

跨摄像机多人目标的跟踪研究



论文作者签名: 王逸贺

指导教师签名: 刘济林

论文评阅人 1: 杨靖宇 教授 南京理工大学

评阅人 2: 陈贺新 教授 吉林大学

评阅人 3: 顾伟康 教授 浙江大学

评阅人 4: 盲审 1

评阅人 5: 盲审 2

答辩委员会主席: 朱淼良 教授 浙江大学计算机学院

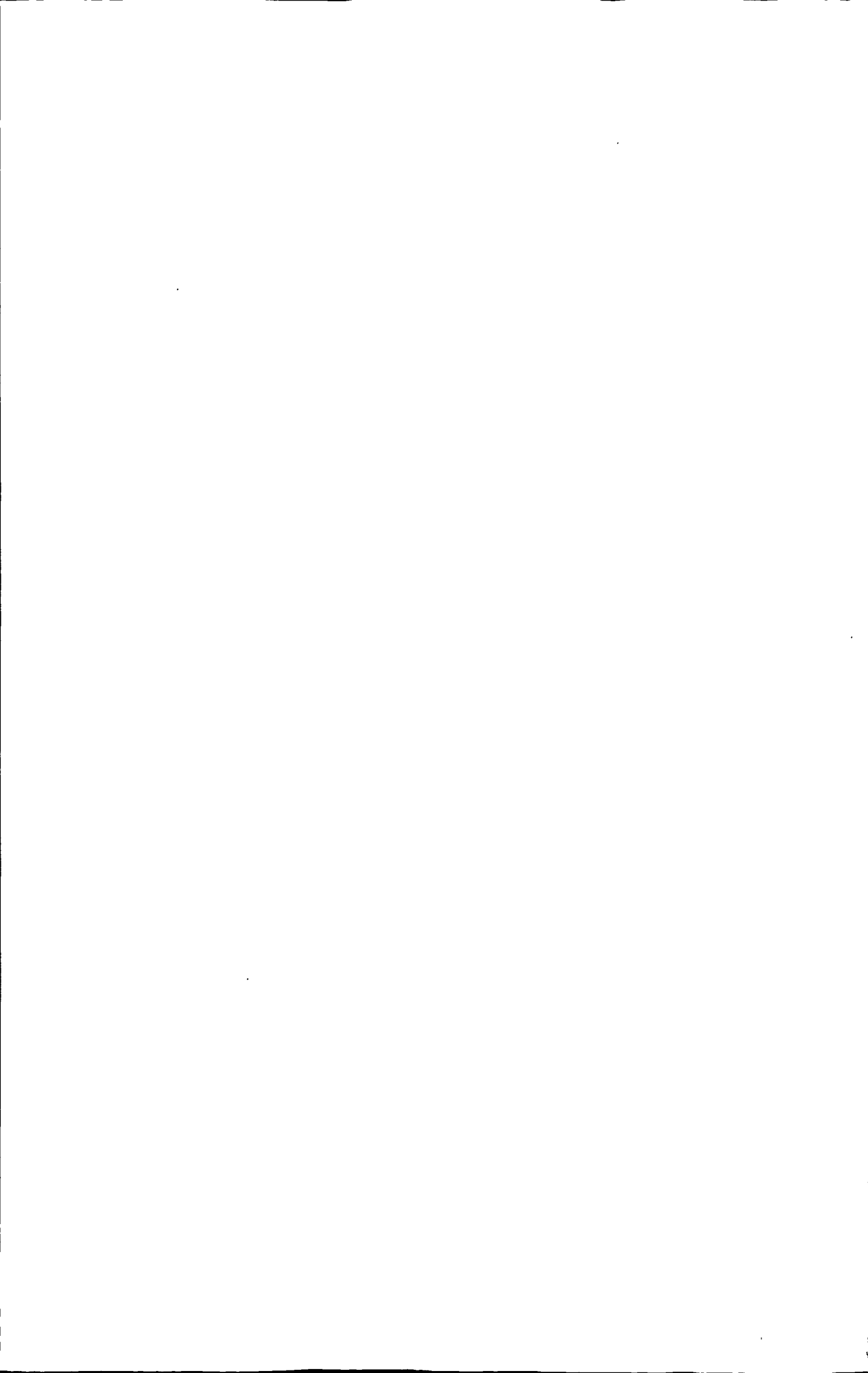
委员 1: 顾伟康 教授 浙江大学信电系

委员 2: 姚庆栋 教授 浙江大学信电系

委员 3: 周兆经 教授 中国计量学院

委员 4: 刘济林 教授 浙江大学信电系

答辩日期: 2011 年 3 月 7 日



Y1863502

Multi-human Tracking across Multi-cameras



Author's signature: Wang Xuanke

Supervisor's signature: [Signature]

Thesis reviewer 1: Prof. Yang Jingyu Nanjing University of Science and Technology

Thesis reviewer 2: Prof. Chen Hexin Jilin University

Thesis reviewer 3: Prof. Gu Weikang Zhejiang University

Thesis reviewer 4: _____

Thesis reviewer 5: _____

Chair: Prof. Zhu Miaoliang Zhejiang University
(Committee of oral defence)

Committeeman 1: Prof. Gu Weikang Zhejiang University

Committeeman 2: Prof. Yao Qingdong Zhejiang University

Committeeman 3: Prof. Zhou Zhaojing China Jiliang University

Committeeman 4: Prof. Liu Jilin Zhejiang University

Date of oral defence: March 7, 2011



独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 浙江大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：王逸贺 签字日期：2011年3月6日

学位论文版权使用授权书

本学位论文作者完全了解 浙江大学 有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权 浙江大学 可以将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、汇编学位论文。

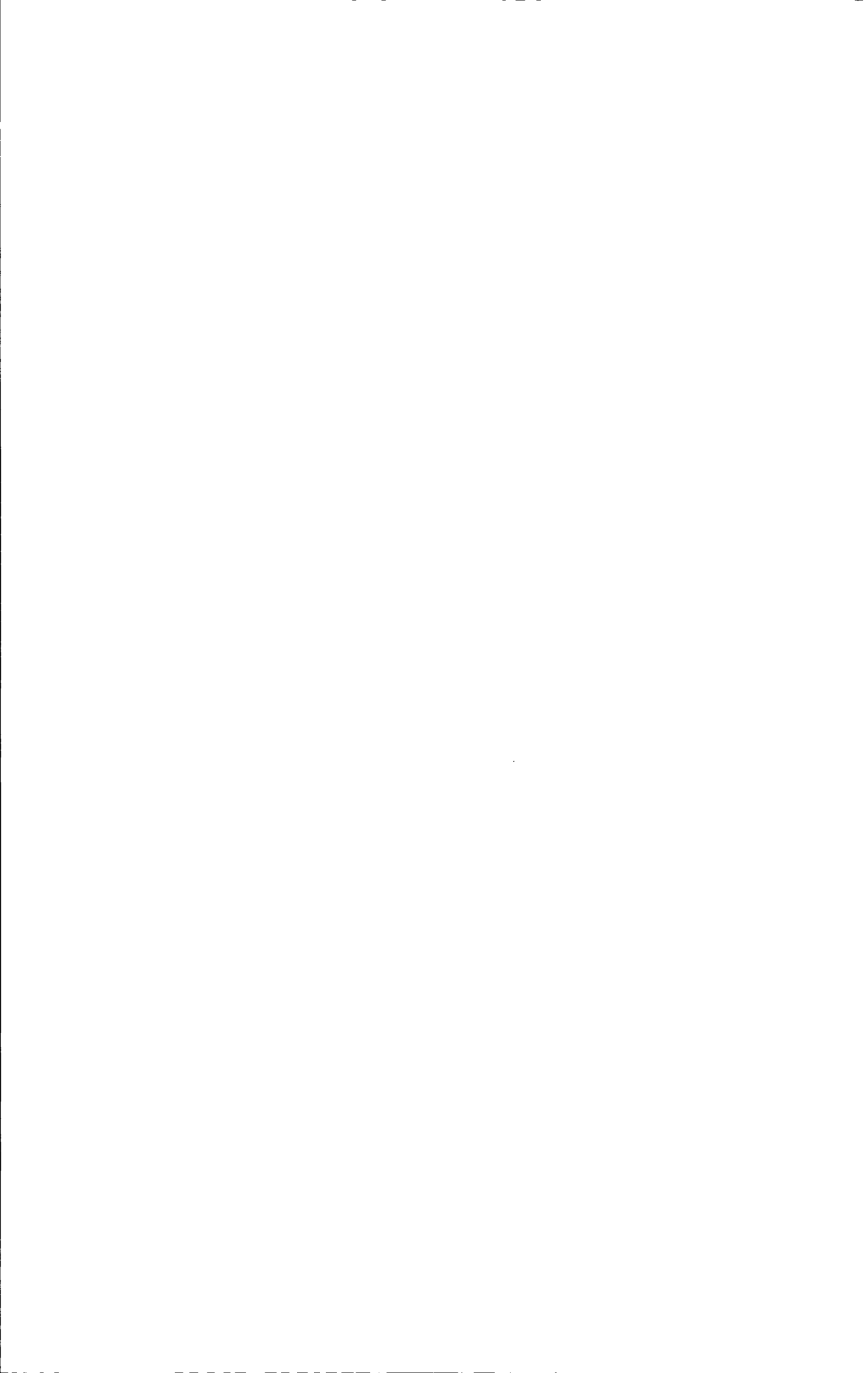
(保密的学位论文在解密后适用本授权书)

学位论文作者签名：王逸贺

导师签名：刘云

签字日期：2011年3月6日

签字日期：2011年3月6日



致谢

时光荏苒岁月如梭，三年半的博士生涯即将结束。在本论文即将完成之际，向所有支持、帮助和关心过我的老师、同学及亲人表示衷心的感谢和崇高的敬意。

首先，对我敬爱的导师刘济林教授致以最诚挚的谢意。感谢您对我孜孜不倦的教诲，感谢您对我成长的关心和爱护。您用渊博的知识将我领进了科学之门，为我指明了奋斗的方向；您广博的理论学识和深厚的理论水平使我知识的田地更加富饶；您循循善诱的精心点拨，帮助我指点了迷津、开拓了思路；您热忱的鼓励，使我快速地走出了人生的低谷；您饱满的工作热情时时激励着我勇于挑战困难；您认真负责的工作态度、严谨的治学精神和人格魅力使我懂得了一个研究者应有的人生态度。恩师之情，终生难忘。

感谢于慧敏教授及实验室的龚小谨老师、项志宇老师、杜歆老师、谢立老师等，对我在科研和指导，帮助我了解了许多与科研相关的知识。

感谢曾经与我共同完成科研课题的科研组成员张莉、陈建娟、陈曦、吴青松、王宁波。他们给我提供了许多宝贵的意见和建议。感谢他们无论酷暑严寒始终都能与我一起到室外进行实验。我今天能够取得一定的科研成果与他们的辛勤付出是密不可分的。

感谢在科研工作中给予我大力支持的实验室的师兄姐妹的热心帮助，他们是：曹腾、陈乘、陈明芽、丁菁汀、丁鑫、杜小丽、韩从道、黄文琦、雷杰、李永佳、林颖、刘洋、卢维、路丹辉、马丽莎、欧阳柳、潘华东、沈敏一、沈焯湖、王其聪、王炜强、王心焕、王延长、武二永、谢斌、徐进、许士芳、徐彦宏、杨镛、杨飞、杨力、杨鹏、杨象军、姚拓中、张雷、郑驰、朱株、左文辉等。我们在一起共同学习、生活中结下的深厚情谊将使我永远铭记在心。

感谢和我一起讨论问题的那些好友：王迅、张飞、陈国赟、叶刚、唐磊、王嘉峰、孙波、颜瑞、梁文锋、齐保振等，我会永远记住与你们在一起度过的快乐时光。

感谢那些即将永久载入我的实验视频当中的“模特演员”——我的师兄姐妹，你们走路的神态和灿烂的笑容将永远地留在了我的记忆深处。没有你们很多次的不辞辛劳地频频上镜，我就不可能取得如此丰硕的科研成果。

真诚的感谢我的父母，他们为了我的成长付出了无数的心血，他们无私的爱一直是我在人生道路上不断向前的力量源泉，父母之恩，天高地厚。感谢我的哥嫂及亲人的全力支持。

再次感谢所有关心、帮助和支持过我的老师、亲人和朋友!

最后,感谢校内外各位教授、专家、评委在百忙之中审阅我的论文和参加我的答辩会。

王选贺

2011年1月于求是园

摘要

长期以来,人体目标跟踪是计算机视觉研究的热点。它是使用计算机及相关设备对生物视觉的一种模拟,是一项非常具有挑战性的研究工作。人体目标跟踪技术目前仍然处于研究和探索阶段,在理论和实际应用中仍存在许多问题没有得到很好地解决。影响目标跟踪算法鲁棒性的原因很大程度上是由于目标运动的复杂性以及光照条件变化等因素造成的。本文主要研究了单摄像机和多摄像机下的人体目标跟踪,单摄像机人体目标跟踪是跨摄像机人体目标跟踪的基础,只有单摄像机人体目标跟踪正确的前提下,跨摄像机跟踪才会有正确的跟踪结果。在单摄像机人体目标跟踪过程中,主要解决遮挡人体目标跟踪。跨摄像机人体目标跟踪分为有公共视场人体目标跟踪和无公共视场人体目标跟踪。

本文的主要创新包括:

1. 针对多人目标跟踪而产生的遮挡问题,提出了一种基于非参数复合模型的粒子滤波方法来描述多个人体目标的情况。该复合模型通过复合预测和复合更新两个步骤交替进行以达到递归跟踪的目的。在复合粒子滤波过程中,通过每个复合粒子的权值大小来衡量粒子的贡献大小。权值的大小是基于 HSV 直方图的多颜色观测模型来实现的。观测模型是通过巴氏距离的核密度估计来建模。实验结果证明复合模型的粒子滤波方法可以很好地实现多人目标跟踪。

2. 充分利用单摄像机的人体目标的运动信息、颜色信息、空间信息,将背景建模、块建模、颜色建模、运动建模和人体的空间信息进行有效地融合,较好地解决了人体目标相互遮挡情况下的人体跟踪的问题。本文利用人体目标的运动信息,采用的混合高斯模型进行背景重建的方法,先提取出运动的目标;利用基于 Epanechnikov 核密度梯度估计算法对存储模型中的人体进行聚类,即对人体的外部颜色相近的像素进行块建模;利用非参数的 Gauss 核密度估计算法对已聚类的块模型和人体的空间信息建立颜色密度函数,同时利用人体运动信息建立运动密度函数,通过颜色密度函数与运动密度函数构成后验概率模型;然后对当前的检测出的人体目标的每个像素计算最大后验概率,得到最大后验概率的颜色图像,通过该图像进行分割,达到对遮挡情况下多人目标跟踪的目的。

- 3 针对有公共视场跨摄像机人体目标跟踪误匹配的问题,提出了一种基于空间映射变换与颜色特征信息相融合的匹配算法。本文先通过跨摄像机的公共视场分界线来初步

确定跨摄像机的人体目标，然后利用变换单应矩阵计算跨摄像机人体目标的对应关系，然而在实际应用中，视频是不完全同步且有交叉遮挡而产生误匹配，因此本文在投影变换单应矩阵的基础上采用基于颜色模型的最大后验概率进行匹配。

4. 本文针对无公共视场跨摄像机人体目标跟踪因颜色差异较大而引起误匹配问题，提出了一种基于颜色转变函数的方法对颜色进行校正，并通过颜色转变函数空间进行概率估计来提高跨摄像机人体目标的匹配的准确率。该算法先在低维度的子空间上对已知的跨摄像机人体目标进行训练，得到颜色转变函数。该方法不用依赖摄像机的内参来计算颜色转变函数的子空间。用基于概率的主要成分分析法对颜色转变函数的子空间进行建模得到概率密度函数，最后利用概率密度函数获取跨摄像机的人体目标的匹配概率。本文在室内外，全部室外环境下对该方法进行了实验，从实验结果上看，本方法能够很好地对室外及室内外环境下的跨摄像机人体目标进行跟踪。

关键词：人体目标跟踪；运动目标检测；粒子滤波；有重叠视场；无重叠视场；

Abstract

Human tracking has been a hot spot in the field of computer vision for a long time. It is to simulate the function of human vision using the computer and relative device. It is an extremely challenging research and comprehensive subject. The technology of human tracking is still in the course of research and exploration at present. There are many problems to be solved in the aspect of theory and practical application. The reasons such as the motion complexity of human, alteration of illumination intensity and others influences the robustness of tracking human to a great extent. The dissertation mainly research on human tracking in a single camera and across multiple cameras. The human tracking across multiple cameras is on the basis of human tracking in a single camera. Only on the accurate premise of tracking human in a single camera, the tracking human across multiple cameras may be correct result. The dissertation mainly solves the tracking human under occlusion in a single camera. There are two cases of tracking human across multi-cameras. One is human tracking across multiple cameras with overlapping field of view. The other is human tracking across multiple cameras with non-overlapping field of view. The two cases will be talked about in the dissertation.

1. The dissertation aims at solving the problems of tracking occluded human object. An algorithm is proposed based on particle filter of non-parameter mixture model which is described the multiple human objects in a single camera. The mixture model carries on two steps of mixture predict and mixture update by turn in order to accomplish recursive human tracking. During the course of mixture particle filter, the contribution of each particle to final object is weighed using weight of each particle. Each particle weight gets attained utilizing multiple observation model of HSV histogram. Observation model gets acquired by kernel estimation of Bhattacharyya distance. The experiment shows that particle filter of mixture model can better track multiple human object under occlusion.

2. The dissertation makes full use of motion information, color information and space information of human object in a single camera. The dissertation effectively solves the problems of human tracking under occlusion each other using background modeling, blob modeling, color modeling, motion modeling and space information of human body. The

dissertation firstly detect the kinetic persons using motion information to reconstruct background of video based on algorithm of mixture Gauss modeling. We will cluster to human body in storage based on Epanechnikov kernel density gradient estimation. Namely, we will obtain blob models, which put the color-similar pixels of human body together. We will establish the color density function based on non-parameter Gauss kernel density estimation and space information of human body, build motion density function based on motion information of human and obtain posterior probability utilizing color density function and motion density function. We apply maximum posterior probability to each pixel of detected human in current frame, obtain color image of maximum posterior probability and segment the color image to aim at tracking multiple human under occlusion in a single camera.

3. The dissertation aims at researching on mis-match problem of tracking human object across cameras with overlapping field of view. A fusion algorithm is proposed based on space map transformation and color feature to accomplish match across cameras. The dissertation firstly ascertains the position of human across camera roughly using the common boundary lines of overlapping field of view. Then the person across cameras gets matched by means of homography transformation matrix. However, in practical application, the videos across cameras is asynchronous and occluded persons, so there exist some mis-matches. The dissertation calibrates the mis-match by maximum posterior probability of color model on the basis of homography matrix.

4. The dissertation aims at solving mis-match problem caused by color differences across cameras with non-overlapping field of view. An algorithm is presented to calibrate the color pixels based on color transfer function and apply probability estimation to the space of color transfer function in order to advance match accuracy of tracking human across cameras. The algorithm firstly trains the given samples of human objects across cameras in a low dimension to acquire color transfer function. The algorithm is not dependent on internal parameter of cameras to acquire color transfer function. Then the probability estimation function is modeled utilizing principal component analysis to sub-space of color transfer function. Finally, match probability of tracking human across cameras using probability estimation function. The experiments in the case of outdoor-outdoor and indoor-outdoor shows the algorithm can better track human object across camera with non-overlapping field of view.

致谢.....	I
摘要.....	I
Abstract.....	III
目次.....	V
图目次.....	IX
1 绪论.....	1
1.1 本文研究的背景和意义.....	1
1.2 视频跟踪国内外现状及所面临的困难.....	2
1.2.1 视频跟踪国内外现状.....	2
1.2.2 视频跟踪所面临的困难.....	3
1.3 人体目标视频跟踪算法概述.....	6
1.3.1 单摄像机人体目标跟踪算法.....	6
1.3.2 多摄像机人体目标跟踪算法.....	11
1.4 本文研究的主要工作和组织结构.....	16
2 无背景检测的单摄像机多目标遮挡人体跟踪.....	19
2.1 引言.....	19
2.2 贝叶斯滤波的蒙特卡罗实现.....	21
2.2.1 最优贝叶斯估计.....	21
2.2.2 重要性采样.....	22
2.2.3 序列重要性采样.....	23
2.2.4 重采样.....	24
2.2.5 基本滤波算法描述.....	25
2.3 多目标多模型的粒子滤波.....	26
2.3.1 多目标跟踪模型.....	26
2.3.2 粒子估计.....	28
2.3.3 遮挡处理.....	30
2.3.4 观测模型.....	31

2.3.5 多目标多模型滤波算法描述.....	32
2.4 实验结果和分析.....	33
2.5 本章小结.....	37
3 基于背景检测的单摄像机多目标遮挡人体跟踪.....	39
3.1 引言.....	39
3.2 改进的混合高斯模型背景重建.....	40
3.2.1.背景模型的建立.....	40
3.2.2 背景更新算法的改进.....	44
3.2.3 背景更新算法的改进结果.....	45
3.3 基于多模型融合的遮挡人体跟踪.....	48
3.3.1 遮挡人体跟踪的基本思想.....	48
3.3.2 块建模.....	51
3.3.3 颜色建模.....	53
3.3.4 运动建模.....	54
3.4 提高算法的运行效率.....	56
3.5 实验结果和分析.....	57
3.6 本章小结.....	60
4 有公共视场的多摄像机人体目标跟踪.....	63
4.1 引言.....	63
4.2 视场分界线的目标交接.....	65
4.3 二维投影变换单应矩阵的计算.....	67
4.5 颜色特征跟踪.....	72
4.6 实验结果和分析.....	75
4.7 本章小结.....	79
5 无公共视场的多摄像机人体目标跟踪.....	81
5.1 引言.....	81
5.2 颜色校正的方法.....	81
5.2.1 基于映射的颜色校正方法.....	81
5.2.2 基于光谱反射率还原的颜色校正方法.....	84

5.2.3 基于图像分析的颜色校正方法.....	86
5.3 亮度转变函数的数学模型.....	88
5.4 亮度转变函数的实际应用.....	92
5.5 对亮度转变函数空间的概率估计.....	93
5.6 实验结果和分析.....	95
5.7 本章小结.....	99
6 总结与展望.....	101
6.1 本文总结.....	101
6.2 工作展望.....	102
参考文献.....	105
作者在攻博期间发表(录用)的论文.....	120

图 2-1 场景 A 的原图.....	34
图 2-2 场景 A 图像的人体目标的粒子.....	35
图 2-3 场景 A 人体目标的跟踪结果.....	35
图 2-4 场景 B 的原图.....	36
图 2-5 场景 B 图像的人体目标的粒子.....	37
图 2-6 场景 B 人体目标的跟踪结果.....	37
图 3-1 改进的混合高斯模型流程图.....	45
图 3-2 场景 A 的运动人体目标检测.....	47
图 3-3 场景 B 的运动人体目标检测.....	48
图 3-4 系统总体流程图.....	48
图 3-5 颜色块建模.....	49
图 3-6 运动建模的向量.....	50
图 3-7 遮挡情况的处理过程.....	51
图 3-8 块建模的实现过程.....	53
图 3-9 基于多人体的运动模型.....	56
图 3-10 隔行抽取后图像遮挡情况处理.....	57
图 3-11 两组视频序列的跟踪过程.....	59
图 3-12 两组视频序列的跟踪轨迹.....	60
图 3-13 Mean shift 算法与本文算法准确率比较.....	60
图 4-1 相邻摄像机视场示意图.....	64
图 4-2 有公共视场的跨摄像机人体目标跟踪框图.....	65
图 4-3 不同摄像机之间公共可视区域的确定示例.....	67
图 4-4 摄像机成像几何关系.....	68
图 4-5 图像坐标系.....	69
图 4-6 单应矩阵变换示意图.....	71
图 4-7 利用单应矩阵变换得到地平面上的点对应.....	72

图 4-8 跨摄像机颜色跟踪的流程图.....	73
图 4-9 两个摄像机情况下的人体目标对应的解空间.....	74
图 4-10 有公共视场的跨摄像机的人体目标跟踪过程.....	77
图 4-11 场景 A 的跨摄像机人体目标跟踪.....	77
图 4-12 场景 B 的跨摄像机人体目标跟踪.....	78
图 4-13 场景 A 的跨摄像机人体目标跟踪轨迹.....	78
图 4-14 场景 B 的跨摄像机人体目标跟踪轨迹.....	79
图 4-15 场景 A 和场景 B 人体目标跟踪的准确率.....	79
图 5-1 基于查找表的颜色校正示意图.....	84
图 5-2 摄像机 A 的像素亮度值 59 与摄像机 B 的像素亮度值 42 对应示意图.....	92
图 5-3 亮度转换函数转换前后的 R 通道直方图对比.....	93
图 5-4 室外环境摄像头位置及索尼监控网络摄像机.....	96
图 5-5 无重叠跨摄像机系统的拓扑结构.....	96
图 5-6 无公共视场的场景 A 情况下跨摄像机人体目标跟踪.....	98
图 5-7 无公共视场的场景 B 室内外情况下跨摄像机人体目标跟踪.....	99

1 绪论

1.1 本文研究的背景和意义

计算机视觉是采用图像处理、模式识别、人工智能技术相结合的手段,着重于一幅或多幅图像的计算机分析。它的主要任务就是通过对采集的图像或视频进行处理以获得相应场景的信息,就像人类和许多其他类生物那样进行感知、解释和理解。计算机视觉是一门新兴的发展迅速的学科,也是工程领域和科学领域中的极其富有挑战性的一门综合性学科,它囊括了计算机科学、信号处理科学、物理学、应用数学和统计学、神经生理学和认知科学等学科,它已经吸引了许多来自各个学科的研究者加入到这门学科的研究之中。

视频监控系统是安全防范系统的重要组成部分,一种防范能力较强的综合系统。视频监控以其直观、准确、及时和信息内容丰富而广泛应用于许多场合。随着网络的迅猛发展、计算机 CPU 处理能力的提高和存储容量的大幅提升,视频信息处理技术的快速发展,视频监控系统的优势越来越明显,使得视频监控系统的处理能力大幅度提升,这样,安防产业就具备了更加广阔的发展潜力,涌现了许多新型的应用模式,其中,智能视频监控就是其中的一项很重要的应用模式。

智能视频监控技术离不开计算机视觉与人工智能的研究,用摄像机和计算机代替人眼来实现对目标进行识别、跟踪。其发展目标在于将图像与事件描述之间建立一种映射关系,使计算机从纷繁的视频图像中分辨、识别出关键目标物体,这一研究应用于安防视频监控系统、将能借助计算机强大的数据处理能力过滤掉图像中无用的或干扰信息、自动分析、抽取视频源中的关键有用信息,从而使传统的监控系统中的摄像机不但成为人的眼睛,也使智能视频分析计算机成为人的大脑,并具有更为聪明的学习思考方式。智能视频监控不仅具有普通的网络视频监控的优点,而且还可以为用户带来更大的好处:可以全天的监控需要监控的场景,不再需要安全工作人员对画面进行监视和分析,可以通过智能视频监控模块对所监控的场景进行分析;视频服务器具有很强的图像处理能力,并运行高级智能算法,可以进一步确认安全威胁的特征,提高报警的准确率;可以提高出现问题的响应速度,尽可能早的通知安全人员,监控场景可能出现安全威胁或可疑活动,并且安全人员可以对出现的状况作出进一步反应,采取措施将事态消灭在萌芽状态,有效防止事态进一步恶化,按照事先准备好的应急方案进行执行,能有效地抑制混乱的发生或产生延误。

由于单个摄像机的视场有限,在监控大范围场景时,单个摄像机很难这样的监控场合,因此,使用多个摄像机监控大范围场景就应运而生。这种需求包括两种情况,一种是对重要的场景点进行监控即无重叠视场的监控,一种是对大范围的场景实现无缝对接即有重叠视场的监控,如何实现跨摄像机的联合协调监控在目标跟踪过程中具有重要的意义。从应用角度来看,跨摄像机的智能视频监控具有广阔的应用前景。然而从技术层面来看,跨摄像机的人体目标跟踪技术研究才刚刚起步,在这种背景下,研究智能视频监控系统中比较棘手的多人目标遮挡及跨摄像机的跟踪问题具有极其重要的意义。本文研究的内容即是利用计算机视觉技术研究多人目标遮挡及跨摄像机跟踪中存在的问题。

1.2 视频跟踪国内外现状及所面临的困难

1.2.1 视频跟踪国内外现状

在智能监控系统方面,当前有许多国际一流期刊为科研工作者提供了很好的相互交流的平台,如 IEEE Transaction on Pattern Analysis and Machine Intelligence (PAMI)、Image and Vision Computing (IVC)、Computer Vision and Image Understanding (CVIU)、International Journal of Computer Vision (IJCV) 等及国际知名的学术会议如 IEEE Computer Society Conference on Computer Vision and Pattern (CVPR)、International Conference on Computer Vision (ICCV)、European Conference on Computer Vision (ECCV)、IEEE International Workshop on Visual Surveillance (IWVS)、International Conference on Image Processing (ICIP) 等。在全世界各国专家学者的共同努力下,智能视频监控技术已经取得了许多创新性的科研成果。

国外有很多科研院所、及企事业单位的科研人员一直从事这方面的研究与开发。1997年,美国国防高级研究计划局(Defense Advanced Research Projects Agency)设立了以卡内基梅隆大学(Carnegie Mellon University)为首联合十几家高等院校和研究机构参加的视频监控重大项目 VSAM(Video Surveillance and Monitoring)[Collins et al,2002;Lipton et al,1998],主要研究对战场及普通民用场景进行监控的自动视频理解技术。2000年,美国国防高级研究计划局远距离人脸识别项目;中央佛罗里达大学(University of Central Florida)的 KNIGHT 智能监控系统利用计算机视觉技术能检测出监视区域目标的变化,并能标注重要事件和描述人的行为,系统对光照变化有较好的鲁棒性。英国雷丁大学(University of Reading)已开展了对车辆和行人的跟踪及其交互作用识别的相关研究

[Remagnino et al,1998]。MaryLand 大学的实时视频监控系统 W4[Haritaoglu et al,2000]不仅能够定位人和分割出人的身体部分，而且通过建立外观模型来实现多人的跟踪，可以检测和跟踪室外环境中的人并对他们之间简单的交互进行监控。IBM 的智能监控系统不仅能够自动监视某个现场，还能够监视数据，执行基于事件的检索，通过标准的网络设施进行实时报警，并且能提取出某项活动的长期的统计模型。

与国外相比，国内智能视频市场还有很大的差距，目前基本上还处于空白状态。一般在监控系统中提到的智能视频监控实际上还停留在普通的网络视频监控（IP 监控、数字化监控）的概念上。但随着市场上开始出现了与国外类似的智能视频应用需求，已经有些国内厂商开始着手引进国外知名厂商的智能视频软硬件产品和技术，计划采用 OEM 的形式在国内推出。尽管国内外市场上有出现了一些产品，在实际的应用中，受到很多限制，还有待于进一步的完善。

无论在国内还是国外，单摄像机的智能视频监控在部分产品上出现了少许应用，但还不成熟不完善，仍作为辅助功能。而跨摄像机的智能视频监控尚未有成型的产品出现，还处在理论研究阶段。

1.2.2 视频跟踪所面临的困难

智能视频人体运动目标跟踪仍然面临着很多难题。主要困难来自于人体的非刚体运动、将物体从三维投影到二维造成深度信息丢失、视频中的人体被遮挡和自遮挡、运动变化的背景、光照变化和运动模糊、人体模型的自动初始化及更新问题、前景检测不完整或误检测、很难实现实时跟踪等，客观存在的诸多因素使得从视频序列中跟踪运动的人体目标至今仍是一个艰难的技术挑战。

(1) 人体的非刚体运动

人体运动是一个比较复杂的关节运动，人体的运动是在骨骼驱动下带动的多个复杂的刚体运动构成的。然而在实际视频跟踪中，人们很难建立一个人体模型来模拟人体的运动，人们穿着的衣服是柔软且可任意变形的，尤其是在风的吹动下，衣服将会随风飘动，而且还有可能人们手持雨伞或肩跨背包，所有这些不可预知的复杂因素都决定了真实世界中的人体运动是高度复杂的非刚体运动。这在人体目标跟踪过程中带来了很大困难，所以必须研究新的非刚体运动分析方法来解决复杂的人体运动。

(2) 将物体从三维投影到二维造成深度信息丢失

将真实世界中的三维物体投影到二维图像中将会丢失深度信息，这是在计算机视觉

里必需面对的困难,尤其是基于单摄像机人体目标跟踪过程中。为了将三维信息重建出来,有不少研究人员采用多个摄像机对同一个场景进行多角度同步拍摄来尽量避免摄像机成像过程中深度信息的丢失,但是,这种方式不仅为实际应用提高了成本,最关键的是,如何对多个同步视频进行有效合理的融合仍是一个很难解决的问题。

(3) 遮挡和自遮挡

由于视频人体目标跟踪过程中所应用的环境是比较复杂,很可能被周围的物体部分遮挡或者全部遮挡以及人体的自身遮挡。这些遮挡将会使人体的部分信息丢失或颜色改变,使抽象出来的人体模型与完整的人体模型不相符合或使当前的颜色模型与未发生遮挡时的外观模型不相符合。

(4) 运动变化的背景

应用视频跟踪的背景多数是复杂且时常变化的,比如背景是郁郁葱葱的树林,在风力的吹拂下来回摇晃,很难进行背景重建;若是下着倾盆大雨,同样会给人体的目标跟踪带来挑战。由于动态背景目标在小范围内的随机移动或定向漂移所造成的同一像素位置上不同物体的更迭与不同位置上物体类似的情形。比如波光粼粼的湖面,又如在风中飘摇不定的树叶,它在场景的不同时刻往往存在于不同的像素位置,从而使得同一像素位置往往为树叶、树枝、天空等目标交替占据。

(5) 光照变化和运动模糊

由环境光照条件变化引起的外观变化,包括渐变、突变、随机干扰等情形。自然环境中日夜更替、四季循环所带来的外观变化,乌云遮挡日光后物体的色彩强度变化,以及跨摄像机跟踪情况下,位置不同的摄像机所处的光照环境差异较大,有的视场在树荫处,有的在室内,有的在炎炎烈日下,有的在窗口处等。环境的光照条件使人体的颜色信息发生变化,给人体的目标跟踪也带来了很大的困难。同时人体的快速运动和摄像机有限的拍摄速率也会不可避免地造成所获取视频中的运动模糊现象,也会给视频人体运动的自动分析带来一定的困难。

(6) 人体模型的自动初始化及更新问题

在人体目标跟踪过程中,人体模型的自动初始化仍是一个比较棘手的问题。自动初始化人体的颜色模型就是应用计算机自动地去除背景信息将单个的人体目标提取出来。在真实的场景中,情况还是比较复杂的。比如进入场景中的人是一群人或是进入场景中的人与离开场景中的人恰好发生遮挡,或者是在艳阳高照的天气,人拖个长长的阴影,

以及人体的部分信息丢失,这些都给自动初始化带来很大的难题。如在自动初始化阶段就有很多噪声信息或不完整的信息,那么跟踪效果可想而知一定会很差。所以如何自动地将一群人、发生遮挡的人、带有阴影的人等情况有效地分割与识别仍需要继续研究与探索。同样,在人体跟踪过程中自动更新模型也会产生非跟踪人体目标的噪声信息和人体的部分信息丢失;如若不更新模型,那么光照条件发生变化,当前人体模型会与初始化模型产生偏差。综合考虑之后,目前绝大多数基于视频的人体运动跟踪系统都依赖于手动初始化人体信息,然后在跟踪过程中并不进行更新人体的颜色模型。这种方法尽管无法避免误差和噪声干扰,但是这种方法可以降低由于更新而产生人体目标误跟踪的问题。

(7) 前景检测不完整或误检测

随着时间变化场景也将随之变动,前景与背景之间的相互转换的情形。比如监视停车场变化一段视频,有个车辆进入停车场后静止下来,然后从车辆中走出的一些人,那么这里停下来的车必须马上变为背景,然而要将车辆变为背景还需要一段时间学习视频进行更新背景模型;还有背景与前景之间相互混淆的情况,主要特点是背景与前景之间的颜色相似比如穿着蓝色服装的行人在蓝色装饰为主的建筑环境中往往会被判为背景的一部份而导致错误;局部背景的显著变化,比如前景的阴影对背景的影响,将突变背景误判为前景而将阴影部分的背景提取出来。

(8) 很难实现实时跟踪

在人体目标跟踪过程中,要对大量的图像数据进行处理,需要用较为复杂的算法和大量的数学运算,如何降低算法的计算时间复杂度,实现人体目标的实时跟踪,在实际应用中也是一个比较棘手的难题。

正是由于对视频中的人体目标跟踪存在这么诸多困难,很多科研人员对各种各样的复杂问题已进行了不同程度的简化,对环境和人体目标等进行了不同的约束和假设,利用不同的知识来辅助问题的解决。本文的基于视频的多目标人体跟踪系统同样也作了一些典型的假设:

1) 针对环境:光照不变,静止背景,背景颜色固定不变,摄像机参数已知,摄像机是固定的等等。

2) 针对目标:手动对目标进行初始化,只对步行的人进行研究,人体目标的分辨率至少应 $20 \text{像素} \times 40 \text{像素}$,在监控过程中人体目标外观是不变的等等;

本文所研究的条件是在符合实际环境且较为宽松情况下进行的,为其能在实际环境中得以应用奠定基础。

1.3 人体目标视频跟踪算法概述

在理论算法研究方面,在过去的几十年里,视觉跟踪的方法有了很大的发展,出现了大量的跟踪算法。这些动态背景条件下适用的跟踪算法大致可以分为两类:确定性方法和统计性方法,两者的代表算法分别是均值漂移和粒子滤波器。确定性算法计算量小,实时性高,适应于实时监控等应用;而统计性算法由于需要计算大量的采样,具有更高的准确性,但实时性低于前者。这些算法从两个方面提高跟踪的性能,一是提高跟踪的速度,二是提高跟踪的精度。

人体目标的视频跟踪是指在视频图像序列之间提取人体目标的位置、速度、形状、纹理、色彩等有关特征并应用这些特征进行匹配,以达到对同一人体目标进行跟踪的目的。目前,人体目标视频跟踪有多种分类方法:就跟踪对象而言,有对手、脸、头、腿等身体部分跟踪与对整个人体的跟踪;就跟踪空间而言,有二维跟踪和三维跟踪;就跟踪环境而言,有跟踪室内跟踪和户外跟踪;就跟踪人数多少,可分为单人跟踪和多人跟踪;就采用的视角而言,将人体目标跟踪分为单视角人体目标跟踪和多视角人体目标跟踪;就摄像机状态而言,分为运动摄像机跟踪和固定摄像机跟踪;就摄像机数目而言,分为单摄像机人体目标跟踪和多摄像机人体目标跟踪,其中多摄像机人体目标跟踪分为有重叠视场的多摄像机的人体目标视频跟踪和无重叠视场的多摄像机的人体目标视频跟踪。本节将按照摄像机数目分别介绍基于单摄像机和基于多摄像机中人体目标跟踪的相关技术。

1.3.1 单摄像机人体目标跟踪算法

近几年,随着图像工程和计算机视觉研究的兴起,人体目标视频跟踪在国际上引起相关科研人员的广泛关注,并将很多重要的数学理论模型应用到视频人体目标跟踪系统。单摄像机人体目标跟踪常用的数学工具有卡尔曼滤波[Isard and Blake,1998;Terzopoulos and Szeliski,1992]、粒子滤波[MacCormick and Blake,2000]、贝叶斯估计理论、高斯模型、非参数核密度估计、均值漂移算法、马尔科夫链等。由于人体跟踪算法的分类方法繁多[王亮等,2002;侯志强和韩崇昭,2006],本文依据不同的跟踪方法,将目前比较常用的几类跟踪方法进行总结,将视觉跟踪方法分为六类,分别是基于区域的跟踪算法,基于轮廓的跟踪,基于特征的跟踪,基于光流的跟踪算法,基于网格的跟踪,基于人体模型的跟踪。下面简

单介绍这些不同类型的视觉跟踪算法。

(1) 基于区域的跟踪算法

基于区域的人体视频跟踪的基本思想就是利用视频或图像的灰度信息、颜色信息、运动信息、帧间变化信息等将每帧图像的前景目标提取出来,接着在相邻帧中进行区域目标匹配,实现人体目标跟踪。基于区域的人体跟踪就是将颜色相同的像素聚类在一个小区域中,使整个人体区域由各个小区域组成,然后对这些小区域块进行匹配,最终实现对整个人体目标进行跟踪。基于区域的跟踪方法有,Wren 等[Wren et al,1996]就是利用小区域特征的匹配实现跟踪人体目标,Wren 等人将人体分成头部、身体、四肢等小块区域,利用高斯分布建立模型,通过匹配各个小区域块达到对人体目标跟踪的目的。处理遮挡的人体目标是区域跟踪的缺点,这个缺点可以通过利用彩色信息来弥补区域跟踪的不足。因此,McKenna 等人[McKenna et al,2000]就是利用色彩和梯度信息,通过背景重建并减背景的方法来检测运动的人体目标;然后,利用人体目标的几何约束条件,将区域合并或分离,因此,在遮挡的情况下,该方法还具有一定的鲁棒性,能够较好地实现人体目标的跟踪。范瑞霞[范瑞霞等,2006]通过背景减除提取人体目标的多个小区域,利用模糊推理评价出它们在相邻两帧间的匹配可靠度;通过跟踪各个可靠度高的小区域,并对可靠度低的小区域进行邻域搜索、位置纠正和估计,完成整个人体目标的跟踪。

