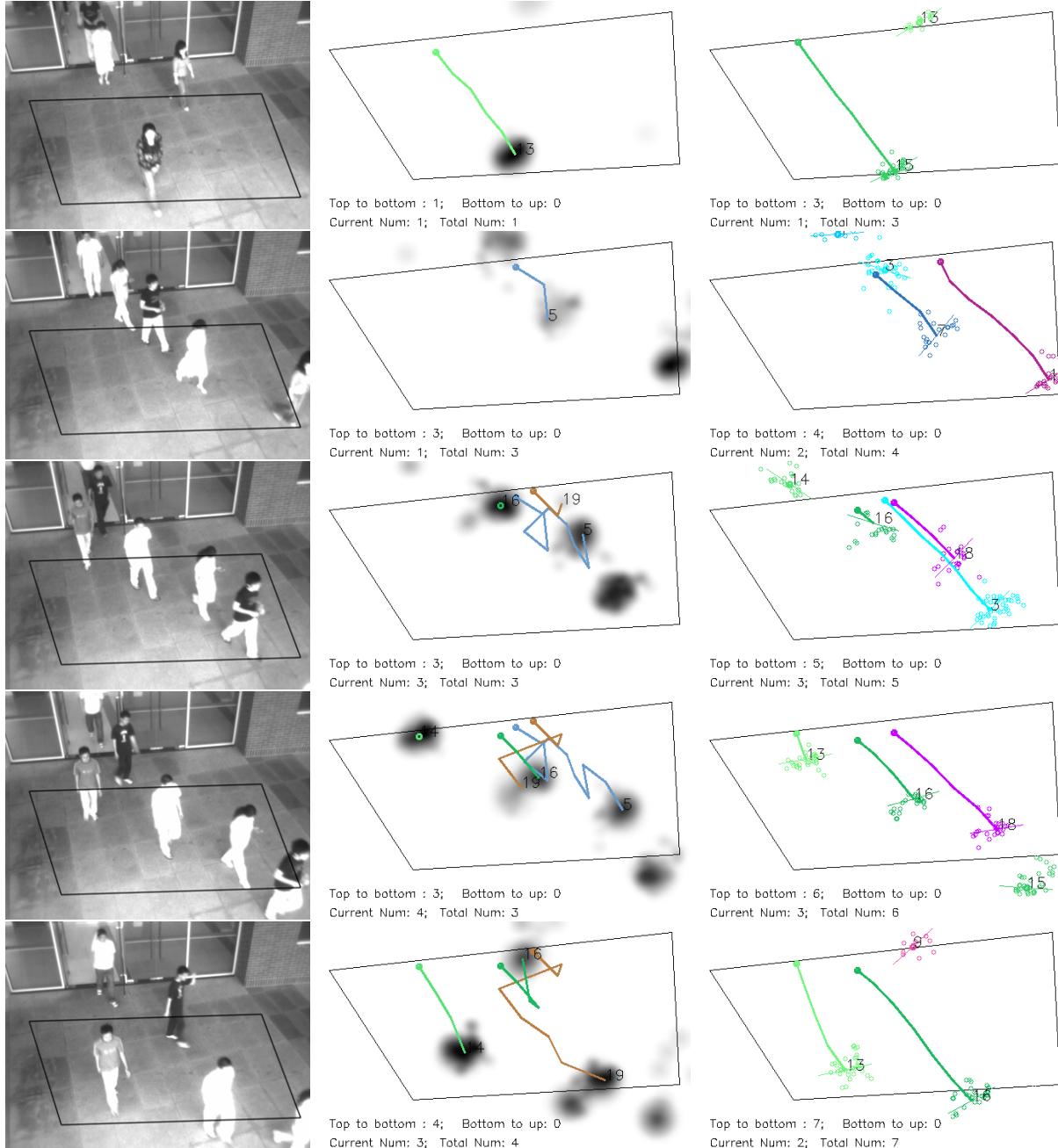


图 4.12 室外环境中的多目标跟踪

4.12 Outdoor multi-object tracking

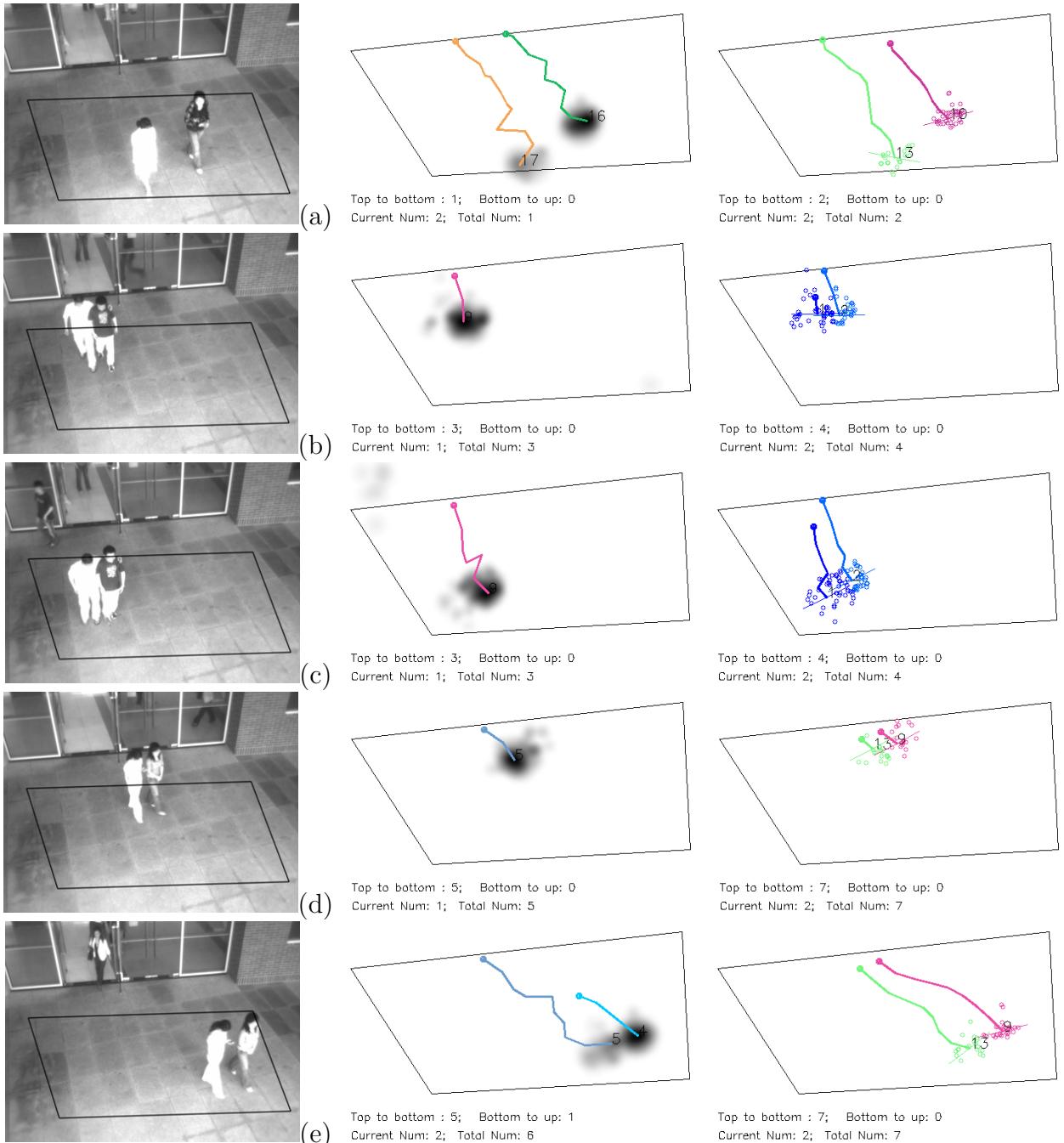


图 4.13 复杂条件下的多目标跟踪（紧密目标）

4.13 Complex multi-object tracking (small distance among objects)

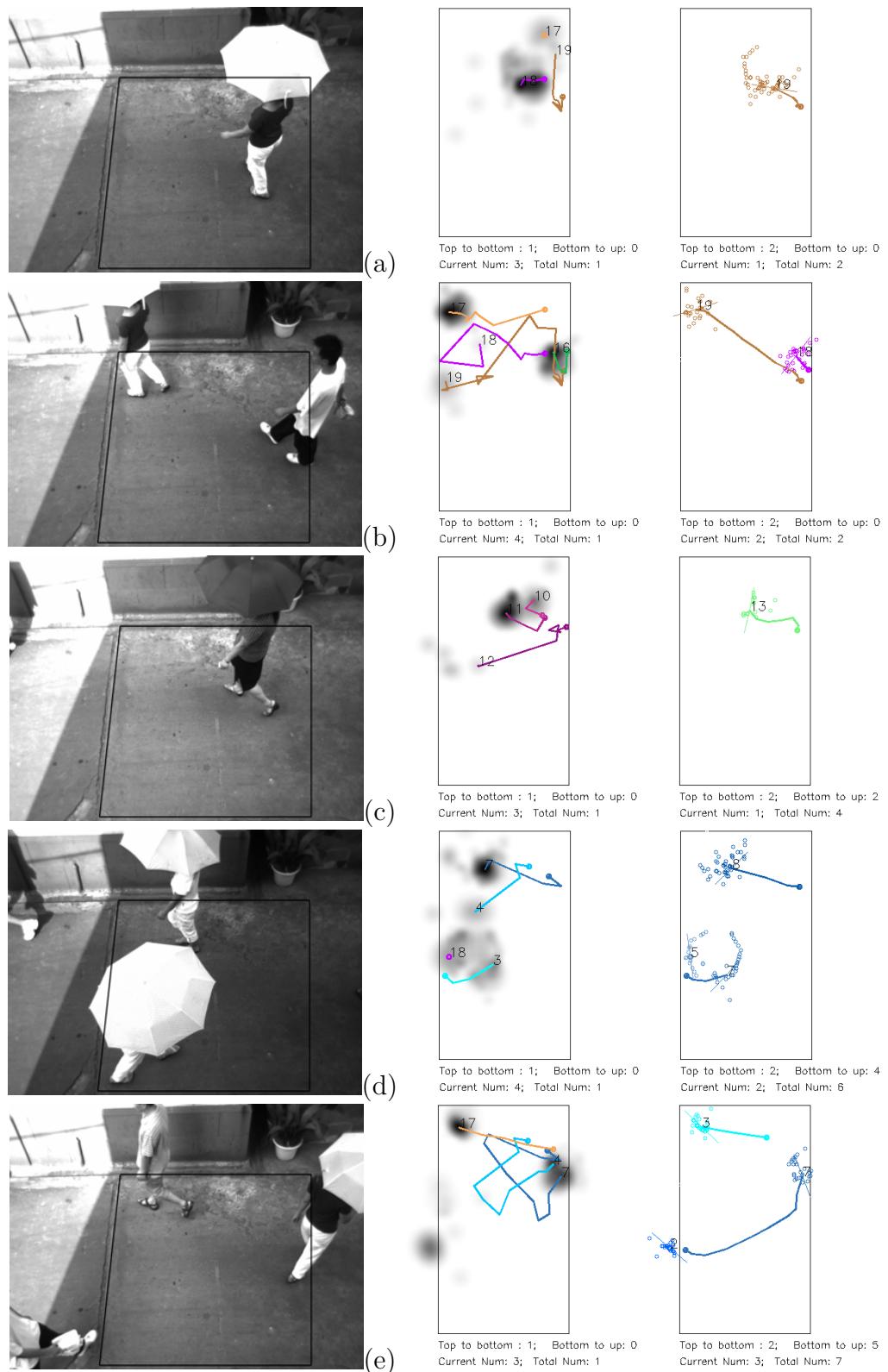


图 4.14 复杂条件下的多目标跟踪 (带伞遮挡)

4.14 Complex multi-object tracking (pedestrians with umbrellas)

第五章 多特征融合的轮廓跟踪

第四章介绍了基于核函数聚类的多目标的检测与跟踪方法，能在线准确地确定监测监控场景中目标的数量、位置和速度等信息。该系统为了能快速地检测并跟踪多个目标，以多个特征点在空间中产生的聚类集合来表达目标。尽管以特征点的方式表达目标作为处理对象，大大减少了监控的计算复杂度，使得系统能在线的处理多个目标；但是基于特征点的方式无法给出精确的目标区域，不利于在更高层次理解监控目标的行为。一些经典的方法在跟踪目标时将目标区域简化为一些近似的几何形状，如：矩形或是椭圆形^[138, 139, 143]，即使这些形状也都只是目标的大概区域而不是其真正精确的目标区域。这些预先设定的固定形状在处理实时目标大小变化时也存在一些问题，同时它们也无助于理解目标姿态和分析目标运动。基于轮廓的目标跟踪方法以目标的精确轮廓或区域为跟踪对象，一旦在第一帧中给出初始的目标轮廓，跟踪算法将在后续帧中持续提取出完整的目标轮廓。

