

密级: _____

中国科学院研究生院

硕士学位论文

基于视觉特征的行人流量统计算法研究

作者姓名: _____ 纪颖夏 _____

指导教师: _____ 焦建彬 教授 中国科学院研究生院 _____

学位类别: _____ 工程硕士 _____

学科专业: _____ 物流工程 _____

培养单位: _____ 中国科学院研究生院工程教育学院 _____

2012年4月

Pedestrian counting based on visual features

By

Yingxia Ji

A Paper Submitted to

Graduate University of Chinese Academy of Sciences

In partial fulfillment of the requirement

For the degree of

Master of Logisticis Engineering

College of Engineering

April, 2012

中国科学院研究生院直属院系

研究生学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

中国科学院研究生院直属院系

学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名：

日 期：

导师签名：

日 期：

摘 要

行人流量统计作为物流工程领域的一个重要研究内容，也是智能视频监控的一个研究热点，具有广泛的应用前景。近年来图像处理、视频分析、模式识别与人工智能等相关领域的研究成果，为行人流量统计提供了一个行之有效的方法。

基于视觉特征的行人流量统计的一般研究方法是通过识别个体目标来达到统计计数的目的，此方法受到遮挡等条件的限制。近来，基于视觉特征的回归方法在人群粗略统计的应用逐渐成为研究热点。因此，视觉特征集和回归方法的选择成为行人流量统计的核心内容，这也是本文研究的重点。

本文提出了一种基于视觉特征的行人流量统计算法，该算法通过提取少量有效的视觉特征集合，利用基于 RBF (Radial Basis Function, 径向基函数)核函数的 SVM(Support Vector Machine, 支持向量机)回归方法以及分段学习回归方法统计行人流量。

本文的主要研究内容和贡献如下：

1、视觉特征集选择上，选择了少量有效的视觉特征集合作为特征向量，提高了时间效率，并且对比了在此特征向量集合上多元线性回归和基于 RBF 核函数的 SVM 回归方法的实验结果。

2、使用分段回归学习的方法，实现不同密度人群分别训练回归模型，从而使人数的统计更加准确，实验结果证明了分段回归方法的有效性。

关键词：视频监控，人群流量统计，特征提取，SVM 回归，分段回归

Abstract

- **Yingxia Ji** (Logistics Engineering)

Directed by: **Jianbin Jiao** (Professor)

Pedestrian counting in public places plays a key role in many applications, such as evacuating from a dense region to a sparse one when an emergency happens, or optimizing the design of traffic infrastructures for providing better transportation services. Furthermore, social security and surveillance strongly depend on the effectiveness of pedestrian counting. In recent years, image processing, video analysis, pattern recognition and artificial intelligence research in related fields provide an effective way for pedestrian counting.

This problem is particularly challenging when the scenes are densely crowded, in which the techniques of tracking a single pedestrian has limited effectiveness. Recently, regression method based on visual features of rough estimates of the crowd is becoming a hot research topic. Therefore, visual feature set and the choice of regression method becomes the core content of the pedestrian counting.

In this paper, in consideration of the fact that there may exist certain relationship between the crowd density and feature, feature-based regression was applied to count number of dense crowd.

Area of blobs, number of contours, ratio between the contour perimeter and area, texture feature based on GLCM (Gray Level Dependence Matrix) are employed. Then pedestrian counting was estimated by a regression function.

We proposed two regression methods and compared experimental results. The experimental results show that a certain role in the area of pedestrian counting based on the RBF (Radial Basis Function) kernel of SVM (Support Vector Machine) regression method, and the use of segmented regression method can improve the accuracy of the experimental results.