(2) 基于轮廓的跟踪

一般来说,运动目标的形状是比较复杂的,很难用一个模型来描述。基于轮廓的跟踪就是通过对目标的轮廓进行描述达到跟踪的目的。目前,比较流行的使用方法是基于形状的匹配跟踪[Yilmaz et al,2004]。基于形状匹配轮廓跟踪是在当前帧搜索目标的轮廓与其模型使其与前一帧形成的假设目标及其模型具有最大的相似度,通过此方法来实现对轮廓的跟踪。在该方法中,非刚体目标的轮廓是变化的,所以不能直接根据上一帧获取最大的相似度。Huttenlocher 等人[Huttenlocher et al,1993]提出了一种基于边缘的形状匹配方法,该方法根据最小相似性原理采用 Hausdorff 距离作为数学测度比较两个集合点。在该文中,Huttenlocher 等人用 Hausdorff 距离来衡量当前帧的人体目标与存储模型中的人体目标的边缘最匹配的人体目标。另一种形状匹配轮廓跟踪方法是在两个连续帧找对应轮廓,然后利用轮廓进行匹配,然而,在实际应用,单一使用轮廓效果并不是很好,一般要和颜色模型、直方图模型相结合的方法效果会更好。Kang 等人[Kang et al,2004]就使用了彩色直方图和边缘相结合的方法。该彩色直方图是对半径不同的同心圆来做统

计生成直方图，这些圆以一些参考圆上的控制点为中心。参考圆是通过目标轮廓的最小外接圆来获取的。匹配个数可通过交叉相关、Bhattacharya 和 Kullback-Leibler 发散距离等距离测度计算出来。

(3) 基于特征的跟踪

基于特征的跟踪方法是从图像中提取目标的典型特征并在图像序列间匹配这些典型特征从而达到对目标进行跟踪的目的。基于特征的跟踪需要通过提取特征和匹配特征两个步骤才能实现。典型特征一般包括目标的特征点、特征线、特征轮廓、位置、速度、形状、纹理、色彩等特征。向桂山等[向桂山等,2007]考虑了人脸的颜色与其他物体的颜色分布一般不同，利用人脸的肤色特征作为人脸检测与跟踪的方法。针对视频图像中的运动人体目标，为了去除近肤色背景的干扰，采用帧差法并结合人体形状信息提取出运动人体目标；在 YCbCr 色彩空间中，利用肤色的 Cb 和 Cr 分量的高斯分布特性，检测出运动人体目标的肤色区域，即为人脸候选区域；判断连通域大小及长宽比，去除明显的非人脸区域，并根据欧拉数及嘴唇的特有颜色特征，最终检测出人脸区域；根据人脸重心位置信息，驱动具有水平和垂直两个自由度的摄像机，把人的头肩部分始终定位于视场中央；在跟踪的过程中，由于摄像机角度变化或者环境光线变化，根据 Y 分量的变化及时更新高斯模型参数，来提高跟踪的准确率和稳定性。以卡尔曼滤波形式的点 K 线特征的跟踪技术在计算机视觉领域中已被很好地开发，如 Jang 和 Choi [Jang and Choi,2000]提出了一种卡尔曼预测、活动模板相融合的方法。活动模板是指充分利用区域的颜色、几何形状、纹理和边缘特征信息，将活动模板的特征用能量函数最小化来完成运动目标的跟踪，由于该算法应用了几何形状，因此该方法只对非刚性物体的跟踪具有较好的跟踪效果。

(4) 基于光流的跟踪算法

光流[Cremers and Schnorr,2003;Mansouri,2002]是在世界坐标系下的运动物体在投影成像面上的像素运动产生的瞬时速度。利用视频中像素值的变化和相关性来确定像素的运动，是研究图像灰度值的变化与空间中的物体的运动之间的关系。摄像机的运动或空间中的目标运动都会产生光流[Mansouri,2002]。可以通过匹配、梯度以及频域三种方法来计算光流。光流匹配的计算方法就是利用特征匹配或区域匹配方法进行定位和跟踪来计算光流的速度。光流梯度的计算方法就是利用视频的时空微分计算二维的光流速度。由于该方法简单有效得到了广泛的研究。光流频域的计算方

法就是速度可调的滤波组输出频率或相位信息。

张泽旭等[张泽旭 等,2003]提出了一种基于特征光流分割和卡尔曼滤波估计的鲁棒性的运动目标跟踪方法,该方法具有很多特点:首先在特征光流的计算中采用由粗到细的层级匹配算法,因而能够计算大的运动速度和具有更好的匹配精度;其次采用了有效的遮挡判决算法,该算法综合利用了先验的信息,对噪声的干扰不敏感;最后建立了线性卡尔曼滤波模型,当特征点被遮挡或丢失时,能够预测它们的位置,这使得跟踪更具有主动性。实验表明该方法具有精度高、跟踪快和鲁棒性好的特点。

(5) 基于网格的跟踪

基于网格的视频目标跟踪是通过目标上的网格节点跟踪来实现的,首先利用多边形得到目标边界上的网格节点,然后根据小波变换得到目标特征能量,通过目标特征来检测角点,并将角点作为网格节点,紧接着利用增量式网格得到目标的网格模型,最后利用网格节点通过仿射变换参数法对全局运动矢量进行估计和通过六边形匹校正法对局部运动矢量进行估计,从而实现对目标进行跟踪[刘重庆,2003]。Brusewitz [Brusewitz,1994]提出基于三角形网格的运动补偿技术,三角形顶点的运动矢量通过光流方程来求解再对每个三角形内部线性插值得到运动矢量场。Sullivan and Baker [Sullivan and Baker,1991]提出采用四边形网格运动补偿,通过最小亮度预测误差搜索节点最佳运动矢量。Nieweglowski 等人[Nieweglowski et al,1993]将网格节点放置在亮度边缘节点运动矢量估计通过块匹配方法实现。Nakaya 和 Harashima [Nakaya and Harashima,1994]提出采用局部最优搜索策略(六边形匹配)进行三角形网格或四边形网格的节点运动估计和补偿。Dudon 等人[Dudon et al,1994]设计了一种自适应网格,利用六边形匹配提高运动补偿预测。Wang 等人[Wang and Lee,1996a;Wang and Lee,1996b]提出基于活动网格的能量优化运动补偿算法。Altunbasak 和 Tekalp 等人[Altunbasak and Tekalp,1997]提出了基于内容的网格设计和跟踪的实用算法网格节点放置在运动目标的边缘及角点等特征区域并使节点的密度与局部运动成比例。[Szeliski and Shum,1996]提出分层网格模型进行节点运动估计。Schroder 和 Mech [Schroder and Mech,1995]提出了一种利用对象的网格模型对视频目标进行跟踪,该算法可以对目标的局部运动进行跟踪,同时也可以对单个的目标进行跟踪。而且该方法可以比较准确地描述目标的边界并利用其进行跟踪。可以避免区域匹配跟踪产生的块效应,提高了算法的准确性。刘李杰和蔡德钧[刘李杰和蔡德钧,2000]提出了一种基于二维网格的运动跟踪算法。该算法首先要对上一帧的边界为基础,然后利用

最大对比度约束条件在上一帧的边界附近查找当前帧的目标边界；如果当前帧存在未覆盖区域的边缘，可以利用二维网格搜索算法重新进行边缘提取；利用运动矢量、颜色信息和边缘信息进行分析并确认是背景、原对象一部分还是新出现的对象，然后对其进行跟踪，同时更新每帧的网格结构进行连续。

(6) 基于人体模型的跟踪

基于人体跟踪的常用人体几何模型主要有主要有骨架模型(Skeleton Model)、纸板模型(Cardboard Model)、圆柱模型(Cylinder Model)、圆台模型(Truncated Cone Model)、超二次曲线(Super Quadric Model)和分层结构化的模型(Hierarchical Model)等。

基于模型的跟踪使用的人体几何模型在细节上从简单到复杂有着很大的差别。在早期研究中，简单的骨架模型[Iwai et al,1999;Karaulova et al,2000;Lee and Chen,1985]经常被使用，在这种模型中，身体的各个部分被表示成简单的棍棒，这些棍棒由关节链接在一起。骨架模型非常简单，但姿态评价时很难度量匹配误差。因为人体的各个部分投影到图像平面之后一般可以近似成规则的四边形，所以一个很自然的想法就是用二维的多边形或曲边形表示这些部分，再由关节连接，纸板人[Gao and Shi,2003;Huang and Huang,2002;Ju et al,1996;Mori and Malik,2002]正是这样一种人体模型。纸板模型避免了模型的三维投影过程，与二维图像有着直接的对应关系，但不能解决三维参数的恢复问题。为此需要更加复杂的立体模型。早在 80 年代初，Hogg [Hogg,1983]在人的运动分析方面做了不少开拓性的工作，他使用三维圆柱模型，人体的每个部分都用一个圆柱体表示。在后来的工作中，圆柱模型被不断地使用[Attwood et al,1989;Demirdjian,2003;Demirdjian and Darrell,2002;Rohr,1994;Rosales and Sclaroff,2000;Sidenbladh et al,2000;Sidenbladh and Black,2001;Wachter and Nagel,1999]。由于圆柱的两端相同大小，而人的手和腿显然不是如此，为此又出现了圆台模型 [Cheng and Moura,1999;Delamarre and Faugeras,1999;Delamarre and Faugeras,2001;Goncalves et al;Ning et al,2008;Ning et al,2002]。为了更加精确地将模型匹配到真实图像中的人，[Gavrila and Davis,1996;Jojic et al,1997;Kakadiaris and Metaxas,2000;Sminchisescu and Triggs,2001a;Sminchisescu and Triggs,2001b]使用超二次曲线，构建了逼真且可变形的模型。[D'Apuzzo et al,1998;Plaenkers and Fua,2002;Plaenkers and Fua,2001]使用分层的结构化的模型，它分为四层：骨架、椭球块表示的肌肉、多边形表示的皮肤、以及最后渲染的结果，这个模型与人体真实的生理结构更加接近。一般来说，模型越复杂越精确，其跟踪结果也就越准

确,但同时需要更多的参数和更多的时间来搜索解空间,如何根据实际问题在这二者之间寻找一个平衡是个必须考虑的问题。

1.3.2 多摄像机人体目标跟踪算法

(1) 有重叠视场多摄像机人体目标跟踪

有重叠视场的多摄像机人体目标跟踪已经成为当前跟踪领域的一个热点问题。许多科研人员一直致力于这方面的研究,已经取得了一些阶段性的成果,一些文献[Chang et al,2000;Hu et al,2006;Khan and Shah,2003]从不同的角度已对这些跟踪方法进行了分类。本节从是否需要标定摄像机的角度,对有重叠视场多摄像机人体目标跟踪的方法分为基于几何约束的跟踪和基于颜色特征的跟踪两大类。

① 基于几何约束的跟踪

近年来,在有重叠视场多摄像机人体目标跟踪技术中,应用比较广泛的一个方案就是基于几何约束的特征点匹配方法。基于几何约束的跟踪可以进一步分为基于三维的几何约束跟踪和基于二维的几何约束跟踪。

基于三维的几何约束跟踪可以通过两种方法实现。一种是将所有的特征点转换到一个相同三维空间,然后根据不同视场内对应特征点映射到三维空间是同一个点的原理进行跨摄像机匹配。在文献[Tsutsui et al,1998;Utsumi et al,1998]将人体目标的质心作为特征点,通过估计这些特征点在真实世界三维的坐标系下的位置建立对应关系以达到匹配的目的。在文献[Kelly et al,1995]中,所有的摄像机被标定并且三维环境模型是事先已知的。然后利用人体目标在三维坐标下具有相似的位置来实现跨摄像机跟踪。另一种三维跟踪的策略是利用三维极线约束来匹配。在文献[Cai and Aggarwal,1999]中,在贝叶斯概率框架下,首先对相邻的两个摄像机做了相对的标定,然后为了减少不同角度的亮度差别,将每次映射的人体目标上半身的中心轴上的多个点作为被跟踪的特征点,提取这些特征点的位置和平均亮度信息,最后根据亮度信息和受极线约束的特征点建立起摄像机间的匹配对应。基于三维的约束跟踪方法要求事先对摄像机进行标定,而且从不同视频序列中提取的对应特征点并不总是很准确的转换到三维真实世界中的点。这是因为对应特征点不是很匹配造成的。

基于二维几何约束跟踪可以克服基于三维几何约束跟踪的一些缺点。利用二维信息建立摄像机间的匹配已在一些文献[Khan et al,2001]中进行了应用。Khan 根据不同视场重叠摄像机间具有共同的地平面的属性利用单应性变换(homography)原理对不同摄像机

中的人脚位置作为特征点进行匹配。然而在实际应用中,视频中人脚的特征点可能由于检测的不准确、检测的鲁棒性差、或被遮挡而不可见等因素的影响,在匹配过程中会产生错误。Black 和 Ellis [Black and Ellis,2006]基于单应性约束提出了转换误差来实现不同视场中质心之间的对应,如果只检测出来部分人体或者部分人体是可见的,那么这种方法的效果也会大打折扣,甚至是误匹配。

② 基于颜色特征的跟踪

颜色特征是跨摄像机跟踪的一个很重要的特征而且无需摄像机标定。利用颜色特征进行跨相机跟踪实质上就是在单个摄像机中将人体目标提取出来,然后利用人体目标的颜色信息进行跨相机间匹配。这种匹配可以通过卡尔曼滤波框架实现[Utsumi and Ohya,2000],贝叶斯模型实现[Chang and Gong,2001],高斯模型实现[Mittal and Davis,2002]或者直方图匹配来实现[Krumm et al,2000]。使用颜色信息来构建跨摄像机匹配是简单而且直接的想法,然而仅仅依靠颜色特征而没有经过任何颜色校正直接进行跨摄像机间人体目标跟踪是不可信赖的。在文献[Azarbayejani and Pentland,1996;Chang and Reid,1996;Stillman et al,1999]中只利用了颜色信息进行匹配从而容易导致误匹配。容易误匹配的原因在于:一方面摄像机视场中不同的人穿着相似的衣服而引发误匹配;另一方面摄像机的视点不同或光照条件不同都会产生同一个人在不同的摄像机当中产生不同的颜色,颜色对光照条件的超敏感性引起误匹配也是在意料之中的。所以,在很多情况下科研人员不仅使用颜色特征还使用许多其它特征。有重叠视场的摄像机跟踪的一个重要的优势就是可以利用摄像机间的几何关系。因此他们使用摄像机的标定信息,根据摄像机之间的几何关系获取额外的约束特征以提高跟踪的准确性和鲁棒性。例如,Chang 和 Gong [Chang and Gong,2001]将基于几何关系模型和基于识别模型也加入贝叶斯模型框架中,其中几何关系的模型包括极线几何和单映性约束条件,识别关系模型包括高度和颜色信息。

(2) 无重叠视场多摄像机人体目标跟踪

无重叠视场多摄像机人体目标跟踪已成为广域视频监控研究中的重要研究内容之一同时也是人体目标跟踪领域最具有挑战性的问题之一。由于观测视场的不重叠,不连续和跨摄像机的颜色特征不一致性,使得解决传统视频跟踪问题的技术方法和研究框架很多都不再适用。孔庆杰和刘允才[孔庆杰和刘允才,2007]对无重叠视场多摄像机所有跟踪对象包括人和车等进行了综述。然而实际上对象不同跟踪方法也不相同。所以,本节对国

内外关于无重叠视场多摄像机人体目标跟踪的研究成果和最新进展进行分析,总结出当前科研人员主要是从三个方面研究该问题:颜色特征目标匹配(Color-based Objects Matching)、拓扑估计(Topology Estimation)和数据关联(Data Association),并对这三部分的研究依次进行介绍:

① 颜色特征目标匹配

颜色特征是无重叠视场多摄像机人体目标的最基本的特征而且具有较好的区分性,因此成为众多特征中被研究最多的一种。Cheng 和 Madden 等人[Cheng et al,2006;Cheng and Piccardi,2006;Cheng and Piccardi,2007]研究室内监控环境下跨摄像机无重叠视场的行人匹配跟踪。他们先通过在线 k 均值聚类的方法对人体目标进行聚类,建立了一种主要颜色光谱直方图表示法(Major color spectrum histogram representation)来构建外观模型,并给出了不同摄像机间基于这种直方图表示法的两个外观模型的相似度匹配估计算法。而且他们还进一步利用基于累积颜色直方图的固定且数据独立的光强度转变算法和对相邻的 3-5 帧图像建立增量的主要颜色光谱直方图表示模型(Incremental Major Color Spectrum Histogram Representation)来克服光强度的变化。除此之外,在利用颜色特征方面,还有许多科研人员提出了一些有效的算法。例如:Orwell 等人[Orwell et al,1999]采用两种方法对目标的颜色直方图进行建模,一种是混合高斯参数化和交叉熵距离测量相结合的方法,另一种则是最大熵编码方法,实验表明这两种方法都具有不错的效果。Bowden 和 KaewTraKulPong [Bowden and KaewTraKulPong,2005]先把目标的颜色直方图进行芒赛尔颜色空间的一致颜色转换(Consensus-Colour Conversion of Munsell colour space) [Sturges and Whitfield,1995],然后进行匹配。该转换是基于人们的视觉效果而对应的一种粗糙的量化过程,对跨摄像机的颜色变化有一定鲁棒性。然而,芒赛尔颜色空间的一致颜色转换是假定摄像机之间对同一颜色有相似的颜色响应的前提之下,然而即便是相同型号的摄像机在光照条件变化的情况下,同一对象的颜色也会存在不同的颜色响应。因此,Gilbert 和 Bowden[Gilbert and Bowden,2005;Gilbert and Bowden,2006]利用对颜色直方图转移矩阵的估计来对颜色进行校准以提高跨摄像机的人体目标跟踪的准确性。Park 等人[Park et al,2006]把目标行人检测出来然后将目标的前景信息从上到下分为三个部分,把下面两部分的颜色作为跟踪的颜色特征,再结合人体目标的高度等特征来参与匹配跟踪。在光照条件变化引起的视频序列图像的颜色变化的问题上,Porikli 和 Divakaran [Porikli and Divakaran,2003]提出了一种基于距离度量和模型函数来实现跨相

机间的颜色校正。在颜色校正过程中,采用相关矩阵分析和动态规划的方法来实现的。通过颜色直方图计算得到相关矩阵,然后利用该矩阵获取最小代价路径来描述转移模型函数。该方法比直接采用距离量度法的优点是能够准确弥补颜色误匹配。Javed 等人 [Javed et al,2003;Javed et al,2008;Javed et al,2005]利用同一人体目标在不同摄像机下拍摄的颜色直方图建立亮度转变函数 BTF (Brightness Transfer Function) 以达到颜色校正的目的,亮度转变函数依赖于一个低维的子空间,他们使用概率主成分分析 PPCA (Probabilistic Principal Component Analysis) [Tipping and Bishop,1999]的方法来实现降维计算。同样 Jeong 和 Jaynes [Jeong and Jaynes,2008]采用颜色转变函数来进行跨摄像机的颜色校正,该方法不再利用目标的亮度信息进行校正而是利用色度信息进行校正的,他们的算法具有更好的匹配效果。

目标匹配是无重叠视场多摄像机跟踪问题的研究基础,匹配算法的效果会直接影响到整个持续跟踪算法的性能,所以,这部分研究是主要的研究方向之一,也是目前的研究热点。

②拓扑关系估计

随着智能视频监控区域的不断扩大,单个摄像机的智能监控已经不能满足人们广域监控的需要,这样多个摄像机之间的联合监控也就应运而生。那么如何自动确定多个摄像机之间的目标出入位置关系,事件连接关系和转移时间关系等拓扑关系已经成为一个重要问题。因此,自动地获取监控摄像机之间的拓扑关系就成为广域视频跟踪研究的一个重要组成部分。Javed 等人 [Javed et al,2003;Javed et al,2008;Javed et al,2005]提出了一个基于 Parzen 窗和高斯核的方法来估计空时概率密度函数,该空时概率密度函数估计了进出视场的位置、出口视场时的运动速度和进出口之间的时间间隔等特征向量的密度函数,整个估计过程需要通过对训练集数据进行学习的方法来实现。Gilbert 等人 [Gilbert and Bowden,2005;Gilbert and Bowden,2006]用一种非监督累加学习的方法自动获取每个摄像机中的人体目标进出口的位置,估计不同摄像机之间的时空关系,由于完全的自动性和非监督性,使得该方法更加适应环境变化,然而该方法的前提条件是假定人体目标在行走的过程中是正常行走,在给定的时间内必须走到某个视场而且要求行走路线也是相对比较固定的,这是该方法的不足之处。Black, Ellis 和 Makris 等人 [Black et al,2005;Ellis et al,2003;Makris et al,2004]利用大量的目标观测数据,通过非监督学习的方法,为一个多摄像机监控网络自动地建立起摄像机之间的时间空间拓扑关系;在此基础上,他们还

给出了验证算法性能的方法并实现了目标在该网络中的跟踪。Niu 和 Tieu 等人[Niu and Grimson,2006;Tieu et al,2005]进一步推广了 Black 和 Makris 等人[Black et al,2005;Makris et al,2004]的方法,利用一个关于统计信任的更一般的信息论思想,把不确定对应和贝叶斯方法相结合,减少了假设条件并体现了更好的性能。Van den Hengel 等人[van den Hengel et al,2006]先假定所有的摄像机的视场都存在着潜在的连接关系,然后去掉那些与观测运动矛盾的拓扑连接,实验证明这种方法能够有效地学习大规模摄像机网络的动态拓扑结构,尤其在学习样本数据量较少的情况下。Rahimi 等人[Rahimi et al,2004a;Rahimi et al,2004b]描述了一种同时获取目标轨迹和无重叠视场的多摄像机的外部参数。为了标定无重叠视场的多摄像机的外部参数,该方法要求每个摄像机的平面坐标系与全局平面坐标系一致。应用运动人体目标的动态先验知识来弥补无重叠视场多摄像机的缺失。该方法允许人体目标以任意速度自由运动。该方法对人体运动轨迹和摄像机参数进行最大后验概率估计。人体运动轨迹的先验概率模型是通过一个线性的高斯马尔科夫链模型来实现的。对目标的位置和速度建模之后,他们用一个非线性最优化的算法计算最相似的目标轨迹,最后得到目标在监控环境下完整的运动轨迹。如果能够建立起一个较好的多摄像机之间的时间和空间拓扑关系,那么就可以利用时空拓扑关系进行约束错误的跟踪情况以提高广域视频跟踪算法准确性和鲁棒性。

③ 数据关联

为了提高无重叠视场多摄像机跟踪的准确性和鲁棒性,在很多情况下,目标的匹配算法和摄像机之间拓扑关系并非独立使用,而是将摄像机的拓扑关系与目标表现模型的匹配算法结合起来。如何将二者的数据有机的关联起来,这个问题实质上就是一个信息融合问题。Kettner 和 Zabih [Kettner and Zabih,1999a;Kettner and Zabih,1999b]建立了一个基于贝叶斯估计的匹配方法来跟踪走廊中活动的人体目标,他们建立了一个由 4 个摄像机组成的无重叠视场监控系统,在这个系统中,目标对象的信息既包括人体目标的外观信息又包括人体目标可能选择的路径信息。他们通过获取当前的目标在已知观测数据目标集合中的最大后验概率的方法来实现人体目标的跟踪,在建立模型之后,他们又给出了如何将最大后验概率估计问题转化成了一个线性规划问题。Javed 等人[Javed et al,2003;Javed et al,2008;Javed et al,2005]不但更加深入细致地给出了表现模型和拓扑关系的估计方法(在前面已经介绍),而且使用有向图的方法计算最优解,计算过程同样符合多项式时间的要求。另外,他们的方法不需要事先标定摄像机,对目标的运动方向也没

有特殊要求,所有这些信息都通过对训练数据集的监督学习得到。[Dick and Brooks,2005]通过减背景法把前景运动目标提取出来,然后利用一个随机转移矩阵来描述被观察的目标人体的运动模式,最后建立一个马尔科夫模型实现了人在不同摄像机之间的实时跟踪。某一特殊环境的模型参数会都是通过观察学习该环境内的一个来回运动的人体目标获得的。Leoputra 等人[Leoputra et al,2006]利用事先已知的人体目标可能经过的轨迹信息地图,此地图包括室内分布的细节和室与室之间的路径信息。当人体目标离开了室内的监控视场,利用地图信息的粒子滤波技术对无重叠视场区域的人体目标进行运动预测。当人体目标再次出现室内监控视场时,该文章利用粒子滤波和直方图匹配权值关联的方法来实现人体目标跟踪。Loke 等人[Loke et al,2006]则采用基于模糊逻辑理论的多摄像机非重叠场景的刚体和非刚体目标跟踪。该方法使用目标的运动、形状和摄像机的拓扑关系等信息进行跟踪。目标的运动和形状信息是通过单个摄像机的 CONDENSATION 算法[Isard and Blake,1996]和 CAMShift 算法[Bradski,1998]跟踪获取的。摄像机的拓扑关系是通过将每个视图映射到共同地平面得到的。该文章通过特征和拓扑关系的融合在一定程度上提高了跨摄像机人体目标跟踪的准确性。

1.4 本文研究的主要工作和组织结构

本文主要研究单摄像机和跨摄像机遮挡人体目标跟踪。单摄像机人体目标跟踪是跨摄像机人体目标跟踪的基础,只有单摄像机人体目标跟踪正确的前提下,跨摄像机跟踪才会有正确的跟踪结果。单摄像机人体目标跟踪主要有两种方法,一种是无背景检测的跟踪方法,一种是有背景检测的跟踪方法;本文的跨摄像机跟踪不是为了解决遮挡问题而对同一视场进行拍摄,而是为了实现扩展视场,所以跨摄像机跟踪分为有公共视场人体目标跟踪和无公共视场人体目标跟踪。

全文共六章,安排如下:

第一章为绪论,介绍了本论文的研究背景和意义,总结了当前的国内外现状及所面临的困难,概述了单摄像机人体目标跟踪算法和多摄像机人体目标跟踪算法。最后,给出了本文研究的主要工作和组织结构。

第二章对无背景检测的单摄像机多目标遮挡人体进行跟踪。本章首先阐述了单目标的贝叶斯滤波的蒙特卡罗实现过程,然后推导出多目标多模型增强的蒙特卡罗滤波算法。为了解决传统粒子滤波跟踪多目标的缺点,本章提出了一种非参数复合模型的方法来描述目标的人体目标的分布,该复合模型包括预测和更新两个阶段,同时提出了在该模型的情况

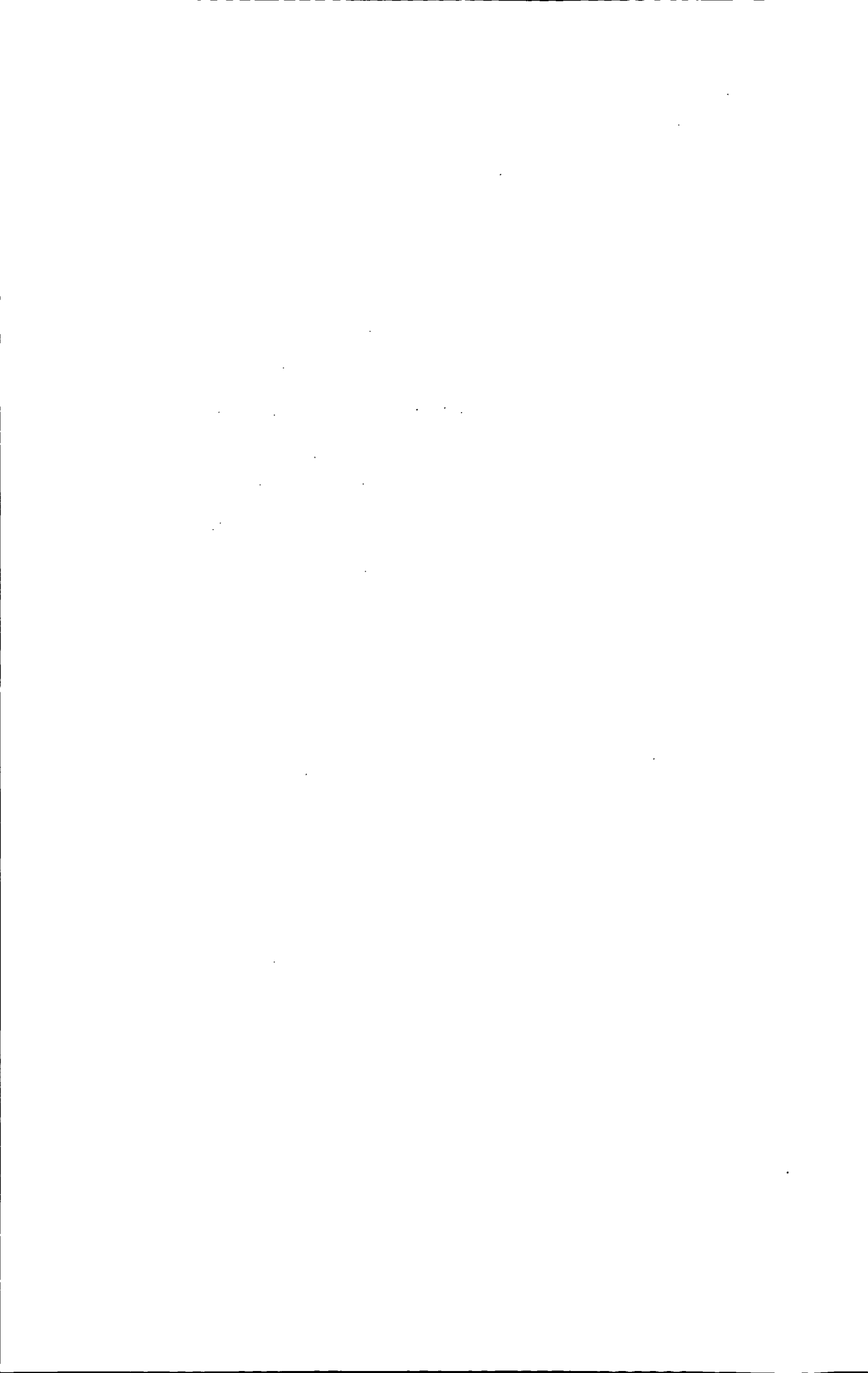
下如何实现递归跟踪。在推导通用递归模型过程中,产生了复合的粒子滤波,通过它们复合权值的大小来衡量粒子的贡献大小。通过两组实验结果证明了该方法可以很好地维持多目标分布的多模块跟踪

第三章采用基于背景检测的单摄像机多目标遮挡人体跟踪。本章充分利用空间信息、颜色信息、运动信息对图像进行块建模、颜色建模和运动建模。首先通过混合高斯建模法,将运动人体的前景信息提取出来;然后基于核密度梯度估计算法,对存储模型中的人体进行聚类,实现块建模;其次利用非参数的核密度估计算法获取颜色密度函数;紧接着利用基于高斯分布的运动建模获取运动密度函数,最后利用颜色密度函数和运动密度函数对当前帧的前景区域进行后验概率估计,以获取后验概率图像,根据该图像遮挡情况的人体进行分割。另外本文介绍了如何提高该算法的运行效率以便更好地将该算法应用于实际中。在实验结果中本文用多组视频对该算法进行测试并进行了比较。

第四章主要解决有公共视场的多摄像机人体目标跟踪。本章首先确定跨摄像机的公共视场分界线,然后利用投影变换单应矩阵计算跨摄像机人体目标的对应关系,然而在实际应用中,视频并不是完全同步或是有遮挡情况,所以在投影变换单应矩阵的基础上采用跨摄像机最大后验概率进行匹配。

第五章主要对无公共视场的多摄像机人体目标进行跟踪。本章在低维度的子空间上对已知的跨摄像机人体目标进行训练,得到颜色转变函数。该方法不用依赖摄像机的内参来计算颜色转变函数的子空间。对颜色转变函数的子空间用基于概率的主要成分分析法对其进行建模得到概率密度函数,然后利用该概率密度函数获取跨摄像机的人体目标的匹配概率。本章在室内外,全部室外环境下对该方法进行了验证并对实验结果进行了分析。

第六章对全文工作进行了总结,并对今后的研究工作进行了展望。



2 无背景检测的单摄像机多目标遮挡人体跟踪

2.1 引言

人体目标跟踪本质就是通过概率估计来判断当前的人体目标在下一帧所处的状态。在基于贝叶斯估计的框架结构中,实际上就是获取在下一帧的很多状态中属于当前人体目标概率最大的状态,即目标状态的最大后验概率。目标状态的最大后验概率取决于状态与观测的似然概率及先验概率。状态的先验概率是通过先验知识学习得到的,似然概率则是通过对状态转移模型和状态测量的观测模型进行估计得到的。状态转移模型描述的是状态随时间的变化情况,状态测量的观测模型是状态转移后测量值。在人体目标跟踪过程中,状态量就是指人体目标的位置信息,观测量就是指状态量所对应的颜色信息。

在过去的几年里,粒子滤波算法得到了广泛的应用。在不同领域,关于粒子滤波算法的表述多种多样,Isard等提出的条件密度传播算法 Condensation (Conditional density propagation) [Isard and Blake,1998;Koller-Meier and Ade,2001]、Gordon等人[Gordon et al,1993]提出的自举滤波算法 (Bootstrap Filter)、Kitagawa等人[Kitagawa,1996]提出的序列蒙特卡罗滤波方法 (Sequential Monte Carlo filter),交互粒子近似(Interacting particle Appropriations)、metropolis-hasting 重要性重采样滤波 (MHIRF)。从本质上说,粒子滤波采用的是序列蒙特卡罗方法(Monte Carlo methods),在序列蒙特卡罗方法中,很重要的两个过程就是重要性采样 IS (Important Sampling) 和序列重要性采样 SIS (Sequential Important Sampling)。

粒子滤波(Particle Filter)算法是一种很实用的算法,它的主要原理就是通过蒙特卡罗法获取贝叶斯概率模型,粒子就是抽取的状态模型,利用一个粒子集来描述概率模型。其主要思路是在后验概率模型中按照一定的分布规律来抽取状态粒子,并用抽取得到的粒子来描述目标。换言之,粒子滤波就是在一组随机的状态样本中获取最优的概率密度,用离散求和运算代替连续积分运算,最终找到方差最小的状态模型进行近似估计。当粒子样本数量趋于无穷大时可以很好地估计任何形式的概率密度分布。

在跟踪情况下,观测值在视频序列中是不断更新的,所以状态后验概率也将随着观测值不断更新,这样才能准确的估计出目标的状态。在应用贝叶斯理论来估计目标状态时,如果已知目标模型和噪声是服从线性和高斯分布的,就可以得到一组最优的解析状

态解,该方法就是卡尔曼滤波算法[Gelb,1974]。然而在实际中,目标模型和噪声可能不是线性和高斯分布的,而是比较复杂非线性、非高斯分布,那么就很难得到一组最优的解析状态解。所以在这种情况下,人们获取的只是各种贝叶斯估计的近似解。求取贝叶斯估计近似解的方法有扩展卡尔曼滤波算法[Jazwinski,1970]、混合高斯滤波算法[Alspach and Sorenson,1972]、基于网格的滤波算法[Bucy and Senne,1971;Julier et al,1995]、Unscented 卡尔曼滤波算法[Julier and Uhlmann,1996;Julier and Uhlmann,1997;Julier et al,1995;Julier et al,2004;Van Der Merwe and Wan,2001]。

当前粒子滤波主要有三种模型:马尔可夫链蒙特卡罗(MCMC)[Chen et al,2001],无迹粒子滤波器(UKF)和 Rao-Blackwellised 粒子滤波器(RBPF)。

(1) 马尔可夫链蒙特卡罗改进策略

马尔可夫链蒙特卡罗[Walsh,2004]基本思想是构造一条具有指定平稳分布的马尔可夫链,即它的转移分布收敛到它的后验分布。然后充分长地运行此链,直到链上取值分布与其平稳分布足够接近时,把来自链上的样本作为指定平稳分布样本,产生来自目标分布的具有很好的收敛样本,基于这些样本进行各种统计推断,求均值、方差等。在每次的样本重要性采样中,采用马尔可夫链法使粒子移动,如果跟踪目标的粒子出现了退化现象,可采用重采样的方法以避免退化现象,提高跟踪过程的准确性。同时,马尔可夫链法能够寻找到最优的粒子的状态概率密度函数,这样,样本的分布会比较有代表性。比较有效的改进的马尔可夫链蒙特卡罗法为: Gibbs 采样器和 MetropolisHasting 方法[Kass et al,1998]。

(2) 无迹粒子滤波器

无迹粒子滤波器[Van Der Merwe et al,2001]是目前被广泛应用的一种目标跟踪算法,能够很好地解决跟踪中的非线性、非高斯问题,无迹粒子滤波器算法是利用无迹卡尔曼滤波器(UKF)产生重要分布函数来改进粒子滤波,从而得到更为接近真实结果的后验概率分布,估计精度更高。无迹卡尔曼滤波器是 Julier 等人提出的。扩展卡尔曼滤波(EKF)使用一阶泰勒展开式逼近非线性项,用高斯分布近似状态分布。无迹卡尔曼滤波器类似于扩展卡尔曼滤波,用高斯分布逼近状态分布,但不需要线性化只使用少数几个称为 Sigma 点的样本。这些点通过非线性模型后,所得均值和方差能够精确到非线性项泰勒展开式的二阶项,从而对非线性滤波精度更高。

(3) Rao-Blackwellised 粒子滤波器

Rao-Blackwellised 粒子滤波器是一种混合滤波器。在高维状态空间采样时一般的粒子滤波的效率很低,对某些状态空间模型,状态向量的一部分在其余部分的条件下的后验分布可以用解析方法求得,例如某些状态是条件线性高斯模型,可用卡尔曼滤波器得到条件后验分布,对另外部分状态用粒子滤波,从而得到一种混合滤波器—Rao-Blackwellised 粒子滤波器。该滤波器可以降低粒子滤波采样空间的维数。

粒子滤波是一种未指定具体表示形式的框架结构,可以灵活地采用不同的状态模型、观测模型或者多种观测模型融合;可以灵活地选择灰度、轮廓、颜色等不同的特征;可以选择灵活地的运动形式,从而根据具体问题采用具体的策略。

2.2 贝叶斯滤波的蒙特卡罗实现

2.2.1 最优贝叶斯估计

本节主要讲述单目标的粒子滤波的跟踪过程,然后在将单目标粒子滤波的跟踪过程推广到多目标粒子滤波跟踪当中。随着时间的推移,视频序列中的人体目标在整个运动过程中可以抽象为数学模型中的两个随机过程:运动随机过程 $(X_t) \in R^{n_x}$ 和观测随机过程 $(Y_t) \in R^{n_y}$ 。这个随机过程是由离散时间的若干个随机变量组成的。任意时刻的随机向量 X_t, Y_t 所对应的向量值分别为 x_t, y_t 。跟踪问题的动态时变系统可以用运动方程(2.1)和观测方程(2.2)来描述。

$$\text{运动方程: } x_t = f_t(x_{t-1}, v_t) \quad (2.1)$$

$$\text{观测方程: } y_t = h_t(x_t, w_t) \quad (2.2)$$

上式中, $f_t: R^{n_x} \times R^{n_v} \rightarrow R^{n_x}$ 和 $h_t: R^{n_x} \times R^{n_w} \rightarrow R^{n_y}$ 可以是线性函数也可以是非线性函数, v_t 代表的含义为运动噪声, w_t 代表的含义为观测噪声,运动噪声和观测噪声是两组相互独立且不相关的序列, n_x 表示状态向量的维数, n_y 表示观测向量的维数, n_v 表示运动噪声向量维数, n_w 表示观测噪声向量的维数。

采用贝叶斯法的跟踪实质就是通过不同时刻的得到的观测颜色信息 $y_{0:t} = \{y_0, \dots, y_t\}$ 来估计 t 时刻目标的状态向量 x_t 的条件概率值 $p(x_t | y_{0:t})$, 若已知初始状态的密度函数为 $p(x_0 | y_0) = p(x_0)$, 则 $p(x_t | y_{0:t})$ 可以通过状态预测方程(2.3)状态更新方程为(2.4)递推得

到。

状态预测方程为:

$$p(\mathbf{x}_t | \mathbf{y}_{0:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1}) d\mathbf{x}_{t-1} \quad (2.3)$$

状态更新方程为:

$$p(\mathbf{x}_t | \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{p(\mathbf{y}_t | \mathbf{y}_{0:t-1})} \quad (2.4)$$

其中,

$$p(\mathbf{y}_t | \mathbf{y}_{0:t-1}) = \int p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t \quad (2.5)$$

公式(2.3)(2.4)(2.5)是从理论的角度推导出贝叶斯估计的概率模型,但式(2.3)中的积分项只能对高斯分布与线性系统获得解析解,对于非高斯非线性的系统,就无法获取准确的积分值,这样,科研人员只能通过近似估计的方法来获取近似解,因而通过采样来实现的粒子滤波估计的方法就应需而生。

2.2.2 重要性采样

基于随机采样运算的蒙特卡罗方法就是将积分运算转化为有限样本点的求和运算。

对于任意 $f_t(\mathbf{x}_{0:t})$ 数学期望可表示为式子(2.6),如果求取 $\mathbf{x}_{0:t}$ 条件均值那么 $f_t(\mathbf{x}_{0:t}) = \mathbf{x}_{0:t}$,

如果求取 $\mathbf{x}_{0:t}$ 的条件方差,则 $f_t(\mathbf{x}_{0:t}) = \mathbf{x}_t \mathbf{x}_t^T - E_{p(\mathbf{x}_t | \mathbf{y}_{0:t})}[\mathbf{x}_t] E_{p(\mathbf{x}_t | \mathbf{y}_{0:t})}^T[\mathbf{x}_t]$ 。

$$E[f_t(\mathbf{x}_{0:t})] = \int f_t(\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t} \quad (2.6)$$