将基于立体视觉的检测与跟踪方法与轮廓跟踪方法有机结合在一起将构成功能全面的监控系统：前者在宏观上负责对场景中所有目标的快速检测和跟踪；后者则在微观上对单个目标实现精确地跟踪、理解其的运动状态和姿态。注重在整体层面观察场景的状态必将牺牲一些细节部分，而分析细节层面也就无法观察到整体的动向，但两个系统能互相配合实现不同层次的监控功能，形成从整体到局部从粗略到细微的全面监控能力。在本章中将着重介绍基于单摄像机的单目标轮廓跟踪算法与系统。

总的来说，单目标跟踪算法的性能与目标特征的分辨能力有十分密切关系，而常用的特征包括：颜色、纹理、边缘、运动和帧差等。Bertalmio等构造了测量两帧之间颜色差的形变等式用于在图像序列中演化目标轮廓。基于颜色帧间不变性的假设前提，文^[144]将目标的轮廓跟踪建模为贝叶斯估计问题。最近的文^[138]用颜色直方图表达移动目标的特征，mean shift方法在可能的目标位置中搜索最可能的位置。对比于颜色特征，纹理特征在光照变化情况下具有更高的稳定性。文^[145]用颜色特征用纹理相似检测与背景最不相似的区域作为背景区域，而文^[146, 147]则根据马尔科夫随机过程沿直线快速检测目标的纹理边

界，然后从中重构出目标的投影轮廓。

单一固定的目标特征在跟踪目标时常常显得力不从心，因此最近的一些新方法尝试在线^[148]或离线^[149]状态下从多个候选特征中选择分辨力最好的目标特征作为跟踪目标重要线索。同时一些方法也通过组合多个特征在不同环境下鲁棒地跟踪目标，例如：Yilmaz^[150]等由独立投票策略依照特征的分辨能力作为权重参照值，整合颜色和纹理特征后统一估计目标的后验概率。在文^[151]中，纹理特征被用于在背景中分离出目标区域，跟踪算法基于对颜色、纹理和运动的统一测量建立帧间的目标对应关系。Paragios和Deriche^[152]组合帧差和移动目标的边界在后续帧中共同推动测量线活动轮廓模型(geodesic active contour)^[153, 154]向目标轮廓运动。此外，他们也在文^[155]中整合边缘、灰度和运动信息为一体使得初始的目标曲线能在偏微分方程的推进下收敛于目标边界。

总结起来，存在的这些方法都是使用区域特征（颜色和纹理）和边界特征（边缘和帧差）跟踪监控场景中的目标，其中：区域特征能快速的、粗略的定位目标位置，但边界特征能提供能在局部提供更精确的目标形状信息。但这些方法没有系统地组合区域和边界信息，以致于无法在一些如建筑阴影下的道路或低对比度和复杂背景下鲁棒地跟踪目标。因为它们遇到目标与背景有相似颜色的情况时系统的跟踪性能会急剧下降，尤其在一些灰度视频监控系统中经常出现这些问题从而造成系统不能处理。

本章中将介绍的新目标跟踪方法通过一个能量泛函系统地整合区域和边界特征。这两类特征通过一个新的权重项系统地组合在一起相互弥补彼此的不足（如区域特征不能精确的提取目标边界而边界特征却常常是不完整的片段轮廓）。区域特征用于计算像素属于目标和背景的后验概率；边界特征同时产生新的测度函数在局部范围内光滑曲线并将其与准确的目标轮廓准确对齐。关于目标曲线的新能量泛函由两类特征的加权和构造而成，从而正确的目标曲线能使提出的能力泛函达到最小。

特征选择在目标跟踪是至关重要的因素之一。如上所述，单一特征是难以鲁棒地在复杂环境下跟踪目标，而系统地融合多个特征能获得更全面的判断与决策。图5.1描述了包含区域特征模块与边界特征模块的系统融合结构。贝叶斯方法结合颜色和纹理估计像素的后验概率，即像素属于目标或者背景的概率；

另一方面，边界能量 $E_{boundary}(C(s))$ 用帧差特征和边缘特征计算边界能量并累积轮廓上的 $C(s)$ 的边界能量总和。

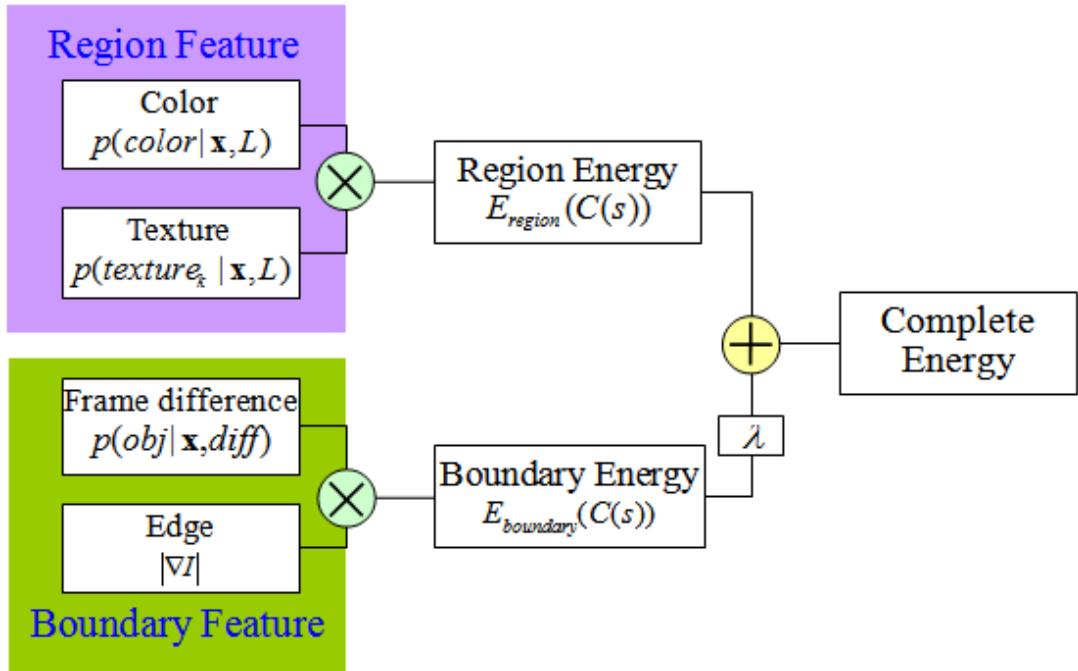


图 5.1 本文所提的模型架构

5.1 The proposed model architecture

5.1 区域特征分析及区域能量泛函

颜色和纹理等区域特征在非参数模型中表达目标和背景的特征。在跟踪的初始条件中第一帧 I_0 的目标区域已被指定，在目标区域内部的像素标记为目标像素而外部的像素则标记为背景像素。目标像素构成正样本集合 $Obj = \{pixel \in object\}$ 而背景像素构成负样本集合 $Bck = \{pixel \in Background\}$ ，从这两个样本集合中提取区域特征如颜色和纹理表达目标和背景的特征。第n帧 I_n 中像素 x 在这些特征上与目标（或背景）的相似性度量 $p(RF|x, obj)$ 能有效地搜索目标的位置。最初的mean shift方法^[138]仅使用正样本集合的颜色直方图近似 $p(color|x, obj)$ 。最近Collins等在49个不同颜色空间中对比 $p(color|x, obj)$ 和 $p(color|x, bck)$ ，挑选出分辨能力最强的颜色空间跟踪目标。由于同时考虑了目标和背景的特征，原始mean shift算法的跟踪性

能被大大提高。不仅是颜色特征的相似性测度，基于纹理特征的相似性测度 $p(texture|\mathbf{x}, obj)$ 和 $p(texture|\mathbf{x}, bck)$ 也常被用于目标跟踪之中^[145, 151]。

在本章所介绍的轮廓跟踪算法利用贝叶斯方法将条件概率 $p(feature|\mathbf{x}, bck)$ 转成像素 \mathbf{x} 属于目标和背景的后验概率。这个概率转换过程能表达为：

$$p(obj|\mathbf{x}, RF) = \frac{p(RF|\mathbf{x}, obj)p(obj|\mathbf{x})}{p(RF|\mathbf{x}, obj)p(obj|\mathbf{x}) + p(RF|\mathbf{x}, bck)p(bck|\mathbf{x})} \quad (5.1)$$

$$p(bck|\mathbf{x}, RF) = \frac{p(RF|\mathbf{x}, bck)p(bck|\mathbf{x})}{p(RF|\mathbf{x}, obj)p(obj|\mathbf{x}) + p(RF|\mathbf{x}, bck)p(bck|\mathbf{x})} \quad (5.2)$$

同等对待目标和背景类型，可令两个后验收概率相等 $p(obj|\mathbf{x}) = p(bck|\mathbf{x})$ ，则等式 (5.1) 和 (5.2) 说明对于像素 \mathbf{x} 的分类主要依靠于区域特征与目标相似性和背景相似性的比值。

像素相似性度量和分类需要确定三个参数： obj , bck and RF 。正负样本集分别代表参数 obj 和 bck 。以前的方法在整个目标区域内收集样本像素点，而这里只在目标轮廓附近区域内收集样本。那些在目标边界领域内目标像素和背景像素分别标记为正样本和负样本，如图5.2所示的红色曲线为目标边界，则在图5.2 (b) 中白色区域的像素构成正样本集并提取 RF 特征描述参数 obj 。同理，黑色区域的像素构成描述参数 bck 的负样本集。在边界之外标记为灰色区域的像素将被忽略不计。因为特征提取的计算复杂度与邻域区域成比例，较少的样本将节约许多特征分析时的计算量。

对于特征相似性度量 RF ，传统的跟踪算法应用标量值作为跟踪特征^[143–145]，可是如颜色和纹理组合成的多特征的向量也能作为相似度量。假设这些特征之间彼此独立^[148, 150–152, 155]，联合分布概率能切分成多个小因子概率函数：

$$p(RF|\mathbf{x}, L) = p(color|\mathbf{x}, L)p(texture_2|\mathbf{x}, L) \cdots p(texture_K|\mathbf{x}, L) \quad (5.3)$$

其中 $L \in \{obj, bck\}$, $texture_k (k = 1, 2 \cdots K)$ 为 k 尺度上的纹理特征。

首先介绍这些特征因子中的颜色概率模型 $p(color|\mathbf{x}, obj)$ 。在跟踪算法中^[150, 156]普遍使用直方图作为颜色特征表达目标特点，但实际上直方图是测

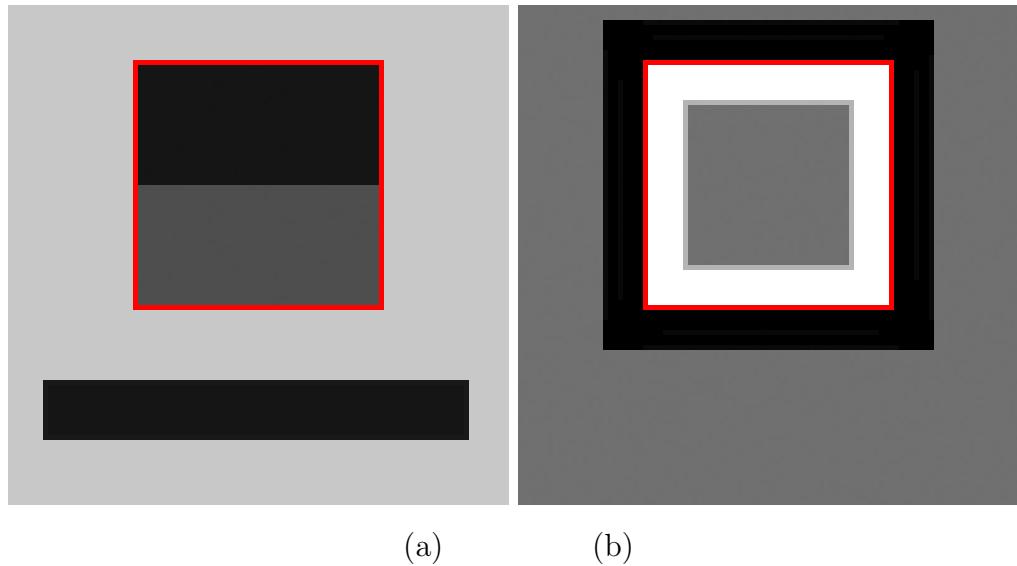


图 5.2 选择正负样本集合

5.2 the positive and negative sample set for selection

量某个像素具有某种特定颜色的概率，即它能泛化为核密度估计方法。因此，核密度估计测量帧 I_n 中像素 \mathbf{x} 在颜色特征上与目标或背景之间相似性度量：

$$p(\text{color}|\mathbf{x}, L) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{|I_n(\mathbf{x}) - I_0(\mathbf{x}_i)|}{h}\right), \quad (5.4)$$

其中 N 为正负集合样本的数量，带宽为 h 的Epanechnikov 核函数可记作为 $K(d) = 0.75(1 - \|d\|)^2$ 。选择Epanechnikov主要考虑其计算复杂性较低，估计方法中也能使用其他类型的核函数，如高斯函数也能作为核函数。