KEY WORDS: Visual Surveillance, Pedestrian Counting, Feature Extraction, SVM Regression, Segmented Regression

目 录

摘 要	I
目 录	III
图目录	V
表目录	VI
第一章 绪论	1
1.1 课题背景和研究意义	1
1.2 国内外研究现状	2
1.3 人群目标检测的主要研究内容和分类	3
1.3.1. 人群目标检测的研究内容	3
1.3.2. 人群目标检测的研究分类	4
1.3.3. 人群目标检测的发展趋势	5
1.4 本文的研究内容	6
1.5 本文的组织结构	7
第二章 行人流量统计的研究基础	9
2.1 引言	9
2.2 运动目标检测方法	9
2.3 视觉特征概念和分类	13
2.4 回归分析基本概念	15
2.5 本章小结	17
第三章 基于视觉特征的行人流量统计	19
3.1 基于视觉特征的行人流量统计算法框架	19
3.2 平均背景建模	19
3.3 视觉特征提取	21
3.3.1. 前景面积特征	21
3.3.2. Canny 边缘特征	22
3.3.3. 轮廓特征	25
3.3.4. 基于灰度共生矩阵的纹理特征	26
3.4 多元线性回归	28
3.5 基于 RBF 核函数的 SVM 回归预测方法	29
3.6 基于视觉特征的行人流量统计算法实验	31
3.6.1. 实验步骤	31
3.6.2. 测试集合及算法评估	32

3.6.3. 实验结果及分析	32
3.6.4. 其他数据集实验	34
3.7 本章小结	36
第四章 基于分段回归的行人流量统计	37
4.1 分段回归的基本理论	37
4.2 基于分段回归的行人流量统计可行性分析	38
4.3 分段线性回归学习方法	38
4.4 基于 SVM 的分段回归方法	40
4.5 基于分段回归的行人流量统计算法实验	43
4.5.1. 实验步骤	43
4.5.2. 测试集合及算法评估	43
4.5.3. 实验结果及分析	44
4.6 本章小结	44
总结与展望	47
参考文献	49
致 谢	55

图目录

图 1-1 行人流量统计的框图	6
图 3-1 人群图像场景监控实例	19
图 3-2 背景建模实例	21
图 3-3 前景面积特征提取实例	21
图 3-4 前景面积特征和人数关系图	22
图 3-5 Canny 边缘特征提取实例	25
图 3-6 Canny 边缘和人数关系图	25
图 3-7 图像轮廓特征提取实例	26
图 3-8 纹理特征和人物关系图	28
图 3-9 实验数据集实例	32
图 3-10 SVM 回归预测人数与真实人数对比图	33
图 3-11 测试集绝对误差变化图	33
图 3-12 累积误差对比图	34
图 3-13 文献[54]数据集实例	35
图 3-14 文献[54]数据集特征提取实例	36
图 4-1 分段回归样本实例	37

表目录

表 3-1 不同参数 c 对应的回归模型预测误差 ($g=1.0$ $p=0.25$)	30
表 3-2 不同参数 g 对应的回归模型预测误差 ($c=1.0$ $p=0.25$)	30
表 3-3 不同方法实验结果对比	34
表 3-4 本文特征集合和回归方法在数据集[54]的实验结果	36
表 4-1 分段线性回归和线性回归实验结果对比图	38
表 4-2 不同核函数和惩罚因子 c 值对应的 accuracy 值 ($g=0.5$ $p=0.25$)	42
表 4-3 使用分段回归的实验对比	44

第一章 绪论

1.1 课题背景和研究意义

近年来，随着国民经济的快速增长、社会的迅速进步和国力的不断增强，视频监控在物流工程领域、生产生活等方面得到了非常广泛的应用。

虽然监控系统已经广泛地存在于银行、商场、车站和交通路口等公共场所，但实际的监控任务仍需要较多的人工完成，而且现有的视频监控系统通常只是录制视频图像，提供的信息是没有经过解释的视频图像，没有充分发挥监控的实时性和主动性，并且在事后查找方面也不具备便利条件。为了能实时分析、跟踪、判别监控对象，并在异常事件发生时提示、上报，为及时决策、正确行动提供支持，视频监控的“智能化”就显得尤为重要。