但在实际应用中由于 $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 可能是非高斯、非线性的概率分布,因而要用采样的方法获取近似解。因此,要引入一个概率密度分布与 $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 相同且已知同时容易从中采样的重要性函数 $q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 。将重要性函数 $q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 代入式(2.6)可得到式(2.7)。

$$\begin{aligned}
E[f_t(\mathbf{x}_{0:t})] &= \frac{1}{p(\mathbf{y}_{0:t})} \int f_t(\mathbf{x}_{0:t}) \frac{p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t}) p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})} q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t} \\
&= \frac{\int f_t(\mathbf{x}_{0:t}) w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t}}{\int p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}) \frac{q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})}{q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})} d\mathbf{x}_{0:t}} \\
&= \frac{\int f_t(\mathbf{x}_{0:t}) w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t}}{\int w_t(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t}} \\
&= \frac{E_{\pi(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})} [f_t(\mathbf{x}_{0:t}) w_t(\mathbf{x}_{0:t})]}{E_{\pi(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})} [w_t(\mathbf{x}_{0:t})]}
\end{aligned} \tag{2.7}$$

其中,

$$w_t(\mathbf{x}_{0:t}) = \frac{p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t}) p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})} \tag{2.8}$$

因此, 对状态后验概率密度分布 $q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 中抽取 N 个相互独立且同分布的样本 $\{\mathbf{x}_{0:t}^{(i)}\}_{i=1}^N$, 则式(2.7)中的 $E[f_t(\mathbf{x}_{0:t})]$ 的就可以式(2.9)近似表示, 该方法就是著名的蒙特卡罗估计。

$$\hat{E}[g(\mathbf{x}_{0:t})] = \frac{\frac{1}{N} \sum_{i=1}^N g(\mathbf{x}_{0:t}^{(i)}) w_t(\mathbf{x}_{0:t}^{(i)})}{\frac{1}{N} \sum_{i=1}^N w_t(\mathbf{x}_{0:t}^{(i)})} = \sum_{i=1}^N f_t(\mathbf{x}_{0:t}^{(i)}) \tilde{w}_t(\mathbf{x}_{0:t}^{(i)}) \tag{2.9}$$

其中,

$$\tilde{w}_t(\mathbf{x}_{0:t}^{(i)}) = \frac{w_t(\mathbf{x}_{0:t}^{(i)})}{\sum_{i=1}^N w_t(\mathbf{x}_{0:t}^{(i)})} \tag{2.10}$$

$\tilde{w}_t(\mathbf{x}_{0:t}^{(i)})$ 为 $w_t(\mathbf{x}_{0:t}^{(i)})$ 的归一化重要性权值。

2.2.3 序列重要性采样

上节的重要性采样方法不足以用来实现递推估计。由式(2.8)可知, 随着时间推移每增加一个新的观测数据 \mathbf{y}_{t+1} 就要重新计算 $q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 、 $p(\mathbf{x}_{0:t})$ 和 $p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t})$ 的概率, 而不是由 $t-1$ 时刻的重要性权值 $w_{t-1}(\mathbf{x}_{0:t-1})$ 递推求取的。为了能够充分利用 $t-1$ 时刻的重要性权值来实现递推, 科研人员通过采用序列重要性采样的方法来解决上述的难点问题, 该

方法可以不改变 t 时刻之前的状态序列样本集来采用递归的形式计算 $t+1$ 时刻的权值。为了实现递推，我们必须先假定当前状态只依赖于上一时刻的状态值，在状态转移过程中认为服从的是马尔可夫分布，只进行滤波而不考虑平滑，则 $q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$ 、 $p(\mathbf{x}_{0:t})$ 、 $p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t})$ 可分别写成式(2.11)、(2.12)、(2.13)的连乘形式：

$$q(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) = q(\mathbf{x}_0) \prod_{j=1}^t q(\mathbf{x}_j | \mathbf{x}_{0:j-1}, \mathbf{y}_{0:j}) \quad (2.11)$$

$$p(\mathbf{x}_{0:t}) = p(\mathbf{x}_0) \prod_{j=1}^t p(\mathbf{x}_j | \mathbf{x}_{j-1}) \quad (2.12)$$

$$p(\mathbf{y}_{0:t} | \mathbf{x}_{0:t}) = \prod_{j=0}^t p(\mathbf{y}_j | \mathbf{x}_j) \quad (2.13)$$

将(2.11)、(2.12)、(2.13)代入式(2.8)得到式(2.14)。

$$\begin{aligned} w_t(\mathbf{x}_{0:t}) &= \frac{p(\mathbf{y}_{0:t-1} | \mathbf{x}_{0:t-1}) p(\mathbf{x}_{0:t-1})}{q(\mathbf{x}_{0:t-1} | \mathbf{y}_{0:t-1})} \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1})}{q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})} \\ &= w_{t-1}(\mathbf{x}_{0:t-1}) \frac{p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1})}{q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})} \end{aligned} \quad (2.14)$$

在给定重要性分布函数 $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})$ 的条件下，根据式(2.14)可以实现用递推的方法来计算各个状态粒子的权值。重要性分布函数 $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})$ 的选择是一个非常关键的问题，选取原则之一就是使得重要性权值的方差最小。Liu 和 Chen [Liu and Chen,1998]指出：如果将 $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) = p(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t})$ ，那么就可以满足重要性权值的方差最小，然而，这种算法在实际应用中实现是比较困难的。为了简单实用，人们通常采用 $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$ ，这种方法得到了广泛的应用。

2.2.4 重采样

在上节提到的序列重要性采样存在一个严重缺陷就是蜕化或权值蜕化问题[Doucet et al,2000;Doucet et al,2001]，即重要性权值的方差随着时间是递增的[Herman,2002;Kong et al,1994]。直接的表现是经过多次迭代后，仅有某一个归一化重要权值趋于1，其它权值都趋于0，其直接后果是，关于粒子集中的绝大多数粒子的计算由于其权值过小而意义甚微。Kong 等人[Kong et al,1994]首先引入了有效样本容量的概念来度量权值蜕化，但该方法难以实际计算。目前常用的准则就是根据有效粒子数。其估计式为

$$N_{eff} = \frac{1}{\sum_{i=1}^N (\tilde{w}_i^{(t)})^2} \quad (2.15)$$

其中, $\tilde{w}_i^{(t)}$ 为归一化重要性权值。可以预先设计一个门限 $N_{threshold}$, 当 $N_{eff} < N_{threshold}$ 时, 则表示蜕化比较严重, 这时候就要用到重采样 [Carpenter et al,1999;Gordon et al,1993;Haug,2005] 方法, 也有文献称为选择 [Van Der Merwe et al,2001] 或再生 [Lee and Chia,2002] 方法的。重采样算法在 1993 年由 Gordon 等人 [Gordon et al,1993] 提出, 其基本思想是抑制或剔除小权值粒子, 对于大权值粒子则依其权值大小进行复制, 从而把处理资源按照粒子权值的大小进行分配; 这有点类似于遗传算法 [Tang et al,1996] 的适者生存原理。重采样算法中, 首先产 N 个满足 $[0,1]$ 均匀分布的随机数 $u_i \in U[0,1]$, 然后通过搜索算法找到满足如下条件的整数 m , 使得 $\sum_{i=1}^{m-1} \tilde{w}_i^j \leq u_i \leq \sum_{i=1}^m \tilde{w}_i^j$ 将粒子 $\tilde{\mathbf{x}}_k^M$ 作为新样本集的样本点, 即将区间 $[0,1]$ 按照 $\lambda^i = \sum_{j=0}^i \tilde{w}_j^i$ 分成 N 个小区间, 当随机数 u_i 落在第 m 个区间时, 对粒子 $\tilde{\mathbf{x}}_k^m$ 进行复制。设定重采样之前的序列为 $\{\mathbf{x}_i^j\}_{i=1}^N$ 和重采样产生的序列 $\{\tilde{\mathbf{x}}_i^j\}_{i=1}^N$, 它们之间理论上满足: $P(\tilde{\mathbf{x}}_i^j = \tilde{\mathbf{x}}_i^i) = w_i^j$ 重采样后粒子集中各粒子的权值均为 $1/N$ 。重采样的方法有多种, 主要有: 多项式重采样 [Chen,2003], 残差重采样 [Liu and Chen,1998], 系统重采样(systematic Resampling)[Arulampalam et al,2002]。

2.2.5 基本滤波算法描述

基于蒙特卡罗方法的粒子滤波, 主要包括初始化、重要性采样、更新权值、重采样输出最终估计的状态等过程, 下面我们将对以上的基本理论进行具体详述实现过程, 给出使用粒子滤波进行人体目标跟踪的一般过程。

$$\text{初始化} \quad \begin{cases} \mathbf{x}_0^i \sim p(\mathbf{x}_0) \\ w_0^i = 1/N \end{cases} \quad i=1, \dots, N$$

for $t=1, \dots, T$

重要性采样

$$\text{For } i=1, \dots, N, \text{ 根据 } \tilde{\mathbf{x}}_i^{(t)} \sim q(\tilde{\mathbf{x}}_i^{(t)} | \tilde{\mathbf{x}}_{i-1}^{(t)}, \mathbf{y}_t) \text{ 采样新粒子 } \tilde{\mathbf{x}}_i^{(t)}$$

更新权值

$$\text{计算未归一化权值 } w_t^{(i)} = w_{t-1}^{(i)} \frac{P(y_t | \tilde{\mathbf{x}}_t^{(i)}) P(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_t)}$$

$$\text{计算归一化权值 } \tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$$

$$\text{计算 } N_{\text{eff}} = \frac{1}{\sum_{i=1}^N (\tilde{w}_t^{(i)})^2}$$

If $N_{\text{eff}} < N_{\text{threshold}}$ 从 $\{\tilde{\mathbf{x}}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i=1}^N$ 重采样得到新的粒子集 $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N$

else $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N = \{\tilde{\mathbf{x}}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i=1}^N$

输出状态估计

$$\hat{\mathbf{x}} = \sum_{i=1}^N \tilde{w}_t^{(i)} \mathbf{x}_t^{(i)}$$

通过以上步骤可以实现基于粒子滤波目标跟踪。上述方法只是一个通用的模型，在不同的场合，要根据不同的情况而具体考虑需要选择服从什么规律的状态概率分布、重采样阈值、观测模型等。基本的粒子滤波算法还有很多问题存在，仍需继续研究。当前，科研人员已经提出了许多改进的粒子滤波算法，在一定程度上提高粒子滤波的鲁棒性与可靠性。该方法主要有：局部线性化方法、辅助变量法、进化滤波法等。通过此类方法来保持粒子的多样性。该内容将作为后续几节多目标多模型粒子滤波算法的基础。

2.3 多目标多模型的粒子滤波

2.3.1 多目标跟踪模型

很多文献讲述的粒子滤波都是单目标粒子滤波过程，本文将介绍一种多目标的增强粒子滤波。本节内容是在单目标粒子滤波的基础上得到的多目标粒子滤波模型。

为了获取多目标的粒子滤波模型，本文用一个权重向量 $\Pi = \{\pi_m\}_{m=1}^M$ 来描述 M 个目标的粒子滤波模型的复合更新权值。 π_m 用来表示当前某个目标与每个人体目标的更新状态概率所占的比重。用式(2.16)来估计当前每个目标的复合更新状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t})$ 。

$$P(\mathbf{x}_t | \mathbf{y}_{0:t}) = \sum_{m=1}^M \pi_{m,t} P_m(\mathbf{x}_t | \mathbf{y}_{0:t}) \quad (2.16)$$

其中, $\sum_{m=1}^M \pi_{m,t} = 1$, $p_m(\mathbf{x}_t | \mathbf{y}_{0:t})$ 表示第 m 个分量的状态更新概率, 这样复合更新状态概率

$P(\mathbf{x}_t | \mathbf{y}_{0:t})$ 可以通过 m 个分量的状态更新概率加权求和获得。单目标的蒙特卡罗粒子滤波采用的是递归算法来求解, 如何采用与单目标相似的递归算法来求取多目标的蒙特卡罗粒子滤波将是接下来要解决的问题, 即如何根据 $t-1$ 时刻第 m 个分量的权值参数 $\pi_{m,t-1}$ 和状态预测概率 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})$ 或状态更新概率 $p_m(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1})$ 求取式(2.16)中 t 时刻第 m 个分量的权值参数 $\pi_{m,t}$ 和状态更新概率 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t})$ 来实现递归算法。

首先假设 $t-1$ 时刻的复合更新状态概率 $P(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1})$ 是已知的, 将该概率代入预测状态概率式(2.3)可得复合预测状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t-1})$:

$$\begin{aligned} P(\mathbf{x}_t | \mathbf{y}_{0:t-1}) &= \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) P(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1}) d\mathbf{x}_{t-1} \\ &= \sum_{m=1}^M \pi_{m,t-1} \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p_m(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1}) d\mathbf{x}_{t-1} \\ &= \sum_{m=1}^M \pi_{m,t-1} P_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) \end{aligned} \quad (2.17)$$

$p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p_m(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1}) d\mathbf{x}_{t-1}$ 表示第 m 个分量的状态预测概率是由第 m 分量 $t-1$ 时刻的状态转移概率 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ 和状态更新概率 $p_m(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1})$ 积分得到的。根据式(2.17)可知, 先通过获取每个分量的状态预测概率, 然后将其加权求和即可求得复合预测状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t-1})$ 。

其次, 用复合预测状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t-1})$ 来表示复合更新状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t})$, 这样将复合预测状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t-1})$ 代入式(2.4)得到式(2.18), 即复合更新状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t})$ 。

$$\begin{aligned} P(\mathbf{x}_t | \mathbf{y}_{0:t}) &= \frac{\sum_{m=1}^M \pi_{m,t-1} p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{\sum_{n=1}^M \pi_{n,t-1} \int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})} \\ &= \frac{\sum_{m=1}^M \pi_{m,t-1} \int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t}{\sum_{n=1}^M \pi_{n,t-1} \int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})} \frac{p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{\int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t} \end{aligned} \quad (2.18)$$

与式(2.16)进行对比得出式(2.19)、(2.20)。

$$p_m(\mathbf{x}_t | \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})}{\int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t} \quad (2.19)$$

$$\pi_{m,t} = \frac{\pi_{m,t-1} \int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t}{\sum_{n=1}^M \pi_{n,t-1} \int p(\mathbf{y}_t | \mathbf{x}_t) p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1}) d\mathbf{x}_t} = \frac{\pi_{m,t-1} p_m(\mathbf{y}_t | \mathbf{y}_{0:t-1})}{\sum_{n=1}^M \pi_{n,t-1} p_m(\mathbf{y}_t | \mathbf{y}_{0:t-1})} \quad (2.20)$$

由式(2.18)可知, 复合更新状态概率 $P(\mathbf{x}_t | \mathbf{y}_{0:t})$ 是由每个分量的更新状态概率加权复合而成, 这是一个很好的结果, 这就意味着只要分别求出每个分量的权值 $\pi_{m,t}$ 和更新状态概率 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t})$ 就可实现递归滤波。由式(2.19)可知, t 时刻的第 m 个分量的更新概率 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t})$ 可由 $t-1$ 时刻的第 m 个分量的预测概率 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t-1})$ 及概率 $p(\mathbf{y}_t | \mathbf{x}_t)$ 获取, 更新概率 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t})$ 的估计值可以由第 2.2.5 节单目标粒子滤波过程获得。由式(2.20)可知, t 时刻的第 m 个分量的权值 $\pi_{m,t}$ 可由 $t-1$ 时刻的 M 个分量的预测概率 $\pi_{m,t-1}$ 和概率 $p_m(\mathbf{y}_t | \mathbf{y}_{0:t-1})$ 来获取, 只要根据式(2.20)更新复合权值参数 $\pi_{m,t}$, 那么就可用递归方式实现多目标的粒子滤波过程。更新复合权值参数 $\pi_{m,t}$ 是经过归一化后的权值参数。

2.3.2 粒子估计

上节介绍的多目标递归跟踪只有在动态系统和似然概率是线性系统且为高斯分布少数情况下才有闭合解。对于非线性非高斯情况下, 给不出一个理论的终极固定表达式, 一般只能以数值估计方法求解。

近年来, 对于非线性非高斯系统比较常见的解决方案是蒙特卡罗粒子滤波算法。由于该方法具有简单性、概括性和广泛的应用性。蒙特卡罗粒子滤波使用一些可以进行传播的且带有权值的样本粒子代表目标的分布。下面我们将详细介绍如何对多目标实现递归的粒子滤波跟踪。

首先, 我们用向量 $\mathbf{V}_t = \{N, M, \Pi_t, \mathbf{X}_t, \tilde{\mathbf{W}}_t, \mathbf{C}_t\}$ 表示复合粒子滤波分布向量。

其中,

N : 表示 t 时刻共有 N 个粒子;

M : 表示 t 时刻共有 M 个目标;

$\Pi_t = \{\pi_{m,t}\}_{m=1}^M$: Π_t 表示 t 时刻所有归一化后的复合权值参数, $\pi_{m,t}$ 表示 t 时刻第 m 个

目标的归一化后的复合权值参数;

$X_t = \{\mathbf{x}_t^{(i)}\}_{i=1}^N$, X_t 表示所有的采样粒子状态集; $\mathbf{x}_t^{(i)}$ 表示 t 时刻第 i 个粒子的状态;

$\tilde{W}_t = \{\tilde{w}_t^{(i)}\}_{i=1}^N$, \tilde{W} 表示 t 时刻所有归一化后的粒子权值集合; $\tilde{w}_t^{(i)}$ 表示 t 时刻第 i 个粒子的归一化权值。

$C_i = \{c_i^{(j)}\}_{j=1}^M$, C_i 表示任意粒子 i 属于某个人体目标的集合。 $c_i^{(j)} \in \{1, \dots, M\}$, $c_i^{(j)}$ 表示粒子 i 是属于某个人体目标的。复合更新状态蒙特卡罗粒子滤波估计可用式(2.21)表示。

$$\hat{p}(\mathbf{x}_t | \mathbf{y}_{0:t}) = \sum_{m=1}^M \pi_{m,t} \sum_{i \in I_m} \tilde{w}_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t) \quad (2.21)$$

其中 $\delta(\square)$ Dirac Delta 函数, $I_m = \{i | i \in \{1, \dots, N\}, c_i^{(j)} = m\}$ 表示所有属于第 m 个目标的粒子集合。注意: 所有目标的权值参数和为 1 和每个目标的粒子权重和为 1, 即:

$$\sum_{m=1}^M \pi_{m,t} = 1$$

$$\sum_{i \in I_m} \tilde{w}_t^{(i)} = 1, \quad m = 1, \dots, M$$

已知上一时刻的复合粒子滤波分布向量 V_{t-1} , 可以求出当前时刻的复合粒子滤波分布向量 V_t 。由式(2.21)可知, 复合更新状态概率估计中的每个目标分量的更新状态概率估计是独立的进行递推的, 复合更新状态概率估计是通过对所有目标分量的更新状态概率进行加权求和得到的, 因此只要独立求出目标分量的更新状态概率, 即可实现复合更新状态概率估计。下面先对 $t-1$ 时刻的第 m 个目标分量进行考虑, 该目标的所有粒子与归一化权值为 $\{\mathbf{x}_{t-1}^{(i)}, \tilde{w}_{t-1}^{(i)}\}_{i \in I_m}$, 则根据式(2.14)(2.10)可以求得 t 时刻的第 m 个目标分量的粒子为 $\{\mathbf{x}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i \in I_m}$,

其中:

$$\mathbf{x}_t^{(i)} \square q(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_t) \quad \tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{j \in I_m} w_t^{(j)}} \quad w_t^{(i)} = \frac{w_{t-1}^{(i)} p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_t)}$$

t 时刻的第 m 个目标分量粒子 $\{\mathbf{x}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i \in I_m}$ 的更新状态概率为 $p_m(\mathbf{x}_t | \mathbf{y}_{0:t})$ $m = 1, \dots, M$ 。为了获得复合权值参数 $\pi_{m,t}$, 根据公式(2.20)必须先计算分量的似然概率 $p_m(\mathbf{y}_t | \mathbf{y}_{0:t-1})$ 根据

式(2.5), 可以得到分量的似然概率为:

$$\begin{aligned}
 p_m(\mathbf{y}_t | \mathbf{y}_{0:t-1}) &= \iint p(\mathbf{y}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{x}_{t-1}) p_m(\mathbf{x}_{t-1} | \mathbf{y}_{0:t-1}) d\mathbf{x}_t d\mathbf{x}_{t-1} \\
 &\approx \sum_{i \in I_m} \tilde{w}_{t-1}^{(i)} \frac{p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) p(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)} | \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_t)} = \sum_{i \in I_m} w_t^{(i)}
 \end{aligned} \tag{2.22}$$

$$\pi_{m,t} \approx \frac{\pi_{m,t-1} \sum_{i \in I_m} w_{t-1}^{(i)}}{\sum_{n=1}^M \pi_{n,t-1} \sum_{i \in I_n} w_{t-1}^{(i)}} \tag{2.23}$$

2.3.3 遮挡处理

到本节为止, 已给出了理想情况下多目标跟踪的复合滤波的实现模型, 该模型可以像单目标滤波一样进行递推方式向下传递。在理想情况下, 人体目标之间没有遮挡或出现新的人体目标, 然而在实际中, 人体目标数目不可能保持固定不变, 如何处理人体目标数变化情况下, 降低对复合更新状态概率的影响将是本节要解决的问题。

在计算复合更新状态概率时, 必须充分考虑人体目标数目变化, 在发生遮挡时合并复合向量中的分量, 在遮挡后, 并拆分复合向量中的分量。用式子 $(X_t, C_t, M') = \xi(X_t, C_t, M)$ 表示一个空间重新聚类的函数。把当前的粒子集、粒子的属性分量、目标数目作为输入, 然后输出一个新的复合向量, 该复合向量与原来的复合向量可能是相同的或也可能是不同的。函数 ξ 的功能包括合并、拆分、聚类等复杂的操作过程。本文采用 K 均值的方法来实现重新聚类函数 ξ , 对初始样本集合进行聚类, 在每次进行迭代过程中, 我们对重叠或分散比较严重的进行合并聚类或拆分聚类。首先通过运行 K 均值算法进行聚类, 然后用合并或拆分程序进行初始化产生新的复合分量。这种简单的策略允许复合分量随着人体目标的遮挡或分开而变化, 在实际应用中也取得了较好的效果。

接下来, 我们将计算新的复合向量的复合权值 Π_t' 和粒子权值 \tilde{w}_t' , 新的复合估计向量 $V_t' = \{N, M', \Pi_t', X_t, \tilde{w}_t', C_t'\}$ 。对式(2.21)进行变换将很容易获得新的复合估计向量式(2.24)。

$$\begin{aligned}
\hat{p}(\mathbf{x}_t | \mathbf{y}_{0:t}) &= \sum_{m=1}^M \pi_{m,t} \sum_{i \in I_m} \tilde{w}_t^{(i)} \delta_{x_t^{(i)}}(\mathbf{x}_t) \\
&= \sum_{m=1}^N \pi_{c_t^{(i)},t} \tilde{w}_t^{(i)} \delta_{x_t^{(i)}}(\mathbf{x}_t) = \sum_{m=1}^{M'} \sum_{i \in I'_m} \pi_{c_t^{(i)},t} \tilde{w}_t^{(i)} \delta_{x_t^{(i)}}(\mathbf{x}_t) \\
&= \sum_{m=1}^{M'} \pi'_{m,t} \sum_{i \in I'_m} \tilde{w}'_t^{(i)} \delta_{x_t^{(i)}}(\mathbf{x}_t)
\end{aligned} \tag{2.24}$$

其中新的复合权重参数 $\pi'_{m,t}$ 和粒子权重 $\tilde{w}'_t^{(i)}$ 分别用式子(2.25)和(2.26)表示。

$$\pi'_{m,t} = \sum_{i \in I'_m} \pi_{c_t^{(i)},t} \tilde{w}_t^{(i)} \tag{2.25}$$

$$\tilde{w}'_t^{(i)} = \frac{\pi_{c_t^{(i)},t} \tilde{w}_t^{(i)}}{\pi'_{m,t}} \tag{2.26}$$

重新计算新的复合向量 V' 与原来的复合向量 V 有相同的分布, 同时并没有影响粒子滤波的聚类属性。

2.3.4 观测模型

在文献[Perez et al,2002]的基础上, 提出基于 HSV (Hue Saturation Value) 直方图的多颜色观测模型, 由于 HSV 将亮度信息 (Value) 与色度 (Hue) 和饱和度 (Saturation) 区别开来, 将会降低对光强度的敏感性。一个 HSV 直方图是由 $N = N_u N_v + N_w$ 个块组成, 用符号 $b_i(d) \in \{1, \dots, N\}$ 表示 t 时刻位置为 d 时对应的颜色向量 $\mathbf{y}_i(k)$ 转换成 HSV 直方图所对应的块索引值。给定 t 时刻一个候选区域, 将该区域的转成直方图形式为 $R(x_t) = l_t + s_t W$, 在 t 时刻的核密度估计可表示为(2.27), $k(n; \mathbf{x}_t)$ 表示 t 时刻第 n 块的颜色概率。

$$K(\mathbf{x}_t) = \{k(n; \mathbf{x}_t)\}_{n=1}^N \tag{2.27}$$

$$k(n; \mathbf{x}_t) = \lambda \sum_{d \in R(\mathbf{x}_t)} \delta[b_i(d) - n] \tag{2.28}$$

其中, λ 为归一化常量, 该参数的目的是确保 $\sum_{n=1}^N k(n; \mathbf{x}_t) = 1$, d 表示在区域 $R(\mathbf{x}_t)$ 内的任意像素位置。当 $t=0$ 时, 则式(2.27)变为 $K(\mathbf{x}_0) = \{k(n; \mathbf{x}_0)\}_{n=1}^N$, 表示人体目标初始化时参考的核密度估计, 用关于 HSV 彩色直方图的 Bhattacharyya 距离 $\psi[K(\mathbf{x}_0), K(\mathbf{x}_t)]$ (式(2.29))来衡量当前 t 时刻的核密度估计 $K(\mathbf{x}_t)$ 与初始化时参考的核密度估计 $K(\mathbf{x}_0)$ 的相似

程度。

$$\psi[K(\mathbf{x}_0), K(\mathbf{x}_t)] = \left[1 - \sum_{n=1}^N \sqrt{k(n; \mathbf{x}_0)k(n; \mathbf{x}_t)} \right]^2 \quad (2.29)$$

当得到了 HSV 彩色直方图的 Bhattacharyya 距离之后, 我们用式(2.30)来描述似然概率 $p(\mathbf{y}_t | \mathbf{x}_t)$ 。

$$p(\mathbf{y}_t | \mathbf{x}_t) \propto e^{-\rho \psi^2[K(\mathbf{x}_0), K(\mathbf{x}_t)]} \quad (2.30)$$

用 HSV 彩色直方图方法来估计两个区域的颜色的相似程度是一种值得信赖的方法。然而当将颜色分布的空间信息考虑进去后, 我们将会得到更好的颜色密度估计。将当前时刻被跟踪的区域分成 B 个子区域的和 $R(\mathbf{x}_t) = \sum_{j=1}^B R_j(\mathbf{x}_t)$, 同理初始时刻的参考直方图也同样被分成了 B 个子区域, 然后对所有相对应的子块的 Bhattacharyya 距离平方求和, 即可将式(2.30)变为式(2.31)。

$$p(\mathbf{y}_t | \mathbf{x}_t) \propto e^{\sum_{j=1}^B -\rho \psi^2[K_j(\mathbf{x}_0), K_j(\mathbf{x}_t)]} \quad (2.31)$$

在选择重要性函数时, 令重要性函数 $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{0:t}) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$, 因此权值计算公式(2.14)将变为式(2.32):

$$w_t^{(i)} = w_{t-1}^{(i)} p(\mathbf{y}_t | \tilde{\mathbf{x}}_t^{(i)}) \quad (2.32)$$

2.3.5 多目标多模型滤波算法描述

1) 初始化

$$\text{粒子状态和权值: } \begin{cases} \mathbf{x}_0^i \sim P(\mathbf{x}_0) \\ w_0^i = 1/N \end{cases} \quad i=1, \dots, N$$

$$\text{复合更新权值: } \pi_{m,0} = 1/M, \quad m=1, \dots, M$$

2) for $t=1, \dots, T$

a) 重要性采样

For $i=1, \dots, N$, 根据 $\tilde{\mathbf{x}}_t^{(i)} \sim q(\tilde{\mathbf{x}}_t^{(i)} | \tilde{\mathbf{x}}_{t-1}^{(i)}, \mathbf{y}_t)$ 采样新粒子 $\tilde{\mathbf{x}}_t^{(i)}$

b) 计算似然概率

$$p(\mathbf{y}_t | \mathbf{x}_t) \propto e^{\sum_{j=1}^B -\rho \psi^2[K_j(\mathbf{x}_0), K_j(\mathbf{x}_t)]}$$

c) 更新权值

$$\text{计算未归一化权值 } w_t^{(i)} = w_{t-1}^{(i)} \frac{p(y_t | \tilde{\mathbf{x}}_t^{(i)}) p(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{x}_{t-1}^{(i)})}{q(\tilde{\mathbf{x}}_t^{(i)} | \mathbf{x}_{0:t-1}, \mathbf{y}_t)}$$

$$\text{计算归一化权值 } \tilde{w}_t^{(i)} = \frac{w_t^{(i)}}{\sum_{i=1}^N w_t^{(i)}}$$

$$\text{d) 计算复合更新权值: } \pi_{m,t} \approx \frac{\pi_{m,t-1} \sum_{i \in I_m} w_{t-1}^{(i)}}{\sum_{n=1}^M \pi_{n,t-1} \sum_{i \in I_n} w_{t-1}^{(i)}}$$

$$\text{e) 计算粒子的复合更新概率: } \hat{p}(\mathbf{x}_t | \mathbf{y}_{0:t}) = \sum_{m=1}^M \pi_{m,t} \sum_{i \in I_m} \tilde{w}_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t)$$

f) 估计每个目标的输出状态

$$\hat{\mathbf{x}}_{t,m} = \arg_{\mathbf{x}} \max \{ \hat{p}(\mathbf{x}_t^{(i)} | \mathbf{y}_{0:t}) \}_{i \in I_m} \quad m = 1, \dots, M$$

$$\text{g) 计算 } N_{eff} = \frac{1}{\sum_{i=1}^N (\tilde{w}_t^{(i)})^2}$$

h) 重采样: If $N_{eff} < N_{threshold}$ 从 $\{\tilde{\mathbf{x}}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i=1}^N$ 重采样得到新的粒子集 $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N$

$$\text{else } \{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N = \{\tilde{\mathbf{x}}_t^{(i)}, \tilde{w}_t^{(i)}\}_{i=1}^N$$

2.4 实验结果和分析

本文用多目标粒子滤波算法对了两组室外场景下拍摄的视频序列进行粒子滤波跟踪。图 2-1 给出了场景 A 情况下视频的第 294、346、361、378 帧且含有遮挡和交叉情况的原始图像, 图 2-2 给出了图 2-1 所给出的图像的人体目标的粒子及粒子的属性(即所属跟踪的目标), 具有相同颜色的矩形框表示跟踪是同一个目标。图 2-3 是对图 2-2 粒子的权值进行加权平均得到场景 A 多人目标跟踪的最终结果; 图 2-4 给出了场景 B 情况下视频的第 75、94、124、144 帧的原始图像, 同样含有遮挡和交叉情况, 图 2-5 给出了图 2-4 所给出的图像的人体目标的粒子及粒子的属性, 图 2-6 是对图 2-5 粒子的权值进行加权平均得到场景 B 多人目标跟踪的最终结果。实验证明本文提出的多目标粒子滤波算法可以对含有遮挡和交叉的多人目标进行跟踪。



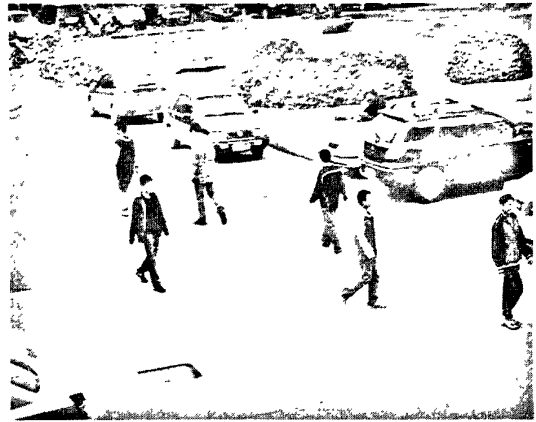
294 帧



346 帧

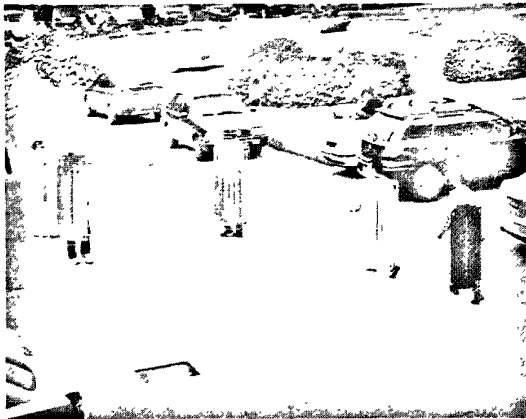


361 帧

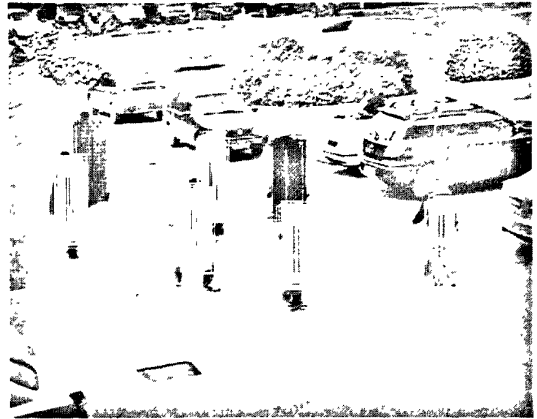


378 帧

图 2-1 场景 A 的原图



294 帧



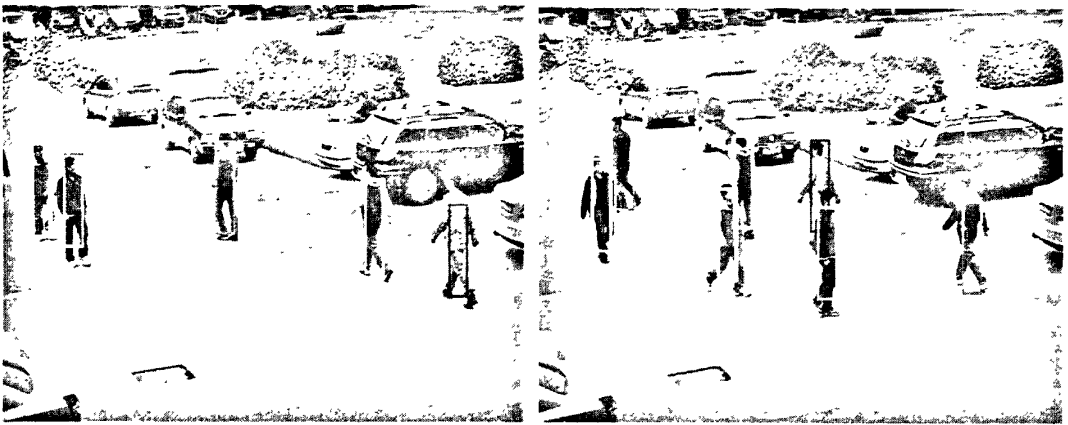
346 帧



361 帧

378 帧

图 2-2 场景 A 图像的人体目标的粒子



294 帧

346 帧



361 帧

378 帧

图 2-3 场景 A 人体目标的跟踪结果



75 帧



94 帧

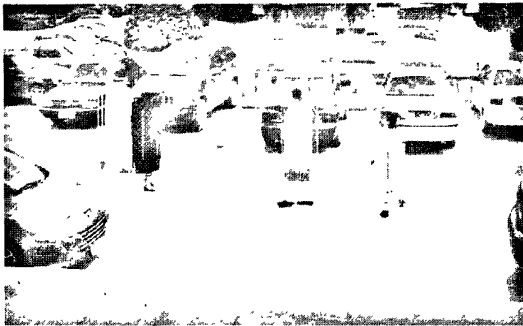


124 帧



144 帧

图 2-4 场景 B 的原图



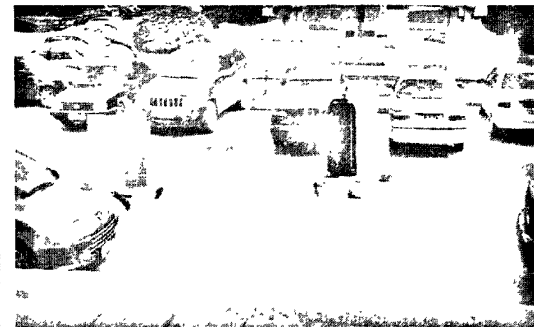
75 帧



94 帧



124 帧



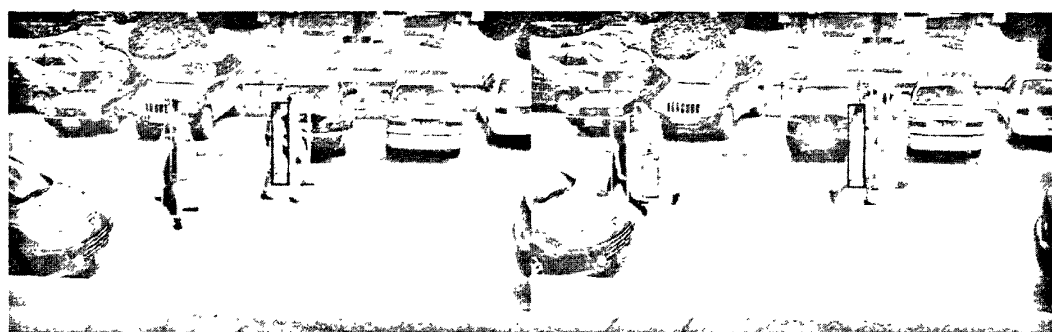
144 帧

图 2-5 场景 B 图像的人体目标的粒子



75 帧

94 帧



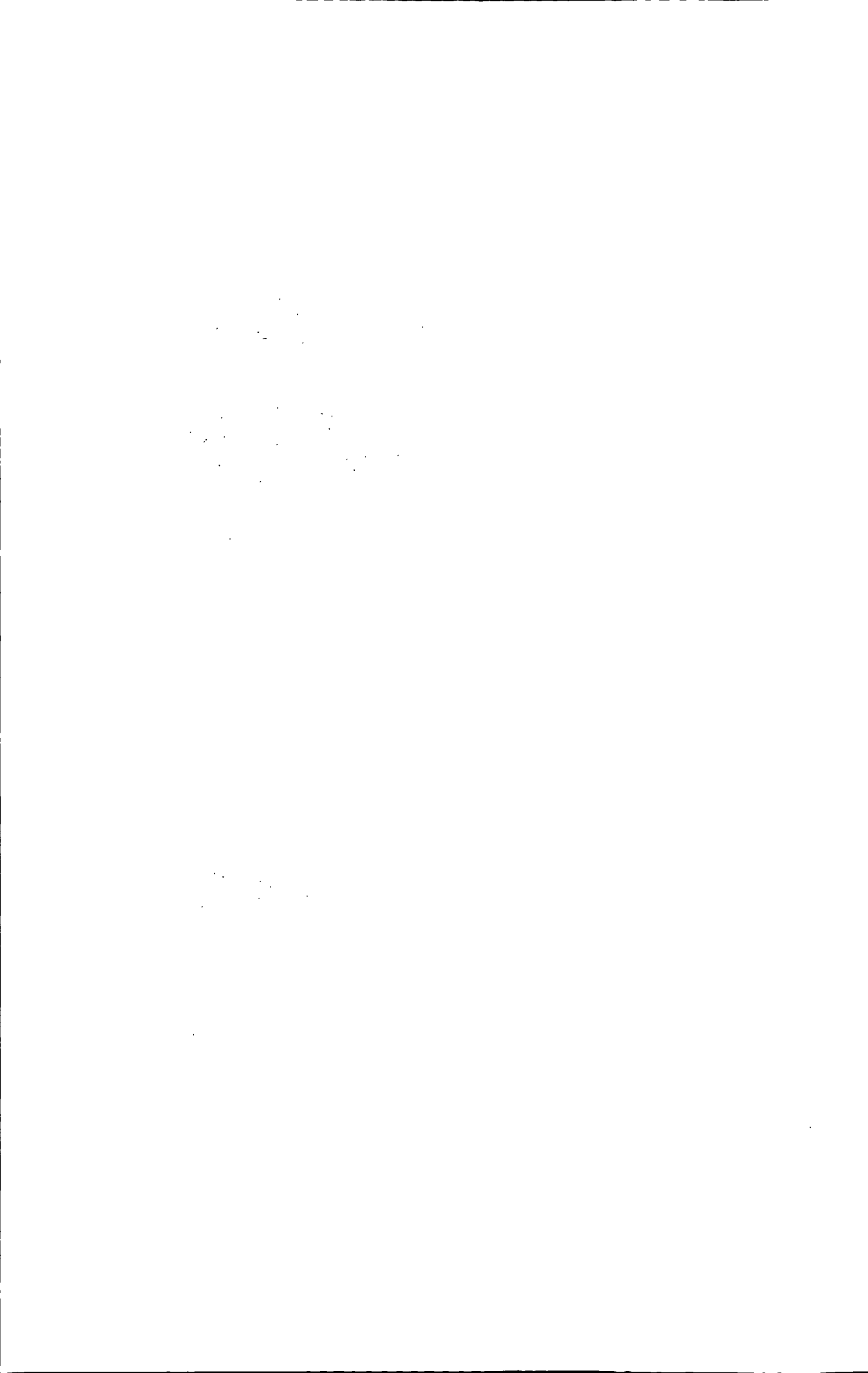
124 帧

144 帧

图 2-6 场景 B 人体目标的跟踪结果

2.5 本章小结

传统的粒子滤波一般是应用于单目标跟踪中，将其应用于多人目标跟踪中效果并不是很理想，因为传统的粒子滤波很难实现维持多目标分布的多模块跟踪。为了解决传统粒子滤波跟踪多目标的缺点，本章提出了一种非参数复合模型的方法来描述目标的人体目标的分布，该复合模型包括预测和更新两个阶段，同时提出了在该模型的情况下如何实现递归跟踪。在推导通用递归模型过程中，产生了复合的粒子滤波，通过它们复合权值的大小来衡量粒子的贡献大小。通过两组实验结果证明了该方法可以很好地维持多目标分布的多模块跟踪。