纹理特征在光照变化时具有比颜色特征更好的稳定性，因此它能辅助颜色特征产生更优的跟踪性能。许多方法^[150, 157, 158]都使用常见的Gabor滤波器^[159, 160] 产生纹理描述子，在一些比较性研究中显示出良好的区分性能，然而Gabor在产生描述子时需要极大计算量无法满足在线跟踪的实时处理要求。

最近，一个新的纹理描述子LBP（local binary pattern）以圆心像素为阈值将圆周上抽样点像素二值化。整个计算过程只需在邻域内进行简单的查表操作即可实现，因此它常被用于实时的目标检测和目标跟踪中^[151, 161, 162]。此外，LBP算法比Gabor滤波器更适用于一些特定区域的处理而无需计算整个图像区

域。LBP_{riu2}在LBP的基础上进一步扩展为旋转不变的纹理描述子，该纹理描述子的数学定义为：

$$LBP_{P,R}^{riu2}(\mathbf{x}, I) = \begin{cases} \sum_{p=0}^{P-1} H(I(\mathbf{x}_p) - I(\mathbf{x})) & U(LBP_{P,R}) \leq 2 \\ P + 1 & otherwise \end{cases} \quad (5.5)$$

其中 \mathbf{x}_p 为以圆心 \mathbf{x} 半径为 R 圆周上第 p 个抽样点， H 为单位阶梯函数。 P 是有尺度系数所确定的抽样点总数，而计算条件中的统一性测度定义为：

$$\begin{aligned} U(LBP_{P,R}) &= |H(I(\mathbf{x}_{P-1}) - I(\mathbf{x})) - H(I(\mathbf{x}_0) - I(\mathbf{x}))| \\ &\quad + \sum_{p=1}^{P-1} |H(I(\mathbf{x}_p) - I(\mathbf{x})) - H(I(\mathbf{x}_{p-1}) - I(\mathbf{x}))| \end{aligned} \quad (5.6)$$

除了以上的LBP描述子，测量抽样点之间的差别分布情况也可认为是纹理特征的组成部分：

$$VAR_{P,R}(\mathbf{x}, I) = round\left(\frac{1}{h'P} \sum_{p=0}^{P-1} (I(\mathbf{x}_p) - \mu)^2\right) \quad (5.7)$$

其中 h' 是尺度因子且均值 $\mu = \frac{1}{P} \sum_{p=0}^{P-1} I(\mathbf{x}_p)$ 。关于LBP的更多理论知识和实验结果可参考文^[161]。

参数 P 和 R 控制描述子LBP 和VAR在多个尺度上计算纹理的特征。对于第 k 尺度上的参数 P_k 和 R_k ，正负样本的纹理特征描述子 $LBP_{P_k,R_k}^{riu2}(\mathbf{x}_i, I_0)$ 和 $VAR_{P_k,R_k}(\mathbf{x}_i, I_0)$ 用核密度计算帧 I_n 中像素 \mathbf{x} 在纹理特征上与目标或背景之间相似性度量：

$$\begin{aligned} p(texture_k | \mathbf{x}, L) &= \frac{1}{N} \sum_{i=1}^N \delta_0(|LBP_{P_k,R_k}^{riu2}(\mathbf{x}, I_n) - LBP_{P_k,R_k}^{riu2}(\mathbf{x}_i, I_0)| \\ &\quad + |VAR_{P_k,R_k}(\mathbf{x}, I_n) - VAR_{P_k,R_k}(\mathbf{x}_i, I_0)|) \end{aligned} \quad (5.8)$$

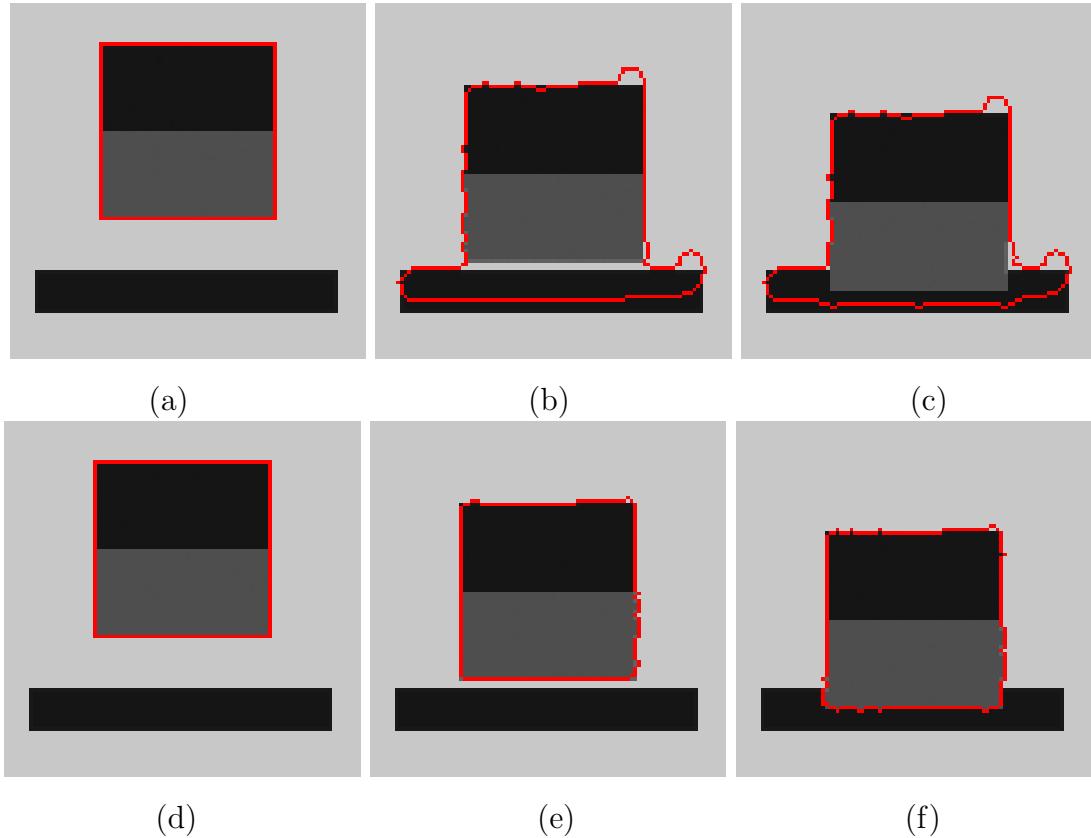
其中 δ_0 为delta函数。事实上纹理特征的概率模型和颜色特征的概率模型相似，唯一的区别在于密度估计中核函数的选择不同。由于纹理特征的分布种类较少（通常三个尺度的样本种类为9, 17和25）而颜色特征的种类较多（通常为灰度颜色的种类为255），因此纹理分析中的窗口核函数比颜色特征的核函数宽度要窄一些。

等式(5.3)组合颜色特征和纹理特征构成区域特征，而等式(5.1)和(5.2)计算像素 \mathbf{x} 的属于目标或背景的后验概率。大多数目标像素属于目标的后验概率 $p(obj|\mathbf{x}, RF)$ 大于属于背景的后验概率 $p(bck|\mathbf{x}, RF)$ ；反之，大多数背景像素属于背景的后验概率 $p(bck|\mathbf{x}, RF)$ 则大于属于目标的后验概率 $p(obj|\mathbf{x}, RF)$ 。从另一方面考虑这些像素的概率值可作为像素所具有的能量，定义目标像素和背景像素的能量分别为 $-\log(p(obj|\mathbf{x}, RF))$ 和 $-\log(p(bck|\mathbf{x}, RF))$ ，但这些目标像素和背景像素的定义依靠于图像区域中目标曲线的表达。这也即是目标轮廓 $C(s)$ 的泛函是区域能量的在图像区域中的总和，其定义为：

$$E_{region}(C(s)) = - \int_{inside(C)} \log(p(obj|\mathbf{x}, RF)) d\mathbf{x} - \int_{outside(C)} \log(p(bck|\mathbf{x}, RF)) d\mathbf{x} \quad (5.9)$$

区域特征被广泛用于在大多数跟踪方法中，但它们在处理与背景有相似颜色或纹理的低对比度目标时存在一些问题。例如图5.3 (a) 中的红色曲线围绕的正方形目标区域有灰色和黑色两个子区域构成，而背景中也含有同样的黑色区域。一旦目标靠近背景中的黑色区域（如图5.3 (b) 所示），传统算法使用的颜色特征将因为目标的黑色区域与背景的黑色区域具有同样的颜色造成其无法从背景中分辨出目标区域，这将产生图5.3 (c) 所示的错误跟踪结果。

不精确的目标轮廓是区域特征存在的另一个问题，即是目标与背景有很大差距时问题也依然不可避免。例如图5.4 (b) - (c) 中跟踪算法基于纹理特征无法确定精确地定位移动目标。虽然颜色和纹理的组合能在一些情况下提供跟踪性能，但这些问题在一些复杂背景下的灰度视频等极端环境下依然存在。为了克服这些区域特征存在的不足，本章所提的方法融合区域特征和边界特征以精确地定位目标轮廓，其结果如图5.3 (d) - (f) 和图5.4 (d) - (f) 所示。

图 5.3 对比区域方法^[150]与本文所提方法

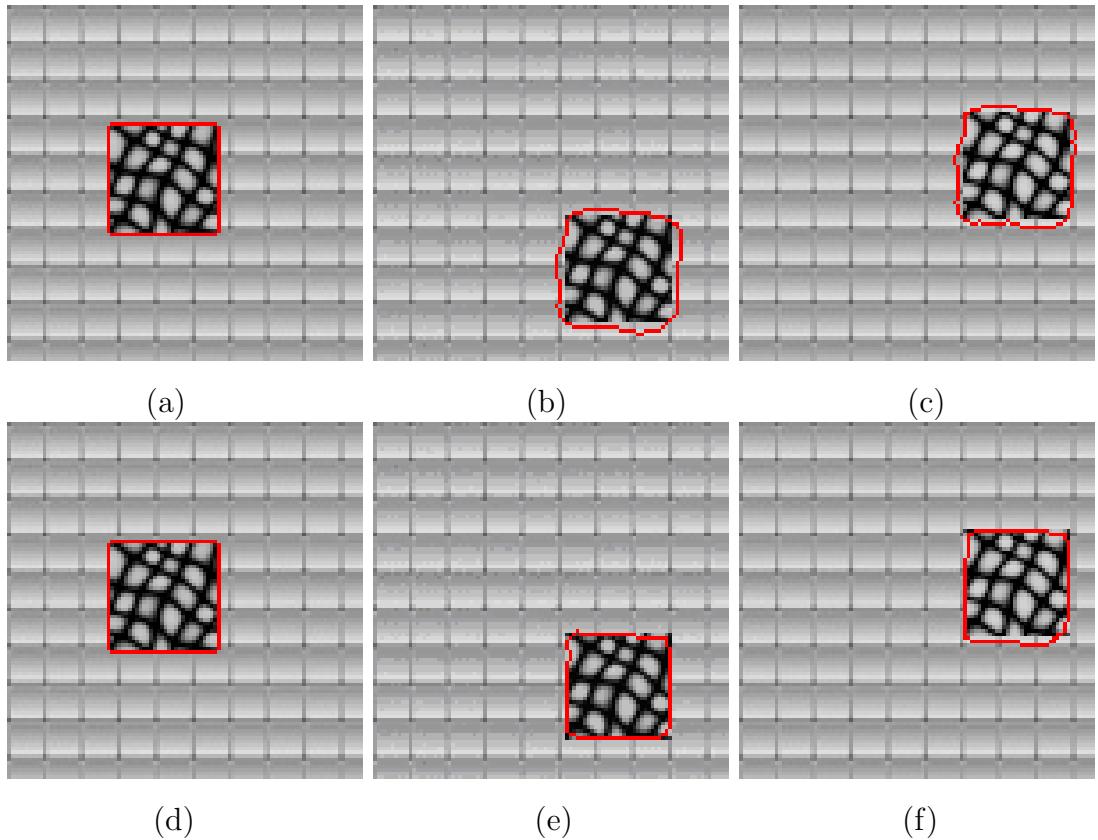
5.3 Comparison of the region-based tracking model in ^[150] and our proposed method

5.2 边界特征分析及边界能量泛函

两个连续帧之间的差值定义为帧差特征，由于其计算复杂度低因此也常被用于检测一些视频信号中的移动目标。帧差特征的数学定义表达为：

$$D(\mathbf{x}) = I_n(\mathbf{x}) - I_{n-1}(\mathbf{x}) \quad (5.10)$$

相邻帧的差别主要是由目标的运动和背景的变化两个部分组成，每个因素又能用高斯分布模型近似表达。整体上帧差特征可用一维高斯混合模型（Gaussian mixture models）拟合帧间差值的统计直方图。混合模型由两个独立的高斯分布分别代表目标运动和背景变化，每个部分有独立的权重值以区别其在模型中的重要程度：

图 5.4 对比区域方法^[150]与本文所提方法

5.4 Comparison of the region-based tracking model in ^[150] and our proposed method

$$Hist(\text{diff}) = w_{obj}p(\text{diff}|\mathbf{x}, obj) + w_{bck}p(\text{diff}|\mathbf{x}, bck) \quad (5.11)$$

$$= w_{obj}p(D(\mathbf{x}); \mu_{obj}, \Sigma_{obj}) + w_{bck}p(D(\mathbf{x}); \mu_{bck}, \Sigma_{bck}) \quad (5.12)$$