智能视频监控是利用计算机视觉技术对视频信号进行处理、分析和理解，在不需要人为干预的情况下，通过对序列图像的自动分析对监控场景中的变化进行目标的定位、识别和跟踪，并在此基础上分析和判断目标的行为，在异常情况发生时及时发出警报或提供有用信息，有效地协助安全人员处理危机，并最大限度地降低误报和漏报现象^[1]。

智能视频监控的主要研究内容包括物体识别、越界识别、轨迹跟踪、遗留或丢失物体识别、车牌识别、车速测量、流量统计、逆行告警、涂鸦行为识别、打架等反常行为识别等。

本文的研究内容主要是监控视频的流量统计，其中以行人流量统计为例。

行人流量统计作为物流工程领域的重要研究内容是在实施采购，生产，市场营销的管理和决策方面不可缺少的数据。对于零售业而言，行人流量更是非常基础的指标。因其与销售量直接的正比关系，人们对行人流量统计数据的高度重视由来已久。

另外，从安全预警的角度考虑，管理人员可以通过观察该区域当前的实际人数等对流量比较大的公共区域采取预防突发事件的措施，如紧急疏散人群等；同时还可以根据公共区域的行人流量统计设置障碍墙等安全措施。

本文通过对固定场景运动人群目标的特征进行提取，然后对其进行回归分析，提出一种行人流量统计估计较好的算法，具有重要的研究意义和实用价值。

在论文工作期间，本文作者参与以下课题：

- 1) “基于多源数据的飞行器进近威胁目标检测跟踪及行为预测”，国家自然科学基金重点项目（课题编号：61039003）；
- 2) “飞行器威胁目标识别与图像鲁棒匹配理论与方法”，国家 973 计划子课题（课题编号：2010CB731804-2）；
- 3) “复杂环境下动态目标跟踪技术”中科院“百人计划”择优支持项目（项目编号：99T3009EA2）。

1.2 国内外研究现状

由于智能视频监控系统较传统监控有很多优势，越来越多的政府、公司、企事业单位以及个人用户倾向于采用智能视频监控系统。

就目前的研究来讲，智能视频监控技术主要包括：运动目标的检测、分类、跟踪以及高层次的行为分析与理解等方面，所涉及到的领域包括计算机视觉、模式识别、图像处理等。智能视频监控是一个得到全世界广大科研人员的关注的具有挑战性的研究课题。

早在 1997 年，美国国防高级研究项目署（DARPA: The U. S. Defense Advanced Research Projects Agency）设立了以卡内基梅隆大学为首联合美国十几所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）^[2]，主要研究用于战场及普通民用场景进行监控的自动视频理解技术。

2006 年，IBM 公司宣布已经开发完成一种用于分析实时视频监控系统数据安全性的软件，这款软件具有发现监控场景中潜在的安全威胁并自动报警。这一产品被 IBM 公司命名为“智能监控系统”（S3: Smart Surveillance System）^[3]，简称“S3”。S3 系统可以将监控系统中各个监控点获取的视频信息通过计算机网络传递到整个系统软件上，进行远程监控，同时 S3 系统是一个开放框架的智能视频监控，能够使各种处理和分析模块方便的嵌入 S3 系统，并提供多摄像机的数据融合和事件关联分析。

此外,由美国马里兰大学开发的实时视频监控系统 W4^[4] 是一个融合了形状分析和跟踪技术的视频监控系统,它可以通过摄像机定位人并分割人的身体部分,还能通过建立人体的外形模型实现复杂环境下的多个人体目标的跟踪,并检测出人体是否携带了物体等;美国的 EMW 公司推出的 IVM 智能视频监控系统采用县级你的视频处理技术,可以实现对场景的实时、自动监控,同时识别、跟踪监控场景内的潜在威胁事件,并能够侦测多种安全威胁事件和行为。日本也开展了用于公共区域及智能小区的视觉监控的计划(CDVP: The Cooperative Distributed Vision Project)^[5]等。