3 基于背景检测的单摄像机多目标遮挡人体跟踪

3.1 引言

在解决人体遮挡问题中,主要采用单摄像机和多摄像机两种方法。多摄像机的方法是指多个摄像机在不同的角度对同一个视场进行拍摄。该方法弥补了一些三维信息的丢失,但是弥补的信息仍然是有限的,而且对计算摄像机间的位置配置的精确度要求非常高,同时场景变化后对原来的算法影响较大。多摄像机的方法在实现过程中主要采用基于颜色的多摄像机方法[Kang et al,2004;Mittal and Davis,2003]和基于模块的多摄像机方法[Black et al,2002;Otsuka and Mukawa,2004];单摄像机的方法主要是通过单摄像机来实现人体的跟踪,该方法能够很容易对位置进行配置,有很好实用价值。解决遮挡问题时还可采用基于颜色的单摄像机的方法和基于模块的单摄像机方法[Han et al,2004;Isard and MacCormick,2001;Khan and Shah,2000;Smith et al,2005]。本文采用的是基于模块与颜色相结合的单摄像机的方法。通过基于 Epanechnikov 核密度梯度估计进行聚类,然后将聚类后的模块通过基于 Gauss 核密度估计获取颜色密度函数。该算法弥补了均值漂移算法 (Mean Shift) [Comaniciu and Meer,1999]的不足。

本文提出的算法不同于均值漂移算法。均值漂移算法是计算核函数加权下的搜索窗口的直方图分布,直方图的分布无法考虑像素的空间信息,无法使多通道的像素信息统一;均值漂移算法是以两个分布的相似性最大为原则,使搜索窗口沿密度增加最大的方向,移动到目标的真实位置,搜索窗口内部的背景颜色对其有很大的干扰,会导致跟踪窗产生抖动甚至飘移,该方法没有很好的利用目标的运动信息。首先本文提出的算法是全自动跟踪算法,而且解决了人体遮挡的问题;其次本文吸取了均值漂移算法利用梯度优化实现聚类的优点,然后充分利用空间信息和多通道的像素信息;本算法不再采取搜索窗口,而是充分利用运动信息,先将运动的前景信息提取出来,然后采用基于多通道 Gauss 核密度估计获取概率密度函数进行估计。

本文将在第3.2节讲述改进的混合高斯模型的背景重建,第3.3节阐述遮挡人体跟踪算法,其中在第3.3.1节总述了遮挡人体跟踪算法的整体实现过程,该算法将块建模、颜色建模、运动建模及人体空间信息有机结合,在3.3.2, 3.3.3, 3.3.4节分别介绍各模块的具体实现过程,然后在第3.4节中讲述了如何提高算法的运行效率,最后给出实验结果与结论。

3.2 改进的混合高斯模型背景重建

3.2.1. 背景模型的建立

在计算机视觉领域中,运动目标的实时检测是一个重要的研究课题。在视频监控、智能导航和人机交互等许多领域中有着广泛的应用。在背景重建过程中,不可不提的就是前景这个概念,前景是指在假设背景为静止的情况下,任何有意义的运动物体。建模的基本思想是从当前帧中提取前景,使背景更接近当前视频帧的背景。即利用当前帧和视频序列中的当前背景帧进行加权平均来更新背景。

混合高斯模型使用 K 个高斯模型来表征图像中各个像素点的特征,在新一帧图像获得后更新混合高斯模型,用当前图像中的每个像素点与混合高斯模型匹配,如果成功则判定该点为背景点,否则为前景点。通观整个高斯模型,他主要是有方差和均值两个参数决定,对均值和方差的学习,采取不同的学习机制,将直接影响到模型的稳定性、精确性和收敛性。由于我们是对运动目标的背景提取建模,因此需要对高斯模型中方差和均值两个参数实时更新。为提高模型的学习能力,改进方法对均值和方差的更新采用不同的学习率;为提高在繁忙的场景下,大而慢的运动目标的检测效果,引入权值均值的概念,建立背景图像并实时更新,然后结合权值、权值均值和背景图像对像素点进行前景和背景的分类。本文在 RGB 色彩空间采用混合高斯的方法对背景进行建模。混合高斯建模是将多个单高斯模型加权相加,然后得到一个混合高斯概率模型,利用该模型对图像中的每个像素描述其变化情况,视频中的每个像素点都需要定义 K 个单高斯分布来表示其颜色变化分布情况,在背景更新过程中,只要当前的像素值符合由 K 个高斯模型组成的混合高斯模型分布,那么就认为该像素是符合背景特征的像素点,这样就会把当前帧的该像素点当做背景点,反之,则把该像素点当做具有前景特征的前景点。背景图像的每个像素 $I(x,y)$ 可由 K 个高斯分布构成的混合高斯模型进行描述。任意时刻 t 像素 $I(x,y)$ 的混合高斯概率可由式(3.1)表示:

$$f(I_{t,xy}) = \sum_{i=1}^K w_{i,t,xy} \varphi(I_{t,xy} | \mu_{i,t,xy}, \Sigma_{i,t,xy}) \quad (3.1)$$

其中, K 为高斯分布的个数,由计算机的计算能力决定,一般取 3-5 个, $\mu_{i,t,xy}$ 、 $\Sigma_{i,t,xy}$ 、 $w_{i,t,xy}$ 分别表示 (x,y) 点处,第 i 个高斯分布在 t 时刻的均值、协方差矩阵、权重,且 $\sum_{i=1}^K w_{i,t,xy} = 1$, $\varphi(I_{t,xy} | \mu_{i,t,xy}, \Sigma_{i,t,xy})$ 表示在点 (x,y) 处,时刻 t 的第 i 个高斯分布的概率函数,其定义如式(3.2)所示。

$$\varphi(I_{t,xy} | \mu_{i,t,xy}, \Sigma_{i,t,xy}) = \frac{1}{(2\pi)^{n/2} |\Sigma_{i,t,xy}|^{1/2}} \exp \left[-\frac{1}{2} (I_{t,xy} - \mu_{i,t,xy})^T \Sigma_{i,t,xy}^{-1} (I_{t,xy} - \mu_{i,t,xy}) \right] \quad (3.2)$$

其中, n 表示向量 $I_{t,xy}$ 的维数。

当利用混合高斯进行背景建模时, 一般假定视频彩色图像的 RGB 三个通道是相互独立的, 而且这三个通道可以分别独立计算, 然后简化成三个独立的高斯模型, 该方法可以提高算法的运行效率, 混合高斯模型的实现过程如下[徐柱,2010]:

(1)初始化: 当前有两种方法可以对混合高斯背景模型进行初始化: 其一是计算视频序列图像中一段时间内每个像素点的像素平均值 $\mu_{0,xy}$ 以及像素点的方差 $\sigma_{0,xy}^2$, 分别用式(3.3)和式(3.4)表示。

$$\mu_{0,xy} = \frac{1}{N} \sum_{i=0}^{N-1} I_{i,xy} \quad (3.3)$$

$$\sigma_{0,xy}^2 = \frac{1}{N} \sum_{i=0}^{N-1} (I_{i,xy} - \mu_{0,xy})^2 \quad (3.4)$$

然后将 $\mu_{0,xy}$ 和 $\sigma_{0,xy}^2$ 分别作为混合高斯模型中 K 个高斯分布的均值与方差。不过该方法需要存储许多帧视频图像, 需要很多较大的存储空间。其二是: 为了降低内存空间的占用, 可采用一种比较简单的初始化方法, 即用视频的第一帧图像的每个像素值作为均值, 对它们赋予较大的方差和较小的权值。因为在初始化过程建立的是一个单峰模型, 而对混合高斯模型而言, 重要的是建立一个多峰分布的且简单的概率模型, 在对后面帧的学习过程中, 混合高斯模型的均值及方差可以随着时间推移对初始化时的均值及方差进行更新调整。

(2)模型学习: 混合高斯背景模型的学习既要更新每个高斯分布的参数, 又要更新各分布的权重, 每个高斯模型的参数, 是由方差和均值两个参数决定, 对均值和方差的学习, 采取不同的学习机制, 将直接影响到模型的稳定性、精确性和收敛性。由于我们是对运动目标的背景提取建模, 因此需要对高斯模型中方差和均值两个参数实时更新。为提高模型的学习能力, 改进方法对均值和方差的更新采用不同的学习率; 建立背景图像并实时更新, 然后结合权值、权值均值和背景图像对像素点进行前景和背景的分类。可以通过下面 3 步来实现模型的学习:

第 1 步: 就当前帧每个像素值 $I_{t,xy}$ 而言, 检查当前的像素值 $I_{t,xy}$ 是否与已建立的混合高

斯模型中的 K 个高斯模型中的某个单高斯模型相匹配, 如果满足式(3.5)中的规则, 则认为当前帧的像素值 $I_{t,xy}$ 和该背景模型匹配, 否则就认为当前帧的像素值 $I_{t,xy}$ 和该背景模型不匹配。

$$|I_{t,xy} - \mu_{i,t-1,xy}| \leq \lambda \delta_{i,t-1,xy} \quad i=1,2,\dots,K \quad (3.5)$$

其中, λ 是为了确定匹配相似度设定的一个经验值, 一般取 2.5 到 3.5 之间。

第 2 步: 执行完第 1 步后, 要考虑以下两种情况, 其一是当前帧的像素值 $I_{t,xy}$ 与已建立的混合高斯模型存在一个或多个高斯分布相匹配的情况, 其二是当前帧的像素值与已建立的混合高斯模型中的单高斯模型不存在相匹配的情况。下面分别对以上两种情况进行处理:

若当前帧的像素值 $I_{t,xy}$ 与已建立的混合高斯模型存在一个或多个高斯分布相匹配时, 则将其归入该模型中, 并对该模型根据当前的像素值对混合高斯背景模型的权值、均值、方差三个参数进行更新。权值的更新如式(3.6)所示:

$$w_{i,t,xy} = (1-\beta)w_{i,t-1,xy} + \beta L \quad i=1,2,\dots,K \quad (3.6)$$

若当前帧的像素值与第 i 个高斯模型匹配, 则令式(3.6)中 L 等于 1; 若当前的像素值与第 i 个高斯模型不匹配, 则令式(3.6)中 L 等于 0。式(3.6)中 β 为权值学习率, 如果权值学习率 β 的值较小, 则式(3.6)中的 $w_{i,t-1,xy}$ 所代表上一帧时的权重就比较大, 那么当前帧 $w_{i,t,xy}$ 的权重变化的就比较小, 需要比较长时间才能适应图像的像素值变化, 也就是说, 当运动的目标停止运动很长时间了, 仍然不能将其当做背景; 反之, 如果权值学习率 β 的值较大, 则式(3.6)中的 $w_{i,t-1,xy}$ 所代表上一帧时的权重就比较小, 那么当前帧 $w_{i,t,xy}$ 的权重变化的就比较大, 需要比较长时间才能适应图像的像素值变化。在实际应用中, 由于图像的像素值存在噪声, 要根据实际情况来设定权值学习率 β 的大小。式(3.6)可以相对准确地描述背景模型的分布情况, 能够有效地更新高斯背景模型的权值。

若当前帧的像素值 $I_{t,xy}$ 与已建立的混合高斯模型存在一个或多个高斯分布相匹配, 则表明当前帧的像素值 $I_{t,xy}$ 符合当前背景模型的分布, 因而要增大该高斯模型的更新权值。当前帧的像素值与混合高斯背景模型中的单高斯模型未匹配的, 要减小未匹配的高

斯模型的更新权值。

若当前帧的像素值与已建立的混合高斯模型中的单高斯模型不存在相匹配的情况时,背景模型的均值和方差的更新分别如式(3.7)和式(3.8)所示:

$$\mu_{i,t,xy} = (1-\rho)\mu_{i,t-1,xy} + \rho I_{t,xy} \quad i=1,2,\dots,K \quad (3.7)$$

$$\sigma_{i,t,xy}^2 = (1-\rho)\sigma_{i,t-1,xy}^2 + \rho(I_{t,xy} - \mu_{i,t-1,xy})^2 \quad i=1,2,\dots,K \quad (3.8)$$

其中更新率 ρ 为:

$$\rho = \beta \varphi(I_{t,xy} | \mu_{i,t-1,xy}, \Sigma_{i,t-1,xy}) \quad (3.9)$$

当新一帧的像素值与混合高斯模型中不匹配的高斯分布,认为这个当前帧的像素值对单高斯模型的均值 $\mu_{i,t,xy}$ 和方差 $\sigma_{i,t,xy}^2$ 没有任何影响,因此并不需要修正这个高斯模型的各个均值 $\mu_{i,t,xy}$ 和方差 $\sigma_{i,t,xy}^2$ 。

若新一帧的像素值与混合高斯模型中任何一个高斯分布都不相匹配,则要在原高斯模型中增加一个新的单高斯分布。这样一来混合高斯模型的总数必将增加一个,混合高斯模型的总数是固定不变的,因而应当将当前的混合高斯模型中权值最小的高斯分布剔除掉,以保证混合高斯模型的总数不变。既然引入了一个新的单高斯分布,那么新的单高斯分布的均值当前的像素值 $I_{t,xy}$,新的单高斯分布的方差为一个较大方差 σ_0 和新的单高斯分布的更新权值为一个较小的权值 w_0 。

第3步:根据更新之后的权值对高斯分布进行归一化处理。为了将更新之后模型的各个高斯分布的权值之和变为1,必须对混合高斯模型中各个单高斯模型的权值做归一化处理,权值的归一化方法可以用式(3.10)来实现。

$$w'_{i,t,xy} = \frac{w_{i,t,xy}}{\sum_{j=1}^K w_{j,t,xy}} \quad (3.10)$$

由于传统的混合高斯模型对复杂多变的场景表述能力较强,同时也存在一些缺点:

(1) 传统混合高斯模型的更新权值是固定的,无法适应摄像机所拍摄的环境的变化而变化。如果更新学习率的值过大,更新过于快,会对运动目标检测造成灵敏度低的影响;如果更新学习率的值过小,更新过于慢,会使已经运动了的目标在背景中仍未得到更新产生时间拖影。如何找到一种能够随着摄像机所拍摄的环境的变化而自动修改更新学习率的自适应法是很多科研人员一直研究问题

(2) 当光照条件突然发生变化时,传统混合高斯模型不能准确地检测出运动的目标,这

也是当前的背景减除法所面临的一个棘手的问题。

(3) 混合高斯背景建模法, 要对每个像素建立 K 个单高斯模型, 要对每个模型计算求和积分, 在图像的分辨率比较高的情况, 做到实时运动目标检测也是很难解决的问题。

针对上述存在的问题, 科研人员对传统的混合高斯模型的背景建模做了许多改进。

3.2.2 改进的混合高斯背景建模

背景的更新速度的快慢是由更新率 ρ 决定的, 当更新率 ρ 值较小时, 能够抑制视频中的噪声干扰和保持模型的稳定性。然而背景模型的均值和方差变化的速度就比较慢且需要较长时间适应环境的变化。因此取一个自适应的更新率 ρ 显得尤为关键。

(1) 引进一个参数 l_k 用其记录当前帧与模型匹配的有效像素的点数 l_k , 令 $\rho = \beta + (1 - \beta) / l_k$ 可以用 l_k 来影响背景更新率 ρ , 设 l_k 初始值为 1。当 l_k 的值比较小时, 则 ρ 值比较大, 会加快更新过程; 当 l_k 的值比较大时, 则 ρ 值比较小, 会减缓更新过程。在一定程度上可以适应摄像机所拍摄的环境的变化。

(2) 为了更好地用混合高斯背景模型来描述当前的背景图像, 为了使背景模型更有效地适应视频图像的背景变化, 本文提出了一个新的参数 T_k , 用式(3.11)来表达该参数:

$$T_k = \begin{cases} 1 & k = \arg \max_j \{p_j\} \\ \frac{p_k}{\sum_{j=0}^K p_j} & \text{其它} \end{cases} \quad (3.11)$$

其中 p 为当前像素值 X_i 属于第 i 个模型的概率:

$$p_i = \begin{cases} w_{i,t,xy} \varphi(X_{i,t,xy}, \mu_{i,t,xy}, \sigma_{i,t,xy}) & |X_{i,t,xy} - \mu_{i,t,xy}| \leq \lambda \sigma_{i,t,xy} \\ 0 & \text{其它} \end{cases} \quad (3.12)$$

对于所有符合式 $|I_{i,t,xy} - \mu_{i,t,xy}| \leq \lambda \delta_{i,t,xy}$ 的模型参数进行更新:

$$\begin{cases} w_{i,t,xy} = (1 - \beta) w_{i,t-1,xy} + \beta T_k \\ \mu_{i,t,xy} = (1 - \rho) \mu_{i,t-1,xy} + \rho T_k I_{i,t,xy} \\ \sigma_{i,t,xy} = (1 - \rho) \sigma_{i,t-1,xy} + \rho T_k I_{i,t,xy} \end{cases} \quad (3.13)$$

(3) 当背景更新速度很快, 将会导致视频图像中运动缓慢的目标很容易融入到背景模型中。为了解决该问题, 本文提出了一个新的参数 l_b , 用参数 l_b 表示某个像素连续为背景的帧数 l_b , 当像素模型的权值增加要转化成背景模型时, 若 $l_b < T_{b1}$, 则仍就认定该像素为运动比较慢的前景像素。若 $l_b > T_{b2}$ ($T_{b2} > T_{b1}$) 则由于该像素模型持续认定为背景模型, 此时才认定该像素为背景像素。即持续为背景的帧数至少要大于 T_{b2} 的情况下, 才可被认

定为背景。

图 3-1 给出了改进的混合高斯模型建模的过程。

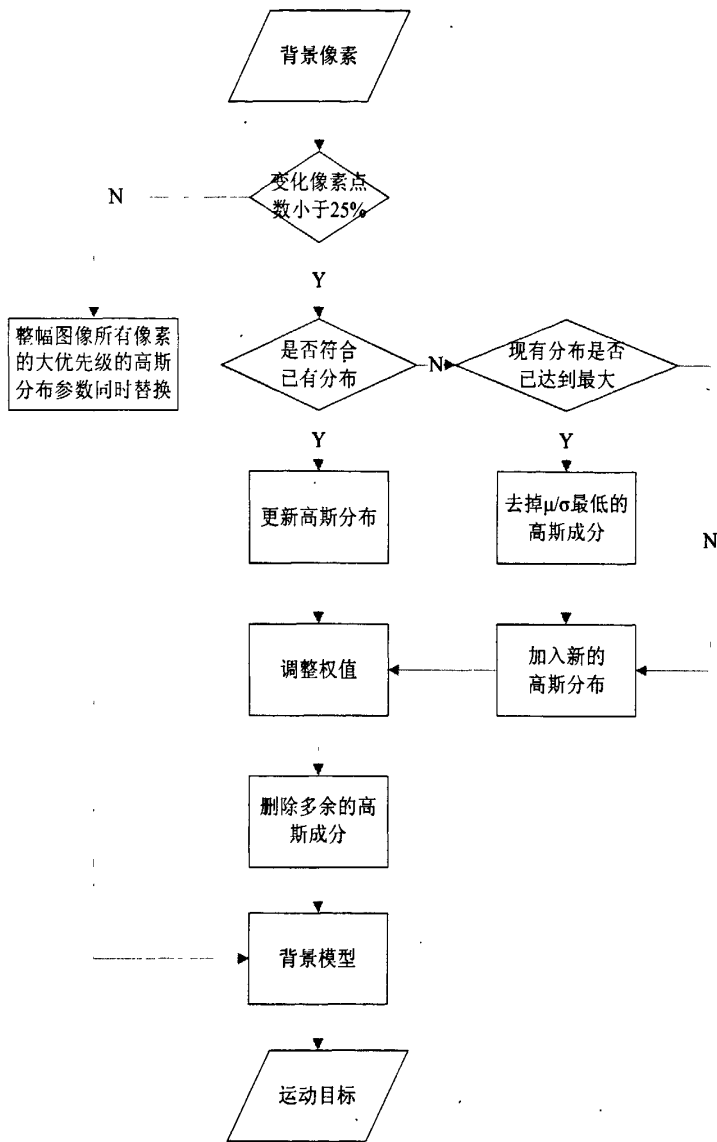


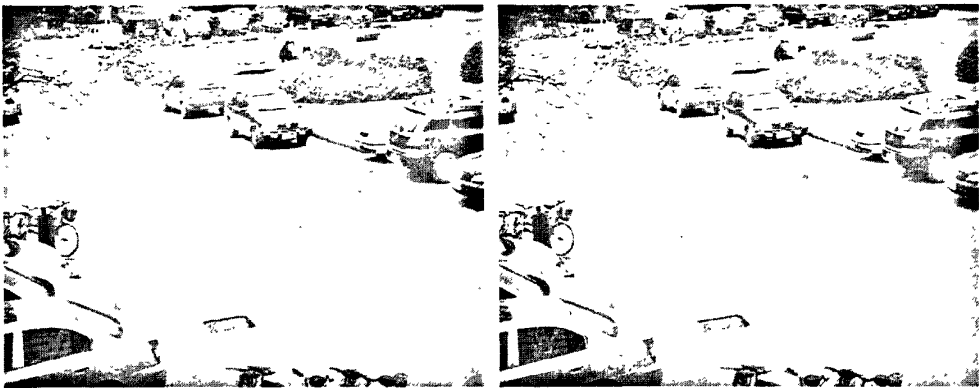
图 3-1 改进的混合高斯模型流程图

3.2.3 背景更新算法的改进结果

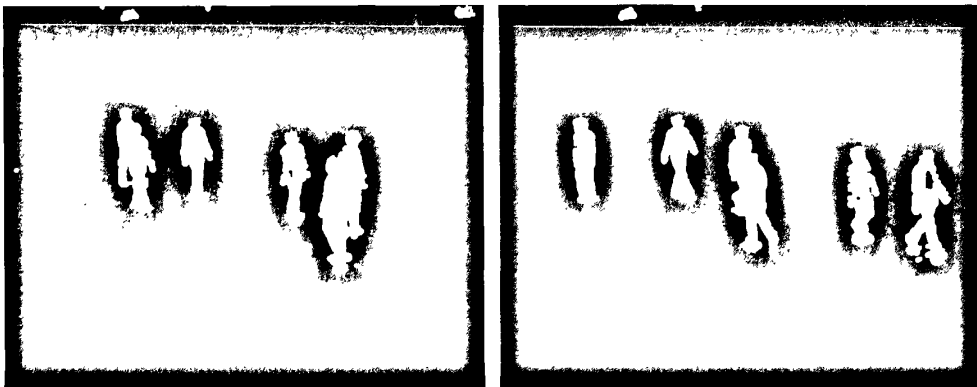
本节将给出改进后的不同场景的两组视频图像。图 3-2 场景 A 的运动人体目标检测和图 3-3 分别给出了场景 A 和场景 B 的运动人体目标检测的结果。(a) 给出了原图，(b) 给出了视频的背景重建的结果图像，(c) 通过减背景法提取出了运动的前景人体目标。



(a) 原图



(b) 背景重建结果



(c) 前景提取

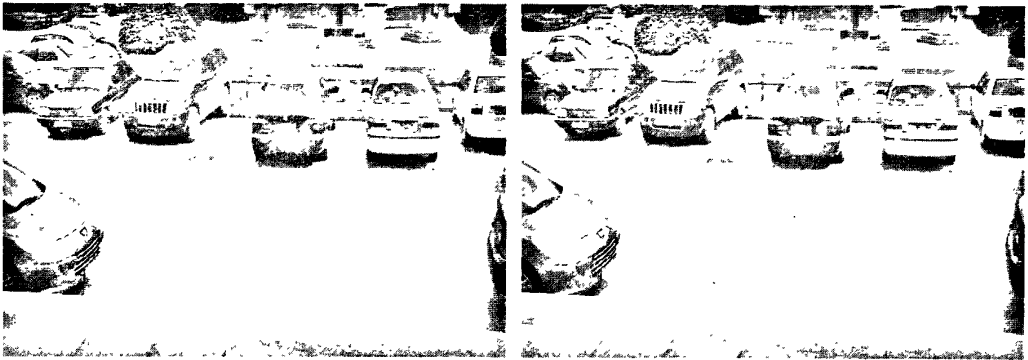


(d) 提取前景人体目标

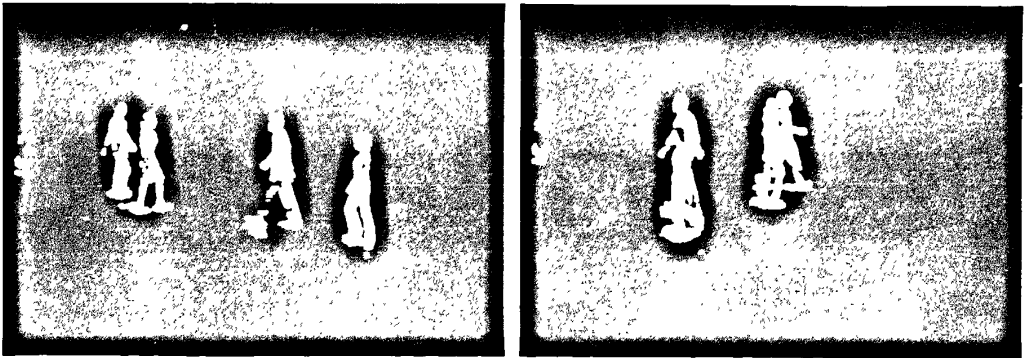
图 3-2 场景 A 的运动人体目标检测



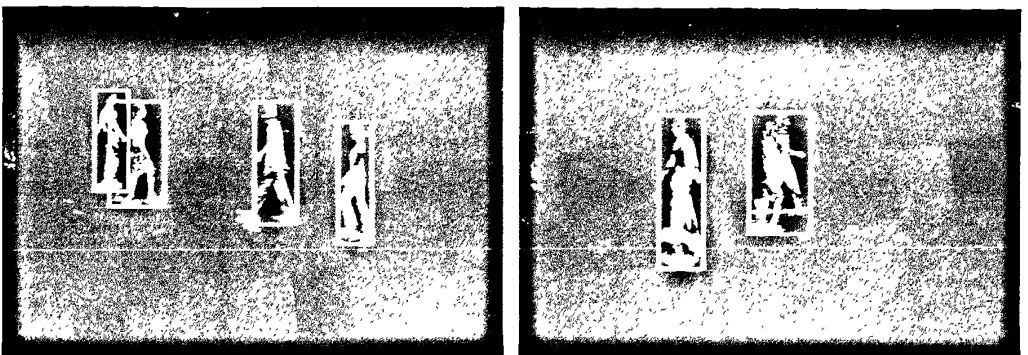
(a) 原图



(b) 背景重建结果



(c) 前景提取



(d) 提取前景人体目标

图 3-3 场景 B 的运动人体目标检测

3.3 基于多模型融合的遮挡人体跟踪

3.3.1 遮挡人体跟踪的基本思想

本文提出了一种基于块建模、运动建模、颜色建模及人体空间信息的算法[王选贺和刘济林]，有效地解决了遮挡人体的跟踪问题，该算法具有广泛的适用性。图 3-4 给出了本文提出的算法的总体流程图。首先，利用上节讲到的改进的混合高斯建模的方法对背景进行重建[Stauffer and Grimson,1999]，利用减背景方法检测出前景运动目标，然后去除阴影；然后，如果这个前景区域的人体目标是第一次出现，通过新人进入模型将其存储起来。如果是再次出现的人体目标，通过更新模型将其进行更新；接着就是解决遮挡的关键一步，本文将块建模、颜色建模、运动建模、及人体空间信息结合来解决这一问题。块建模主要根据每个人的衣服颜色信息把存储的人体模型分割成许多颜色相近的区域；颜色建模是在块建模的基础上根据块的颜色信息和块在人体中分布的空间位置进行建模来获取颜色概率密度函数；运动建模是将连续的视频序列中运动位置比较近的人体看作是同一个人的概率的方法来获取运动概率密度函数。通过上述建模将得到后验概率图像，然后对后验概率图像进行分割；最后，将跟踪后的人体位置信息和颜色信息通过更新模型将其存储起来以达到连续跟踪的目的。

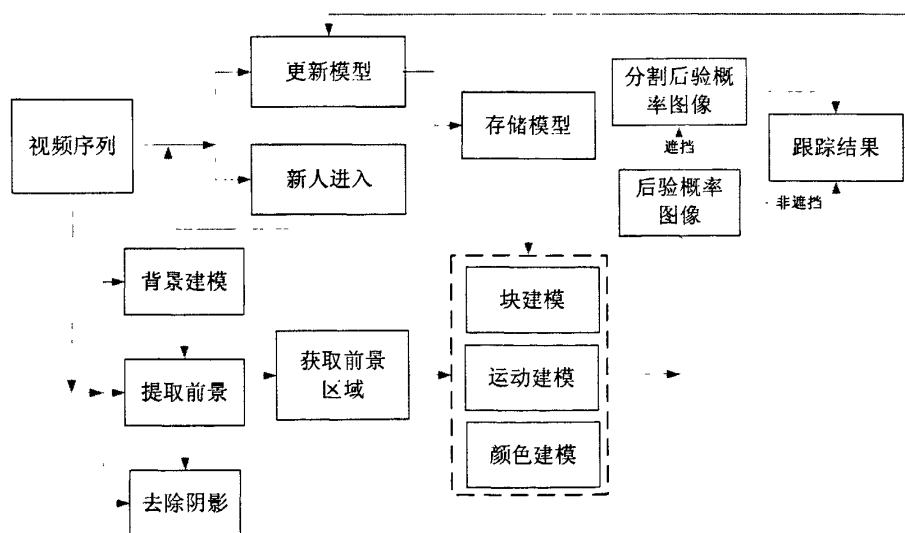


图 3-4 系统总体流程图

在块建模中，首先根据人体衣服的颜色信息，通过均值漂移的聚类算法进行聚类，将存

储模型中的人体聚类成许多颜色相近的模块 $M' = \{A_i\} \quad i = \dots n$, n 是利用均值漂移算法对人体进行聚类的块数, 分割的块数是由人体的衣服颜色决定的, 块建模会根据衣服颜色不同而分割成不同的区域 (在图 3-5 中, 为了说明的方便, 假设通过块建模将人体被分割成了六个模块)。聚类算法在 3.2.2 节中详细介绍。同时, 本文在获取颜色概率密度函数过程中, 也充分考虑了人体的块 A_i 空间信息的分布。因为人的头部的颜色信息不可能与存储模型中的人体下半身的颜色信息进行匹配。考虑人体的空间信息既可以提高跟踪的准确性, 又可以提高算法的运行效率。假设根据当前帧的待识别的前景区域中的像素 $c_{t,x,y}$ 位置获取另一个集合 M'' (如图 3-5 所示), 该集合的表达式如下:

$$M'' = \{(x, y) | 0 < x < w_{t-1}, h_{t-1}(\frac{y}{h_t} - \alpha) < y < h_{t-1}(\frac{y}{h_t} + \alpha)\} \quad (3.14)$$

其中 w_{t-1}, h_{t-1} 分别表示已存储模型中人体的宽度和高度信息, h_t 表示当前像素 $c_{t,x,y}$ 所在的人体的高度信息, y_t 表示当前像素 $c_{t,x,y}$ 的纵坐标 y_t , α 是个小数, 表示在该像素的上下活动的范围内进行统计。如果当前像素 $c_{t,x,y}$ 改变, 那么集合 M'' 也将随之改变, 是个动态过程。最后将根据均值漂移算法得到模块 $M' = \{A_i\} \quad i = 1 \dots n$ 和位置信息得到的集合 M'' 取交集得到的集合 $M = \{B_1, B_2, B_3, B_4, \dots\}$ 。

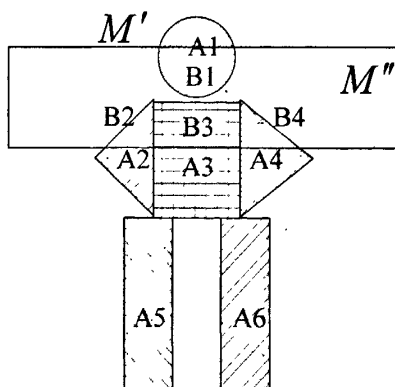


图 3-5 颜色块建模

如果已知建模得到的人体模型 $M_{t-1,k} = \{B_1, B_2, B_3, B_4, \dots\}$, 将得到当前像素 $c_{t,x,y}$ 的属于人体模型 $M_{t-1,k}$ 的概率 $P_{M_{t-1,k}}(c_{t,x,y})$, 即:

$$\begin{aligned}
 P(\mathbf{c}_{t,x,y} | M_{t-1,k}) &= \max \{P(\mathbf{c}_{t,x,y} | B_1), P(\mathbf{c}_{t,x,y} | B_2), P(\mathbf{c}_{t,x,y} | B_3) \dots\} \\
 &= \max \{P_{B_1}(\mathbf{c}_{t,x,y}), P_{B_2}(\mathbf{c}_{t,x,y}), P_{B_3}(\mathbf{c}_{t,x,y}) \dots\}
 \end{aligned}
 \tag{3.15}$$

其中 $P_{B_i}(\mathbf{c}_{t,x,y})$ 为颜色密度函数，由公式 (3.15) 看出，像素 $\mathbf{c}_{t,x,y}$ 属于模型 $M_{t-1,k} = \{B_1, B_2, B_3, B_4, \dots\}$ 的概率是通过最大后验概率得到的，即通过获取所有 B_i 中颜色密度函数 $P_{B_i}(\mathbf{c}_{t,x,y})$ 中概率最大的值作为该像素属于该模型的概率。颜色密度函数 $P_{B_i}(\mathbf{c}_{t,x,y})$ 将在 3.3.3 节中详细介绍。

同时，本文也考虑了人体的运动信息，当前帧与上一帧的坐标变化情况。一般来说，同一个人体在 $t-1$ 时刻和 t 时刻两帧之间不会移动很大的距离。所以该算法选取了前景区域外接矩形的三点中心的绝对坐标作为这个人体在图像中的运动位置，并将三个特征点作为运动建模的向量 $\mathbf{v}_t = (x_{t,1}, y_{t,1}, x_{t,2}, y_{t,2}, x_{t,3}, y_{t,3})$ 。如图 3-6 所示。这个向量可以充分描述左右相互遮挡和上下相互遮挡人体的运动信息。当前帧的每个人体都是通过运动密度函数 $P_{M_{t-1,k}}(\mathbf{v}_t)$ 进行描述的（在 3.3.4 节中详细介绍）。该运动建模的思想是前后两帧的人体距离越接近，那么是同一个人的概率越大。但是当发生遮挡时，遮挡人体目标的距离都很近，此时运动密度函数的概率值也非常接近，所以此时颜色建模将在遮挡中起主要作用。

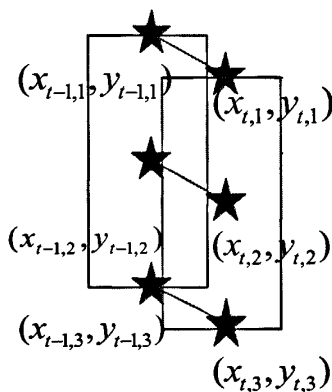


图 3-6 运动建模的向量

如果已知当前帧某像素 $\mathbf{c}_{t,x,y}$ 对应的人体模型 $M_{t-1,k} = \{B_1, B_2, B_3, B_4, \dots\}$ ，并且运动模型的运动密度函数和颜色模型的颜色密度函数是相互独立的，假定颜色与运动都具有相同的先验概率，则当前像素 $\mathbf{c}_{t,x,y}$ 属于人体模型 $M_{t-1,k}$ 的概率 $P(\mathbf{c}_{t,x,y}, \mathbf{v}_t | M_{t-1,k})$ 可表示为：

$$\begin{aligned}
 P(\mathbf{c}_{i,x,y}, \mathbf{v}_i | M_{t-1,k}) &= P(\mathbf{c}_{i,x,y} | M_{t-1,k})P(\mathbf{v}_i | M_{t-1,k}) \\
 &= \max\{P(\mathbf{c}_{i,x,y} | B_1), P(\mathbf{c}_{i,x,y} | B_2), P(\mathbf{c}_{i,x,y} | B_3) \cdots\}P(\mathbf{v}_i | M_{t-1,k}) \\
 &= \max\{P_{B_1}(\mathbf{c}_{i,x,y}), P_{B_2}(\mathbf{c}_{i,x,y}), P_{B_3}(\mathbf{c}_{i,x,y}) \cdots\}P_{M_{t-1,k}}(\mathbf{v}_i)
 \end{aligned} \quad (3.16)$$

给定当前帧的一个人体的样本集合 $S = \{\mathbf{c}_i\} \quad i=1 \cdots N$ 和前景区域的运动建模向量 $\mathbf{v}_i = (x_{i,1}, y_{i,1}, x_{i,2}, y_{i,2}, x_{i,3}, y_{i,3})$, 则可以根据公式(3.17)计算出当前样本集合 $S = \{\mathbf{c}_i\}$ 中的每个像素 $\mathbf{c}_{i,x,y}$ 的属性, 即确定每个像素属于 n 个存储模型中某个人体模型的概率。

$$\mathbf{c}_{i,x,y} \in S \quad s.t. \quad k = \arg_k \max P(\mathbf{c}_{i,x,y}, \mathbf{v}_i | M_{t-1,k}) \quad k=1 \cdots n \quad (3.17)$$

其中 n 代表存储模型的个数。最后, 当前帧的一个人体的样本集合被分成许多个子集, 其中子集中样本数目最多的将决定集合 S 的属性。即找到了该集合与上一帧相对应的原型。当前帧的其它模型与此类推。

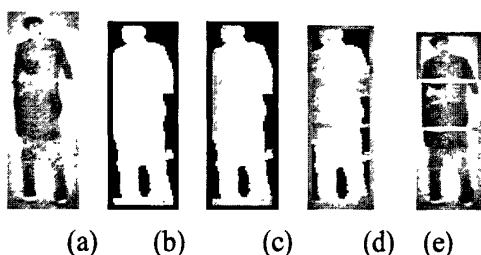


图 3-7 遮挡情况的处理过程

- (a) 原始图像 (b) 检测出的前景图像 (c) 后验概率图像
(d) 根据颜色分割后验概率图像(e) 最终的分割结果

图 3-7 给出了解决人体遮挡的跟踪算法的实现过程。首先对视频序列的背景进行重建, 运用背景相减的方法提取出前景区域如图 3-7 (b) 所示。然后计算前景区域每个像素属于存储模型中某个人体模型的概率 $P(\mathbf{c}_{i,x,y}, \mathbf{v}_i | M_{t-1,k})$, 再根据公式(3.17)找到该像素属于某个模型的概率最大, 即将该模型的代表颜色赋予该像素如图 3-7 (c)所示。然后根据颜色的分布对当前的前景区域进行分割如图 3-7 (d)。最后, 在原图上将目标分开如图 3-7(e)所示。

3.3.2 块建模

块建模是通过均值漂移聚类的算法来实现的。均值漂移算法是一种密度函数梯度估计的非参数方法, 该算法不需要对目标聚类数目有先验知识, 只需通过迭代寻优找到概率分布的极值来进行聚类。本文根据衣服的颜色信息及空间信息对存储模型中的人体进

行聚类 $M = \{A\} \quad i = 1 \cdots n$ 。像素的位置信息被考虑成定义域，而像素的颜色像素值被考虑成值域。这样，空间向量和颜色向量被连接成一个联合向量 $\mathbf{x} = (\mathbf{x}^s, \mathbf{x}^r)$ ，其中 \mathbf{x}^s 和 \mathbf{x}^r 分别表示空间和颜色特征向量，它们的核带宽参数 h_s 和 h_r [Comanicu et al, 2001]。多向量的核函数被定义成空间向量核函数和颜色向量核函数乘积的形式。

$$K_{h_s, h_r}(\mathbf{x}) = \frac{C}{h_s^2 h_r^2} k\left(\left\|\frac{\mathbf{x}^s}{h_s}\right\|^2\right) k\left(\left\|\frac{\mathbf{x}^r}{h_r}\right\|^2\right) \quad (3.18)$$

其中 $k(x)$ 是服从 Epanechnikov

$$k(x) = \begin{cases} 1-x & x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.19)$$

该核函数的导数是：

$$g(x) = -k'(x) = \begin{cases} 1 & x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (3.20)$$

根据单核向量的推导过程 [Comanicu and Meer, 2002]，我们可以推导出联合向量核函数的均值漂移递推公式：

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^n \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i^s}{h_s}\right\|^2\right) g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i^r}{h_r}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i^s}{h_s}\right\|^2\right) g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i^r}{h_r}\right\|^2\right)} \quad (3.21)$$

$\mathbf{x}_i = (\mathbf{x}_i^s, \mathbf{x}_i^r) \quad i = 1, \cdots, n$ 为输入存储的人体模型的联合作用域向量（如图 3-8 (a) 所示）。

Minisize: 分成模块的最小的像素。 h_s, h_r : 分别表示空间作用域和颜色作用域的核带宽参数。图 3-8 采用的参数为 $(h_s, h_r, M) = (7, 11, 60)$ 。该聚类过程只是将人体颜色相近的信息进行粗分类。所以该参数的选取具有较大的范围，参数的变化对整个算法的影响并不十分敏感。聚类过程的算法如下：

(1) for $i = 1$ to n :

(a) 对 $\mathbf{y}_{i,1}$ 进行赋初始值 即, $\mathbf{y}_{i,1} := \mathbf{x}_i$ 。

$$(b) \quad j:=1, \quad \text{计算} \quad \mathbf{y}_{i,j+1} := \frac{\sum_{i=1}^n \mathbf{x}_i \mathbf{g}\left(\left\|\frac{\mathbf{y}_{i,j}^s - \mathbf{x}_i^s}{h_s}\right\|^2\right) \mathbf{g}\left(\left\|\frac{\mathbf{y}_{i,j}^r - \mathbf{x}_i^r}{h_r}\right\|^2\right)}{\sum_{i=1}^n \mathbf{g}\left(\left\|\frac{\mathbf{y}_{i,j}^s - \mathbf{x}_i^s}{h_s}\right\|^2\right) \mathbf{g}\left(\left\|\frac{\mathbf{y}_{i,j}^r - \mathbf{x}_i^r}{h_r}\right\|^2\right)}$$