其中 $p(D(\mathbf{x}); \mu_{obj}, \Sigma_{obj})$ 和 $p(D(\mathbf{x}); \mu_{bck}, \Sigma_{bck})$ 分别是目标和背景的高斯分布模型。 μ_L 和 Σ_L 是高斯模型中的均值和方差系数，而 w_L 是混合模型的权重系数且满足 $w_{obj} + w_{bck} = 1$ 的条件约束。模型含有的这六个未知系数可用期望最大算法Expectation-Maximization 求解出最优系数解^[163, 164]。

混合模型分离出的目标高斯分布成份进一步给出像素 \mathbf{x} 在帧差特征上属于目标的后验概率为：

$$p(obj|\mathbf{x}, diff) = \frac{w_{obj}p(diff|\mathbf{x}, obj)}{w_{obj}p(diff|\mathbf{x}, obj) + w_{bck}p(diff|\mathbf{x}, bck)} \quad (5.13)$$

两个高斯成份中的均值参数 μ_{obj} 和 μ_{bck} 在文^[152]中被设置为0以简化整个未知参数的估计过程，然而这个简化处理在背景亮度发生剧烈变化时却会产生巨大的错误。为了解决该问题，等式（5.12）将高斯均值也作为自由参数加入到参数估计过程中。此外，背景的高斯分布模型对应于光照的变化或相机抖动等外部因素影响。例如，对比两个系统的背景高斯分布，其中一个摄像机在光照稳定的室内工作而另一个在户外工作中受到阳光和大风的影响。图5.5 和图5.6 分别比较室内高斯分布和户外高斯分布在连续60帧中均值和方差的变化情况，图中红色曲线显示的室内背景高斯十分稳定，而蓝色曲线表示的户外高斯模型却波动变化比较剧烈。

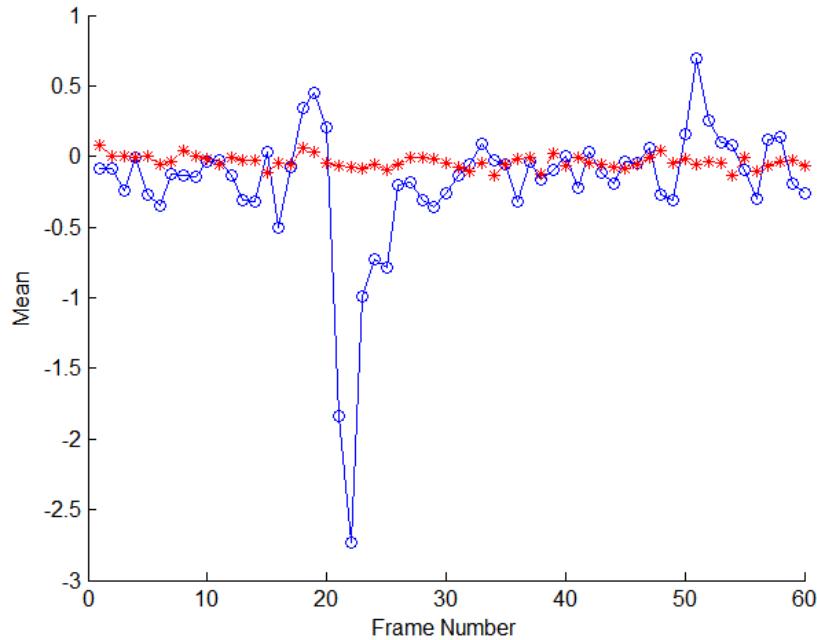


图 5.5 室内外相机的高斯均值变化

5.5 The mean of indoor and outdoor camera systems

剧烈变化的背景不符合等式（5.13）中不变背景的假设条件。在这种情况下，帧差不能稳定精确地估计目标运动。为了量化帧差特征在目标检测中的可

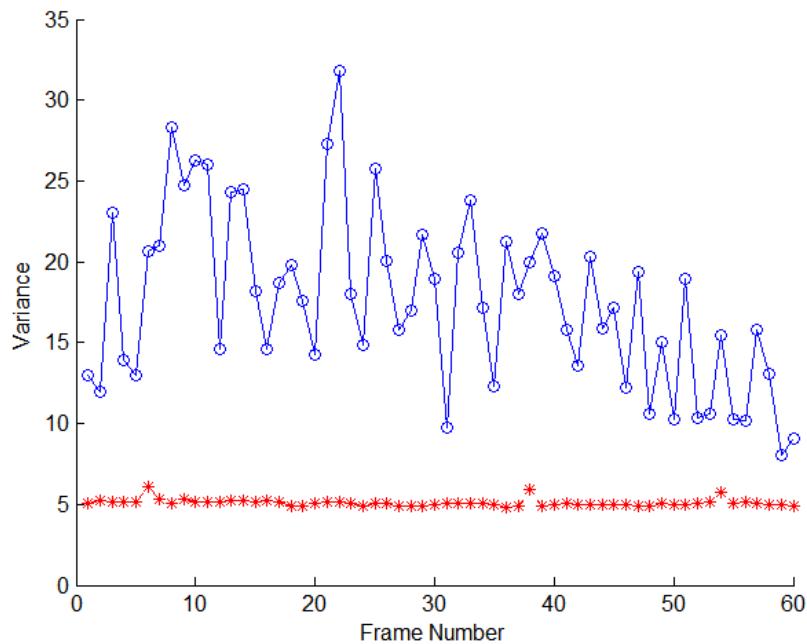


图 5.6 室内外相机的高斯方差变化

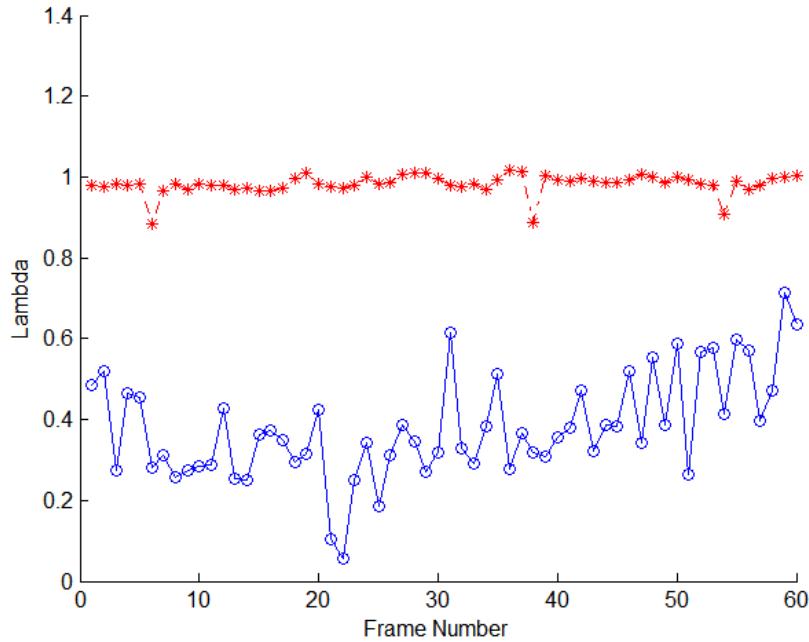
5.6 The variance of indoor and outdoor camera systems

靠性，定义权重函数为：

$$\lambda = 1.5 * 0.5^{\sum_{bck}/5} + 0.25 * 0.5^{|\mu_{bck}|} \quad (5.14)$$

背景中的剧烈变化将产生比较大的 μ_{obj} 和 μ_{bck} ，导致权重值 λ 比较小（参见图5.7中的蓝色曲线）；当背景变化较小时权值 λ 比较大且稳定（参见图5.7中的红色曲线），这也表明边界特征中的帧差可信度较高。

帧差特征提取的像素为移动目标在两帧中的区域并集，因此仍然无法确定目标在当前帧中的精确边界。边缘作为一种边界特征只与当前帧相关，并能与帧差联合确定目标的边界。另一方面，边缘对图像中的噪音比较敏感。为了只增强目标边缘而抑制背景边缘，图像的梯度与帧差后验概率融合后可以精确地确定目标的边缘位置。换言之，同时具有较大后验概率和较大梯度幅值的像素才能被认为是边界像素。以图像梯度和帧差的乘积为参数定义边界测度函数为：

图 5.7 室内外相机的权重值 λ 变化5.7 The λ value of indoor and outdoor camera systems

$$g(\mathbf{x}) = \frac{1}{1 + |\nabla I_n(\mathbf{x})| p(obj|\mathbf{x}, diff)} \quad (5.15)$$

当像素 \mathbf{x} 在移动目标的边缘时, 测度函数 $g(\mathbf{x})$ 趋近于0; 反之, 当它不位于移动目标边缘时则对应的函数值为1。同时在边界能量中引入轮廓长度项^[154]控制曲线的光滑程度, 整体的边界能量泛函以测地活动模型定义为:

$$E_{boundary}(C(s)) = \int g(C(s)) |\dot{C}(s)| ds \quad (5.16)$$

5.3 曲线演化及最小化能量泛函

轮廓的区域能量泛函和边界能量泛函在权重系数的组合下构成完成的能量泛函, 其完整的曲线能量泛函定义可表达为:

$$E(C(s)) = \lambda E_{boundary}(C(s)) + E_{region}(C(s)) \quad (5.17)$$

其中 λ 是由(5.14)定义的权重函数负责平衡两个能量子模块的作用。当光照或背景发生剧烈变化时，边界特征可信度降低对应的边界能量泛函也被削弱。

代入式(5.9)和(5.16)到式(5.17)中，完整的能量泛函可重写为：

$$\begin{aligned} E(C(s)) = \lambda \int g(C(s)) |\dot{C}(s)| ds - \int_{inside(C)} \log(p(obj|\mathbf{x}, RF)) d\mathbf{x} \\ - \int_{outside(C)} \log(p(bck|\mathbf{x}, RF)) d\mathbf{x} \quad (5.18) \end{aligned}$$

为了最小化该能量泛函，参数函数 $C(s)$ 需依照泛函的梯度下降流进行形变变化。通过计算能量泛函的变分，曲线 $C(s)$ 演化的等式可表达为：

$$\frac{\partial C}{\partial t} = (\lambda \kappa g(C) - \lambda \nabla g \bullet N - \log(p(obj|\mathbf{x}, RF)) + \log(p(bck|\mathbf{x}, RF))) N \quad (5.19)$$

值得注意的是，等式(5.19)中的曲线 $C(s)$ 的显式表达方式在处理形变过程中拓扑变化能力显得极其不足。为了克服拓扑变化时局限性，利用距离函数的零水平集能隐式表达活动曲线：

$$\Phi(\mathbf{x}) = \begin{cases} Dist(\mathbf{x}, C) & outside(C) \\ -Dist(\mathbf{x}, C) & inside(C) \\ 0 & \mathbf{x} \in C \end{cases} \quad (5.20)$$

其中 $Dist(\mathbf{x}, C)$ 是像素 \mathbf{x} 与曲线 $C(s)$ 最近点的欧式距离。微分几何理论已经证明了显式表达曲线演化流与隐式表达曲线演化流有如下关系：

$$\frac{\partial C}{\partial t} = FN \Rightarrow \frac{\partial \Phi}{\partial t} = F |\nabla \Phi| \quad (5.21)$$

将隐式曲线表达(5.20)代入到演化等式(5.19)中替代显示曲线的表达方式，以隐式曲线表达方式重写演化流为：

$$\begin{aligned} \frac{\partial \Phi}{\partial t} = & (\lambda \operatorname{div}(g(\mathbf{x}) |\nabla \Phi|) - \log(p(\operatorname{obj}|\mathbf{x}, RF))) \\ & + \log(p(\operatorname{bck}|\mathbf{x}, RF))) |\nabla \Phi| \quad (5.22) \end{aligned}$$

5.4 运动特征与曲线初始化

已有的跟踪算法通常假设目标的运动是光滑轨迹，在连续两帧之间不会发生大变化。这样的运动约束模型在跟踪目标突然改变其运动轨迹，监控系统常会丢失所跟踪的目标。为了更普适、更鲁棒的跟踪性能，目标的运动特征不应当作为能量泛函中的函数项控制曲线的演化过程、影响收敛结果。

上节介绍的方法在前帧中提取目标轮廓在等式 (5.22) 演化推进作用下收敛到目标在当前帧的位置。在开始演化曲线之前，水平集函数需要初始化过程设置曲线的起始位置，而初始值的设定与曲线的收敛结果和速度有密切关系。恰当的初始化设置能减少计算迭代次数，使曲线快速收敛到目标区域位置。传统的跟踪方法只是简单的用上一帧收敛结果初始化下一帧，而目标运动特征能帮助设置更好的初始化曲线以致于曲线可以快速地收敛到精确目标轮廓。令 $\Phi_{n-1,T}$ 为第 $n-1$ 帧中最后收敛的水平集函数， $\Phi_{n,0}$ 为第 n 帧中初始化的水平集函数。基于目标运动特征，当前帧的初始化设置为：

$$\Phi_{n,0}(\mathbf{x}) = \Phi_{n-1,T}(\mathbf{x} + \Delta\mathbf{x}) + \Delta B \quad (5.23)$$

其中 ΔB 是扩展曲线搜索空间的常量。 $\Delta\mathbf{x}$ 是由 $\Phi_{n-1,T}$ 和 $\Phi_{n-2,T}$ 估计得到的平移向量，其定义为：

$$\Delta\mathbf{x} = \frac{\sum_{\Phi_{n-1,T}(\mathbf{x}) < 0} \mathbf{x}}{\sum_{\Phi_{n-1,T}(\mathbf{x}) < 0} 1} - \frac{\sum_{\Phi_{n-2,T}(\mathbf{x}) < 0} \mathbf{x}}{\sum_{\Phi_{n-2,T}(\mathbf{x}) < 0} 1}$$

估计目标中心在前两帧估计的平移向量 $\Delta\mathbf{x}$ ，前帧的初始目标曲线可设置为前一帧的水平集函数 $\Phi_{n-1,T}$ 和 $\Delta\mathbf{x}$ 之和。该方法不是由迭代演化前一帧曲线，而是将前一帧的曲线整体平移到当前帧中最可能位置之上。通过这样的方式，迭代过程只需要几十次迭代即可到达目标区域从而节约大量计算开销。

整个目标轮廓跟踪算法可参考如下流程图：

5.5 单目标系统的实验及评价

本章介绍的跟踪算法主要解决单目标的精确跟踪问题，在所有测试算法性能的测试实验中选择纹理特征的尺度系数为(P, R)为(8, 2), (16, 4)和(24, 6)，分别在三个不同的尺度上提取LBP纹理描述子。运动特征中的曲线扩展系数 ΔB 为1.5，由文中^[165, 166]提出的窄带法求解等式（5.22），并在其邻域中选择正负样本提取颜色特征和纹理特征。所提算法与文^[150]所介绍的区域特征算法进行对比实验测试算法在一些复杂环境下的跟踪性能。为了测试的公平性，测试中在第一帧中给予两个算法同样的初始化轮廓。

图5.8比较所提算法与对比算法^[150]之间的跟踪结果。第一列结果图中方法^[150]只考虑区域特征得到的跟踪结果，它只能检测一些与背景有强烈对比的黑色区域同时将静止的背景区域错误地标示为目标区域，而一些低对比度的目标区域在跟踪过程中被丢失导致了破损的目标边界。测试视频来自于室内摄像机且背景相对静止，这样的外部条件能获得高可靠性的边界特征，即能量泛函中的边界特征权重系数 λ 较大从而加强边界特征在跟踪中的作用。所提算法的跟踪结果显示在图5.8的第二列中。利用上节所介绍的组合方法融合边界特征和区域特征，跟踪算法能在连续图像序列中提取完整、精确的目标边界。

“shop center corridor view”是视频测试集CAVIAR中常用的标准测试数据之一，在测试中所提方法时预先被转成灰度视频以增加跟踪的难度。在实验图5.9的第三帧中跟踪的女性目标被另一个男性目标遮挡后，图中第一列的两个目标具有相似的颜色和纹理特征以至于这些区域特征无法区分跟踪目标和遮挡物。但新模型在区域特征的帮助下除了遮挡帧外都可精确地跟踪目标。在实验图5.9的第二列中当两个行人互相重叠时，遮挡物的部分边界被错误地当成目标边界，但是它内部的边缘却阻止了轮廓将其纳入其中。一旦遮挡物离开跟踪目标之后，跟踪轮廓立刻恢复到正常状态仅包括目标轮廓。

粗糙的初始化轮廓也能获得令人满意的性能，这种灵活的初始化能力也是所提方法的一个优点。在实际应用中，初始化轮廓越不确定意味需要收集更多的正负样本。不同的初始化方式可能在开始的几帧中产生不同的结果，但是多帧过后不同的初始化也会产生同样的跟踪结果，即粗糙的初始化能获得与精确

初始化一样的最终结果。图5.10中显示的是矩形轮廓作为初始化的跟踪结果。因为粗糙的初始化轮廓在正样本中产生了大量噪音（背景像素），所以从该区域提取特征不能准确的描述目标的特点。图5.10中的第一列显示文^[150]的方法在粗糙初始化的情况下混淆了目标和背景，然后所提的方法利用边界能量泛函克服噪音的问题显示出另人满意的跟踪结果（参加结果图第二列）。

算法同样也测试对于刚体目标在室外光照波动和相机抖动情况下的跟踪性能。在这样的情况下，权重系数 λ 减少边界特征的作用同时颜色特征的分辨能力也因为建筑物的阴影而发生弱化。在图5.10的第一列中黑色汽车与灰色道路的差别很小，单一特征难以在这样困难条件下跟踪目标。通过组合区域特征和边界特征，所提方法在第二列测试结果排除了外部干扰准确的跟踪目标区域。

最后，算法在移动摄像机条件下跟踪目标轮廓，在图5.12中摄像机移动着记录穿过广场的行人目标。所有帧中的背景都不同以致于权重系数 λ 产生较少的边界能量，由区域特征主要负责跟踪任务。

所提算法除了具有精确和鲁棒的跟踪性能，还具有高效的处理过程、较少的计算开销。该算法不需要在整个图像域中提取特征，而是在目标邻域中提取目标特征。低开销的LBP纹理描述子在计算纹理特征时进一步减少计算开销。此外，运动特征被用于在当前帧中初始化新的水平集函数以快速收敛到目标区域。表5.2列出以上实验的处理时间，它证明所提算法能用于在线目标跟踪任务。

表 5.1 完整的轮廓跟踪算法

Algorithm	Tracking loop
Requires:	Initial object contour: ϕ_0
Initialization	
	$I_0 \leftarrow$ Read the first frame from input video
	$\mathbf{x}_i \leftarrow$ Draw samples in I_0 according to Φ_0 (see Fig. 5.2)
	$LBP_{P_k, R_k}^{riu2}(\mathbf{x}_i), VAR_{P_k, R_k}(\mathbf{x}_i) \leftarrow$ Generate the texture descriptor for samples (Eq. (5.5, 5.7))
Tracking	
	$t \leftarrow 1$
loop	
	$I_t \leftarrow$ Read a frame from input video
	$LBP_{P_k, R_k}^{riu2}(\mathbf{x}), VAR_{P_k, R_k}(\mathbf{x}) \leftarrow$ Generate the texture descriptor for pixels (Eq. (5.5, 5.7))
	$p(color \mathbf{x}, L) \leftarrow$ Calculate similarity probability of color feature(Eq. (5.4))
	$p(texture_k \mathbf{x}, L) \leftarrow$ Calculate similarity probability of texture feature(Eq. (5.8))
	$p(RF \mathbf{x}, L) \leftarrow$ Combine color feature and texture feature (Eq. (5.3))
	$p(L \mathbf{x}, RF) \leftarrow$ Calculate the posterior probability given region feature (Eq. (5.1))
	$p(obj \mathbf{x}, diff) \leftarrow$ Calculate the posterior probability given frame difference (Eq. (5.13))
	$g(\mathbf{x}) \leftarrow$ Calculate the metric function of boundary feature (Eq. (5.15))
	$\Phi_{t,0}(\mathbf{x}) \leftarrow$ Transform $\Phi_{t-1,T}$ to $\Phi_{t,0}$ according to the motion feature(Eq. (5.23))
	$\Phi_{t,T} \leftarrow$ Evolve $\Phi_{t,0}$ until convergence (Eq. (5.22))
	$t \leftarrow t + 1$
end loop	