直到 $\mathbf{y}_{i,j+1}$ 收敛于 $\mathbf{y}_{i,c}$, (如图 3-8 (b)所示)。

(c) 将收敛向量 $\mathbf{y}_{i,c}$ 存储到向量 \mathbf{z}_i , 即 $\mathbf{z}_i := \mathbf{y}_{i,c}$

(2)根据空间核带宽和颜色核带宽 h_s 和 h_r 的范围, 对所有存储在 \mathbf{z}_i 中的向量进行分组

$$\{C_p\} \quad p=1 \cdots k$$

(3)for $i=1$ to n :

将 \mathbf{z}_i 中属于同一组 C_p 赋成 p 值 $L_i := \{p | \mathbf{z}_i \in C_p\}$ (如图 3-8 (c))

(4)合并分成的模块中小于 Minsize。

(5)将属于同一组 L_j 的原像素 \mathbf{x}_j 构成新的集合 A_i , 即 $A_p = \{\mathbf{x}_j | L_j = p, j=1 \cdots n\}$

$p=1 \cdots k$ 所以 $M' = \{A_p\}$ (如图 3-8 (d)). 输入存储的人体模型 M' 被分成 k 个子集 A_p

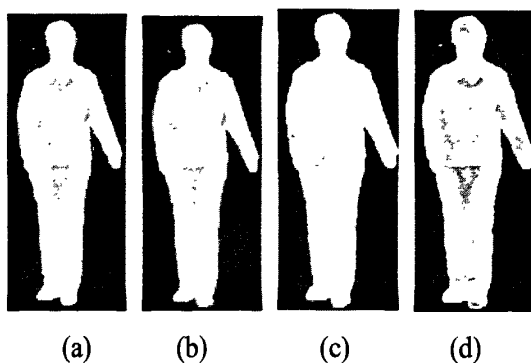


图 3-8 块建模的实现过程

(a) 存储模型中的前景图像 (b) 对原始图像聚类后的过滤图像

(c) 对过滤图像进行分组 (d) 将原始图像分成了许多模块子集

3.3.3 颜色建模

在解决人体遮挡过程中, 颜色信息起着很重要的作用。在 3.3.2 节中, 已经将人体分成很多块, 本节将介绍如何使用核密度估计的方法来估计每个块的颜色密度分布情况。

核密度估计属于非参数密度估计的一种, 非参数估计的优点就是没有固定的结构, 它的估计取决于所有的数据点。在许多文献中, 将核密度估计用于背景重建中, 本文将其用来解决遮挡人体跟踪。用 K 表示核函数, 用 σ 表示带宽, 样本数据 $S = \{x_i\} i = 1, 2, \dots, N$, 任意点 x 的密度估计可表示为:

$$\hat{f}(x) = \frac{1}{N\sigma} \sum_{i=1}^N K\left(\frac{x-x_i}{\sigma}\right) \quad (3.22)$$

如果 x 是一个 d 维的向量, 则核密度估计可以通过多元核乘积估计 (multivariate product kernel estimate) [Scotta and Sainb, 2005] 来进行估计。所以公式(3.22)可改变成公式(3.23)的形式, 其中每维向量都有自己的带宽 σ_k 。

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{\sigma_1 \dots \sigma_d} \prod_{k=1}^d K\left(\frac{x^{(k)} - x_i^{(k)}}{\sigma_k}\right) \right] \quad (3.23)$$

本文的颜色建模就是通过多元核乘积估计进行建模的。为了光亮度改变具有较强的鲁棒性, 本文对每个像素的色度进行规范化处理, 表示成一个向量 $\mathbf{x} = (r, g, s)$, 其中 $r = \frac{R}{R+G+B}$, $g = \frac{G}{R+G+B}$, $s = \frac{R+G+B}{3}$ 。在公式(3.23)中, 如果给定样本空间 $S_B = \{\mathbf{x}_i = (r_i, g_i, s_i)\}$, $d = 3$, $\sigma_r, \sigma_g, \sigma_s$ 分别表示 r, g, s 的带宽, 则颜色建模的颜色密度函数既可表示为:

$$P_S(r, g, s) = \frac{1}{N\sigma_r\sigma_g\sigma_s} \sum_{i=1}^N K\left(\frac{r-r_i}{\sigma_r}\right) K\left(\frac{g-g_i}{\sigma_g}\right) K\left(\frac{s-s_i}{\sigma_s}\right) \quad (r_i, g_i, s_i) \in S \quad (3.24)$$

其中, 核函数服从高斯分布:

$$K(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2} \quad (3.25)$$

核密度估计的方法使每个样本数据 $S = \{\mathbf{x}_i\}$ 起到了平滑的作用, 对某一个数据 $\mathbf{x} = (r, g, s)$ 概率的大小主要取决于它与样本数据 $S = \{\mathbf{x}_i\}$ 的色度空间的距离、带宽 σ 及所采用的核函数。本文对带宽 σ 的选取是用点估计法求取给定样本空间 $S_B = \{\mathbf{x}_i = (r_i, g_i, s_i)\}$ 的方差来实现。关于多元密度估计带宽的选取可参考文献 [Duong and Hazelton, 2005]。

3.3.4 运动建模

运动建模是基于连续两帧图像中前景人体的位置坐标进行估计获取概率。运动建模的目的在于建立运动密度函数,在实际中,有些人穿着相似的外衣,但通过位置可以将其识别出来。在这种情况下运动建模就起到了重要作用,而且有了运动信息也会提高跟踪的准确性。运动密度函数可表示成 $P_{M_{t-1}}(\mathbf{v}_t)$,本文选择每个前景人体区域的外接矩形的中点坐标作为人体的运动特征点(如图6所示)。运动向量为 $\mathbf{v}_t = (x_{t,1}, y_{t,1}, x_{t,2}, y_{t,2}, x_{t,3}, y_{t,3})$,图3-9给出了三个人在连续两帧的前景区域外接矩形的中点坐标。运动密度函数 $P_{M_{t-1}}(\mathbf{v}_t)$ 可以通过联合高斯概率分布函数来描述。运动向量的三点坐标是相互独立的。所以运动密度函数 $P_{M_{t-1}}(\mathbf{v}_t)$ 可以表示成:

$$P_{M_{t-1}}(\mathbf{v}_t) = g_{\sigma_1}(x_{t,1} - x_{t-1,1}, y_{t,1} - y_{t-1,1}) g_{\sigma_2}(x_{t,2} - x_{t-1,2}, y_{t,2} - y_{t-1,2}) g_{\sigma_3}(x_{t,3} - x_{t-1,3}, y_{t,3} - y_{t-1,3}) \quad (3.26)$$

其中:

$$g_{\sigma}(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x^2+y^2}{\sigma^2}\right)} \quad (3.27)$$

在当前帧中,两个人的距离较近时,颜色建模将会起到主要作用,相对远时,运动建模会起主要作用。其中 σ 是由同一个人在前后两帧运动的距离决定的。从图6可以看出,当前帧person A的三个中点坐标分别与上一帧Person A, Person B, Person C的中点坐标进行相减,就当前帧的Person A而言,可以获得三个运动密度函数 $P_{AA'}(\mathbf{v}_t)$, $P_{AB'}(\mathbf{v}_t)$ 和 $P_{AC'}(\mathbf{v}_t)$,最终 $P_{AA'}(\mathbf{v}_t)$ 值是最大;就当前帧的Person B而言,三个运动密度函数 $P_{BA'}(\mathbf{v}_t)$, $P_{BB'}(\mathbf{v}_t)$ 和 $P_{BC'}(\mathbf{v}_t)$ 中, $P_{BB'}(\mathbf{v}_t)$ 值是最大;就当前帧的Person C而言,三个运动密度函数 $P_{CA'}(\mathbf{v}_t)$, $P_{CB'}(\mathbf{v}_t)$ 和 $P_{CC'}(\mathbf{v}_t)$ 中, $P_{CC'}(\mathbf{v}_t)$ 的值是最大;因此,在连续两帧中距离较近的人体被看作是同一个人的概率较大。

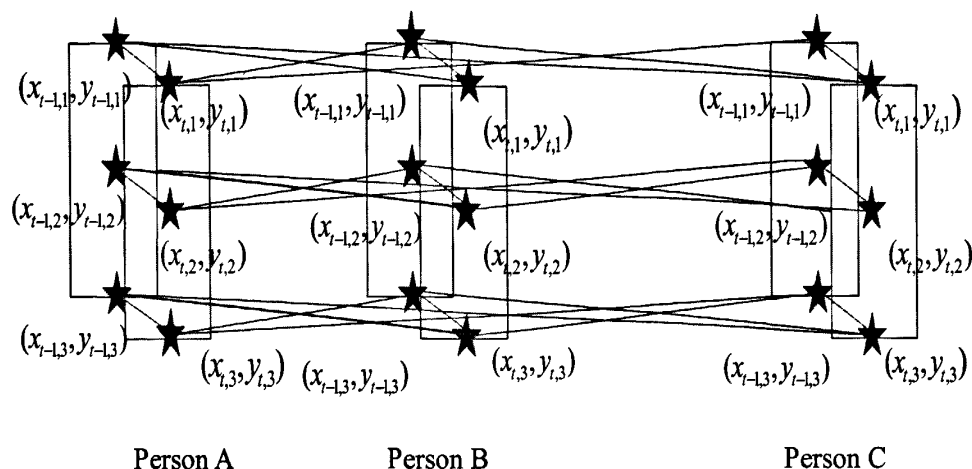


图 3-9 基于多人体的运动模型

3.4 提高算法的运行效率

本文采用的是非参数的核密度估计方法，很好地描述了样本的概率密度，同时对当前的前景图像的每个像素获取后验概率。尽管该算法获得了较理想的效果，但是在实际应用中由于运行的时间较长，降低了该算法的实际应用性，为了解决该问题，在本节主要讲述了如何提高该算法的运行效率，使其在实际应用中得以实现。

本文提出的遮挡人体跟踪算法是对当前的前景图像的每个像素逐一获取后验概率的方法来实现的。本文首先利用改进的快速高斯变换法[Yang et al,2008]对颜色密度函数公式(3.24)的核函数展成厄密特函数(Hermite)，提高了算法的运行效率，然后考虑降低运算的数据量来提高算法的运算效率。减少数据量有两种方法：一种方法是对当前的前景图像进行聚类，然后对聚类后的每个模块选出一个值作为该模块的代表来获取后验概率。尽管该方法减少了数据量，但是该方法对聚类的代表值选取要求很高，如果代表值选错就会影响整个模块的后验概率的判定，就会影响最终的判定结果；另一种方法是对当前的前景图像和存储模型中图像进行采取隔行抽取的方法来减少数据量，在这里如果分辨率已经很小的图像就不必采用隔行抽取，以保证图像不失真，只有分辨率较大的图像(如图 3-10(a))采用隔行抽取的方法得到图 3-10.(b) 图像。对隔行抽取后的图像的每个像素逐一获取后验概率(图 3-10(c))，然后对该概率图进行分割，将分割后的矩形框乘以二倍的方式反推到原来图像上(图 3-10(d))。该方法不仅达到了原来的分割效果而且大大提高了算法的运行效率。

本文对运行效率算法改进前与算法改进后进行了比较。在配置为 CPU Pentium 2.8G，内存为 1G 的环境下进行测试的，将图 3-11 中的序列 1 作为被测试的对象，在算法未

改进前的运行速度为每秒 2.63 帧, 算法改进后运行速度为每秒 10.57 帧, 基本满足了实时的要求, 而且达到了原来的跟踪效果。

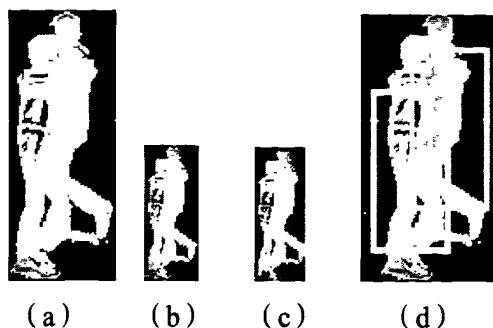


图 3-10 隔行抽取后图像遮挡情况处理

(a) 前景图像 (b) 隔行抽取后图像 (c) 后验概率图像 (d) 跟踪结果

3.5 实验结果和分析

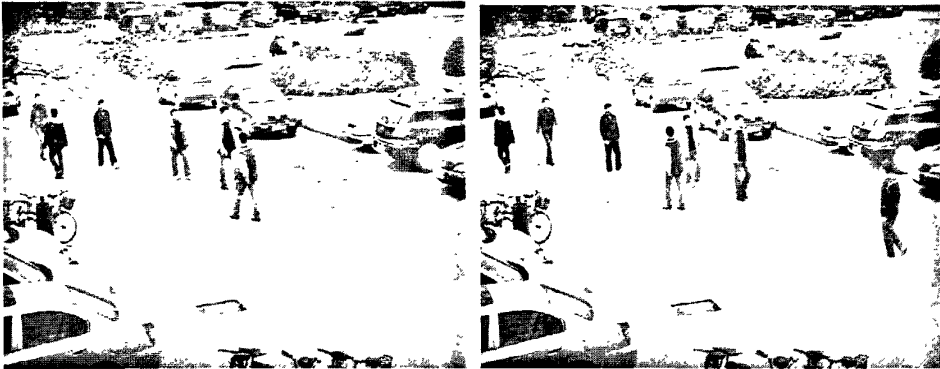
实验的视频是在室外场景的环境下拍摄的, 摄像机是固定的, 该摄像机置于 4 米左右的高度, 视频序列的分辨率为 640×480 像素, 视频中共有些人从左向右行走, 有些人从右向左行走, 使其发生充分遮挡。算法是通过 C++ 来实现了模块与颜色相结合的单摄像机方法。图 3-11 显示了两组视频序列, 多处发生遮挡情况下多运动目标跟踪结果, 这里给出了该测试序列中比较有代表性的跟踪结果。在实验中, 被跟踪的人是通过不同颜色的外接矩形框标识的, 不同的颜色代表不同的人。

在图 3-11 中我们给出了人体遮挡的跟踪算法的处理过程。首先, 重建视频序列的背景, 利用与背景相减法及形态学方法检测出前景信息, 去除阴影(如图 3-11 (b))所示。然后利用本文提出的算法获取后验概率图(图 3-11 (b)中绘制出来的颜色图), 该图像中每个像素颜色表示该像素属于存储模型中某个人的概率最大。从图 3-11 (b)可以看出, 具有同一颜色的区域表明该区域与存储模型中的某个人体相匹配, 这样就实现了遮挡人体跟踪的目的。最后根据相同颜色的区域代表同一个人体目标, 将对遮挡情况进行分割, 如图 3-11 (b) 所示。从图中我们也可以看出当两人的距离较远时, 运动建模起了主要作用; 当运动目标较近时颜色建模起了主要作用。当有的人在摄像机视场内静止不动, 然后又开始运动, 这种情况本算法也能够很好地适应, 达到较好的跟踪效果, 因为存储模型中的人体尚未被删除。另外, 即便是某一帧的图像跟踪错误, 但是在下一帧仍有可能将其纠正过来。图 3-12 给出了两组视频序列的跟踪轨迹。为了证明本文提出的跟踪算法的鲁棒性, 本文用六组视频对本文提出的基于核密度估计方法与 Comaniciu 提出的基

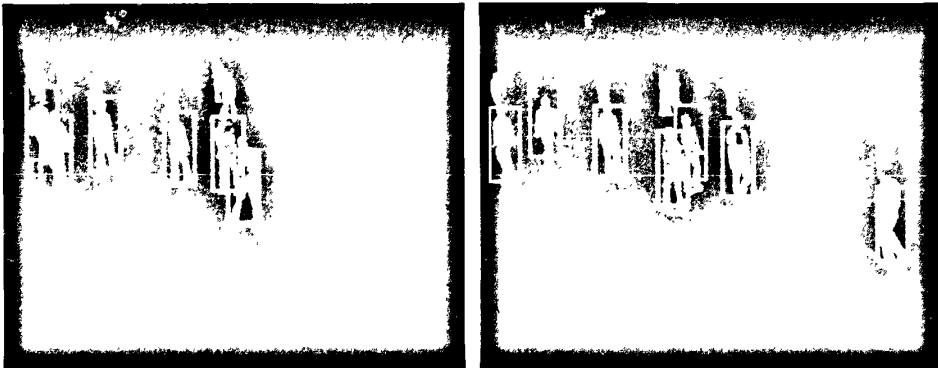
于 Mean shift[Comaniciu and Ramesh,2003]的跟踪方法进行了测试,并用准确率这个参数对六组视频进行比较(如图 3-13 所示)。准确率是指某个视频中所有人跟踪正确的次数与所有人出现的次数的比值。实验结果表明本文提出的方法能够在人体遮挡情况下实现多运动目标的有效跟踪而且具有较高的准确率,可以满足视频跟踪的可靠性和鲁棒性。

视频1 391帧

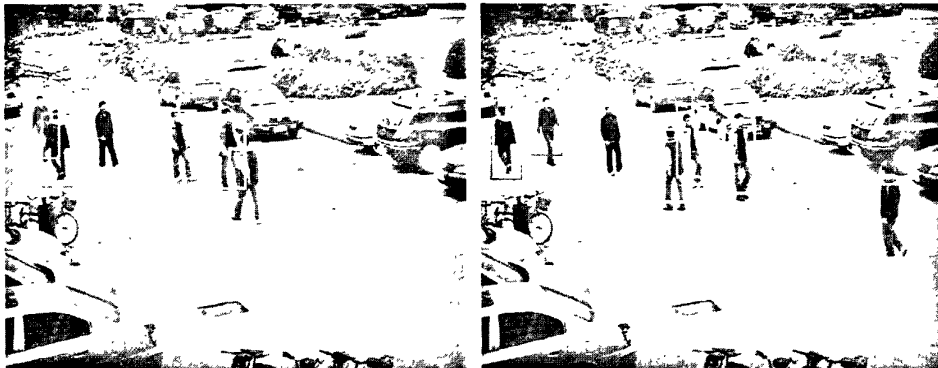
视频1 422帧



(a) 原始图像



(b) 检测出前景图像并获得后验概率图像最后对概率图像分割



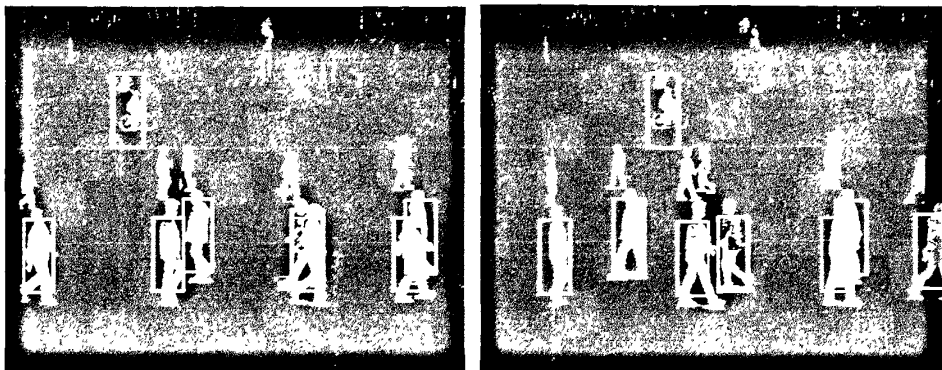
(c) 最终的跟踪结果

视频 2 127 帧

视频 2 146 帧



(a) 原始图像



(b) 检测出前景图像并获得后验概率图像最后对概率图像分割



(c) 最终的跟踪结果

图 3-11 两组视频序列的跟踪过程



(a) 视频 1 跟踪轨迹

(b) 视频 2 跟踪轨迹

图 3-12 两组视频序列的跟踪轨迹

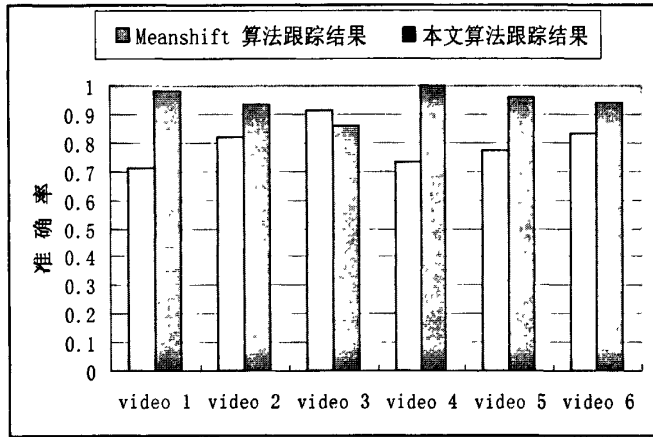


图 3-13 Mean shift 算法与本文算法准确率比较

3.6 本章小结

本文基于核密度估计的方法解决了遮挡情况下的人体跟踪的问题。本文未使用基于多相机的映射法信息，而是充分利用单相机的所有信息将背景建模、块建模、颜色建模、运动建模和人体的空间信息有机结合，然后得到最大后验概率的颜色图像，通过该图像进行分割，达到对遮挡情况下多人目标进行跟踪的目的。该算法不用限制衣服必须是同一颜色，相反，可以是块状混合颜色的衣服。本文提出的算法是全自动跟踪算法，有效地利用了均值漂移算法的梯度优化、空间信息和多通道像素信息来进行聚类的优点；本算法不再采取搜索窗口，而是充分利用运动信息，先将运动的前景信息提取出来，防止跟踪到静态的背景中。本文首先通过基于 Epanechnikov 核密度梯度估计对存储模型中的人体进行聚类，然后对当前的前景图像通过基于非参数的 Gauss 核密度估计颜色密度

函数和运动密度函数进行估计,在颜色建模过程充分利用了人体的空间信息,该算法有效地解决了遮挡情况下的人体跟踪的问题,提高了视频序列跟踪的可靠性。

同时基于核密度估计算法对某些特殊情况有一定的局限性,如在遮挡时两个人的衣服颜色是完全相同的,或者刚进入摄像机视场时,遮挡情况就已经发生了等,这些情况还需进一步解决。



4 有公共视场的多摄像机人体目标跟踪

4.1 引言

目前在视频目标跟踪方面的研究多集中在单摄像机环境下的单目标或多目标跟踪,然而由于单摄像机的视场范围有限,难以实现大范围、多角度、长时间对人的跟踪,这使得多摄像机的使用成为必然。为了扩大视频监控视场范围,人们将多个摄像机进行部分重叠的方式构建宽视场系统。多摄像机监控应用可以很好地解决监控中的阻塞问题,确保对目标长时间多角度的跟踪,可以弥补单摄像机监控的不足。多摄像机具有监控范围大,视场宽阔,全方位视点等优点,使得多摄像机环境下的人体跟踪成为计算机视觉领域内的一个主要研究方向。在多个摄像机组成的视频系统中,当人体目标从一个摄像机视场当中进入到另一个摄像机的视场时,如何实现人体目标的跨摄像机继续跟踪是跨摄像机系统中要解决的关键问题。

多摄像机跟踪中的目标交接是建立在单摄像机的检测和跟踪基础之上。在前面章节中已经实现了单摄像机的人体目标跟踪,本章要解决的是多摄像机之间重叠视场区域内目标交接问题,即在下一个摄像机视场的多个目标中准确识别出前一个摄像机视场中的跟踪的目标,从而得到同一个人体目标正确的标号。图 4-1 给出了相邻摄像机视场示意图,用 C1 和 C2 分别表示两个摄像机,每个摄像机的视场范围用两条射线所夹的部分表示,阴影部分是摄像机 C1 和摄像机 C2 的重叠视场,假设两个摄像机是同步的,那么在该重叠视场内的人体目标将同时被摄像机 C1 和 C2 所拍摄到, L2 为 C2 在 C1 中的视场分界线, R1 为 C1 在 C2 中的视场分界线, L2 和 R1 所夹的部分为重叠区域的公共视场。下面我们分析一下人体目标穿过整个摄像机系统的视场所需要的处理过程:假设人体目标从 A 点进入从左向右行走,最后从 G 点离开整个跨摄像机系统的过程,当人体目标由 A 点运动到 B 点时,由于该人体目标第一次出现在这个跨摄像机系统中,所以摄像机系统将为此个人体目标分配一个唯一的标号假设为 1 号目标;1 号目标由 B 点运动到 C 点过程中,只在摄像机 C1 的视场当中可见,此时的跟踪过程是以前章节讲述的单摄像机人体目标跟踪过程;当 C1 中 1 号目标由 C 点运动到 D 点时,该目标刚刚进入摄像机 C2 的视场并发现该目标,并视为一个新目标,必需给目标一个标号,由于该目标与 C1 中的 1 号目标是同一目标,所以该目标的标号为 1,与 C2 中的 1 号目标具有一致的标号;当 1 号目标处在视场重叠区域时,由 D 点运动到 E 点时,对摄像机 C1, C2 来说,都分别进行单个摄像机人体目标跟踪;当 1 号目标由 E 点运动到 F 点时,对摄像机 C1 来说,

该人体目标离开了摄像机 C1 的视场，要去除人体目标的人体模型；当 1 号目标由 F 点运动到 G 点时，对摄像机 C2 来说，继续对 1 号人体目标进行单摄像机进行跟踪；当 1 号目标由 G 点运动到 H 点时，该人体目标离开了摄像机 C2 的视场去除人体目标在摄像机 C2 中的目标模型，整个跟踪过程全部结束。用同样的操作流程，可以处理人体目标从 H 点进入从右向左行走，最后从 B 点离开整个跨摄像机系统的过程。

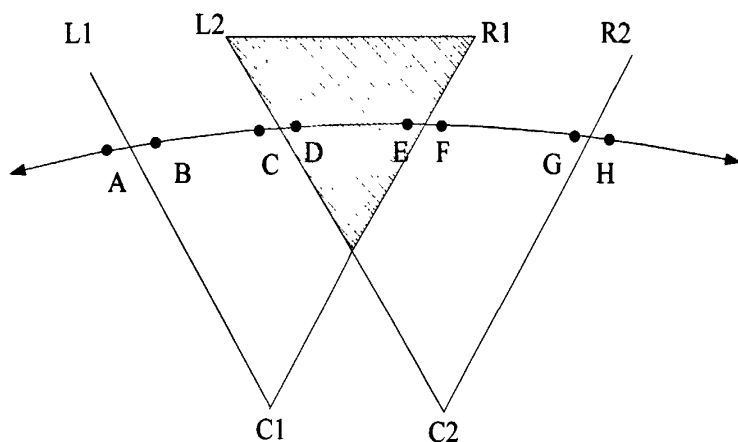


图 4-1 相邻摄像机视场示意图

当前的主要跨摄像机的人体目标跟踪算法有基于特征匹配的跟踪算法、基于单应矩阵的跟踪算法和基于视场分界线的跟踪算法。基于特征匹配的跟踪算法是解决跨摄像机人体目标跟踪中最基本的匹配算法，然而颜色特征受光照条件变化的影响比较严重，同时，不同摄像机的光感设备有差异，拍摄的视频也会收到光感设备噪声的影响。由于跨摄像机匹配要将多个摄像机的目标同时匹配，计算量也比较大。基于单应矩阵的跨摄像机跟踪算法是根据多个摄像机间的空间几何关系，在不同摄像机间进行映射对应。该算法比较准确，同时不用所有目标一一匹配，然而当人体目标相互比较接近时又容易产生误匹配。基于视场分界线的跟踪算法比较容易实现，然而跨摄像机人体目标跟踪无论是拍摄还是检测很难做到同步，这样一来也很容易产生误匹配的现象。

针对上述跨摄像机人体目标跟踪算法存在的问题，本文提出了一种基于颜色特征、视场分界线和单应矩阵变换相融合的跟踪算法[Wang and Liu,2009]，使跨摄像机人体目标跟踪结果更为准确、可靠。本文首先给出了如何在视频图像中划定不同摄像机之间的公共可视区域的方法，以确定某一摄像机在另一摄像机中的视场边界[Khan and Shah,2006]，然后当人体目标在摄像机的视场边界时进行跨摄像机跟踪或去除人体模型，

为准确对跨摄像机的人体目标进行准确定位，接下来本文采用了投影变换单应矩阵 (Homography) [Black and Ellis,2006;Yue et al,2004]的目标交接方法，最后重点阐述用颜色特征的跟踪方法来准确实现有重叠区域的跨摄像机人体目标跟踪。本文将该方法应用到真实场景中进行实验并对实验结果进行了分析。

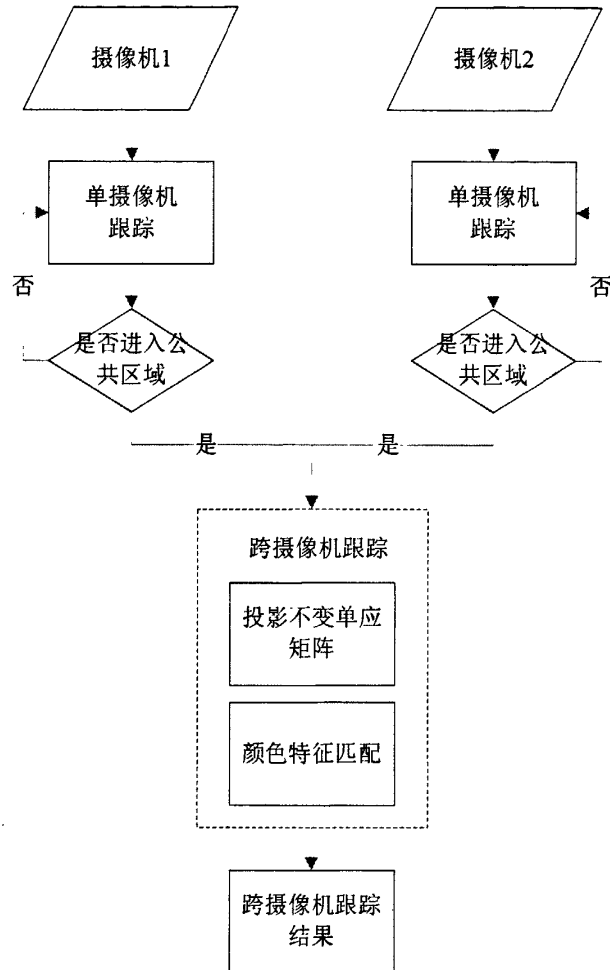


图 4-2 有公共视场的跨摄像机人体目标跟踪框图

4.2 视场分界线的目标交接

当一个人体目标刚进入一个摄像机的视场，如果这个人体目标在另一个摄像机的视场中是可见的，那么这个公共视场中的人被认为是同一个人，刚进入摄像机视场的人体目标的标号与另一个摄像机中的人体目标的标号是相同的；如果这个人体目标在另一个摄像机视场中不可见，那么这个人体目标不在公共视场中，刚进入摄像机视场的人体目标被赋予一个新的标号。

考虑这样一个场景，该场景是由两台同步拍摄的摄像机构成，它们对场景进行拍摄

时具有公共视场，然后有两个人在这个场景中行走。在 t_1 时刻，有两个人体目标刚好在摄像机 C_1 中都是可见的，因为这两个人体目标在摄像机 C_1 中是第一次出现，所以分别给他们分配个标号 P_1 、 P_2 ，在 t_2 时刻，有一个人体目标进入了摄像机 C_2 的视场，此刻我们如何给该目标分配标号，此刻该目标有三种可能性，该目标既可能是 P_1 也可能是 P_2 还有可能是新进来的目标 P_3 。如何确定这个刚进入的人体目标的标号是我们接下来要解决的问题。

用符号 L_i^x 表示将摄像机 C_j 中地平面中的第 x 根线映射到摄像机 C_i 中的线，其中 i ， j 分别用来表示摄像机的标号， $x \in \{l, r\}$ ， l 和 r 分别表示摄像机视场的左边界线和右边界线。 L_i^l 表示将摄像机 C_i 中地平面中的左边界线 l 映射到摄像机 C_i 中的线，其实仍是摄像机 C_i 中的左边界线 l ，可简化即作 L_i^l ，同样，摄像机 C_i 中的右边界线表示为 L_i^r 。

图 4-3 给出了一组室外情况下具有公共视场的跨摄像机人体目标交接图像。图像有一个人体目标从摄像机 C_2 的左侧进入该摄像机的视场，那么摄像机 C_2 的左边界线 L_2^l 映射到摄像机 C_1 的边界线为 L_1^{2l} ；一个人体目标从摄像机 C_1 的右侧进入该摄像机的视场，那么摄像机 C_1 的右边界线 L_1^r 映射到摄像机 C_2 的边界线为 L_2^{1r} 。在这里一共有四条直线，在每个摄像机中有两条直线。当人体目标跨入摄像机 C_2 时，在另一摄像机 C_1 正好能看到该人体目标，同样当人体目标跨入摄像机 C_1 时，在另一摄像机 C_2 也能看到该人体目标。在这种情况下，当人体目标从左侧线 L_2^l 跨入到摄像机 C_2 时，与此同时在摄像机 C_1 中该人体目标也一定正好跨过与线 L_2^l 所映射到 C_1 的线 L_1^{2l} 。同理，当人体目标从右侧线 L_1^r 跨入到摄像机 C_1 时，与此同时在摄像机 C_2 中该人体目标也一定恰好跨过与线 L_1^r 所映射到 C_2 中的线 L_2^{1r} 。

摄像机 C_1

摄像机 C_2

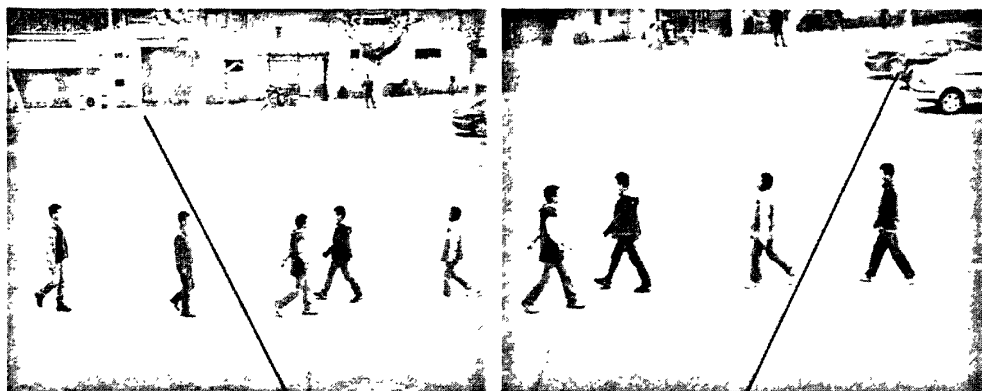


图 4-3 不同摄像机之间公共可视区域的确定示例

4.3 二维投影变换单应矩阵的计算

单应矩阵是指两个平面之间通过空间几何关系进行相互变换，它在摄像机标定、视觉测量、图像拼接、三维重建等多视几何中得到了广泛的应用。为实现有重叠的多摄像机下的人体目标跟踪，首先利用单应矩阵变换将某一摄像机中的人体目标映射到其它摄像机中所对应的区域。二维平面投影变换进行配准，其主要适应两种情况，一是，摄像机任意运动(旋转和平移)来拍摄同一平面图像；二是，摄像机绕其光心旋转或缩放来拍摄。本文是两个摄像机对同一平面进行拍摄的场景，在此过程中，两个摄像机中地平面的图像关系可以看作是一个摄像机的旋转和平移得到另一摄像机的图像。下面将推导两个图像的地平面是如何映射的。

在摄像机拍摄过程中通常要涉及到三个坐标系：图像坐标系、摄像机坐标系和世界坐标。图 4-4 给出了摄像机成像的几何模型。在该模型中，以摄像机的光心 O 为摄像机的原点， $X_c - Y_c - Z_c$ 为摄像机的三维坐标系坐标，其中 Z_c 与光轴重合， $x - y$ 为摄像机成像的图像坐标系， f 为摄像机的焦距， $P(X_c, Y_c, Z_c)$ 为摄像机坐标系下的空间一点，该摄像机在世界坐标系 $X_w - Y_w - Z_w$ 中。

线性摄像机的针孔模型可以通过图 4-4 表示，点 O 为摄像机光心，点 O 与 X 轴、 Y 轴和 Z 轴组成的坐标系称为摄像机的空间直角坐标系；点 O_2 为图像坐标系的原点，点 O_2 与 x 轴、 y 轴组成的坐标系称为图像平面直角坐标系。摄像机的空间直角坐标系 X 轴和 Y 轴分别与图像平面直角坐标系的 x 轴和 y 轴平行， Z 轴为摄像机的光轴，与图像平面垂直。光轴与图像平面的交点为图像坐标系的原点 O_2 。

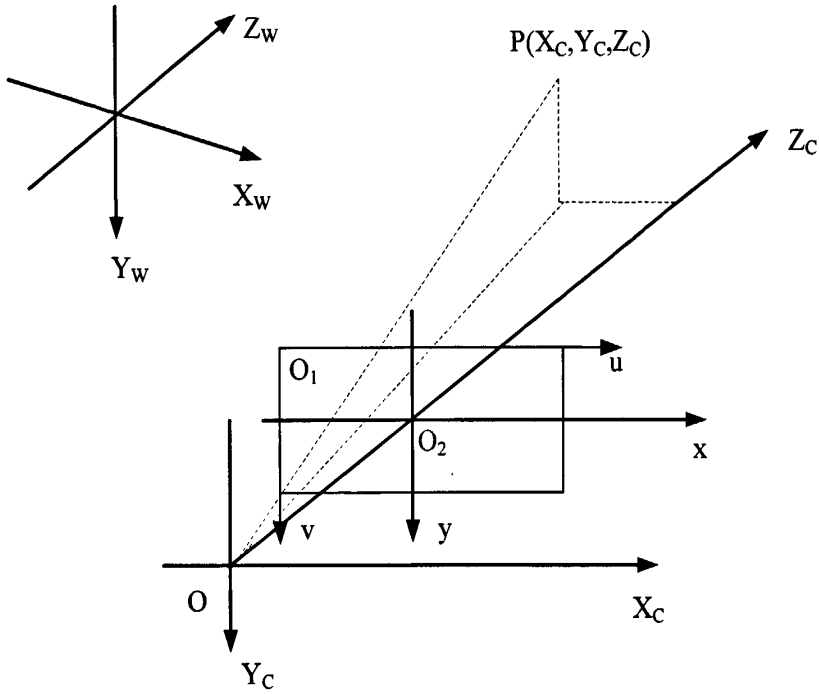


图 4-4 摄像机成像几何关系

在图 4-4 的图像坐标系 $x-y$ 中， (u, v) 是以摄像头的左上角的点 O_1 为原点，以像素为单位的图像坐标系的坐标，因为通常情况下，摄像机采集的图像是以 M 行 N 列的像素方式进行存储的， (x, y) 表示以光轴与像平面的交点 O_2 为原点，以毫米为单位的图像坐标系的坐标。为了用存储方式的像素坐标进行表示，所以必须将毫米为单位的图像坐标系的坐标转换为以像素为单位的图像坐标系的坐标。原点 O_1 在 $u-v$ 坐标系下的坐标为 (u_0, v_0) ，每个像素在 x 轴与 y 轴方向物理尺寸为 dx, dy ，则图像坐标系 $x-y$ 中任意一点 (x, y) 转换为图像坐标系下 $u-v$ 像素 (u, v) 的关系如式 (4.1) 所示。

$$u = \frac{x}{dx} + u_0, \quad v = \frac{y}{dy} + v_0 \quad (4.1)$$

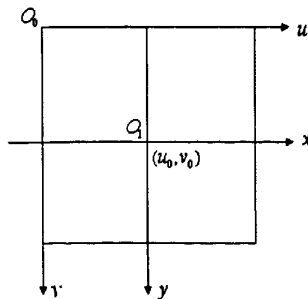


图 4-5 图像坐标系

为以后推导方便, 用齐次坐标与矩阵形式(4.2)来表示式(4.1)

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4.2)$$

在图 4-4 所示的摄像机几何模型中, 在摄像机坐标系下, P 的点坐标为 $X = (X_c, Y_c, Z_c)^T$, 将其映射到图像坐标系 O_2-xy 下点 p 的坐标为 $s = (x, y)^T$, 根据几何的比例关系可得到式(4.3)这种关系。

$$x = \frac{fX_c}{Z_c}, \quad y = \frac{fY_c}{Z_c} \quad (4.3)$$

用齐次坐标与矩阵表示上述透视投影关系可得式(4.4):

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (4.4)$$

世界坐标系是在环境中选择的一个基准坐标系, 可以根据描述和计算方便等原则自由选取。摄像机坐标系与世界坐标系之间的关系可以通过旋转矩阵 $R_{3 \times 3}$ 与平移向量 $t_{3 \times 1}$ 来描述。因此, 可以通过式(4.5)将世界坐标系下的空间中任意一点 P 的齐次坐标 $(X_w, Y_w, Z_w, 1)^T$ 转换为摄像机坐标下的齐次坐标 $(X_c, Y_c, Z_c, 1)^T$ 。

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4.5)$$

由式(4.2), 式(4.4), 式(4.5)联合可得到式(4.6):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_c} \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4.6)$$

将式(4.6)化简整理得式(4.7)。A矩阵完全由摄像机的内部参数结构决定，称这些参数为摄像机内部参数； $[R_{3 \times 3} \ t_{3 \times 1}]$ 是由摄像机与世界坐标系的空间关系决定的，将该参数定义为摄像机外部参数。

$$U = \frac{1}{Z_c} A[R_{3 \times 3} \ t_{3 \times 1}]x = \frac{1}{Z_c} A(R_{3 \times 3}x + t_{3 \times 1}) \quad (4.7)$$