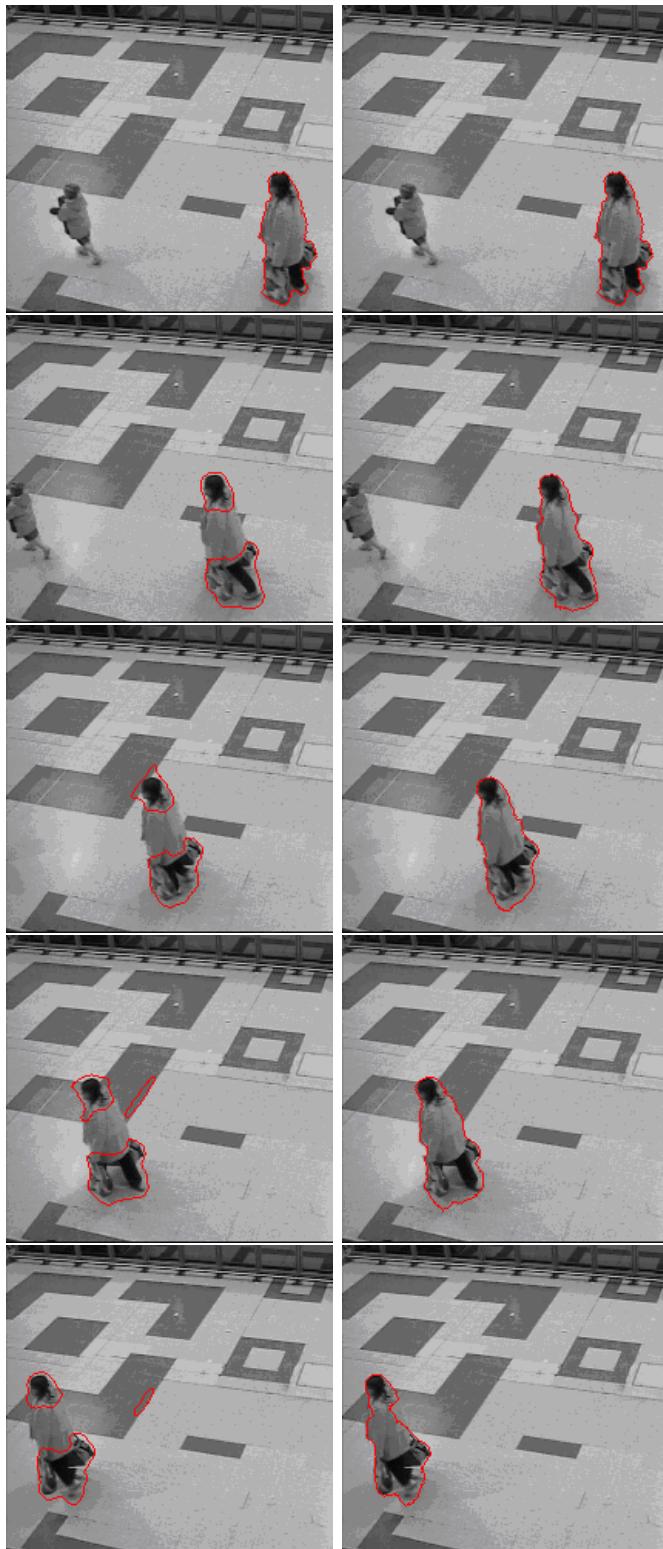


图 5.8 灰度视频中低对比度目标的跟踪结果

5.8 Tracking results for low-contrast object in monochrome video



图 5.9 灰度视频中遮挡目标的跟踪结果

5.9 Tracking results for occlusion in monochrome video

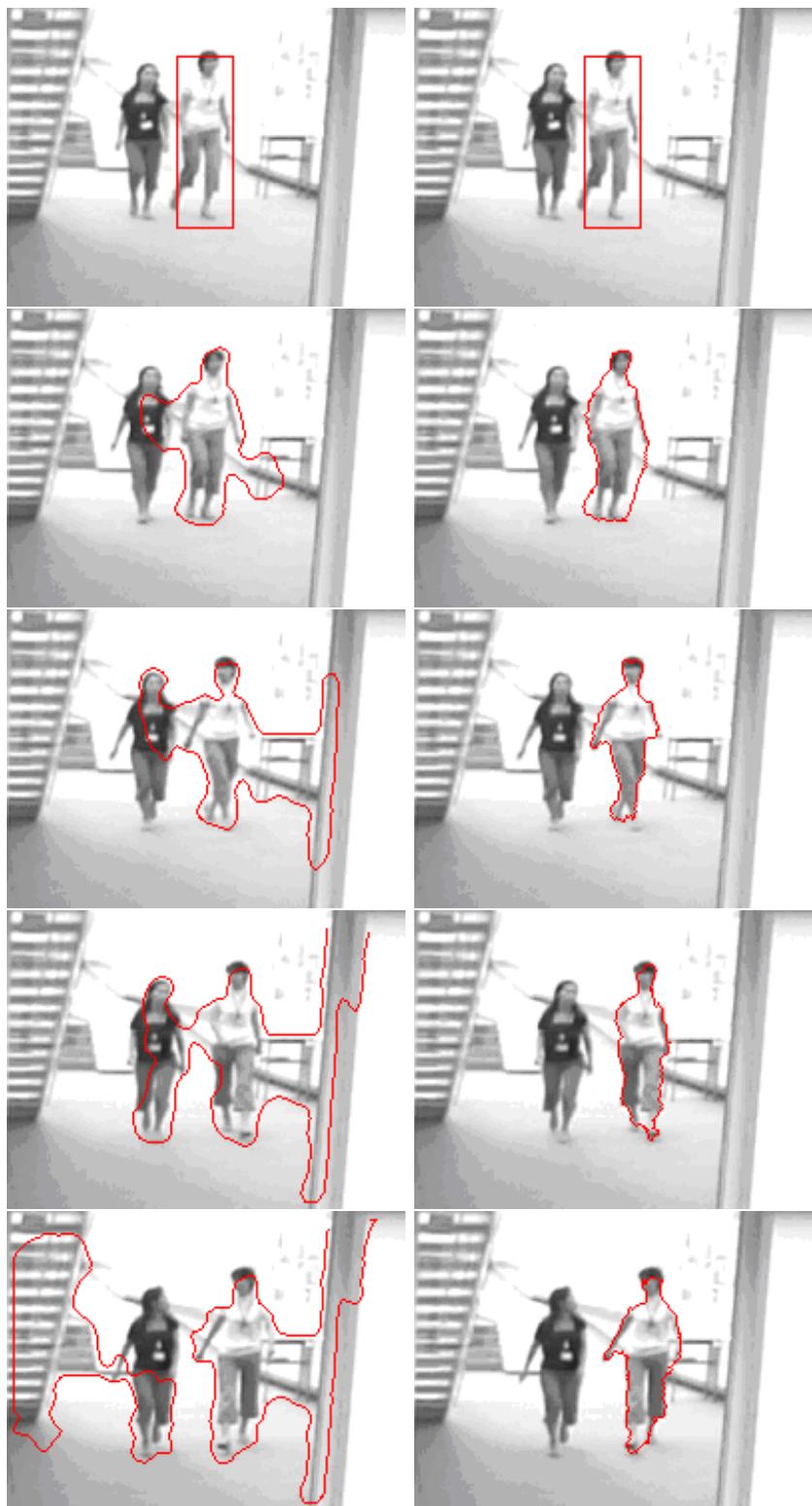


图 5.10 粗糙初始化下的跟踪结果

5.10 Tracking results for flexible initialization

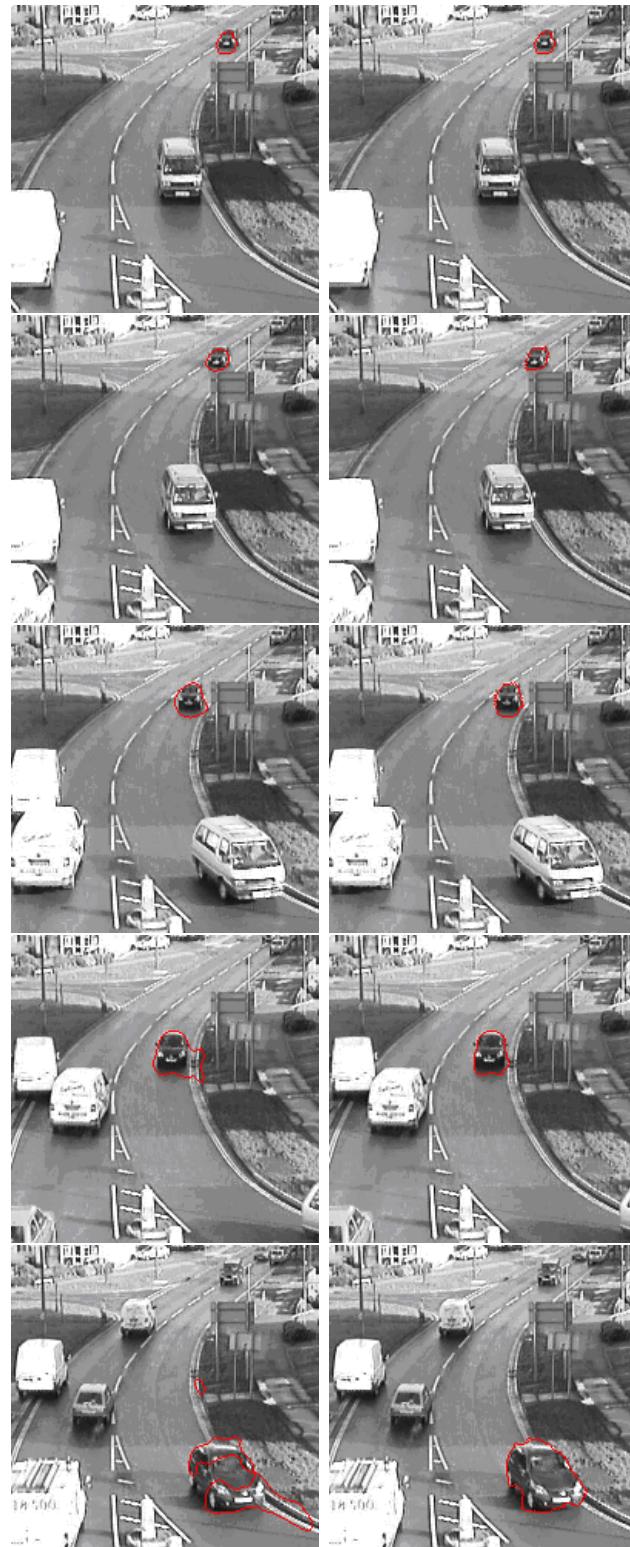


图 5.11 刚性目标的跟踪结果

5.11 Tracking results for rigid object

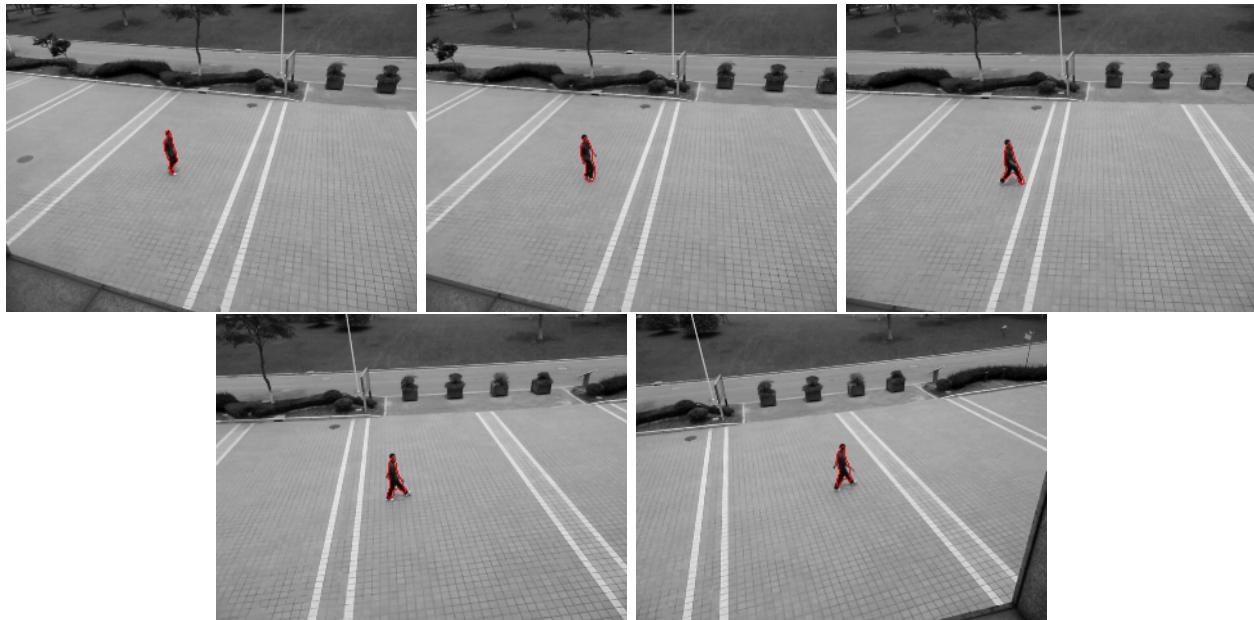


图 5.12 移动摄像机的跟踪结果

5.12 tracking results for mobile camera

表 5.2 实验的处理时间对比

5.2 The comparison of processing time cost in experiments

No	[150]的计算时间 (ms)	所提算法的计算时间 (ms)	帧数	平均的目标边界长度 (pixel)	视频尺寸 (pixel)
图5.8	556.51	106.34	60	218.08	212*200
图5.9	1188.39	226.42	50	227.93	384*288
图5.10	358.28	83.86	15	224.53	200*150
图5.11	1158.33	176.32	80	137.71	360*288

(以上为ACER的Aspire 2920Z (Pentium T3200 2.00G, 2G RAM) 笔记本上运行时间)

第六章 总结

本文以实现监控场景中多目标的在线检测跟踪为目的，全面研究了计算机视觉领域中所涉及的单相机的内外参数标定估计，双目摄像机的极线矫正与视察匹配，多目标的检测与跟踪等课题。论文的创新点体现在所提出的核函数的聚类算法，多特征的分层逐步组合模型以及区域测地能量泛函的偏微分优化。

利用搭建的双目立体视觉平台将监控场景中二维特征点还原出三维相机坐标，并投影到地面平面上得到特征点的世界坐标系下位置。新颖的核函数聚类算法依据投影点的高度和位置形成聚类集合，确定目标的数量、位置以及方向。该方法与传统基于颜色特征的方法相比在光照变化条件下更稳定，同时靠近地板平面的阴影的特征点由于高度较低而被过滤，而俯视角度下不存在目标之间的遮挡问题。

目标的轮廓跟踪模型则组合区域特征中的颜色和纹理计算图像像素属于目标和背景的后验概率，组合边界特征中的帧差和边缘构成测度函数确定移动目标的边缘轮廓。另外，运动特征给出每帧图像中的最优初始目标轮廓以减少迭代计算的次数，快速收敛到精确的目标区域。水平集框架被用于处理曲线拓扑变化问题，在区域测地框架的偏微分方程作用下不断演化最小能量曲线直至收敛。

尽管大量的实验证明了所系统在不同极端条件下的稳定性，但仍然有许多不足和有待提高之处。这也为今后的工作指明了研究方向，具体可开展的工作有：

1. 采用场景中目标的运动轨迹自动标定场景地板平面方程。监控场景中的目标运动都位于平面之上，并在此平面目标的投影后点集聚合度最高，以此两点作为约束条件构建目标函数计算空间中最优的投影平面为地板平面方程。
2. 目前系统采用同一方向两点逐步切分策略切分高密度目标群体下聚类集合，当多个目标聚合在一起时切分结果不够理想。而多方向多位置切分

策略以非极大值抑制删除过多的中间点集，以更小的核函数寻找最优的切分结果。

3. 轮廓跟踪算法接受目标轮廓或粗糙形状作为初始化轮廓，但立体视觉基于特征点检测空间目标，两者之间的匹配还存在一定差距。将空间中的特征点标记在图像中后直接作为初始化条件，由ballon模型从目标内部直接扩充至目标边界。

参考文献

- [1] Collins R. L.A.K.T.F.H.D.D.T.Y.T.D.E.N.H.O.B.P. and Wixson L., “A system for video surveillance and monitoring, VSAM final report”, Applied Physics Letters, 2000, CMU-RI-TR-00-12.
- [2] “<http://www.cs.cmu.edu/vsam/>”, .
- [3] Bogaert M. C.N.C.P.R.C.S.T.A. and Thonnat M., “The PASSWORDS project”, *IEEE International Conference on Image Processing*, volume 3, IEEE Computer Society, 1996, 675–678.
- [4] Wren C. R. A.A.D.T. and Pentland A.P., “Pfinder: Real-time tracking of the human body”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19(7)**, 780–785.
- [5] “<http://vismod.media.mit.edu/vismod/demos/pfinder/>”, .
- [6] Haritaoglu I Harwood D D.L., “W4: Real-time surveillance of people and their activities”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **22(8)**, 809–830.
- [7] Olson T. and Brill F., “Moving Object Detection and Event Recognition Algorithms for Smart Cameras”, Proceedings DARPA Image Understanding Workshop, 1997, 157–175.
- [8] Siebel N. and Maybank S., “Fusion of Multiple Tracking Algorithms for Robust People Tracking”, *The 7th European Conference on Computer Vision (ECCV)*, volume 4, 2002, 373–387.
- [9] “<http://www.cvg.rdg.ac.uk/projects/reason/index.html>”, .
- [10] “<http://tml.ece.ohio-state.edu/TML/>”, .

- [11] Coifman B. B.D.M.P. and Malik J., “A real-time computer vision system for vehicle tracking and traffic surveillance”, *Transportation Research Part C*, 1998, **6**(4), 271–288.
- [12] Zhu Z. X.G.Y.B.S.D. and Lin X., “VISATRAM: A real-time vision system for automatic traffic monitoring”, *Image and Vision Computing*, 2000, **18**(10)(10), 781–794.
- [13] “<http://www-cs.engr.ccny.cuny.edu/zhu/VISATRAM.html>”, .
- [14] Masoud O. and Papanikolopoulos N.P., “A novel method for tracking and counting pedestrians in real-time using a single camera”, *IEEE Transactions on Vehicular Technology*, 2001, **50**(5), 1267–1278.
- [15] “<http://www.cs.umn.edu/research/airvl/its/index.html>”, .
- [16] Tai J. T.S.L.C. and Song K., “Real-time image tracking for automatic traffic monitoring and enforcement application”, *Image and Vision Computing*, 2004, **22**(6), 485–501.
- [17] “<http://isci.cn.nctu.edu.tw/ResearchArea/Task4tm>”, .
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.

- [23] 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.
- [24] 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.
- [25] 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.
- [26] 李戈赵杰., “基于立体视觉平台的彩色图像视觉跟踪”, 人工智能及识别技术, 2004, **39**(6), 932–935.
- [27] 张汝波, 张亮, 张子迎, “基于立体视觉的机器人目标识别与跟踪”, 中南大学学报(自然科学版), 2007, **38**(1), 553–557.
- [28] 郑小东, 赵杰文, 刘木华, “基于双目立体视觉的番茄识别与定位技术”, 计算机工程, 2004, **30**(22), 155–157.
- [29] “<http://www.cvg.rdg.ac.uk/PETS2000/>”, .
- [30] “<http://www.cvg.cs.rdg.ac.uk/PETS2001/>”, .
- [31] “<http://www.cvg.cs.rdg.ac.uk/PETS2001/>”, .
- [32] “<http://www.elec.qmul.ac.uk/staffinfo/andrea/avss2007.html>”, .
- [33] “<http://www.avss09.org>”, .
- [34] Wang L. S.J.S.G. and Shen I., “Object Detection Combining Recognition and Segmentation”, *Asian Conference On Computer Vision*, 2007, 189–199.
- [35] Dar-Shyang L., “Effective Gaussian Mixture Learning for Video Background Subtraction”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, **27**(5), 827–832.

- [36] Yaser S. and S M., “Bayesian Modeling of Dynamic Scenes for Object Detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(11), 827–832.
- [37] Heikkila M. and Pietikainen M., “A texture-based method for modeling the background and detecting moving objects”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**, 657–662.
- [38] Papageorgiou C. O.M. and Poggio T., “A general framework for object detection”, *IEEE International Conference on Computer Vision*, 1998, 555–562.
- [39] Viola P. J.M. and Snow D., “Detecting pedestrians using patterns of motion and appearance”, *IEEE International Conference on Computer Vision*, 2003, 734–741.
- [40] Dalal N. and Triggs B., “Histograms of Oriented Gradients for Human Detection”, *IEEE International Conference on Computer Vision Pattern Recognition*, volume 2, 2005, 886–893.
- [41] Isard M. and Blake A., “Condensation - conditional density propagation for visual tracking”, *International Journal of Computer Vision*, 1998, **29**(1), 5–28.
- [42] Nikos P. and Rachid D., “Geodesic active contours and level sets for the detection and tracking of moving objects”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(3)(3), 266–280.
- [43] Crowley J. and Berard F., “Multi-modal tracking of faces for video communications”, *IEEE International Conference on Computer Vision Pattern Recognition*, 1997, 640–645.
- [44] Moreno-Noguer F. S.A. and Samaras D., “A Target Dependent Colorspace for Robust Tracking”, *IEEE International Conference on Pattern Recognition*, 2006, 43–46.

- [45] Collins R. Liu Y. and Leordeanu M., “On-Line Selection of Discriminative Tracking Features”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10), 1631–1643.
- [46] Horprasert T. Harwood D. and Davis L., “A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection”, *IEEE International Conference on Computer Vision*, 1999, 1–19.
- [47] Mikic I. C.P.C.K.G.T. and Trivedi M., “Moving Shadow and Object Detection in Traffic Scenes”, *IEEE International Conference on Pattern Recognition*, 2000, 321–324.
- [48] Stauder J. Mech R. and Ostermann J., “Detection of Moving Cast Shadows for Object Segmentation”, *IEEE Transactions on Multimedia*, 1999, **1**(1), 65–76.
- [49] Prati A. M.I.T.M.M. and Cucchiara R., “Detecting Moving Shadows: Algorithms and Evaluation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(7), 918–923.
- [50] Nadimi S. and Bhanu B., “Physical models for moving shadow and object detection in video”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8), 1079–1087.
- [51] Paragios N. and Deriche R., “Detecting Multiple Moving Targets using Deformable Contours”, *IEEE International Conference on Image Processing*, volume 2, 2000, 183.
- [52] Paragios N. and Deriche R., “A PDE-based Level Set Approach for Detection and Tracking of Moving Objects”, *IEEE International Conference on Computer Vision*, 1998, 1139.
- [53] Darrell T. G.G.H.M. and Woodfill l., “Integrated Person Tracking Using Stereo, Color, and Pattern Detection”, *International Journal of Computer Vision*, 2000, **37**(2), 175–185.

- [54] Darrell T. D.D.C.N. and Felzenszwalb P., “Plan-view trajectory estimation with dense stereo background models”, *IEEE International Conference on Computer Vision*, 2001, 628–635.
- [55] Mittal A. and Davis L., “M2Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo”, *European Conference on Computer Vision*, 2002, 189–203.
- [56] Huang X. L.L. and Sim T., “Stereo-Based Human Head Detection from Crowd Scenes”, *IEEE International Conference on Image Processing*, 2004, 1353–1356.
- [57] 陈棣湘罗飞路., “立体视觉测量中的图像匹配策略的研究”, 光学技术, 2002, **32**(5), 392–394.
- [58] 潘华., “立体视觉研究的进展”, 计算机测量与控制, 2002, **22**(12), 1121–1124.
- [59] 吴福朝李华., “基于主动视觉系统的摄像机自标定研究”, 自动化学报, 2001, **27**(6), 752–762.
- [60] Abdel-Aziz Y. and Karara H., “Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry”, *Proceedings of the Symposium on Close-Range Photogrammetry*, American Society of Photogrammetry, 1971, 1–18.
- [61] Brown D., “Close-Range Camera Calibration”, *Photogrammetric Engineering*, 1971, **37**, 855–866.
- [62] Faig W., “Calibration of Close-Range Photogrammetry Systems: Mathematical Formulation”, *Photogrammetric Engineering and Remote Sensing*, 1975, **41**, 1479–1486.
- [63] Faugeras O. L.T. and Maybank S., “Camera Self-Calibration: Theory and Experiments”, *European Conference on Computer Vision*, 1992, 321–334.

- [64] Faugeras O. and Toscani G., “The Calibration Problem for Stereo”, *IEEE Conference Computer Vision and Pattern Recognition*, 1986, 15–20.
- [65] Ganapathy S., “Decomposition of Transformation Matrices for Robot Vision”, *Pattern Recognition Letters*, 1984, **2**, 401–412.
- [66] Gennery D., “Stereo-Camera Calibration”, *10th Image Understanding Workshop*, 1979, 101–108.
- [67] Tsai R., “A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses”, *IEEE Journal of Robotics and Automation*, 1987, **3**(4), 323–344.
- [68] Wei G. and Ma S., “A Complete Two-Plane Camera Calibration Method and Experimental Comparisons”, *IEEE International Conference on Computer Vision*, 1993, 439–446.
- [69] Weng J. C.P. and Herniou M., “Camera Calibration with Distortion Models and Accuracy Evaluation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, **14**(10), 965–980.
- [70] Zheng-you Z., “A Flexible New Technique for Camera Calibration”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2000, **22**(11), 1330–1334.
- [71] Zheng-you Z., “Flexible Camera Calibration By Viewing a Plane From Unknown Orientations”, *International Conference on Computer Vision*, 1999, 666–673.
- [72] 李华吴福朝., “一种新的线性摄像机自标定方法”, *计算机学报*, 2000, **23**(11), 1121–1129.
- [73] 杨长江孙凤梅., “基于二次曲线的纯旋转摄像机自标定”, *自动化学报*, 2001, **27**(3), 310–317.

- [74] 吴福朝., “线性确定无穷远平面的单应矩阵和摄像机自标定”, 自动化学报, 2002, **28**(4), 488–495.
- [75] 黄凤荣孙凤梅., “基于条件数的摄像机自标定方法的鲁棒性分析”, 自动化学报, 2006, **32**(3), 337–344.
- [76] Meng X. and Hu Z., “A New Camera Calibration Technique based on Circular Points”, Pattern Recognition, 2003, **36**(5), 1155–1164.
- [77] 孟晓桥., “摄像机自标定方法的研究与进展”, 自动化学报, 2003, **29**(1), 1155–1164.
- [78] 张广军, 机器视觉, 北京: 科学出版社, 2005.
- [79] 袁亚湘, 孙文瑜, 最优化理论与方法, 北京: 科学出版社, 1999.
- [80] R. H. and A Z., *Multiple View Geometry in Computer Vision*, MIT Press, 2003.
- [81] Laurentini A., “How far 3D shapes can be understood from 2D silhouettes”, IEEE Transaction on Pattern Analysis and Machine Intelligence, 1995, **17**(2), 188–195.
- [82] TOMASI C. and KANADE T., *Shape and Motion from Image Streams: a Factorization Method*, Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.
- [83] Huang T. and Netravali A., “Motion and structure from feature correspondences: A review”, Proceedings of the IEEE, 1994, **82**(2), 252–268.
- [84] Torresani L. H.A. and Bregler C., “Learning Non-Rigid 3D Shape from 2D Motion”, *Neural Information Processing Systems(NIPS)*, 2003, 8–13.
- [85] Xiao J. and Kanade T., “Non-Rigid Shape and Motion Recovery: Degenerate Deformations”, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 2004, 668 – 675.

- [86] Ruo Zhang Ping-Sing Tsai J.C. and Shah M., “Shape from Shading: A Survey”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1999, **21**(8), 690–706.
- [87] Horn B.K.P. and Brooks M.J., *Shape From Shading*, MIT Press, 1989.
- [88] Prados E. and Faugeras O., “Shape From Shading”, Y.C. N. Paragios and O. Faugeras (eds.), *Handbook of Mathematical Models in Computer Vision*, chapter 23, Springer, 2006, 375–388.
- [89] Devries S. C. K.A.M.L. and J. K.J., “Shape from stereo : a systematic approach using quadratic surfaces”, *Perception psychophysics*, 1993, **53**(1), 71–80.
- [90] Devernay F. and Faugeras O., “ Computing differential properties of 3-D shapes from stereoscopic images without 3-D models”, *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, 1994, 208–213.
- [91] J. Banks M. Bennamoun K.K. and Corke P., “An accurate and reliable stereo matching algorithm incorporating the rank constraint”, *Symposium on Intelligent robotic systems*, 1999, 23–32.
- [92] Faugeras O., *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.
- [93] 许岚兵桂国富., “一种图像矫正算法”, *图象识别与自动化*, 2004, **16**(1), 86–90.
- [94] 许岚兵., “基于匹配变换对的图像矫正算法”, *现代制造工程*, 2005, **12**(5), 65–69.
- [95] Torr P. Z.A.a.y.S., “ Robust detection of degenerate configurations for the fundamental matrix”, *The 5th International Conference on Computer Vision*, volume 1, IEEE Computer Society Press, 1995, 1037–1042.

- [96] Zhang Z. D.R.L.Q.T. and Faugeras O., “A robust approach to image matching: Recovery of the epipolar geometry”, *International Symposium of Young Investigators on Information, Computer, Control*, 1994, 7–28.
- [97] Deriche R. Z.Z.L.Q.T. and Faugeras O., “Robust recovery of the epipolar geometry for an uncalibrated stereo rig”, *The 3rd European Conference on Computer Vision*, volume 1, 1994, 567–576.
- [98] Zhang Z. D.R.F.O. and Luong Q.T., “A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry”, *Artificial Intelligence Journal*, 1995, **78**, 87–119.
- [99] Dhond U. and Aggarwal J., “Structure from stereo - a review”, *IEEE Transaction on System Man and Cybernetics*, 1989, **19**(6), 1489–1510.
- [100] Ayache N. and Lustman F., “Trinocular stereo vision for robotics”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1991, **13**, 73–85.
- [101] Papadimitriou D.V. and Dennis T.J., “Epipolar line estimation and rectification for stereo images pairs”, *IEEE Transaction on Image Processing*, 1996, **3**(4), 672–676.
- [102] Hartley R. and Gupta R., “Computing matched-epipolar projections”, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 1993, 15–17.
- [103] Hartley R., “Computing matched-epipolar projections”, *International Journal of Computer Vision*, 1999, **35**(2), 1–16.
- [104] H. H.H.Y. and Wu P., “Projective rectification with reduced geometric distortion for stereo vision and stereoscopic video”, *Journal of Intelligent and Robotic Systems*, 2005, **42**(1), 71–94.
- [105] Fusiello A. T.E. and Verri A., “A compact algorithm for rectification of stereo pairs”, *Machine Vision and Applications*, 2000, **12**(1), 16–22.