其中：

$$U = (u, v, 1)^T$$

$$x = (X_w, Y_w, Z_w)^T$$

$$A = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

因为平面场景中的所有点满足平面方程 $\bar{n}^T x = d$ ，所以 $\bar{n}^T x \frac{1}{d} = 1$ 则式(4.7)可表示成式(4.8)。

$$U = \frac{1}{Z_c} A(R_{3 \times 3}x + t_{3 \times 1}) = \frac{1}{Z_c} A(R_{3 \times 3}x + \frac{1}{d}t_{3 \times 1}\bar{n}^T x) = \frac{1}{Z_c} A(R_{3 \times 3} + \frac{1}{d}t_{3 \times 1}\bar{n}^T)x \quad (4.8)$$

同理，对相同世界坐标系中的地平面场景中的同一点，用内参相同的另一摄像机进行拍摄，可得该图像坐标系与世界坐标系的关系可用式(4.9)表示：

$$U' = \frac{1}{Z'_c} A(R'_{3 \times 3} + \frac{1}{d}t'_{3 \times 1}\bar{n}^T)x \quad (4.9)$$

根据式(4.8)可得式(4.10)，将其代入式子(4.9)可得式(4.11)：

$$x = Z_c (R_{3 \times 3} + \frac{1}{d}t_{3 \times 1}\bar{n}^T)^{-1} A^{-1}U \quad (4.10)$$

$$U' = HU \quad (4.11)$$

其中：

$$H = \frac{Z_c}{Z'_c} A(R'_{3 \times 3} + \frac{1}{d}t'_{3 \times 1}\bar{n}^T)(R_{3 \times 3} + \frac{1}{d}t_{3 \times 1}\bar{n}^T)^{-1} A^{-1}$$

由(4.11)可得：

$$U = H^{-1}U' \quad (4.12)$$

由此可以看到,在已知单应矩阵 H 的情况下,可以由视角1中图像平面的点计算出其在视角2图像平面中的对应位置。图4-2给出了两个摄像机对地平面上的投影产生的图像可以通过矩阵单应性矩阵 H 来进行映射。一般的场景中,目标均以地面运动为主,如果已知多摄像机系统中所有视角平面之间的单应矩阵 H ,就可以由目标在某一视角的地面位置计算其在其他视角中的位置。计算homography所需要的对应点对可以通过人工标定也可以采用在场景中放置特征标记的方法获得。

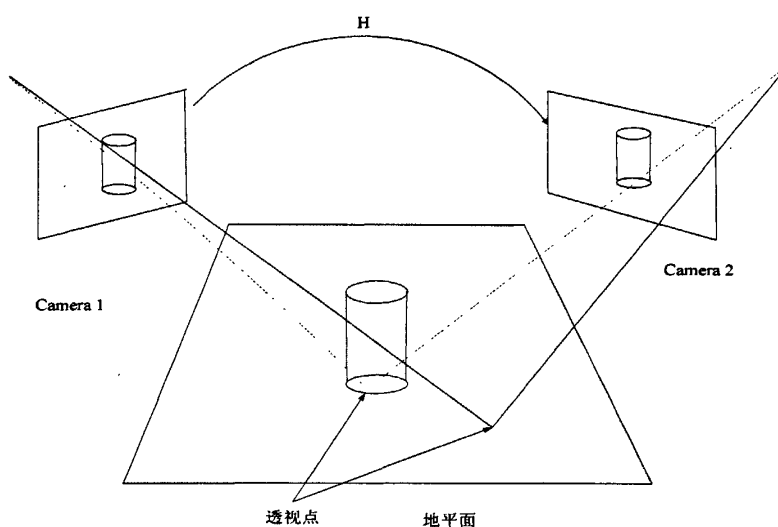


图 4-6 单应矩阵变换示意图

当前共有四种方法可以实现对单应矩阵进行估计:代数法、几何法、鲁棒估计法及统计估计法。代数法就是通过代数距离的最小方法来线性估计进行求解单应矩阵的方法;几何法是通过非线性迭代计算来获取单应矩阵的方法。鲁棒估计法是通过M-估计法、RANSAC(Random sample consensus)算法等方法对包含异常数据的样本数据集进行估计,剔除异常样本数据的影响来估计单应矩阵。统计估计法是用概率模型的方法进行估计单应矩阵。代数法与几何法只能应用于测量数据比较准确的情况下,鲁棒估计法一般应用于包含错误对应点的情况,统计估计法一般实现起来比较复杂。这四种方法中线性方法是比较常用的,因为该方法易于实现并且效率高、准确度也较高。

下面定义一个 3×3 的单应性矩阵,这个矩阵就是我们要求得的矩阵。

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \quad (4.13)$$

(u'_i, v'_i) 与 (u_i, v_i) 是分别是两个摄像机对地平面上的同一点拍摄得到的一组对应点, 可以通过单应矩阵 H 进行转换, 得到式(4.14).

$$\begin{bmatrix} u'_i \\ v'_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \quad (4.14)$$

将式(4.14)写成方程组可得到式(4.15),

$$\begin{cases} u'_i = u_i h_{11} + v_i h_{12} + h_{13} \\ v'_i = u_i h_{21} + v_i h_{22} + h_{23} \\ 1 = u_i h_{31} + v_i h_{32} + 1 \end{cases} \quad (4.15)$$

式(4.15)是线性相关矩阵, 可将其化简成式(4.16)

$$\begin{cases} u'_i u_i h_{31} + u'_i v_i h_{32} + u'_i = u_i h_{11} + v_i h_{12} + h_{13} \\ v'_i u_i h_{31} + v'_i v_i h_{32} + v'_i = u_i h_{21} + v_i h_{22} + h_{23} \end{cases} \quad (4.16)$$

一组对应点可生成 2 个方程, 那么 N 组对应点可生成 $2N$ 个方程, 那么至少需要 4 组对应点才可求出单应矩阵的解。图 4-7 给出了利用单应矩阵变换得到地平面上的点对应结果。

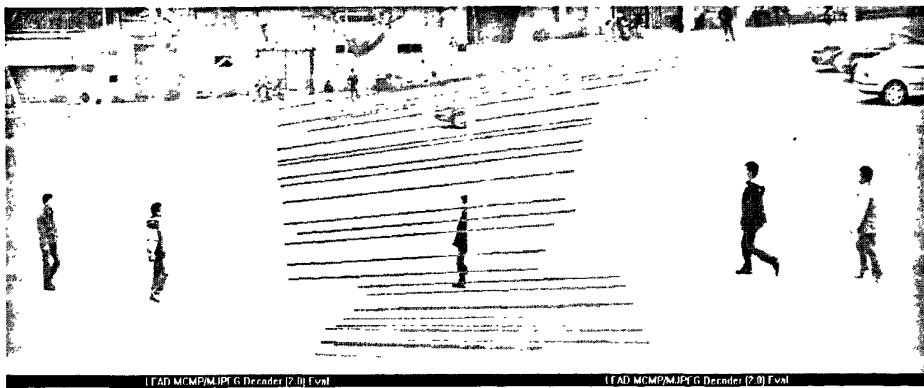


图 4-7 利用单应矩阵变换得到地平面上的点对应

4.5 颜色特征跟踪

在实际应用中, 由于两个摄像机不能完全同步, 所以只依赖于单应矩阵的方法所得到的效果并不是很好, 为了解决该问题, 本文提出了基于颜色特征的匹配跟踪作为进一步判断的依据。图 4-8 给出了跨摄像机人体目标颜色跟踪的流程图。

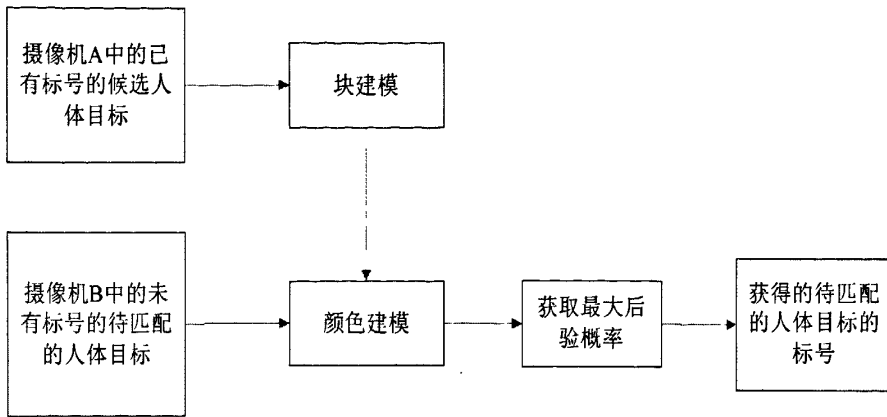


图 4-9 跨摄像机颜色跟踪的流程图

假设由 r 台摄像机 C_1, C_2, \dots, C_r 构成一个有重叠视场的系统, 在该环境中, 有 n 个人体目标在该系统中出现 (人体目标的个数是未知的), 在不同的时刻, 人体目标出现在不同的摄像机视场中。用 $O_j = \{O_{j,1}, O_{j,2}, \dots, O_{j,m}\}$ 表示第 j 个摄像机拍摄到的 m 个人体目标集合。 $O_{j,a}$ 表示进入摄像机 C_j 视场到离开该摄像机视场的第 a 个人体目标的表面模型。跨摄像机跟踪实质就是在不同的摄像机中找到属于同一个人体目标的跨摄像机匹配。用 $k_{r,a}^{s,b}$ 表示在第 r 个摄像机中的第 a 个人体目标 $O_{r,a}$ 与在第 s 个摄像机中的第 b 个人体目标 $O_{s,b}$ 进行匹配。这样, 这个多摄像机系统就会产生一个对应集合 $K = \{k_{r,a}^{s,b}\}$, 当且仅当人体目标模型 $O_{r,a}$ 与人体目标模型 $O_{s,b}$ 对应同一个人体模型时, $k_{r,a}^{s,b} \in K$ 。用 Ω 表示多摄像机跟踪的解空间。多摄像机跟踪的解空间 Ω 中后验概率最大的作为最终的正确的解空间 K' , 用公式(4.17)表示。

$$K' = \arg \max_{K \in \Omega} \prod_{k_{i,a}^{j,b} \in K} P(O_{i,a}, O_{j,b} | k_{i,a}^{j,b}) \quad (4.17)$$

为了更好的阐述上述算法, 图 4 给出了两个摄像机 r 和 s 中的人体目标, 在摄像机 r 中有 3 个已有标号的候选跨摄像机的人体目标 1, 2, 3, 在摄像机 s 有 2 个未被标号的待匹配的跨摄像机人体目标, 多摄像机跟踪的解空间为 $\Omega = \{K_1, K_2, K_3, K_4, K_5, K_6\}$ 其中 $K_1 = \{k_{r,1}^{s,1}, k_{r,2}^{s,2}\}$, $K_2 = \{k_{r,1}^{s,1}, k_{r,2}^{s,3}\}$, $K_3 = \{k_{r,1}^{s,2}, k_{r,2}^{s,1}\}$, $K_4 = \{k_{r,1}^{s,2}, k_{r,2}^{s,3}\}$, $K_5 = \{k_{r,1}^{s,3}, k_{r,2}^{s,1}\}$, $K_6 = \{k_{r,1}^{s,3}, k_{r,2}^{s,2}\}$ 。通过式(4.17)可以得到最大后验概率解为 $K' = K_1$ 。

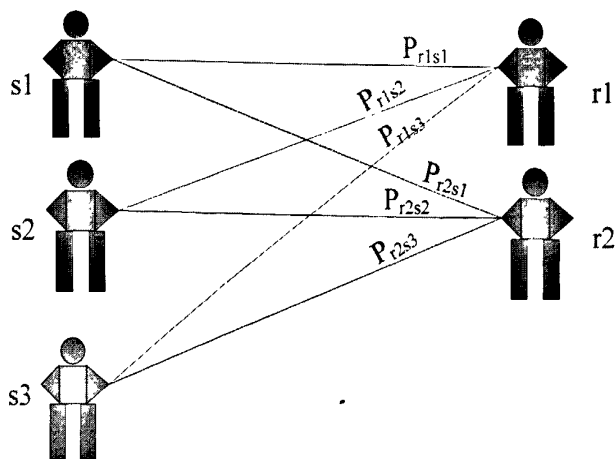


图 4-10 两个摄像机情况下的人体目标对应的解空间

$$P(O_{i,a}, O_{j,b} | k_{i,a}^{j,b}) = \frac{1}{N} \sum_{\mathbf{c}_{x,y} \in O_{i,a}} P(\mathbf{c}_{x,y} | M_b) \quad (4.18)$$

其中, $\mathbf{c}_{x,y}$ 表示第 i 个摄像机第 a 个人体目标 $O_{i,a}$ 的像素值, N 表示人体目标 $O_{i,a}$ 的像素值的数目, M_b 是通过 3.3.2 节的块建模方法得到的人体模型 $M_{t-1,k} = \{B_1, B_2, B_3, B_4, \dots\}$, 将得到当前像素 $\mathbf{c}_{t,x,y}$ 的属于人体模型 $M_{t-1,k}$ 的概率 $P_{M_{t-1,k}}(\mathbf{c}_{t,x,y})$, 即:

$$\begin{aligned} P(\mathbf{c}_{x,y} | M_k) &= \max\{P(\mathbf{c}_{x,y} | B_1), P(\mathbf{c}_{x,y} | B_2), P(\mathbf{c}_{x,y} | B_3), \dots\} \\ &= \max\{P_{B_1}(\mathbf{c}_{x,y}), P_{B_2}(\mathbf{c}_{x,y}), P_{B_3}(\mathbf{c}_{x,y}), \dots\} \end{aligned} \quad (4.19)$$

其中 $P_{B_i}(\mathbf{c}_{t,x,y})$ 为颜色密度函数, 由公式 (3.15) 看出, 像素 $\mathbf{c}_{t,x,y}$ 属于模型 $M_{t-1,k} = \{B_1, B_2, B_3, B_4, \dots\}$ 的概率是通过最大后验概率得到的, 即通过获取所有 B_i 中颜色密度函数 $P_{B_i}(\mathbf{c}_{t,x,y})$ 中概率最大的值作为该像素属于该模型的概率。

本文的颜色建模是颜色密度函数 $P_{B_i}(\mathbf{c}_{t,x,y})$ 已在 3.3.3 节中详细介绍。本文的颜色建模就是通过多元核乘积估计进行建模的。为了光亮度改变具有较强的鲁棒性, 本文对每个像素的色度进行规范化处理, 表示成一个向量 $\mathbf{x} = (r, g, s)$, 其中 $r = \frac{R}{R+G+B}$, $g = \frac{G}{R+G+B}$, $s = \frac{R+G+B}{3}$ 。在公式(3.23)中, 如果给定样本空间 $S_b = \{\mathbf{x}_i = (r_i, g_i, s_i)\}$, $d=3$, $\sigma_r, \sigma_g, \sigma_s$ 分别表示 r, g, s 的带宽, 则颜色建模的颜色密度函数既可表示为:

$$P_s(r, g, s) = \frac{1}{N\sigma_r\sigma_g\sigma_s} \sum_{i=1}^N K\left(\frac{r-r_i}{\sigma_r}\right) K\left(\frac{g-g_i}{\sigma_g}\right) K\left(\frac{s-s_i}{\sigma_s}\right) \quad (r_i, g_i, s_i) \in S \quad (4.20)$$

其中,核函数服从高斯分布:

$$K(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2} \quad (4.21)$$

核密度估计的方法使每个样本数据 $S = \{\mathbf{x}_i\}$ 起到了平滑的作用,对某一个数据 $\mathbf{x} = \{r, g, s\}$ 概率的大小主要取决于它与样本数据 $S = \{\mathbf{x}_i\}$ 的色度空间的距离、带宽 σ 及所采用的核函数。本文对带宽 σ 的选取是用点估计法求取给定样本空间 $S_b = \{\mathbf{x}_i = (r_i, g_i, s_i)\}$ 的方差来实现。关于多元密度估计带宽的选取可参考文献 [Duong and Hazelton, 2005]。

4.6 实验结果和分析

本文用视场分界线、二维投影变换单应矩阵和颜色特征相融合的算法对室外场景下拍摄的有重叠区域的视频序列进行测试。本文给出了两组不同环境下拍摄的跨摄像机有遮挡的多人体目标跟踪结果。视频序列的分辨率为 640×480 像素,视频中共有些人从左向右行走,有些人从右向左行走,使其发生充分遮挡。图 4-11 显示了有公共视场的跨摄像机的人体目标跟踪过程,多处发生遮挡情况下多运动目标跟踪结果,这里给出了该测试序列中比较有代表性的跟踪结果。在跟踪过程中,本文用不同颜色的外接矩形框标识不同人体目标,不同的颜色代表不同的人。

在图 4-11 中给出了多人跟踪算法的处理过程。有重叠区域的多摄像机人体目标跟踪是在单摄像机跟踪的基础上来实现的。图 4-11(a)给出有重叠区域的两个摄像机的原始图像;图 4-11(b)给出了利用背景重建并与背景相减法及形态学方法检测出前景信息;图 4-11(c)显示了利用第 3 章提出的算法获取后验概率图,该图像中每个像素颜色表示该像素属于存储模型中这个颜色的人的概率最大;图 4-11(d)给出了根据相同颜色的区域代表同一个人体目标的分割图;图 4-11(e)给出了在单摄像机跟踪基础上得到的跨摄像机跟踪结果。图 4-12 给出了场景 A 的跨摄像机人体目标跟踪结果,图 4-13 显示了场景 B 的跨摄像机人体目标跟踪结果。图 4-14 场景 A 的跨摄像机人体目标跟踪轨迹;图 4-15 场景 B 的跨摄像机人体目标跟踪轨迹。图 4-16 场景 A 和场景 B 人体目标跟踪的准确率。准确率是指在不同场景下的整个摄像机系统的所有人跟踪正确的次数与所有人出现的次数的比值。实验结果表明本文提出的视场分界线、二维投影变换单应矩阵和颜色特征相融合的算法能够实现多运动目标的有效跟踪而且具有较高的准确率。

摄像机 1



摄像机 2



(a) 原图



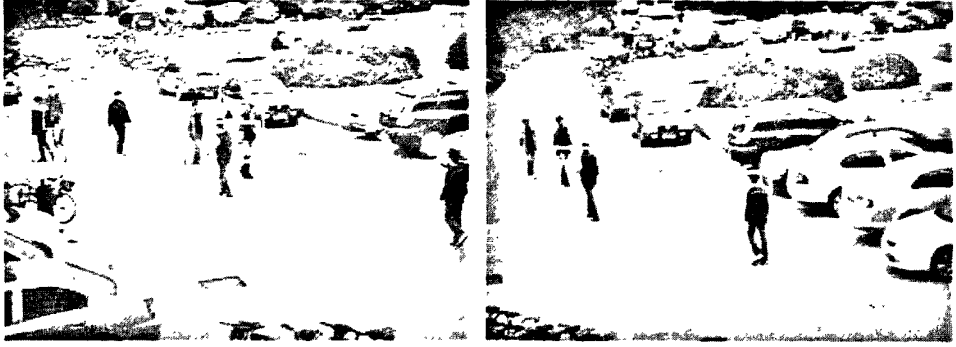
(b) 运动人体目标检测



(c) 检测出前景图像并获得后验概率图像



(d) 对最大后验概率图像分割



(e) 跨摄像机的跟踪结果

图 4-11 有公共视场的跨摄像机的人体目标跟踪过程

摄像机 1

摄像机 2

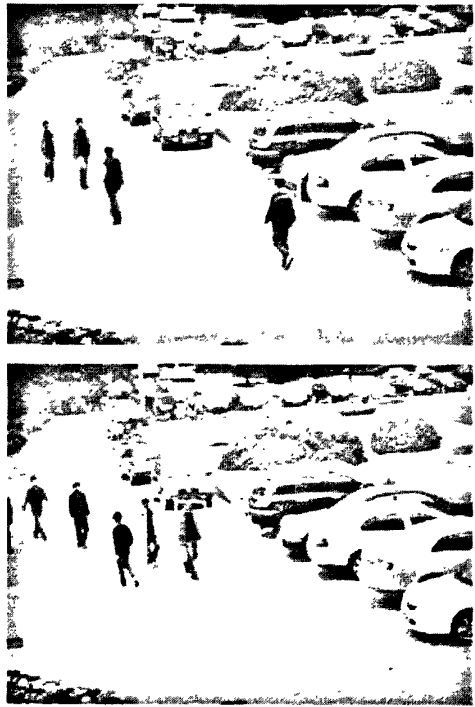
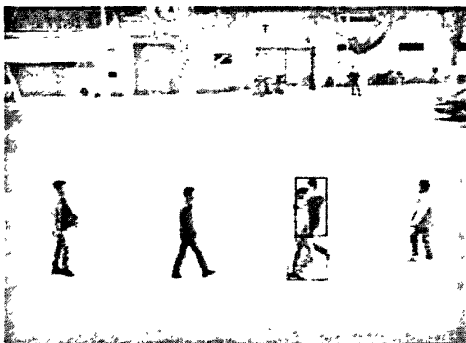


图 4-12 场景 A 的跨摄像机人体目标跟踪

摄像机 1

摄像机 2



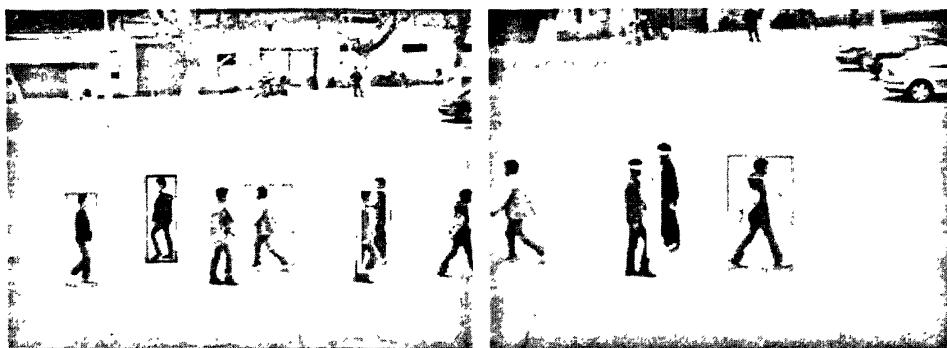
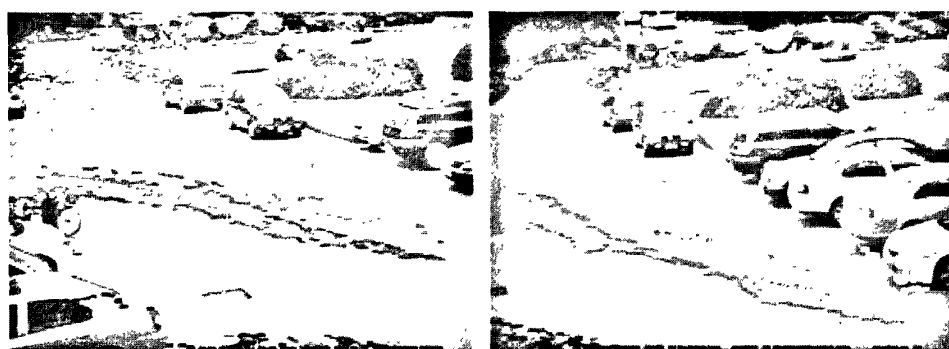


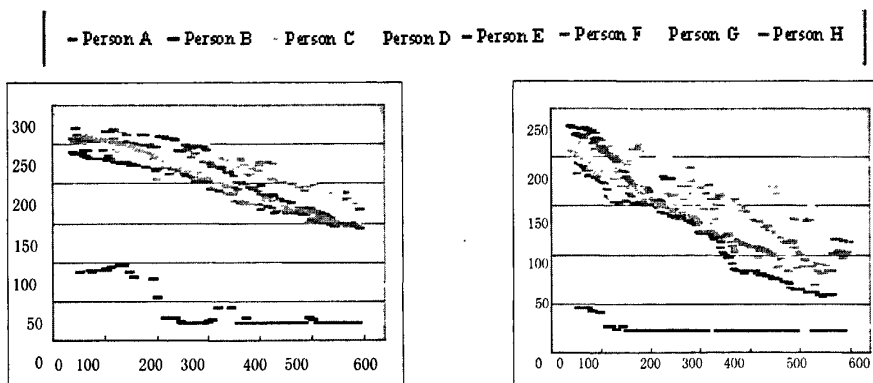
图 4-13 场景B的跨摄像机人体目标跟踪

摄像机 1

摄像机 2



(a) 真实环境下的跨摄像机人体目标跟踪轨迹

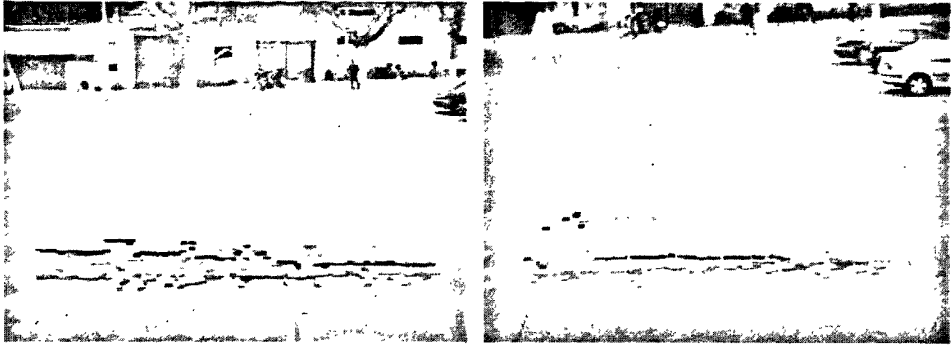


(b) 跨摄像机人体目标跟踪轨迹的平面坐标

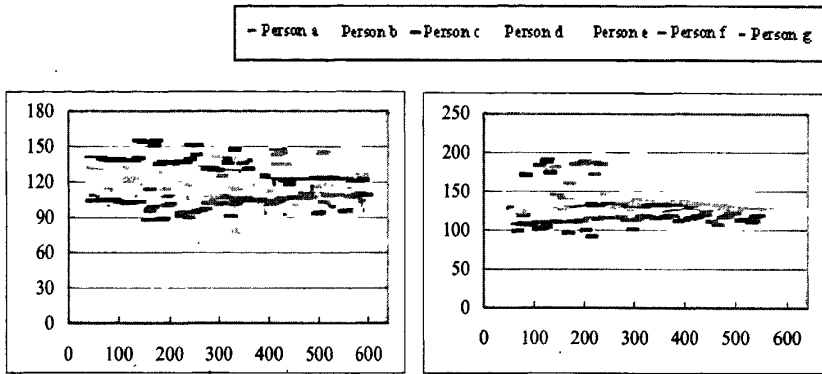
图 4-14 场景 A 的跨摄像机人体目标跟踪轨迹

摄像机 1

摄像机 2



(a) 真实环境下的跨摄像机人体目标跟踪轨迹



(b) 跨摄像机人体目标跟踪轨迹的平面坐标

图 4-15 场景 B 的跨摄像机人体目标跟踪轨迹

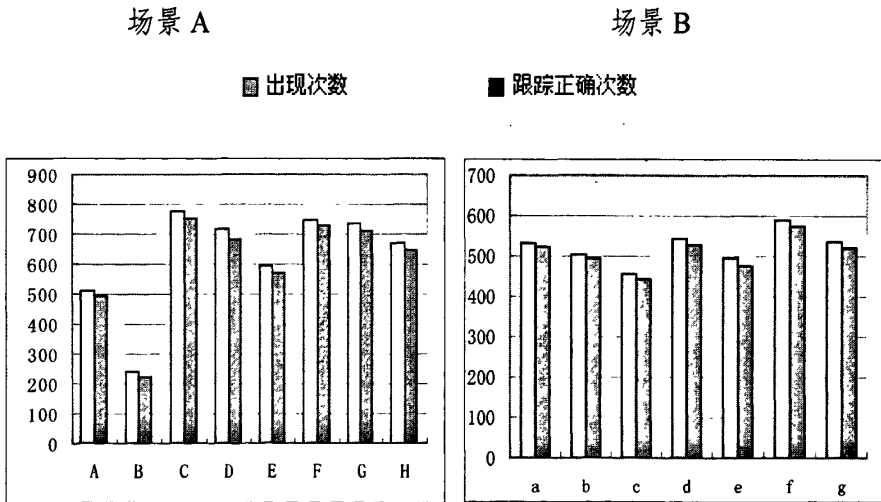


图 4-16 场景 A 和场景 B 人体目标跟踪的准确率

4.7 本章小结

传统方法是在摄像机重叠视场区域内采用特征匹配方法、基于视场分界线的目标交

接算法, 利用投影不变量等。然而, 特征匹配计算量较大, 受摄像机参数设置因素影响较大, 实用性不高; 而基于视场分界线的目标交接算法由于新目标的检测存在延时, 仅利用该方法来实现一致性标号容易出错; 计算目标在相邻摄像机中的投影位置来进行目标交接的方法, 在多个目标相距较近的情况下, 投影过程中产生的误差会造成目标对应错误。

本章针对上述问题, 提出了解决有公共视场的多摄像机人体目标跟踪的方法。本章首先确定跨摄像机的公共视场分界线, 然后利用投影变换单应矩阵计算跨摄像机人体目标的对应关系, 然而在实际应用中, 视频并不是完全同步或是有遮挡情况, 所以在投影变换单应矩阵的基础上采用跨摄像机最大后验概率进行匹配。文中提出的视场分界线、二维投影变换单应矩阵和颜色特征的最大后验概率相融合的算法, 通过构建多个场景的 2 台摄像机组成的平台进行实验, 实验表明能较好地解决传统方法存在的问题。

5 无公共视场的多摄像机人体目标跟踪

5.1 引言

在视频监控中,单摄像机的视场是有限的,而且由于经济和计算量等方面的原因,用大量重叠视场的摄像机把所有监控区域都覆盖也是不现实的。因此,人们不得不用无重叠视场的摄像机进行视频监控。在这种情况下,无重叠视场摄像机视频监控就成为了近年来的研究热点。其中,不同场景图像间的目标匹配则是一个亟待解决的关键问题。然而在跨摄像机匹配过程中,颜色特征是匹配跟踪过程中很重要的一个特征。在实际拍摄中,图像颜色失真严重影响了跟踪的准确性。图像生成的必要条件是光源、被拍摄的物体与图像采集设备。因此,在跨摄像机跟踪过程中,影响跟踪效果主要是图像颜色的失真,失真主要原因是由光源、图像采集设备和被拍摄的物体产生的。

(1) 光源对颜色失真的影响

光照是图像生成的灵魂,在室外情况下,由于受天气的影响,太阳的光照条件是变化的;在室内情况,也会因人体目标离光源的远近影响光照条件;彩色物体在不同的光源照射下显示的颜色是有偏差的。同一种颜色,在自然光、卤素灯、钨丝灯等不同光源下所呈现的颜色是不一样的,这种差异是由光源的色温[汤顺青,1990]不同造成的。

(2) 图像采集设备对颜色失真的影响

数码摄像机是一种新型的图像采集设备[王永刚和卫保国,2000],其颜色还原效果与光学系统和光电传感器件 CCD 和 CMOS 的性能密切相关,数码摄像机的参数设置也会对图像颜色造成影响。由于光学镜头会吸收颜色光的能量,并且吸收多少受颜色的差异而不同。这样也就必然造成了图像的失真。同时,由于镜头有剩余像差,这样也会导致图像失真。光电传感器的性能也会导致颜色失真,不同的波长在传感器上的光电转换也是不同的。

5.2 颜色校正的方法

5.2.1 基于映射的颜色校正方法

(1) 基于多项式回归的颜色校正

多项式回归的颜色校正基本思想就是利用一组方程组,对输入的 r, g, b 三个变量通过选择项组合,生成更高的 d 维($d \geq 3$)的输入数据,这些数据通过线性组合对目标数据进行拟合,同时再求出方程组的系数。

一阶多项式颜色校正就是通过线性变换将源颜色空间的像素值转换成目标颜色空间像素值的方法。下面用矩阵的方式来描述一阶多项式构成的方程组。设校正前信号的三维彩色空间像素值向量 $m = \{r_i, g_i, b_i\}^T$ ，校正后信号的像素值为向量 $m' = \{r'_i, g'_i, b'_i\}^T$ ，通过颜色变换矩阵 $R_{3 \times 3}$ 及向量 $C = \{c_1, c_2, c_3\}^T$ 对三维空间彩色像素值向量 M 进行校正为 $m' = R_{3 \times 3}m + C$ 即式(5.1)。

$$\begin{bmatrix} r'_i \\ g'_i \\ b'_i \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} r_i \\ g_i \\ b_i \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \quad (5.1)$$

给出 n 组校正前后像素的匹配对 $M_{3 \times n}$ ， $M'_{3 \times n}$ 。 $M_{3 \times n}$ 是校正前信号的三维空间彩色像素的输入的 $3 \times n$ 矩阵， $M'_{3 \times n}$ 校正目标的三维空间彩色像素的输出的 $3 \times n$ 矩阵。用矩阵表示为式(5.2)，根据最小二乘法求解该方程组即可求出颜色变换矩阵 $R_{3 \times 3}$ 的值。

$$M'_{3 \times n} = RM_{3 \times n} + C \quad (5.2)$$

基于一阶多项式颜色校正是校准方法中较普遍的一种方法，实现简单易效率高等特点，然而由于转换矩阵维数较小，一般效果并不是很理想。对于更复杂的高阶多项式，可以在一阶多项式的基础上加入 $rg, rb, gb, r^2, g^2, b^2, r^2g, rgb, r^2b, rg^2, r^3 \dots$ 等项构成。尽管多项式的项数增加可以提高算法的精度，然而并不是项数增加，校正效果就一定好。因为项数的增加得越多，广泛应用性就越低。为了综合衡量计算量、复杂度以及校正后的图像质量，一般采用项数为 $r, g, b, rg, rb, gb, r^2, g^2, b^2$ 模型。采用二阶多项式表示为 $v' = R_{3 \times 9}v + D$ 。

其中

$$v = (r_i, g_i, b_i, r_i g_i, r_i b_i, g_i b_i, r_i^2, g_i^2, b_i^2)^T$$

$$v' = (r'_i, g'_i, b'_i, r'_i g'_i, r'_i b'_i, g'_i b'_i, r_i'^2, g_i'^2, b_i'^2)^T$$

$$D = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9)^T$$

$$R_{3 \times 9} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{19} \\ r_{21} & r_{22} & \cdots & r_{29} \\ r_{31} & r_{32} & \cdots & r_{39} \end{bmatrix}$$

若令

$$\begin{aligned}x_{i,1} &= r_i, & x_{i,2} &= g_i, & x_{i,3} &= b_i, \\x_{i,4} &= r_i g_i, & x_{i,5} &= r_i b_i, & x_{i,6} &= g_i b_i, \\x_{i,7} &= r_i^2, & x_{i,8} &= g_i^2, & x_{i,9} &= b_i^2\end{aligned}$$

则二阶多项式方程又转换为线性方程, 则可用一阶多项式相同的方法进行求解矩阵 $R_{3 \times 9}$ 和向量 D 。

Hung 和 Kang 等人 [Hung,1991;Kang,1992;王永刚 和 王爱民,2001]采用多项式来描述图像的像素值和期望的像素值之间的关系。该算法的优点是, 源颜色空间与目的颜色空间转换比较方便, 交换输入数据与输出数据即可实现转换, 对未学习到的数据可以通过多项式估计得到, 可以弥补间隔采样的不足。然而, 多项式校正的准确性取决于源颜色空间和目的颜色空间的映射及多项式的选择有关系。由于多项式估计是全局估计, 难免局部估计会失真, 所以局部估计最好有更细致的局部函数来估计。

(2) 查找表法

查找表法是一种基于映射的颜色校正方法, 该方法准确率高, 实用性强, 能应用于非线性转换。常用的有两种查找表法: 直接查找表和映射函数查找表。直接查找表的优点是使用简单, 应用广泛; 缺点是要统计大量的数据, 对表更新缺乏实时性。映射函数查找表先通过对源数据空间进行量化, 通过已知对应的样本找到对应关系。对于任意给定的一个颜色值, 通过插值法计算输入值对应的输出颜色值。该方法的映射函数的选择显得尤为重要。通过该方法可以推测出未被学习到的数据的输出值。

映射函数查找表法来实现颜色校正可通过三步骤:

1. 分割: 按照一定的采样间隔对源空间进行量化;
2. 重构: 根据学习样本数据对, 通过插值、映射等算法获得源空间量化块对应的目标颜色值;
3. 建表: 将源空间的颜色值和目标空间的颜色值对应的数据对有顺序地放入表格中, 形成查找表。

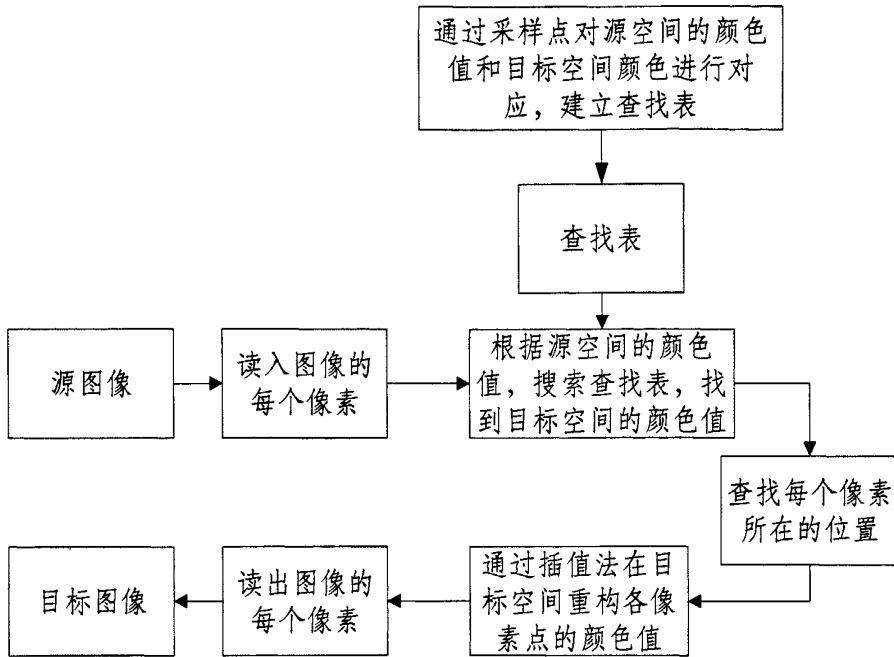


图 5-1 基于查找表的颜色校正示意图

(3) 神经网络法

神经网络也是彩色校正常用的方法 [Chang and Chang,2002;Kang and Anderson,1992;Kocheisen et al,2002;Schettini et al,1995;Vrhel and Trussell,2002;赵忠旭和沈兰荪,2000], 间接查找表法也可以用神经网络法来实现。由于神经网络的具有使用内存空间小且得到高精度外插的优点, 同时可以逼近任意的非线性映射, 然而缺点是学习训练时间长, 数据量大等问题。前馈网络(Back Propagation) 多层感知器在实际中应用比较广泛 [Kang and Anderson,1992;Schettini et al,1995;赵忠旭 和 沈兰荪,2000]。CMAC 网络也有应用, 因为该网络易于实现且具有较快的收敛速度。Kang 和 Anderson [Kang and Anderson,1992]通过神经网络来构造函数模型达到彩色校正的目的, 在某种程度上, 要比多项式法准确度高。Vrhel 和 Trussell [Vrhel and Trussell,2002]通过对多项式颜色校正、全局线性法颜色校正、局部线性法校正及神经网络法颜色校正进行比较, 认为全局线性法颜色校正和多项式颜色校正的效果没有局部线性法颜色校正和神经网络法颜色校正的效果好, 而且神经网络法的颜色校正效果最佳。

5.2.2 基于光谱反射率还原的颜色校正方法

基于光谱反射率还原的颜色校正方法的核心是将采集得到的图像颜色数据还原为被摄物体对应点的光谱反射率, 进而可以确定物体在某一特定的光照下的颜色值。根据数码摄像机的成像过程, 具有一定光谱分布的光照射在被摄物体上, 被摄物体依据自身的

光谱反射率特性反射照明光,形成入射数码相机光谱辐射,再由数码相机各通道的光谱灵敏度得到相机各通道的颜色值输出。应用更为广泛的则是采用近似线性函数的成像模型,不考虑数码相机中的信号处理和设备噪声,用求和来近似表达积分,因此摄像机的第 k 个通道的成像模型可表示为

$$I_{k(x,y)} = \sum_{\lambda=\lambda_1}^{\lambda=\lambda_2} P(\lambda)R(\lambda)C_k(\lambda) \quad k=1,2,3 \quad (5.3)$$

式中 λ 是波长, λ_1 和 λ_2 分别是波长离散求和的下限和上限区间, $P(\lambda)$ 是照明的光谱功率分布, $R(\lambda)$ 是物体的光谱反射率, $C_k(\lambda)$ 表示摄像机的第 k 个通道的光谱敏感系数。 $k=1,2,3$ 分别表示摄像机的红、绿、蓝三个通道。

光谱反射率是通过一组波长离散的数据来表示,包含数据的维数比图像颜色维数多。由于在数学中很难实现通过高维计算光谱反射率,因此通常要采用降维的方法减少数据量,用少量的低维的方法来表示光谱反射率。光谱反射率数据降维的方法一般有非线性和线性两种方法。

光谱反射率数据降维的一种途径是采用非线性系统对光谱数据进行精简。Usui 等人 [Usui et al,1992] 认为非线性系统可以将光谱反射率数据用三个参数编码,并且能够有效还原。他们采用五层神经网络(前面三层为编码网络,后面三层为解码网络,中间层采用三个隐层单元),对 1280 个 Munsell 色块的表面光谱反射数据进行映射编码及解码。Owens 等人 [Owens et al,2000] 将上述研究进一步扩展到自然光照下的 1269 个 Munsell 色块的反射光谱信号。

光谱反射率降维最常用的方法是有限维线性模型。在光谱反射率分析方面, Cohen 等人 [Cohen,1964] 采用主元分析方法,分析了从一组 433 个 Munsell 颜色片中随机选取的 150 个颜色片进行表面反射率实验,得出的结论是用前 3 个基函数的线性模型可以近似表示这些颜色数据。Maloney 和 Wandell [Maloney and Wandell,1986] 将 Cohen 的分析进行了进一步扩展至 462 个颜色块,实验表明采用 5-7 个基函数可以很好地近似这些颜色数据。另外, Maloney 也证明大部分天然物体的光谱反射率函数是带宽受限函数,用 3 个特征向量可比较精确地来描述光谱反射率。Stiles 等人 [Stiles et al,1977] 认为一般需要 3—5 个基函数的线性模型可以较好地描述光谱反射率函数,并且基函数的数量与光谱反射率带宽是一定成比例。Buchsbaum 和 Gottschalk [Buchsbaum and Gottschalk,1984] 证明了用 3 个基函数产生的光谱反射率函数能够近似大部分自然光谱反射。近年来,有限

维模型已经成为机器颜色恒常性研究中的主要手段[Takebe et al,2002;刘关松和徐建国,2002]。基于以上结论, $R(\lambda)$ 可以用一组基函数的线性组合来表示:

$$R(\lambda) = \sum_{j=1}^n w_j r_j(\lambda) \quad (5.4)$$

其中, $r_j(\lambda)$ 表示光谱反射基函数, 可以通过对大量的光谱反射率进行特征向量分解得到。

w_j 为第 j 个反射基函数对应的权值。将式(5.4)代入(5.3), 得式(5.5)。

$$I_{k(x,y)} = \sum_{j=1}^n w_j \sum_{\lambda=\lambda_1}^{\lambda=\lambda_2} P(\lambda) r_j(\lambda) C_k(\lambda) \quad k=1,2,3 \quad (5.5)$$

根据式(5.5), 首先测得 $P(\lambda)$ 和 $C_k(\lambda)$ 函数, 然后根据图像像素的 R、G、B 三通道颜色值获得的数据 $I_{k(x,y)}$ 来估计出反映物体颜色属性的 $R(\lambda)$, 在标准光照 $P(\lambda)$ 和 $C_k(\lambda)$ 情况下, 根据式(5.5)计算得到设备的响应的图像像素值 $I_{k(x,y)}$ 以达到颜色校正的目的。

基于有限维模型的颜色校正过程为:

1. 经过测量获取光照环境光谱功率分布函数 $P(\lambda)$ 和摄像机的光谱响应函数 $C_k(\lambda)$;
2. 测量多种目标的光谱反射率, 用 SVD 方法分析得到有限维模型基函数 $R(\lambda)$;
3. 对摄像机所拍摄的彩色图像进行学习, 运用式(5.5)计算三通道 RGB 的每个像素值 $I_{k(x,y)}$ 所对应的有限维的权重系数 w_j 。
4. 根据式(5.5)计算得到设备的响应的图像像素值 $I_{k(x,y)}$ 以达到颜色校正的目的, 此时(5.5)式中的 $P(\lambda)$ 、 $C_k(\lambda)$ 分别是指标准环境中光照光谱功率分布及数码相机的光谱响应函数, $I_{k(x,y)}$ 是有限维模型校正后的第 k 通道的标准颜色数据。

5.2.3 基于图像分析的颜色校正方法

光源颜色和图像采集设备会影响拍摄目标的色彩信息, 使整体色彩偏暗或偏明。当光源比较强时, 图像会出现偏亮的情况, 当光源比较暗时, 图像会出现偏暗的情况, 当光源位置不当时, 会出现侧光或偏光现象。亮度偏差和色温偏差会使图像的颜色跟原有的色调不同而产生色偏, 所以有必要对光照进行补偿。色温补偿法即白平衡法, 白平衡就是针对不同色温条件下, 无论环境光线如何, 通过对图像的色彩进行调整使图像抵消偏色。白平衡可以简单地理解为在任意色温条件下, 标准白色在任何条件拍摄下仍然为白色。彩色图像采集设备工作之前, 应该校准设备的白平衡。但是一般设备的校准功能有限, 只有有限几个典型的色温等级, 而实际拍摄环境的色温千差万别, 即使能够手

动校准白平衡,也存在一定的误差。所以在需要色彩定量分析的场合,应对图像进行色彩校正。传统的白平衡方法对图像的调整都是先找到整幅图像的调整系数,再同时对整幅图像的所有像素点进行全局调整。基于图像分析的常用的方法为灰平衡法和白平衡法,白平衡法又可细分为全局白平衡法和局部白平衡法[谷元保和付宇卓,2005]。

灰平衡法是对灰度世界使用的一种检测方法,灰度世界是指RGB色度空间中的RGB各通道的值平均值相接近的所产生的图像,对其进行色度空间转换,转换到Lab色度空间中,计算与中性点的距离,判断是否发生了偏色。然而当拍摄环境过暗或过亮时,图像的颜色不能覆盖整个颜色空间,不能满足灰度世界这个前提条件,所以灰平衡法得到的图像结果就会错。

白平衡法是对有镜面反射的图像进行判断是否存在偏色的一种检测方法。光源的色度是通过镜面反射表现出来的,分别统计RGB通道的极大亮度值,然后转换到Lab色度空间中,获得相对均匀的Lab坐标,通过计算理想光源与相对均匀的Lab之间色度距离。

在一幅彩色图像中,如果将像素 $\{r_w, g_w, b_w\}$ 被认为是白色的物体,并且将255被认为是校正成的目标白色,在理论上的一种方法是三个通道上的每个分量都分别乘以一个系数 $255/r_w, 255/g_w, 255/b_w$,将此类型的转换过程写成矩阵形式为:

$$\begin{bmatrix} r' \\ g' \\ b' \end{bmatrix} = \begin{bmatrix} 255/r_w & 0 & 0 \\ 0 & 255/g_w & 0 \\ 0 & 0 & 255/b_w \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \quad (5.6)$$

其中: r', g', b' 为经过白平衡转换后三个通道的颜色值, r, g, b 为白平衡前图像的像素值。

r_w, g_w, b_w 表示在白平衡前图像中认为白色像素的颜色值。

另一种通常用在摄像机中的自动白平衡方法数学表示如(5.7):

$$\begin{bmatrix} r' \\ g' \\ b' \end{bmatrix} = \begin{bmatrix} \frac{(\bar{r} + \bar{g} + \bar{b})/3}{\bar{r}} & 0 & 0 \\ 0 & \frac{(\bar{r} + \bar{g} + \bar{b})/3}{\bar{g}} & 0 \\ 0 & 0 & \frac{(\bar{r} + \bar{g} + \bar{b})/3}{\bar{b}} \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} \quad (5.7)$$

其中: $\bar{r}, \bar{g}, \bar{b}$ 分别为图像的各通道的颜色值的均值, r, g, b 分别为源颜色空间图像的像素

值, r',g',b' 分别为经过白平衡后的各通道的颜色值。自动白平衡是一个很复杂的问题, 目前还没有一个鲁棒性很强的方法可以解决所有场景的白平衡问题。上面的方法只是一种粗糙的估计值, 当颜色种类少的时候就会出现错误调整。Lam 等人[Lam et al,2004]提出了一种基于标准偏差加权灰度域的方法来自动解决白平衡的问题。

参考白方法(reference white)是由 Anil K. Jain, Rein. Lien Hsu 等人[Finlayson et al,1997;Rein-Lien et al,2002]提出的, 该方法实际是一种局部白平衡。“参考白”映射法就是设法使被重现的图像中的白点与理想的白点的颜色一致。线性调整公式为:

$$\begin{bmatrix} I'_r(x,y) \\ I'_g(x,y) \\ I'_b(x,y) \end{bmatrix} = \begin{bmatrix} \frac{255}{refW_r - refB_r} & 0 & 0 \\ 0 & \frac{255}{refW_g - refB_g} & 0 \\ 0 & 0 & \frac{255}{refW_b - refB_b} \end{bmatrix} \begin{bmatrix} I_r(x,y) - refB_r \\ I_g(x,y) - refB_g \\ I_b(x,y) - refB_b \end{bmatrix} \quad (5.8)$$

其中 $I_r(x,y), I_g(x,y), I_b(x,y)$ 和 $I'_r(x,y), I'_g(x,y), I'_b(x,y)$ 分别为调整前后 (x,y) 位置各通道的像素值; $refW_r, refW_g, refW_b$ 分别为图像中各通道的白色参考值; $refB_r, refB_g, refB_b$ 分别为图像中各通道的黑色参考值。

5.3 亮度转变函数的数学模型

研究人员已经采用了多种外观特征去建立匹配模型, 例如: 颜色直方图特征、UV 色度特征、尺度不变特征 (SIFT) 等; 但是, 没有一种特征可以不受大部分环境、条件变化 (如光照、观测角度、摄像机参数等) 的影响。因此, 一些学者尝试采用特征融合的方法来克服这些条件变化, 并获得了更高的匹配准确率。

在本节中, 我们将给出一对基于子空间的跨摄像机的亮度转变函数 (Bright transfer function), 这个子空间的关系是通过训练数据学习得到的, 然后使用这个子空间的关系对一组新拍摄的跨摄像机人体目标进行基于颜色的匹配。必须注意, 通过训练得到的从一个摄像机到另一个摄像机的映射必须每个像素都存在一一映射。

用 $L_i(p,t)$ 表示在 t 时刻人体目标上的点 p (在真实世界系中) 的在白色光照条件下被摄像机 C_i 捕获到的场景辐射值 (scene radiance)。假设人体目标没有发生镜面反射,

我们可以将 $L_i(p, t)$ 写成与人体目标的衣物本身的材料属性函数 $M_i(p, t) = M(p)$ 和与摄像机的几何关系及目标形状相关函数 $G_i(p, t)$ 的乘积的形式(5.9)。

$$L_i(p, t) = M(p)G_i(p, t) \quad (5.9)$$

鉴于以上给定的双向反射分布函数模型 BRDFM (Bi-directional Reflectance Distribution Function) 被广泛而有效地使用, 例如: 在朗伯模型 (Lambertian model) 和广义朗伯模型 (generalized Lambertian model) [Oren and Nayar, 1995] 中都被使用。在给定的摄像机的条件下, 对于对象上的任意两个点 p, q 都有相同的摄像机的几何关系及目标形状相关函数 $G_i(p, t) = G_i(q, t) = G_i(t)$, 因此(5.9)式可改写成式(5.10)。

$$L_i(p, t) = M(p)G_i(t) \quad (5.10)$$

根据文献[Horn, 1986]可知, 图像的辐照度(image irradiance) $E_i(p, t)$ 与场景的场景辐射值 $L_i(p, t)$ 是成比例的关系, 可用式(5.11)表示。

$$E_i(p, t) = L_i(p, t)Y_i(t) = M(p)G_i(t)Y_i(t) \quad (5.11)$$

其中,

$$Y_i(t) = \frac{\pi}{4} \left(\frac{d_i(t)}{h_i(t)} \right)^2 \cos^4 \alpha_i(p, t) = \frac{\pi}{4} \left(\frac{d_i(t)}{h_i(t)} \right)^2 c \quad (5.12)$$

式(5.12)表示在 t 时刻的相机参数的函数, 其中 $h_i(t)$ 和 $d_i(t)$ 分别表示摄像头的焦距和直径, $\alpha_i(p, t)$ 是世界坐标系中的点 p 的主要光线与光轴所成的夹角。在敏感性要求不是很高的情况下, 可以将 $\cos^4 \alpha_i(p, t)$ 写成一个常量 c [Horn, 1986]。

如果 g_i 是摄像机 C_i 辐射度函数, $E_i(p, t)$ 为图像的辐照度, $X_i(t)$ 是曝光时间, 则世界坐标系下的 p 点在时刻 t 的图像亮度为 $B_i(p, t)$,

$$B_i(p, t) = g_i(E_i(p, t)X_i(t)) \quad (5.13)$$

将式(5.11)代入(5.13)得到式(5.14)。

$$\begin{aligned} B_i(p, t) &= g_i(E_i(p, t)X_i(t)) \\ &= g_i(M(p)G_i(t)Y_i(t)X_i(t)) \end{aligned} \quad (5.14)$$

图像亮度 $B_i(p, t)$ 是材料属性 $M(p)$ 、几何属性 $G_i(t)$ 、摄像机参数 $Y_i(t)$ 和曝光时间 $X_i(t)$ 乘积的非线性函数。

下面我们考虑两个摄像机的情况，假定两台摄像机分别在 t_i 、 t_j 时刻对世界坐标中的同一点 p 进行拍摄。由于对同一目标进行拍摄，所以该目标被这两台摄像机拍摄过程中影响成像的材料属性 $M(p)$ 是相同的，通过两台摄像机成像函数式(5.14)分别推导出材料属性 $M(p)$ 的表达式为式(5.15)。

$$M(p) = \frac{g_i^{-1}(B_i(p, t_i))}{G_i(t_i)Y_i(t_i)X_i(t_i)} = \frac{g_j^{-1}(B_j(p, t_j))}{G_j(t_j)Y_j(t_j)X_j(t_j)} \quad (5.15)$$

根据式(5.15)可以推导出式(5.16)，该式表示从摄像机 C_i 在 t_i 时刻成像的亮度值 $B_i(p, t_i)$ 到摄像机 C_j 在 t_j 时刻成像的亮度值 $B_j(p, t_j)$ 的转换。

$$\begin{aligned} B_j(p, t_j) &= g_j \left(\frac{G_j(t_j)Y_j(t_j)X_j(t_j)}{G_i(t_i)Y_i(t_i)X_i(t_i)} g_i^{-1}(B_i(p, t_i)) \right) \\ &= g_j(w(t_i, t_j)g_i^{-1}(B_i(p, t_i))) \end{aligned} \quad (5.16)$$

其中 $w(t_i, t_j)$ 表示两台摄像机 C_j 、 C_i 分别在 t_j 、 t_i 时刻的场景参数、几何参数及曝光参数乘积的比值。对于两台摄像机所拍摄同一目标的任意点 p ，式(5.16)都是有效的，所以从公式中去掉参数 p ，同时，为了简化起见，我们认为两台摄像机的拍摄是基本同步的，所以不在考虑不同时刻的 t_j 、 t_i ，认为它们发生在同一时刻。用 f_{ij} 表示从摄像机 C_i 到摄像机 C_j 的亮度转换函数，因此，式(5.16)可以写成式(5.17)。

$$B_j = g_j(wg_i^{-1}(B_i)) = f_{ij}(B_i) \quad (5.17)$$

通过对 f_{ij} 在固定的且增加的亮度值进行采样得到亮度转换函数的非参数形式，

$B_i(1) < B_i(2) < \dots < B_i(d)$ ，并将其看作一个向量，用式(5.18)表示。

$$(B_j(1), \dots, B_j(d)) = (f_{ij}(B_i(1)), \dots, f_{ij}(B_i(d))) \quad (5.18)$$

用 Γ_{ij} 表示从摄像机 C_i 到摄像机 C_j 亮度转变函数的空间。从式(5.18)可以看出空间 Γ_{ij} 的维数最多就是离散亮度值的数目 d ，在大多数图像系统中 $d = 256$ 。然而，下面将证明亮度

转变函数事实上可以由 d 维空间的一个子空间即可确定。由式(5.17)与式(5.18)可得式(5.19)

$$\begin{aligned} f_{ij} &= g_j(wg_i^{-1}(B_i)) \\ &= [g_j(wg_i^{-1}(B_i(1))), \dots, g_j(wg_i^{-1}(B_i(n)))]^T \end{aligned} \quad (5.19)$$

对任意的 $a, x \in R$, 都有:

$$g_j(ax) = \sum_{u=1}^m r_u(a)s_u(x) \quad (5.20)$$

其中, r_u 、 s_u 是任意的且固定的函数, $1 \leq u \leq m$, 将式(5.20)代入式(5.19)

$$\begin{aligned} f_{ij} &= \sum_{u=1}^m r_u(w)[s_u(g_i^{-1}(B_i(1))), \dots, s_u(g_i^{-1}(B_i(n)))]^T \\ &= \sum_{u=1}^m r_u(w)s_u(g_i^{-1}(B_i)) \end{aligned} \quad (5.21)$$

因此, 每个亮度转变函数 $f_{ij} \in \Gamma_{ij}$ 可以用 m ($m \leq d$) 个子空间向量元素的函数 $s_u(g_i^{-1}(B_i))$ 求和形式表示。尽管不同的摄像机辐射度的响应函数是不同的且非线性的, 但可以通过一个简单的参数模型很好地估计不同的摄像机辐射度的响应函数。许多研究人员通过式(5.22) γ 函数[Farid,2002;Mann et al,1995]来估计摄像机的辐射度响应函数。

$$g(x) = \lambda x^\gamma + \mu \quad (5.22)$$

对于任意的 $a, x \in R$, 可得式(5.23)

$$g(ax) = \lambda(ax)^\gamma + \mu = \lambda a^\gamma x^\gamma + \mu = r_1(a)s_1(x) + r_2(a)s_2(x) \quad (5.23)$$

其中, $r_1(a) = a^\gamma$, $s_1(x) = \lambda x^\gamma$, $r_2(a) = 1$, $s_2(x) = \mu$ 。因此, 如果摄像机的辐射度响应函数是 γ 函数, 那么亮度转变函数空间用二维形式表示即可。

除了 γ 函数法来估计摄像机的辐射度响应函数外, 还有多项式法是一种比较普遍的方法来估计摄像机的辐射度响应函数。

$$g(x) = \sum_{u=0}^q \lambda_u x^u \quad (5.24)$$

对于任意的 $a, x \in R$, 可得式

$$g(ax) = \sum_{u=0}^q \lambda_u a^u x^u = \sum_{u=0}^q r_u(a) s_u(x) \quad (5.25)$$

其中 $r_u(a) = a^u$, $s_u(x) = \lambda x^u$, $0 \leq u \leq q$, Grossberg 和 Nayar[Grossberg and Nayar,2003] 已经证明: 在实际中, $q \leq 10$ 就能很好地估计摄像机的辐射度响应函数。

5.4 亮度转变函数的实际应用

首先考虑一对摄像机 C_1 和 C_2 , 对同一目标进行跨摄像机对应可以确定摄像机间的亮度转变函数。一种实现的方法是估计跨摄像机间同一目标的像素点对应来计算亮度转变函数, 可以利用公式(5.17)来实现, 然而在实际应用中, 由于同一目标的自遮挡或不同的姿态, 实现跨摄像机的像素对应是很难实现的。这样, 我们采用归一化的亮度直方图来计算亮度转变函数。该方法对遮挡与姿态变化具有较强的鲁棒性。为了寻找跨摄像机亮度值的对应点, 我们采用下面的方法: 如果摄像机 C_1 拍摄出的共同视场前景目标的亮度值小于等于 B_1 的像素数所占整个人体目标的比值与摄像机 C_2 拍摄出的共同视场前景目标的亮度值小于等于 B_2 的像素数所占整个人体目标的比值相同, 即两个比值相同, 我们就认定摄像机 C_1 的像素亮度值 B_1 与摄像机 C_2 的像素亮度值 B_2 是相对应的如图 5-2所示图像A中亮度值小于等于59的数目等于图像B亮度值小于等于42数目。



图像 A (阈值 59)

图像 B 阈值 (42)

图 5-2 摄像机 A 的像素亮度值 59 与摄像机 B 的像素亮度值 42 对应示意图

Grossberg 和 Nayar [Grossberg and Nayar,2003]在不同光照条件下, 用同一摄像机对相同

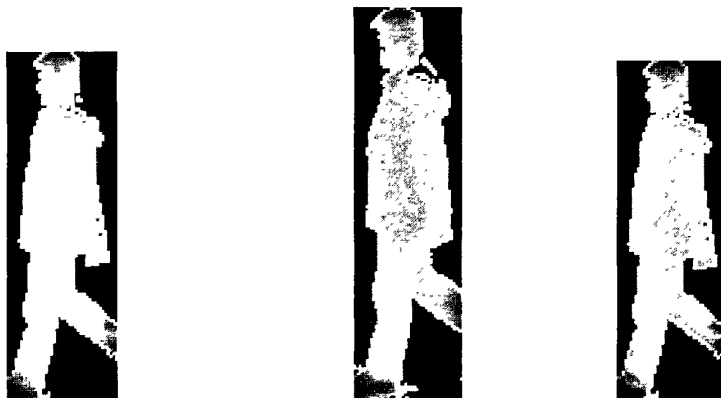
视场的同一目标进行拍摄, 采用与上面相似的策略来获取亮度转变函数。用 H_i 和 H_j 分别表示对象 O_i 和 O_j 的被归一化的累积直方图(normalized cumulative histogram), 那么

$$H_i(B_i) = H_j(B_j) = H_j(f_{ij}(B_i)) \quad (5.26)$$

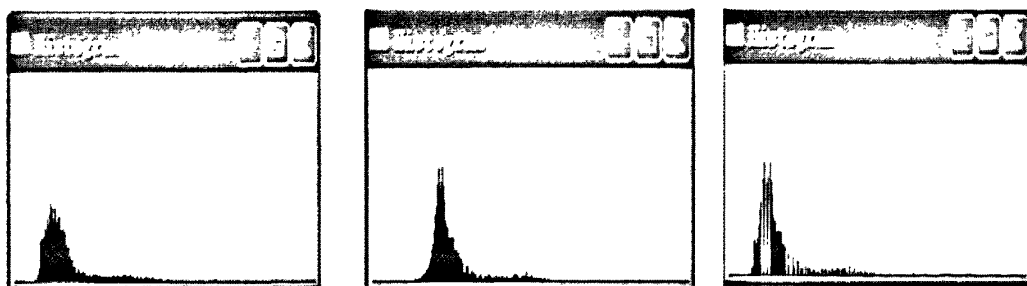
因此,

$$f_{ij}(B_i) = H_j^{-1}(H_i(B_i)) \quad (5.27)$$

其中, H^{-1} 为逆累积直方图(inverted cumulative histogram)。根据上述亮度转变方法对一组图像进行实验。图 5-3 亮度转换函数转换前后的 R 通道直方图对比。通过比较图 5-3 可以看出, 对摄像机 1 拍摄的图像 (a) 进行校正使之与摄像机 2 拍摄的图像 (b) 具有相同的效果, 图 (c) 是对图像 (a) 校正后的图像。通过比较图像 (b) 的直方图 (e) 和图像 (c) 的直方图 (f) 不难发现, 它们的直方图是比较相似的。



(a) 摄像机 1 拍摄的图像 (b) 摄像机 2 拍摄的图像 (c) 对图像 (a) 校正后的图像



(d) 图像 (a) 的直方图

(e) 图像 (b) 的直方图

(f) 图像 (c) 的直方图

图 5-3 亮度转换函数转换前后的 R 通道直方图对比

5.5 对亮度转变函数空间的概率估计

我们将用式(5.27)对观测对集合进行训练来估计亮度转变函数 f_{ij} 。用该方法我们将

获得亮度转变函数集 $F_{ij} = \{f_{(ij)_1}, f_{(ij)_2}, \dots, f_{(ij)_n}\}$, 为了学习这个集合的子空间, 本文将采用概率主要成分分析法(probabilistic Principal Component Analysis)[Tipping and Bishop,1999]来建模。根据这个模型, 一个 d 维的亮度转变函数 f_{ij} 可以用式(5.28)表示。

$$f_{ij} = \mathbf{W}y + \bar{f}_{ij} + \varepsilon \quad (5.28)$$

其中, y 是一个正态分布的 q 维子空间向量, $q < d$, \mathbf{W} 是一个使子空间向量与观测的亮度转变函数相关的 $d \times q$ 维的映射矩阵, \bar{f}_{ij} 是亮度转变函数集合的均值, ε 是服从正态分布的高斯噪声即, $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$, 由于 y 和 ε 都是一个正态分布, 所以可以得到式也是服从正态分布的。

$$f_{ij} \sim N(\bar{f}_{ij}, \mathbf{Z}) \quad (5.29)$$

其中, $\mathbf{Z} = \mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}$, 根据文献[Tipping and Bishop,1999], 映射矩阵 \mathbf{W} 可以通过式进行估计。

$$\mathbf{W} = \mathbf{U}_q (\mathbf{E}_q - s^2 \mathbf{I})^{\frac{1}{2}} \mathbf{R} \quad (5.30)$$

其中, \mathbf{U}_q 是由 F_{ij} 的样本协方差矩阵的特征向量构成的 $d \times q$ 维矩阵, \mathbf{E}_q 是与特征向量对应的特征值 $\lambda_1, \dots, \lambda_q$ 构成的 $q \times q$ 维对角阵, \mathbf{R} 是任意的正交旋转矩阵, 可以将一个单位赋值给这个矩阵。 s^2 的值可以通过式(5.31)进行计算。

$$\sigma^2 = \frac{1}{d-q} \sum_{v=q+1}^d \lambda_v \quad (5.31)$$

如果 \mathbf{W} 和 σ^2 的值被确定, 就可以通过式(5.29)计算属于被学习的亮度转变函数子空间的亮度转变函数的概率。

到目前为止, 我们已经给出了亮度值和计算亮度转变函数的方法。为了除了彩色图像, 我们将独立地处理图像的每个通道 R 、 G 、 B 。每个通道的亮度转变函数用上面同样的方法进行处理, 子空间的参数 \mathbf{W} 和 σ^2 也是每个通道单独处理的。在获取亮度转变函数和子空间过程中, 没有对摄像机的参数和响应函数做任何假定, 这是该方法的优点。

接下来, 我们将解决无重叠视场的跨摄像机跟踪的问题, 阐述如何基于子空间的颜色模型匹配。假设由 r 台摄像机 C_1, C_2, \dots, C_r 构成一个无重叠视场的系统, 在该环境中, 有 n 个人体目标在该系统中出现 (人体目标的个数是未知的), 在不同的时刻, 人体目标出现在不同的摄像机视场中。人体目标在单摄像机中跟踪过程是通过前几章节的方法来

实现的。用 $O_j = \{O_{j,1}, O_{j,2}, \dots, O_{j,m}\}$ 表示第 j 个摄像机拍摄到的 m 个人体目标集合。 $O_{j,a}$ 表示进入摄像机 C_j 视场到离开该摄像机视场的第 a 个人体目标的表面模型。跨摄像机跟踪实质就是在不同的摄像机中找到属于同一个人体目标的跨摄像机匹配。用 $k_{r,a}^{s,b}$ 表示在第 r 个摄像机中的第 a 个人体目标 $O_{r,a}$ 与在第 s 个摄像机中的第 b 个人体目标 $O_{s,b}$ 进行匹配。这样，这个多摄像机系统就会产生一个对应集合 $K = \{k_{r,a}^{s,b}\}$ ，当且仅当人体目标模型 $O_{r,a}$ 与人体目标模型 $O_{s,b}$ 对应同一个人体模型时， $k_{r,a}^{s,b} \in K$ 。用 Ω 表示多摄像机跟踪的解空间。多摄像机跟踪的解空间 Ω 中后验概率最大的作为最终的正确的解空间 K' ，用公式(5.32)表示。

$$K' = \arg \max_{K \in \Omega} \prod_{k_{i,a}^{j,b} \in K} P(O_{i,a}, O_{j,b} | k_{i,a}^{j,b}) \quad (5.32)$$

如果表面模型的概率密度函数 P 是已知的，用图形理论方法得到后验概率最大的情况即为最优解。最大策略法通过文献[Javed and Shah,2008]可知实现的细节。现在将考虑如何获取表面模型的概率密度函数 P 。

训练阶段已经提供了跨摄像机的颜色转变函数的子空间模型，该模型描述了对象的颜色在跨摄像机时的变化情况。在训练过程中，如果被学习的子空间能很好地描述跨摄像机的两个模型的颜色映射，那么是同一个的人体目标产生的模型的概率是比较大的。具体的讲就是，两个观测模型 $O_{i,a}$ 与 $O_{j,b}$ 的颜色转变函数的分布 $\mathbf{f}_{i,j}^R, \mathbf{f}_{i,j}^G, \mathbf{f}_{i,j}^B$ 是由公式(5.29)决定的，我们定义式

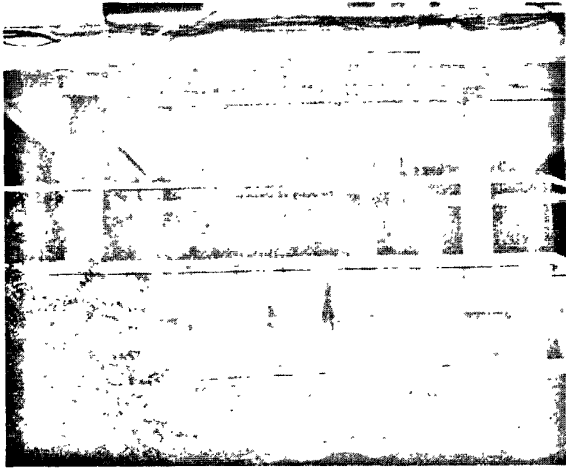
$$P_{i,j}(O_{i,a}, O_{j,b} | k_{i,a}^{j,b}) = \prod_{\alpha \in \{R,G,B\}} \frac{1}{(2\pi)^{d/2} |\mathbf{Z}^\alpha|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{f}_{ij}^\alpha - \bar{\mathbf{f}}_{ij}^\alpha)^T (\mathbf{Z}^{-\alpha}) (\mathbf{f}_{ij}^\alpha - \bar{\mathbf{f}}_{ij}^\alpha)^T\right) \quad (5.33)$$

其中， $\mathbf{Z} = \mathbf{W}\mathbf{W}^T + \sigma^2\mathbf{I}$ ，下标 α 表示颜色空间的 R 、 G 、 B 三个通道中的每个通道。 \mathbf{W} 和 σ^2 的各通道的值分别通过式(5.30)和式(5.31)计算求得。

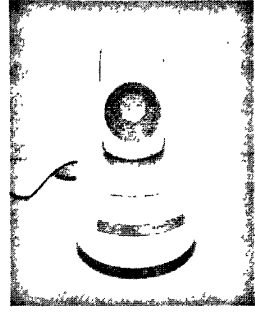
5.6 实验结果和分析

本节将用本章提出的方法对室内-室外、室外-室外两组无重叠的跨摄像机系统进行实验并给出实验结果。每次实验都包括训练阶段和测试阶段，在单摄像机中对人体目标的检测与跟踪都是采用第二章的基于背景检测的单摄像机多人目标跟踪方法。在训练阶段，跨摄像机的人体目标对应是已知的。在测试阶段，采用本章提出的方法对无重叠的跨摄像机人体目标进行跟踪。在室外环境下，实验所需要的图像数据是通过索尼监控网络摄像机 SNC-RZ25N/RZ25P（如图 5-4（b）所示）在图 5-4（a）的位置采集得到

的。



(a) 室外环境摄像头位置



(b) 索尼监控网络摄像机

图 5-4 室外环境摄像头位置及索尼监控网络摄像机

第一组实验是在室外环境下，用两台摄像机搭建了一个简单系统平台，其拓扑结构如图 5-5 所示。在室外环境下，亮度变化并不是很强烈。通过训练得到颜色转变函数，实验结果如图 5-6 所示。

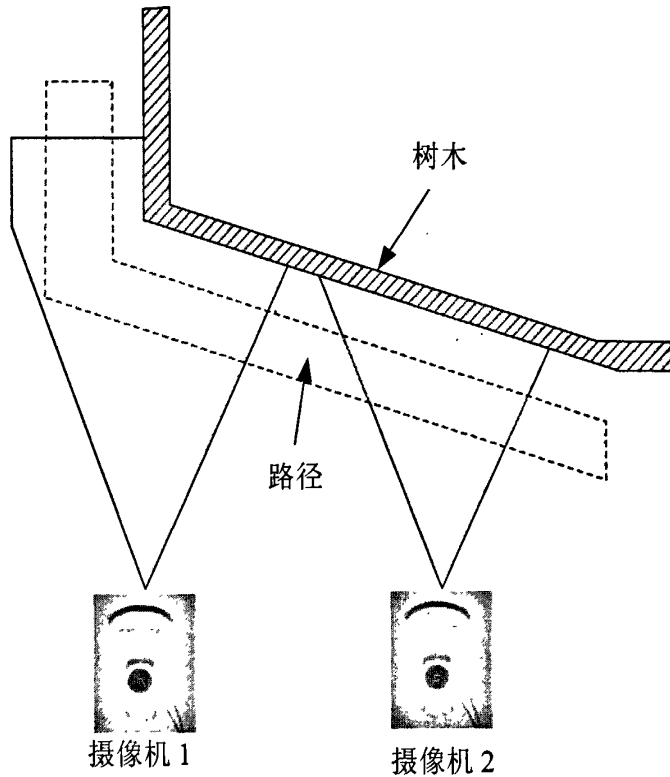
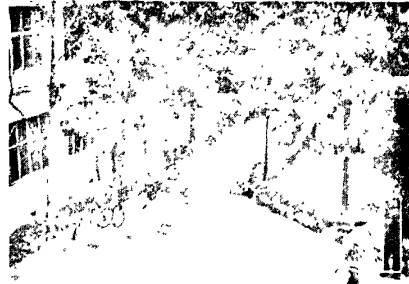


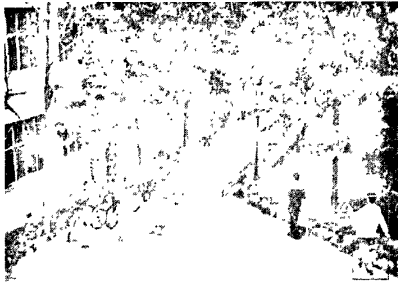
图 5-5 无重叠跨摄像机系统的拓扑结构



2937 帧



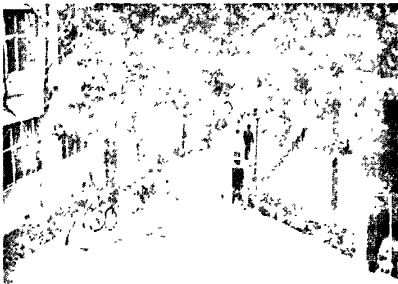
3124 帧



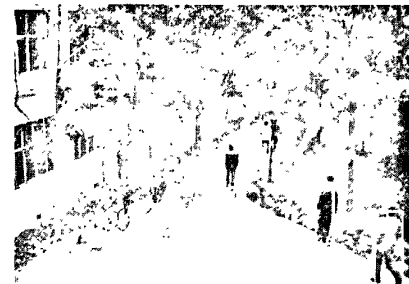
3164 帧



3228 帧



3334 帧



3302 帧

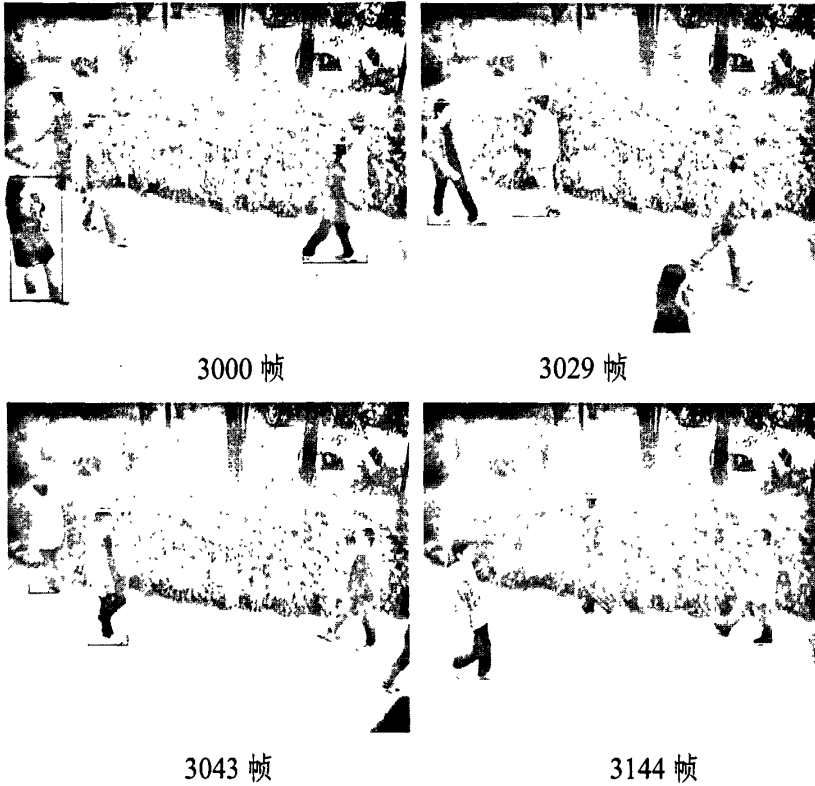
(a) 摄像机 1 人体目标跟踪



2948 帧



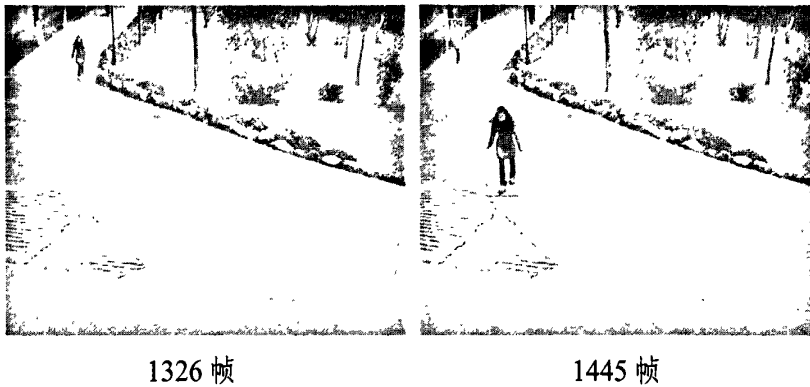
2988 帧

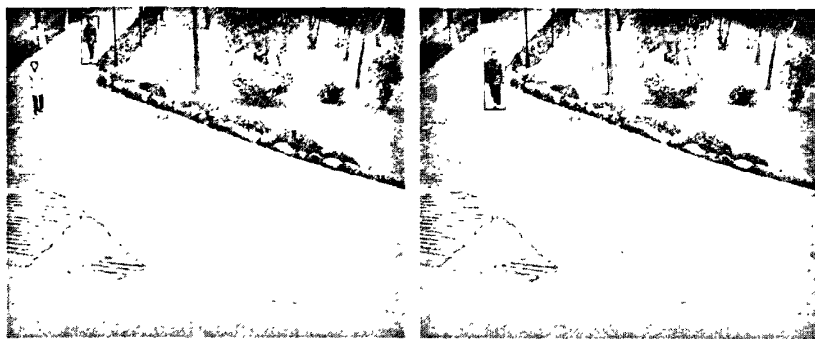


(b) 摄像机 2 人体目标跟踪

图 5-6 无公共视场的场景 A 情况下跨摄像机人体目标跟踪

第二组实验是在室内室外情况下，一台摄像机拍摄实验楼的入口处，另一台摄像机拍摄该实验楼走廊的视场。一个是自然光的条件下，另一个是在室内灯光的条件下，光照条件变化情况较大。通过训练得到室内室外情况下颜色转变函数，实验结果如图 5-7 所示。

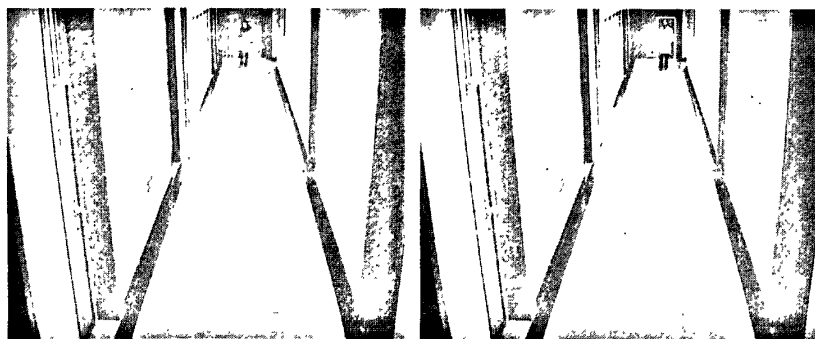




1519 帧

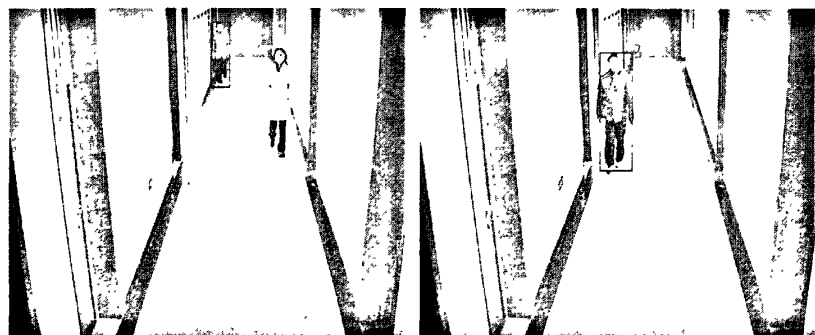
1592 帧

(a) 室外环境下的跨摄像机人体目标跟踪结果



2136 帧

2389 帧



2484 帧

2560 帧

(b) 室内环境下的跨摄像机人体目标跟踪结果

图 5-7 无公共视场的场景 B 室内外情况下跨摄像机人体目标跟踪

5.7 本章小结

本章在低维度的子空间上对已知的跨摄像机人体目标进行训练,得到颜色转变函数。该方法不用依赖摄像机的内参来计算颜色转变函数的子空间。对颜色转变函数的子空间用基于概率的主要成分分析法对其进行建模得到概率密度函数,然后利用该概率密度函数获取跨摄像机的人体目标的匹配概率。本章在室内外,全部室外环境下对该方法进行了验证并给出了实验结果,从实验结果上来看,本方法能够很好地对室外环境下的跨摄