- [106] Fusello A. and Irsara L., “Quasi-euclidean Uncalibrated Epipolar Rectification”, *International Conference on Pattern Recognition*, volume 12, 2000, 16–22.
- [107] Cox I. H.S. and Rao S., “A maximum likelihood stereo algorithm”, *Computer Vision and Image Understanding*, 1996, **63**(3), 542–567.
- [108] Szeliski R. and Zabih R., “An Experimental Comparison of Stereo Algorithms”, *Vision Algorithms: Theory and Practice*, volume 1, 1997, 1–19.
- [109] Belhumeur P.N. and Mumford D., “A Bayesian treatment of the stereo correspondence problem using half-occluded regions”, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 1992, 506–512.
- [110] Barnard S., “Stochastic stereo matching over scale”, *International Journal of Computer Vision*, 1989, **3**(1), 17–32.
- [111] Scharstein D. and Szeliski R., “Stereo matching with nonlinear diffusion”, *International Journal of Computer Vision*, 1998, **28**(2), 155–174.
- [112] Kim J. K.V. and Zabih R., “Visual Correspondence using Energy Minimization and Mutual Information”, *The 9th International Conference on Computer Vision*, volume 1, IEEE Computer Society Press, 2003, 1–8.
- [113] Wei Xiong H.S.C. and Jia J., “Fractional Stereo Matching Using Expectation-Maximization”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2009, **31**(3), 428–443.
- [114] Andersen R., *Modern Methods for Robust Regression*, Sage University Press, 2008.
- [115] Wilcox R.R., *Applying contemporary statistical techniques*, Academic Press, 2003.
- [116] Lowe D.G., “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, 2004, **60**(2), 91–110.

- [117] Ke Y. and Sukthankar R., “PCA-SIFT: A More Distinctive Representation for Local Image Descriptors”, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, IEEE Computer Society Press, 2004, 506–513.
- [118] Herbert Bay T.T. and Gool L.V., “Surf: Speeded up robust features”, *The Ninth European Conference on Computer Vision*, 2006, 404–417.
- [119] Herbert Bay Andreas Ess T.T.L.V.G., “SURF: Speeded Up Robust Features”, *Computer Vision and Image Understanding (CVIU)*, 2008, **110**(3), 346–359.
- [120] Mikolajczyk K. and Schmid C., “A performance evaluation of local descriptors”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10), 1615–1630.
- [121] Birchfield S. and Tomasi C., “Depth Discontinuities by Pixel-to-Pixel Stereo”, *International Journal of Computer Vision*, 1999, **35**(3), 269–293.
- [122] Birchfield S. and Tomasi C., “ Depth Discontinuities by Pixel-to-Pixel Stereo”, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 1998, 1073–1080.
- [123] Birchfield S. and Tomasi C., “A Pixel Dissimilarity Measure that is Insensitive To Image Sampling”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, **20**(4), 401–406.
- [124] Dockstader S. and Murat T.A., “Multiple camera tracking of interacting and occluded human motion”, *Proceedings of IEEE*, 2001, 1441–1455.
- [125] Mittal A. and Davis L., “M2Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo”, *European Conference on Computer Vision*, 2002, 18–36.

- [126] Huang X. L.L. and Sim T., “Stereo-Based Human Head Detection from Crowd Scenes”, *the International Conference on Image Processing*, 2004, 1353–1356.
- [127] Darrell T. D.D.C.N. and Felzenszwalb P., “Plan-view trajectory estimation with dense stereo background models”, *International Conference on Computer Vision*, 2001, 628–635.
- [128] Harris C. and Stephens M., “A combined corner and edge detector”, *Alvey Vision Conference*, 1998, 147–152.
- [129] Smith S. and Brady J., “SUSAN - a new approach to low level image processing”, *International Journal of Computer Vision*, 1997, **23**(1), 45–78.
- [130] Smith S., “Flexible filter neighbourhood designation”, *The 13th International Conference on Pattern Recognition*, 1996, 206–212.
- [131] Lowe D.G., “Distinctive Image Features from Scale-Invariant Keypoints”, *International Journal of Computer Vision*, 2004, **60**(2), 91–110.
- [132] Ke Y. and Sukthankar R., “PCA-SIFT: A More Distinctive Representation for Local Image Descriptors”, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, IEEE Computer Society Press, 2004, 506–513.
- [133] Herbert Bay T.T. and Gool L.V., “Surf: Speeded up robust features”, *The Ninth European Conference on Computer Vision*, 2006, 404–417.
- [134] Herbert Bay Andreas Ess T.T.L.V.G., “SURF: Speeded Up Robust Features”, *Computer Vision and Image Understanding (CVIU)*, 2008, **110**(3), 346–359.
- [135] Fukunaga K. and Hostetler L., “The estimation of the gradient of a density function, with applications in pattern recognition”, *IEEE Transaction on Information Theory*, 1975, **22**(1), 32–40.

- [136] Cheng Y., “Mean Shift, Mode Seeking, and Clustering”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, **17**(8), 790–799.
- [137] Comaniciu D. and Meer P., “Mean shift: A robust approach toward feature space analysis”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5), 603–619.
- [138] Comaniciu D. Ramesh V. and Meer P., “Kernel-based object tracking”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5), 564–575.
- [139] Collins R., “Mean-shift blob tracking through scale space”, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 2003, 234–240.
- [140] Veenman C. R.M. and Backer E., “Resolving motion correspondence for densely moving points”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(1), 54–72.
- [141] Vaswani N. R.A. and Chellappa R., “Activity recognition using the dynamics of the configuration of interacting objects”, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 2003, 633–640.
- [142] Bradski G., “Real Time Face and Object Tracking as a Component of a Perceptual User Interface”, *IEEE Workshop on Application of Computer Vision*, IEEE Computer Society Press, 1998, 214–219.
- [143] Jepson A. F.D. and Elmaraghi T., “Robust online appearance models for visual tracking”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(10), 1296–1311.
- [144] Mansouri A.R., “Region Tracking via Level Set PDEs without Motion Computation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(7), 947–961.

- [145] Heikkila M. and Pietikainen M., “A texture-based method for modeling the background and detecting moving objects”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(4), 657–662.
- [146] Shahrokni A. D.T. and Fua P., “Texture Boundary Detection for Real-Time Tracking”, *The 8th European Conference on Computer Vision*, 2004, 566–577.
- [147] Shahrokni A. D.T. and Fua P., “Fast Texture-Based Tracking and Delinement Using Texture Entropy”, *IEEE International Conference on Computer Vision*, volume 2, 2005, 1154–1160.
- [148] Collins R. L.Y. and Leordeanu M., “On-Line Selection of Discriminative Tracking Features”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10), 1631–1643.
- [149] Allili M.S. and Ziou D., “Object of Interest segmentation and Tracking by Using Feature Selection and Active Contours”, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 2007, 1–8.
- [150] Yilmaz A. Li X. and Shah M., “Contour based object tracking with occlusion handling in video acquired using mobile cameras”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(11), 1531–1536.
- [151] V. T. and Pietikinen M., “Multi-object tracking using color, texture and motion”, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 2007, 1–7.
- [152] Paragios N. and Deriche R., “Geodesic active contours and level sets for the detection and tracking of moving objects”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(3), 266–280.

- [153] Kichenassamy S. K.A.O.P.T.A. and Yezzi A., “Gradient flows and geometric active contour models”, *IEEE International Conference on Computer Vision*, volume 1, 1995, 810–815.
- [154] Caselles V. K.R. and Sapiro R., “Geodesic active contours”, *International Journal of Computer Vision*, 1997, **22**(1), 61–79.
- [155] Paragios N. and Deriche R., “Geodesic active regions for motion estimation and tracking”, *IEEE International Conference on Computer Vision*, volume 1, 1999, 688–694.
- [156] Birchfield S., “Elliptical Head Tracking Using Intensity Gradients and Color Histograms”, *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 1998, 232–237.
- [157] Kruger V. H.A. and Sommer G., “Affine Real-time tracking using Gabor wavelet networks”, *International Conference on Computer Vision Workshop Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, IEEE Computer Society Press, 1999, 26–27.
- [158] He C. Z.Y.F. and Ahalt S.C., “Object Tracking Using the Gabor Wavelet Transform and the Golden Section Algorithm”, *IEEE Transactions on Multimedia*, 2002, **4**(4), 528–538.
- [159] Manjunath B. and Ma W., “Texture features for browsing and retrieval of image data”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, **18**(8), 837–842.
- [160] Grigorescu S. P.N. and Kruizinga P., “Comparison of texture features based on gabor filters”, *IEEE Transactions on Image Processing*, 2002, **11**(10), 1160–1167.
- [161] Ojala T. Pietikainen M. and Maenpaa T., “Multiresolution gray scale and rotation invariant texture analysis with local binary patterns”, *IEEE Trans-*

- actions on Pattern Analysis and Machine Intelligence, 2002, **24**(7), 971–987.
- [162] Heikkil M. and Pietikinen M., “A texture-based method for modeling the background and detecting moving objects”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, **28**(4), 657–662.
- [163] Dempster A. P. L.N.M. and Rubin D.B., “Maximum-likelihood from incomplete data via the EM algorithm”, Journal of Royal Statistical Society Series B, 1977, **39**(1), 1–38.
- [164] Wu C.F., “On the convergence properties of the EM algorithm”, The Annals of Statistics, 1983, **11**(1), 95–103.
- [165] Adalsteinsson D. and Sethian J.A., “A fast level set method for propagating interfaces”, Journal of Computational Physics, 1995, **118**(1), 269–277.
- [166] Malladi R. S.J.A. and Vemuri B.C., “Shape modeling with front propagation: a level set approach”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, **17**(1), 158–175.

致 谢

至此，基本介绍完博士期间所作大部分研究工作，同时也意味着学生生涯临近结束。回想十年的求学生涯，从农村到城市、从内陆到沿海、从幼稚到成熟、从求学到自学。其中虽经历许多艰难困苦，但同时也促进了自己一步一步地蜕变、成长。所取得的成绩，除了艰苦的努力和一部分幸运外，与老师以及同学、朋友的帮助是密不可分的。

四年前，杨新老师在众多面试学生中选择了名次居中的自己，才使得我有机会进入交通大学，并最终完成博士学业。同时，也是他从始至终的教育、督促与点拨才使得自己在学习时有所感悟、在倦怠时及时振作、在迷茫时逐渐明朗。不仅如此，他在讨论时的宽容大度，讲课时的激情四溢，生活中的处处关怀，让包括自己在内的所有师兄弟都终生难忘。当然也要感谢赵宇明老师在论文选题，研究方法以及论文撰写过程中给予的悉心指导和帮助，在平常生活中给予的关心、照顾。尤其是当遇到许多挫折和打击时，总是她以真心的谈话、爽朗的笑声、豁达的心态，把我心中的重重阴霾一扫而空，赋予我重新向前的希望和勇气。非常感谢胡福乔老师在哲学高度上帮助我领悟人生的真谛，启迪我寻求理想的勇气以及克服重重困难的决心与意志。非常感谢美国NIH的何磊老师多次不辞辛苦地为我纠正英语论文中的语法错误，毫不保留地传授我写作的技巧与重点。当然，OMRON公司吴越博士在摄像机硬件上给予的大力支持也本文的重要基础，在此也他的热心帮助、幽默睿智表示真挚的感谢。

非常感谢多年来一起在实验室里朝夕相处的师弟徐轶人、葛诚、付文林、邓秋平、宋金龙等同学在学习、生活和娱乐中给我带来的精神支持和物质支持。衷心感谢所有帮助、关心过我的老师、同学和朋友。在此，为他们送上我最真诚的祝福。

最后，感谢我的亲人在我攻读博士期间默默的支持和鼓励。是父母含辛茹苦地将我一步一步培养长大，希望自己以后能帮助他们改善生活。感谢妻子多年对自己学业默默的支持、无私的奉献，承担了家庭大部分的责任让我能专注于学业与研究。在此，对他们表达深深的致意和诚挚的感谢！

谨以此文献给我贤惠的妻子和我敬爱的父母亲。

攻读学位论文期间发表的学术论文目录

- [1] **Ling Cai**, Lei He, Yamasita Takayoshi, Yiren Xu, Yuming Zhao and Xin Yang, “Region and Boundary-based Object Tracking”, IEEE Transactions on Circuits and Systems for Video Technology. (已接受)
- [2] **Ling Cai**, Lei He, Yiren Xu, Yuming Zhao and Xin Yang, “Multi-object detection and tracking by stereo vision”, Pattern Recognition, 2010, 43(12), 4028-4041. (SCI收录,影响因子: 3.279)
- [3] **Ling Cai**, Cheng Ge, Yiren Xu, Yuming Zhao and Xin Yang, “Fast tracking of object contour based on Color and Texture”, International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23(07), 1421 - 1438. (SCI收录,影响因子: 0.660)
- [4] **Ling Cai**, Yiren Xu, Yuming Zhao and Xin Yang, “Gamma Mixture Model Based Gradient Function for Noisy Image Segmentation”, Signal Processing. (2审中)
- [5] **Ling Cai**, Lei He, Yiren Xu, Yuming Zhao and Xin Yang, “An effective Segmentation for Noise based Image Verification using Gamma Mixture Models”, The 9th Asian Conference on Computer Vision (ACCV) 2009, LNCS 5996(3), 21-32. (EI收录)
- [6] **Ling Cai**, Lei He, Yiren Xu, Yuming Zhao and Xin Yang, “Texture Image Segmentation by Active Bayesian Contour”, International Conference on System Design and Data Processing (ICSDDP) 2011. (已接受, EI收录)