像机人体目标很好地跟踪，而在光照条件变化比较大的室内外环境下的效果不如室外环境下效果好。不过，利用本章提出的方法要比直接利用颜色进行匹配的效果要好的多。

6 总结与展望

6.1 本文总结

人体目标跟踪一直是计算机视觉研究的热点，是一项非常具有挑战性的研究工作，是涉及很广泛的一门综合性学科，是使用计算机及相关设备对生物视觉的一种模拟。但是人体目标跟踪技术仍然处于研究和探索阶段，在理论和实际应用中仍存在许多问题没有得到很好地解决。目前，影响目标跟踪算法鲁棒性的原因很大程度上是由于目标运动的复杂性以及光照条件变化等因素造成的。本文主要研究了单摄像机和多摄像机下的人体目标跟踪，单摄像机人体目标跟踪是跨摄像机人体目标跟踪的基础，只有单摄像机人体目标跟踪正确的前提下，跨摄像机跟踪才会有正确的跟踪结果。在单摄像机人体目标跟踪过程中，主要是解决遮挡人体目标跟踪，本文主要采用两种方法来解决单摄像机遮挡人体目标跟踪，一种是基于无背景检测的粒子滤波跟踪方法，一种是基于背景检测的非参数核密度估计跟踪方法。本文的跨摄像机跟踪是为了实现扩展视场，所以跨摄像机跟踪分为有公共视场人体目标跟踪和无公共视场人体目标跟踪。

1. 本文针对多人目标跟踪而产生的遮挡问题，提出了一种基于非参数复合模型的粒子滤波方法来描述多个人体目标的情况。该复合模型通过复合预测和复合更新两个步骤交替进行以达到递归跟踪的目的。在复合粒子滤波过程中，通过每个复合粒子的权值大小来衡量粒子的贡献大小。权值的大小是基于 HSV 直方图的多颜色观测模型来实现的。观测模型是通过巴氏距离的核密度估计来建模。实验结果证明复合模型的粒子滤波方法可以很好地实现多人目标跟踪。

2. 本文充分利用单摄像机的人体目标的运动信息、颜色信息、空间信息将背景建模、块建模、颜色建模、运动建模和人体的空间信息进行有效融合有效地解决了人体目标相互遮挡情况下的人体跟踪的问题。本文利用人体目标的运动信息，采用的混合高斯模型进行背景重建的方法，先提取出运动的目标；利用基于 Epanechnikov 核密度梯度估计算法对存储模型中的人体进行聚类，即对人体的外部颜色相近的像素进行块建模；利用非参数的 Gauss 核密度估计算法对已聚类的块模型和人体的空间信息建立颜色密度函数，同时利用人体运动信息建立运动密度函数，通过颜色密度函数与运动密度函数构成后验概率模型；然后对当前的检测出的人体目标的每个像素计算最大后验概率，得到最大后验概率的颜色图像，通过该图像进行分割，达到对遮挡情况下多人目标进行跟踪的目的。同时本文采用隔行列抽取的方法来减少图像的数据量以提高算法的运行效率。实验

证明,本算法提高了遮挡人体目标跟踪的可靠性,收到了很好的效果。

3. 本文针对有公共视场跨摄像机人体目标跟踪误匹配的问题,提出了一种基于空间映射变换与颜色特征信息相融合的匹配算法。本文先通过跨摄像机的公共视场分界线来初步确定跨摄像机的人体目标,然后利用变换单应矩阵计算跨摄像机人体目标的对应关系,然而在实际应用中,视频是不完全同步且有交叉遮挡而产生误匹配,因此本文在投影变换单应矩阵的基础上采用基于颜色模型的最大后验概率进行匹配。

4. 本文针对无公共视场跨摄像机人体目标跟踪因颜色差异较大而引起误匹配问题,提出了一种基于颜色转变函数的方法对颜色进行校正,并通过颜色转变函数空间进行概率估计来提高跨摄像机人体目标的匹配的准确率。该算法先在低维度的子空间上对已知的跨摄像机人体目标进行训练,得到颜色转变函数。该方法不用依赖摄像机的内参来计算颜色转变函数的子空间。用基于概率的主要成分分析法对颜色转变函数的子空间进行建模得到概率密度函数,最后利用概率密度函数获取跨摄像机的人体目标的匹配概率。本文在室内外,全部室外环境下对该方法进行了实验,从实验结果上看,本方法能够很好地对室外及室内外环境下的跨摄像机人体目标跟踪。

6.2 工作展望

本文提出的多摄像机下的人体检测与跟踪算法目前处于实验阶段,实现一个鲁棒性强、准确性高和运行速度快的智能视频监控跟踪系统是非常困难的,因为有很多不确定性因素存在,光照条件变化、深度信息丢失、人体模型的初始化等,这些都是视觉跟踪技术发展的难点问题。本文针对多摄像机人体跟踪中常常会遇到的一些情况,作了一些工作,但并没有最终解决人体跟踪问题。对于本课题来说仍然还有很多有待进一步研究和改进的问题:

1. 基于无背景检测的粒子滤波算法对某些特殊情况有一定的局限性,如,人体目标的颜色与背景颜色很接近时,此时无法使用运动信息来进行跟踪,就有可能跟踪到背景上;还有当一个目标被另一个目标较长时间遮挡,要对人进行连续跟踪十分困难;如何有效地更新人体目标模型也是必须要研究的内容;当相互遮挡的两个或多个人体目标的颜色特征非常接近,如何能够在这种情况下仍能有效跟踪各个个体目标是需要进一步研究的内容。另外,由于粒子滤波的计算的数据量比较大,如何实现快速粒子滤波的人体目标跟踪也是值得研究的一个课题。

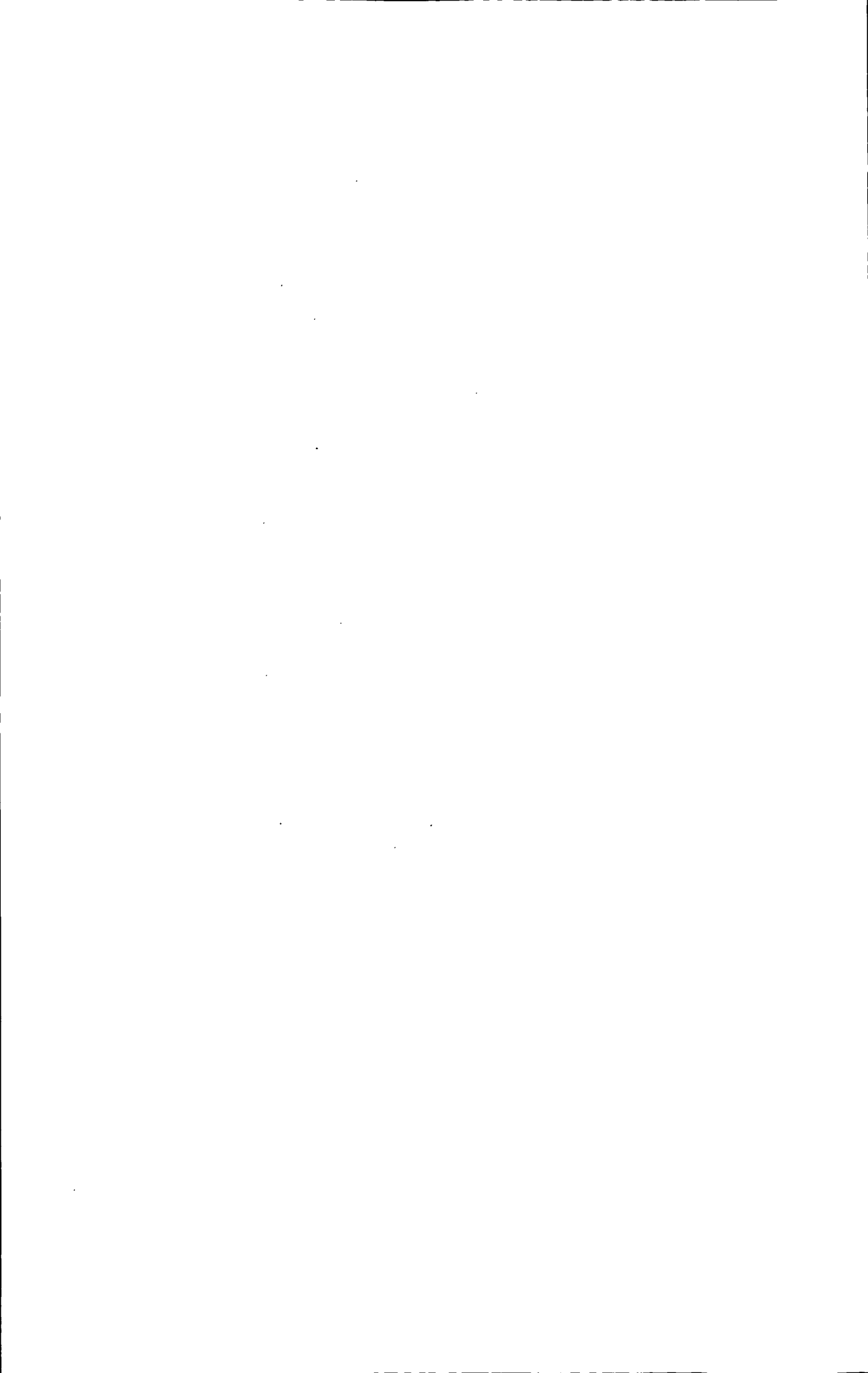
2. 基于背景检测的核密度估计算法对某些特殊情况有一定的局限性,在背景比较复杂的情况下,如背景是郁郁葱葱的树林,在风力的吹拂下来回摇晃,很难进行背景重建,或

在人体衣服的颜色相近情况下,如在遮挡时两个人的衣服颜色是完全相同的等,这些情况还需进一步解决。

3. 在有公共视场的跨摄像机人体目标跟踪情况下,开发出一种较鲁棒性的人体跟踪算法仍存在一些困难,当几个相互遮挡的人体目标进入一个摄像机视场时,无法对这些人体目标在该摄像机中进行自动建模。如何判断进来的目标是单个目标还是一个互相遮挡的人群,如果是一个互相遮挡的人群,如何对该人群的个体目标进行区分并进行标号,这些问题仍然是目前人体跟踪技术研究的难点问题。

4. 无公共视场的多摄像机人体目标跟踪在颜色变化剧烈的情况下比较敏感。当运动的人体目标所处环境的光照发生改变时,采用图像的颜色特征进行跟踪就会产生错误,尽管本文采用了亮度转变函数,但是该转换的前提条件是处在不同环境下的摄像机的光照条件是不变的。

本文提出的多摄像机下的人体检测与跟踪算法目前处于实验阶段,随着科学技术的不断发展和相关学科研究的不断深入,单摄像机人体目标跟踪和跨摄像机的人体目标跟踪的技术难点必将会得以解决,视觉跟踪技术必将得到强有力的发展并对人们未来的生活产生深远的影响。



参考文献

- [Alspach and Sorenson,1972]Alspach D. and Sorenson H., Nonlinear Bayesian estimation using Gaussian sum approximations. IEEE transactions on automatic control, 1972. 17(4): p. 439-448.
- [Altunbasak and Tekalp,1997]Altunbasak Y. and Tekalp A., Occlusion-adaptive, content-based mesh design and forward tracking. IEEE Transactions on Image Processing, 1997. 6(9): p. 1270-1280.
- [Arulampalam et al,2002]Arulampalam M., Maskell S., Gordon N., et al., A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing, 2002. 50(2).
- [Attwood et al,1989]Attwood C., Sullivan G. and Baker K. Model-based recognition of human posture using single synthetic images. 1989.
- [Azarbayejani and Pentland,1996]Azarbayejani A. and Pentland A. Real-time self-calibrating stereo person tracking using 3-D shapeestimation from blob features. 1996.
- [Black and Ellis,2006]Black J. and Ellis T., Multi camera image tracking. Image and Vision Computing, 2006. 24(11): p. 1256-1267.
- [Black et al,2002]Black J., Ellis T. and Rosin P., Multi view image surveillance and tracking. 2002.
- [Black et al,2005]Black J., Makris D. and Ellis T. Validation of blind region learning and tracking. 2005.
- [Bowden and KaewTraKulPong,2005]Bowden R. and KaewTraKulPong P., Towards automated wide area visual surveillance: tracking objects between spatially-separated, uncalibrated views. IEE Proc.-Vis. Image Signal Process, 2005. 152(2): p. 213.
- [Bradski,1998]Bradski G., Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal, 1998. 2(2): p. 12-21.
- [Brusewitz,1994]Brusewitz H., Motion compensation with triangles. Proceedings of IEEE International Conference on Image Processing, Austin, TX:IEEE Signal Processing Society, 1994: p. 433-437.

- [Buchsbaum and Gottschalk,1984]Buchsbaum G. and Gottschalk A., Chromaticity coordinates of frequency-limited functions. *Journal of the Optical Society of America A*, 1984. 1(8): p. 885-887.
- [Bucy and Senne,1971]Bucy R. and Senne K., Digital synthesis of non-linear filters. *Automatica*, 1971. 7(3): p. 287-298.
- [Cai and Aggarwal,1999]Cai Q. and Aggarwal J., Tracking human motion in structured environments using a distributed-camera system. *Ieee Transactions on Pattern Analysis and Machine Intelligence*, 1999: p. 1241-1247.
- [Carpenter et al,1999]Carpenter J., Clifford P. and Fearnhead P., Improved particle filter for nonlinear problems. *IEE Proceedings-Radar, Sonar and Navigation*, 1999. 146(1): p. 2-7.
- [Chang and Chang,2002]Chang P. and Chang C. Color correction for scanner and printer using B-spline CMAC neural networks. 2002: IEEE.
- [Chang and Gong,2001]Chang T. and Gong S., Tracking multiple people with a multi-camera system. *womot*, 2001: p. 0019.
- [Chang et al,2000]Chang T., Gong S. and Ong E. Tracking multiple people under occlusion using multiple cameras. 2000.
- [Chang and Reid,1996]Chang Y. and Reid J., RGB calibration for color image analysis in machine vision. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 1996. 5(10): p. 1414.
- [Chen et al,2001]Chen Y., Rui Y. and Huang T., JPDAF Based HMM or Real-Time Contour Tracking. 2001.
- [Chen,2003]Chen Z., Bayesian filtering: From Kalman filters to particle filters, and beyond. 2003, Technical report, Adaptive Systems Lab, McMaster University, 2003.
- [Cheng et al,2006]Cheng E., Madden C. and Piccardi M., Mitigating the effects of variable illumination for tracking across disjoint camera views. 2006.
- [Cheng and Piccardi,2006]Cheng E. and Piccardi M. Matching of objects moving across disjoint cameras. 2006.
- [Cheng and Piccardi,2007]Cheng E. and Piccardi M., Disjoint track matching based on a

- major color spectrum histogram representation. *Optical Engineering*, 2007. **46**: p. 047201.
- [Cheng and Moura,1999]Cheng J. and Moura J., Capture and representation of human walking in live video sequences. *IEEE Transactions on Multimedia*, 1999. **1**(2).
- [Cohen,1964]Cohen J., Dependency of the spectral reflectance curves of the Munsell color chips. *Psychon. Sci*, 1964. **1**: p. 369-370.
- [Collins et al,2002]Collins R. T., Lipton A. J. and Kanade T., A system for video surveillance and monitoring. VSAM final report Technical Report: CMU-RI-TR-00-12, Carnegie Mellon University, Pittsburgh, Peen, America, , 2002.
- [Comaniciu and Meer,1999]Comaniciu D. and Meer P. Mean shift analysis and applications. 1999.
- [Comaniciu and Meer,2002]Comaniciu D. and Meer P., Mean shift: A robust approach toward feature space analysis. *Ieee Transactions on PAMI*, 2002. **24**(5): p. 603.
- [Comaniciu and Ramesh,2003]Comaniciu D. and Ramesh V., Real-time tracking of non-rigid objects using mean shift. 2003, Google Patents.
- [Comaniciu et al,2001]Comaniciu D., Ramesh V. and Meer P. The variable bandwidth mean shift and data-driven scale selection. 2001.
- [Cremers and Schn rr,2003]Cremers D. and Schn rr C., Statistical shape knowledge in variational motion segmentation. *Image and Vision Computing*, 2003. **21**(1): p. 77-86.
- [D'Apuzzo et al,1998]D'Apuzzo N., Plankers R., Fua P., et al., Modeling human bodies from video sequences. *Videometrics Vi*, 1998. **3641**: p. 36-47 300.
- [Delamarre and Faugeras,1999]Delamarre Q. and Faugeras O. 3D articulated models and multi-view tracking with silhouettes. 1999.
- [Delamarre and Faugeras,2001]Delamarre Q. and Faugeras O., 3D Articulated Models and Multiview Tracking with Physical Forces* 1. *Computer Vision and Image Understanding*, 2001. **81**(3): p. 328-357.
- [Demirdjian,2003]Demirdjian D., Enforcing constraints for human body tracking. 2003.
- [Demirdjian and Darrell,2002]Demirdjian D. and Darrell T., 3-d articulated pose tracking for untethered diectic reference. Fourth Ieee International Conference on Multimodal

Interfaces, Proceedings, 2002: p. 267-272 543.

[Dick and Brooks,2005]Dick A. and Brooks M., A stochastic approach to tracking objects across multiple cameras. AI 2004: Advances in Artificial Intelligence, 2005: p. 53-84.

[Doucet et al,2000]Doucet A., Godsill S. and Andrieu C., On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and computing, 2000. 10(3): p. 197-208.

[Doucet et al,2001]Doucet A., Gordon N. and Krishnamurthy V., Particle filters for state estimation of jump Markov linear systems. IEEE Transactions on Signal Processing, 2001. 49(3): p. 613.

[Dudon et al,1994]Dudon M., Avaro O. and Eude G., Object-oriented motion estimation. Picture Coding Symposium, California, 1994: p. 284-287.

[Duong and Hazelton,2005]Duong T. and Hazelton M., Cross-validation bandwidth matrices for multivariate kernel density estimation. Scandinavian Journal of Statistics, 2005. 32(3): p. 485-506.

[Ellis et al,2003]Ellis T., Makris D. and Black J. Learning a multi-camera topology. 2003: Citeseer.

[Farid,2002]Farid H., Blind inverse gamma correction. Image Processing, IEEE Transactions on, 2002. 10(10): p. 1428-1433.

[Finlayson et al,1997]Finlayson G., Finlayson C. and Drew M., White-point preserving color correction. 1997.

[Gao and Shi,2003]Gao J. and Shi J., Inferring human upper body motion. IEEE Computer Vision and, 2003.

[Gavrila and Davis,1996]Gavrila D. and Davis L. 3-D model-based tracking of humans in action: a multi-view approach. 1996.

[Gelb,1974]Gelb A., Applied optimal estimation. 1974: The MIT press.

[Gilbert and Bowden,2005]Gilbert A. and Bowden R., Incremental modelling of the posterior distribution of objects for inter and intra camera tracking. Proc. of BMVC05, 2005. 1: p. 419-428.

[Gilbert and Bowden,2006]Gilbert A. and Bowden R., Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity.

- Computer Vision—ECCV 2006, 2006: p. 125-136.
- [Goncalves et al]Goncalves L., Di Bernardo E., Ursella E., et al. Monocular tracking of the human arm in 3D: Citeseer.
- [Gordon et al,1993]Gordon N., Salmond D. and Smith A. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. 1993.
- [Grossberg and Nayar,2003]Grossberg M. and Nayar S., Determining the camera response from images: what is knowable? Ieee Transactions on PAMI Intelligence, 2003: p. 1455-1467.
- [Han et al,2004]Han M., Xu W., Tao H., et al. An algorithm for multiple object trajectory tracking. 2004: IEEE.
- [Haritaoglu et al,2000]Haritaoglu I., Harwood D. and Davis L., W4: Real-Time Surveillance of People and Their Activities. Ieee Transactions on PAMI, 2000. **22**(8): p. 809.
- [Haug,2005]Haug A., A tutorial on Bayesian estimation and tracking techniques applicable to nonlinear and non-Gaussian processes. The MITRE Corporation, USA, Tech. Rep, 2005.
- [Herman,2002]Herman S., A particle filtering approach to joint passive radar tracking and target classification. University of Illinois at Urbana-Champaign, 2002.
- [Hogg,1983]Hogg D., Model-based vision: a program to see a walking person. Image and Vision Computing, 1983. **1**(1): p. 5-20.
- [Horn,1986]Horn B., Robot vision. 1986: McGraw-Hill Higher Education.
- [Hu et al,2006]Hu W., Hu M., Zhou X., et al., Principal axis-based correspondence between multiple cameras for people tracking. Ieee Transactions on PAMI, 2006: p. 663-671.
- [Huang and Huang,2002]Huang Y. and Huang T. S., Model-based human body tracking. 16th International Conference on Pattern Recognition, Vol I, Proceedings, 2002: p. 552-555 834.
- [Hung,1991]Hung P. Colorimetric calibration for scanners and media. 1991.
- [Huttenlocher et al,1993]Huttenlocher D., Klanderman G. and Rucklidge W., Comparing images using the Hausdorff distance. Ieee Transactions on PAMI, 1993: p. 850-863.
- [Isard and Blake,1996]Isard M. and Blake A., Contour tracking by stochastic propagation of

conditional density. *Computer Vision—ECCV'96*, 1996: p. 343-356.

[Isard and Blake,1998]Isard M. and Blake A., Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, 1998. **29**(1): p. 5-28.

[Isard and MacCormick,2001]Isard M. and MacCormick J., BraMBLe: A Bayesian multiple-blob tracker. 2001.

[Iwai et al,1999]Iwai Y., Ogaki K. and Yachida M. Posture estimation using structure and motion models. 1999: Published by the IEEE Computer Society.

[Jang and Choi,2000]Jang D. and Choi H., Active models for tracking moving objects. *Pattern Recognition*, 2000. **33**(7): p. 1135-1146.

[Javed et al,2003]Javed O., Rasheed Z., Alatas O., et al. KNIGHT(tm): a real time surveillance system for multiple and non-overlapping cameras. 2003.

[Javed et al,2008]Javed O., Shafique K., Rasheed Z., et al., Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views. *Computer Vision and Image Understanding*, 2008. **109**(2): p. 146-162.

[Javed et al,2005]Javed O., Shafique K. and Shah M., Appearance modeling for tracking in multiple non-overlapping cameras. 2005 IEEE Computer Society Conference on CVPR, Vol 2, Proceedings, 2005: p. 26-33 1216.

[Javed and Shah,2008]Javed O. and Shah M., Tracking in multiple cameras with disjoint views. *Automated Multi-Camera Surveillance: Algorithms and Practice*, 2008: p. 1-26.

[Jazwinski,1970]Jazwinski A., Stochastic processes and filtering theory. 1970: Academic Pr.

[Jeong and Jaynes,2008]Jeong K. and Jaynes C., Object matching in disjoint cameras using a color transfer approach. *Machine Vision and Applications*, 2008. **19**(5): p. 443-455.

[Jojic et al,1997]Jojic N., Gu J., Shen H., et al., 3-D reconstruction of multipart self-occluding objects. *Computer Vision—ACCV'98*, 1997: p. 455-462.

[Ju et al,1996]Ju S. X., Black M. J. and Yacoob Y., Cardboard people: A parameterized model of articulated image motion. *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, 1996: p. 38-44 386.

[Julier and Uhlmann,1996]Julier S. and Uhlmann J., A general method for approximating nonlinear transformations of probability distributions. Dept. of Engineering Science,

University of Oxford, Tech. Rep, 1996.

- [Julier and Uhlmann,1997]Julier S. and Uhlmann J. A new extension of the Kalman filter to nonlinear systems. 1997: Citeseer.
- [Julier et al,1995]Julier S., Uhlmann J. and Durrant-Whyte H. A new approach for filtering nonlinear systems. 1995: American Automatic Control Council, Evanston, IL.
- [Julier et al,2004]Julier S., Uhlmann J., Ind I., et al., Unscented filtering and nonlinear estimation. Proceedings of the IEEE, 2004. 92(3): p. 401-422.
- [Kakadiaris and Metaxas,2000]Kakadiaris I. and Metaxas D., Model-based estimation of 3D human motion. Ieee Transactions on PAMI, 2000. 22(12): p. 1453-1459.
- [Kang,1992]Kang H., Color scanner calibration. 1992.
- [Kang and Anderson,1992]Kang H. and Anderson P., Neural network applications to the color scanner and printer calibrations (Journal Paper). Journal of Electronic Imaging, 1992. 1(02): p. 125-135.
- [Kang et al,2004]Kang J., Cohen I. and Medioni G. Tracking people in crowded scenes across multiple cameras. 2004: Citeseer.
- [Karaulova et al,2000]Karaulova I., Hall P. and Marshall A. A hierarchical model of dynamics for tracking people with a single video camera. 2000: Citeseer.
- [Kass et al,1998]Kass R., Carlin B., Gelman A., et al., Markov chain monte carlo in practice: A roundtable discussion. The American Statistician, 1998. 52(2).
- [Kelly et al,1995]Kelly P., Katkere A., Kuramura D., et al. An architecture for multiple perspective interactive video. 1995: ACM.
- [Kettner and Zabih,1999a]Kettner V. and Zabih R. Bayesian multi-camera surveillance. 1999a: Citeseer.
- [Kettner and Zabih,1999b]Kettner V. and Zabih R., Counting people from multiple cameras. Ieee International Conference on Multimedia Computing and Systems, Proceedings Vol 2, 1999b: p. 267-271 1127.
- [Khan et al,2001]Khan S., Javed O. and Shah M. Tracking in uncalibrated cameras with overlapping field of view. 2001: Citeseer.
- [Khan and Shah,2000]Khan S. and Shah M. Tracking people in presence of occlusion. 2000.

- [Khan and Shah,2003]Khan S. and Shah M., Consistent labeling of tracked objects in multiple cameras with overlapping fields of view. *Ieee Transactions on PAMI*, 2003: p. 1355-1360.
- [Khan and Shah,2006]Khan S. and Shah M., A multiview approach to tracking people in crowded scenes using a planar homography constraint. *Computer Vision–ECCV 2006*, 2006: p. 133-146.
- [Kitagawa,1996]Kitagawa G., Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics*, 1996. 5(1): p. 1-25.
- [Kocheisen et al,2002]Kocheisen M., Muller U. and Troster G. A neural network for grey level and color correction used in photofinishing. 2002: IEEE.
- [Koller-Meier and Ade,2001]Koller-Meier E. and Ade F., Tracking multiple objects using the condensation algorithm. *Robotics and Autonomous Systems*, 2001. 34(2-3): p. 93-105.
- [Kong et al,1994]Kong A., Liu J. and Wong W., Sequential imputations and Bayesian missing data problems. *Journal of the American statistical association*, 1994. 89(425): p. 278-288.
- [Krumm et al,2000]Krumm J., Harris S., Meyers B., et al., Multi-camera multi-person tracking for easyliving. vs, 2000: p. 3.
- [Lam et al,2004]Lam H., Au O. and Wong C. Automatic white balancing using standard deviation of RGB components. 2004: IEEE.
- [Lee and Chia,2002]Lee D. and Chia N., A particle algorithm for sequential Bayesian parameter estimation and model selection. *IEEE Transactions on Signal Processing*, 2002. 50(2).
- [Lee and Chen,1985]Lee H. and Chen Z., Determination of 3D human body postures from a single view. *Computer Vision, Graphics, and Image Processing*, 1985. 30(2): p. 148-168.
- [Leoputra et al,2006]Leoputra W., Tan T. and Lim F., Non-overlapping Distributed Tracking using Particle Filter. *Pattern Recognition*, 2006. 3: p. 181-185.
- [Lipton et al,1998]Lipton A., Fujiyoshi H. and Patil R. Moving target classification and

- tracking from real-time video. 1998: Citeseer.
- [Liu and Chen,1998]Liu J. and Chen R., Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association*, 1998. **93**(443): p. 1032-1044.
- [Loke et al,2006]Loke Y., Kumar P., ranganath S., et al., Object matching across multiple non-overlapping fields of view using fuzzy logic. *ACTA Automatica Sinica*, 2006.
- [MacCormick and Blake,2000]MacCormick J. and Blake A., A probabilistic exclusion principle for tracking multiple objects. *International Journal of Computer Vision*, 2000. **39**(1): p. 57-71.
- [Makris et al,2004]Makris D., Ellis T. and Black J., Bridging the gaps between cameras. 2004.
- [Maloney and Wandell,1986]Maloney L. and Wandell B., Color constancy: a method for recovering surface spectral reflectance. *JOSA A*, 1986. **3**(1): p. 29-33.
- [Mann et al,1995]Mann S., Picard R. and Section M. I. o. T. P. C., On being'undigital'with digital cameras: extending dynamic range by combining differently exposed pictures. 1995: Citeseer.
- [Mansouri,2002]Mansouri A., Region tracking via level set PDEs without motion computation. *Ieee Transactions on PAMI*, 2002. **24**(7): p. 947.
- [McKenna et al,2000]McKenna S., Jabri S., Duric Z., et al., Tracking groups of people. *Computer Vision and Image Understanding*, 2000. **80**(1): p. 42-56.
- [Mittal and Davis,2003]Mittal A. and Davis L., M 2 Tracker: a multi-view approach to segmenting and tracking people in a cluttered scene. *International Journal of Computer Vision*, 2003. **51**(3): p. 189-203.
- [Mittal and Davis,2002]Mittal A. and Davis L. S., M2Tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo. *Computer Vison - Eccv 2002, Pt 1*, 2002. **2350**: p. 18-33.
- [Mori and Malik,2002]Mori G. and Malik J., Estimating human body configurations using shape context matching. *Computer Vision - Eccv 2002 Pt Iii*, 2002. **2352**: p. 666-680.
- [Nakaya and Harashima,1994]Nakaya Y. and Harashima H., Motion compensation based on spatial transformations. *IEEE Transactions on circuits and systems for video technology*, 1994. **4**(3): p. 339-356.

- [Nieweglowski et al,1993]Nieweglowski J., Campbell T. and Haavisto P., A novel video coding scheme based on temporal prediction using digital image warping. *IEEE transactions on consumer electronics*, 1993. **39**(3): p. 141-150.
- [Ning et al,2008]Ning H., Wang L., Hu W., et al. Model-based tracking of human walking in monocular image sequences. 2008: Citeseer.
- [Ning et al,2002]Ning H. Z., Wang L., Hu W. M., et al., Articulated model based people tracking using motion models. *Fourth IEEE International Conference on Multimodal Interfaces, Proceedings*, 2002: p. 383-388 543.
- [Niu and Grimson,2006]Niu C. and Grimson E. Recovering non-overlapping network topology using far-field vehicle tracking data. 2006.
- [Oren and Nayar,1995]Oren M. and Nayar S., Generalization of the Lambertian model and implications for machine vision. *International Journal of Computer Vision*, 1995. **14**(3): p. 227-251.
- [Orwell et al,1999]Orwell J., Remagnino P. and Jones G., Multi-camera color tracking. vs, 1999: p. 14.
- [Otsuka and Mukawa,2004]Otsuka K. and Mukawa N. Multiview occlusion analysis for tracking densely populated objects based on 2-d visual angles. 2004: IEEE.
- [Owens et al,2000]Owens H., Westland S. and Doherty L., Representation of natural reflectance spectra by auto-associative neural network". *Colour Image Science* (Derby, UK: Derby University) pp, 2000: p. 304-311.
- [Park et al,2006]Park U., Jain A., Kitahara I., et al., ViSE: visual search engine using multiple networked cameras. *Pattern Recognition*, 2006. **3**: p. 1204-1207.
- [Perez et al,2002]Perez P., Hue C., Vermaak J., et al., Color-based probabilistic tracking. *Computer Vision—ECCV 2002*, 2002: p. 661-675.
- [Plaenkers and Fua,2002]Plaenkers R. and Fua P., Model-based silhouette extraction for accurate people tracking. *Computer Vision - Eccv 2002, Pt Ii*, 2002. **2351**: p. 325-339.
- [Plankers and Fua,2001]Plankers R. and Fua P., Tracking and modeling people in video sequences. *Computer Vision and Image Understanding*, 2001. **81**(3): p. 285-302.
- [Porikli and Divakaran,2003]Porikli F. and Divakaran A. Multi-camera calibration, object

- tracking and query generation. 2003.
- [Rahimi et al,2004a]Rahimi A., Dunagan B. and Darrell T., Simultaneous calibration and tracking with a network of non-overlapping sensors. 2004a.
- [Rahimi et al,2004b]Rahimi A., Dunagan B. and Darrell T., Tracking people with a sparse network of bearing sensors. Computer Vision-ECCV 2004, 2004b: p. 48-58.
- [Rein-Lien et al,2002]Rein-Lien H., Abdel-Mottaleb M. and Jain A. K., Face detection in color images. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. **24**(5): p. 696-706.
- [Remagnino et al,1998]Remagnino P., Tan T. and Baker K., Multi-agent visual surveillance of dynamic scenes. Image and Vision Computing, 1998. **16**(8): p. 529-532.
- [Rohr,1994]Rohr K., Towards Model-Based Recognition of Human Movements in Image Sequences. Cvgip-Image Understanding, 1994. **59**(1): p. 94-115.
- [Rosales and Sclaroff,2000]Rosales R. and Sclaroff S., Learning and synthesizing human body motion and posture. fg, 2000: p. 506.
- [Schettini et al,1995]Schettini R., Barolo B. and Boldrin E., Colorimetric calibration of color scanners by back-propagation. Pattern Recognition Letters, 1995. **16**(10): p. 1051-1056.
- [Schroder and Mech,1995]Schroder K. and Mech R. Combined description of shape and motion in an object based coding scheme using curved triangles. 1995: Published by the IEEE Computer Society.
- [Scotta and Sainb,2005]Scotta D. and Sainb S., Multi-dimensional density estimation. Data mining and data visualization, 2005: p. 229.
- [Sidenbladh et al,2000]Sidenbladh H., Black M. and Fleet D., Stochastic tracking of 3D human figures using 2D image motion. Computer Vision—ECCV 2000, 2000: p. 702-718.
- [Sidenbladh and Black,2001]Sidenbladh H. and Black M. J., Learning image statistics for Bayesian tracking. Eighth Ieee International Conference on Computer Vision, Vol Ii, Proceedings, 2001: p. 709-716 758.
- [Sminchisescu and Triggs,2001a]Sminchisescu C. and Triggs B., Covariance scaled sampling

for monocular 3D body tracking. 2001 IEEE Computer Society Conference on CVPR, Vol 1, Proceedings, 2001a: p. 447-454 1186.

[Sminchisescu and Triggs,2001b]Sminchisescu C. and Triggs B., A robust multiple hypothesis approach to monocular human motion tracking. RAPPORT DE RECHERCHE-INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE, 2001b.

[Smith et al,2005]Smith K., Gatica-Perez D. and Odobez J., Using particles to track varying numbers of interacting people. 2005.

[Stauffer and Grimson,1999]Stauffer C. and Grimson W. Adaptive background mixture models for real-time tracking. 1999: Citeseer.

[Stiles et al,1977]Stiles W., Wyszecki G. and Ohta N., Counting metameric object-color stimuli using frequency-limited spectral reflectance functions. JOSA, 1977. 67(6): p. 779-784.

[Stillman et al,1999]Stillman S., Tanawongsuwan R. and Essa I. A system for tracking and recognizing multiple people with multiple cameras. 1999: Citeseer.

[Sturges and Whitfield,1995]Sturges J. and Whitfield T., Locating basic colours in the Munsell space. Color Research & Application, 1995. 20(6): p. 364-376.

[Sullivan and Baker,1991]Sullivan G. J. and Baker R. L. Motion compensation for video compression using control grid interpolation. in Acoustics, Speech, and Signal Processing, 1991. ICASSP-91., 1991 International Conference on. 1991.

[Szeliski and Shum,1996]Szeliski R. and Shum H., Motion estimation with quadtree splines. IEEE Transactions on PAMI, 1996. 18(12): p. 1199-1210.

[Takebe et al,2002]Takebe K., Ito N., Nakauchi S., et al. A digital imaging system with color constancy from a couple of images. 2002: IEEE.

[Tang et al,1996]Tang K., Man K., Kwong S., et al., Genetic algorithms and their applications. IEEE Signal Processing Magazine, 1996. 13(6): p. 22-37.

[Terzopoulos and Szeliski,1992]Terzopoulos D. and Szeliski R., Tracking with Kaiman snakes. Active vision, 1992: p. 3-20.

[Tieu et al,2005]Tieu K., Dalley G. and Grimson W., Inference of non-overlapping camera network topology by measuring statistical dependence. 2005.

- [Tipping and Bishop,1999]Tipping M. and Bishop C., Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 1999. **61**(3): p. 611-622.
- [Tsutsui et al,1998]Tsutsui H., Miura J. and Shirai Y., Optical flow-based person tracking by multiple cameras. *Proc. MVA*, 1998. **98**: p. 418-421.
- [Usui et al,1992]Usui S., Nakauchi S. and Nakano M., Reconstruction of Munsell color space by a five-layer neural network. *Journal of the Optical Society of America A*, 1992. **9**(4): p. 516-520.
- [Utsumi et al,1998]Utsumi A., Mori H., Ohya J., et al., Multiple-human tracking using multiple cameras *Proc. IEEE International Conference Automatic Face and Gesture Recognition*, 1998: p. 498-503.
- [Utsumi and Ohya,2000]Utsumi A. and Ohya J. Multiple-camera-based human tracking using non-synchronous observations. 2000.
- [van den Hengel et al,2006]van den Hengel A., Dick A. and Hill R., Activity topology estimation for large networks of cameras. 2006.
- [Van Der Merwe et al,2001]Van Der Merwe R., Doucet A., De Freitas N., et al., The unscented particle filter. *Advances in Neural Information Processing Systems*, 2001: p. 584-590.
- [Van Der Merwe and Wan,2001]Van Der Merwe R. and Wan E. The square-root unscented Kalman filter for state and parameter-estimation. 2001: Citeseer.
- [Vrhel and Trussell,2002]Vrhel M. and Trussell H. Color scanner calibration via a neural network. 2002: IEEE.
- [Wachter and Nagel,1999]Wachter S. and Nagel H. H., Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, 1999. **74**(3): p. 174-192.
- [Walsh,2004]Walsh B., Markov chain monte carlo and gibbs sampling. 2004.
- [Wang and Liu,2009] Wang X., Liu J.,Tracking multiple people under occlusion and across cameras using probabilistic modeles.*Journal of Zhejiang University Sci A* .2009 **10**(7)985-996 [doi:10.163/jzus.A0820474
- [Wang and Lee,1996a]Wang Y. and Lee O., Use of two-dimensional deformable mesh structures for video coding, Part II-the analysis problem and a region-based coder

- employing an active mesh representation. IEEE Trans. Circuits and Systems for Video Technology, 1996a. 6(6): p. 647-659.
- [Wang and Lee,1996b]Wang Y. and Lee O., Use of two -dimensional deformable mesh structures for video coding Part 1-the synthesis problem:mesh-based function approximation and mapping. IEEE Trans. Circuits and Systems for Video Technology, 1996b. 6(6): p. 636-646.
- [Wren et al,1996]Wren C., Azarbayejani A., Darrell T., et al., Pfunder: Real-time tracking of the human body. Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996: p. 51-56 386.
- [Yang et al,2008]Yang C., Duraiswami R., Gumerov N., et al. Improved fast gauss transform and efficient kernel density estimation. 2008: IEEE.
- [Yilmaz et al,2004]Yilmaz A., Li X. and Shah M., Contour-based object tracking with occlusion handling in video acquired using mobile cameras. Ieee Transactions on PAMI, 2004: p. 1531-1536.
- [Yue et al,2004]Yue Z., Zhou S. and Chellappa R. Robust two-camera tracking using homography. 2004: Citeseer.
- [孔庆杰 和 刘允才,2007]孔庆杰 和 刘允才, 面向广域视频监控的无重叠视域多摄像机目标跟踪. 2007.
- [王永刚 和 王爱民,2001]王永刚 和 王爱民, 舌象分析仪中舌色重现方法的研究. 照明工程学报, 2001. 12(002): p. 4-10.
- [王永刚 和 卫保国,2000]王永刚 和 卫保国, 基于数码相机的图像采集系统. 测控技术, 2000. 19(5): p. 17-19.
- [王亮 等,2002]王亮, 胡卫明 和 谭铁牛, 人运动的视觉分析综述. 计算机学报, 2002. 25(3): p. 225-237.
- [向桂山 等,2007]向桂山, 王宣银 和 梁冬泰, 基于人脸肤色和特征的实时检测跟踪算法. 光电工程, 2007. 34(004): p. 44-48.
- [谷元保 和 付宇卓,2005]谷元保 和 付宇卓, 一种基于灰度世界模型自动白平衡方法. 计算机仿真, 2005. 22(009): p. 185-188.
- [侯志强 和 韩崇昭,2006]侯志强 和 韩崇昭, 视觉跟踪技术综述. 自动化学报, 2006.

32(004): p. 603-617.

[范瑞霞等,2006]范瑞霞,任复明和方浩,基于小区域匹配的运动人体跟踪.北京理工大学学报,2006. 26(11).

[刘李杰和蔡德钧,2000]刘李杰和蔡德钧,一种面向运动的视频对象分割算法.计算机学报,2000. 23(012): p. 1326-1331.

[刘重庆,2003]刘重庆,基于网格模型的视频对象跟踪研究与应用.博士学位论文 上海交通大学,2003.

[刘关松和徐建国,2002]刘关松和徐建国,基于有限维线性模型的监督颜色回归恒常性算法.华东理工大学学报:自然科学版,2002. 28(004): p. 423-425.

[王选贺和刘济林,2011]王选贺和刘济林,基于核密度估计的遮挡人体跟踪.浙江大学学报

[徐柱,2010]徐柱,运动目标检测与阴影消除的研究,硕士学位论文,江苏大学硕士学位论文 2010.

[张泽旭等,2003]张泽旭,李金宗和李冬冬,一种运动目标多特征点的鲁棒跟踪方法研究.数据采集与处理,2003. 18(004): p. 423-428.

[汤顺青,1990]汤顺青,色度学.北京:北京理工大学出版社,1990.

[赵忠旭和沈兰荪,2000]赵忠旭和沈兰荪,基于人工神经网络的彩色校正方法研究.中国图象图形学报:A辑,2000(009): p. 785-789.

作者在攻博期间发表（录用）的论文

- [1] Xuanhe Wang, Jilin Liu, Tracking multiple people under occlusion and across cameras using probabilistic models. Journal of Zhejiang University Sci A .2009 10(7)985-996 [doi:10.163/jzus.A0820474].
- [2] 王选贺, 刘济林, 基于核密度估计的遮挡人体跟踪. 浙江大学学报: 工学版 (已录用)