虽然智能化视频监控技术在国内起步较晚,但得到了国家的高度重视,863 项目为此开展了重大项目研究,公安部也为此在全国开展城市报警与监控系统建设“3111”试点工程^[6]。中科院自动化所里的生物识别与安全技术研究中心 CBSR (Center for Biometrics and Security Research) 研究开发的智能视频监控系统^[7]作为这方面技术与研究成果的集中展示,其包括的功能有:多目标遮挡跟踪,目标检测、跟踪以及分类,目标异常行为的识别与报警等。

近年来,视频监控及相关技术的迅速发展,为从视频分析的角度进行人群的管理和监控奠定了基础。由此,关于人群运动分析的研究逐渐得到人们的关注。

在物流工程领域,通过零售 ERP 系统或 POS 系统结合,可以获取单位时间成交量、客单价、各部类商品销量等有用的信息,再加上客流方面的统计数据,许多管理上的盲点会变的清晰。

1.3 人群目标检测的主要研究内容和分类

1.3.1. 人群目标检测的研究内容

由于研究人员往往把监控视频中行人目标作为监控对象,所以,行人检测成为了一个研究的热点。而实际应用中,对人体目标的个体检测并不能适用于各种环境,尤其是在人数较多的公共场所,由于人群密度过大产生的遮挡等问题使对个体目标的检测十分困难。因此,对人群目标检测的研究十分必要。

其中人群的密度和数量是人群特性的两个重要指标。人群密度估计主要是通过模式分类的相关方法对提取到的人群密度特征进行分类,人数估计则通过

个体分割识别或对提取到的特征，如运动像素面积、边缘、纹理等，进行回归分析实现计数。

1.3.2. 人群目标检测的研究分类

人群的密度和数量是人群特性的两个重要指标。

人群密度估计的算法通常是先提取人群密度相关的图像特征，然后通过模式分类的相关方法其按照密度级别进行分类。其中以 Marana^[8-11]为代表，对人群大小进行密度估计。Marana 通过纹理特征对人群图像进行人群密度估计，他认为人群图像具有纹理图像的特征，低密度的图像具有较粗的纹理，而高密度的图像具有较细的纹理，根据这个特点，统计人群纹理图像的灰度共生矩阵（Grey Level Dependence Matrix）的统计值和纹理图像的闵科夫斯基分形维数（Minkowski Fractal Dimension）作为特征，利用 Kohonen 的神经网络自组织映射模型（Self Organising Map）进行分类，将人群分为“很低”、“低”、“中”、“高”、“很高”五个密度等级。Cho^[12]将图像二值化处理，并提取边缘，前景面积特征，用最小二乘法和全局搜索混合算法学习，然后用神经网络建模估计人群密度。Rahmalan^[13]将人群图像看成纹理图像，使用了尺度不变正交切比雪夫矩（Invariant Orthonormal Chebyshev Moments）做特征，用自组织映射学习和分类算法进行密度等级分类，并在分别比较其与灰度共生矩阵特征和闵科夫斯基分形维数特征方法在不同时间，天气，人群密度不同情况下的实验结果。其中对结果影响较大的因素主要是阴影和二值噪声。Wu^[14]将图像划分为一个个小的单元，用透视投影模型校正，从每个单元提取纹理特征矢量，用 SVM 进行分类和密度异常检测。Ma^[15]采用多尺度局部纹理信息和基于置信区间的软分类方法统计人群密度。

对人群计数的研究主要分成三类：

传统意义上的人群计数都是通过个体分割识别来实现的，是人数统计最直观的方法。通过利用视频或图像当中的形状、颜色和运动信息来实现人的分割和识别，进而达到人数统计的目的。常用的特征有人头，四肢，人体轮廓等。Sacchi^[16]针对旅游景点偏僻地区设计了人群计数系统，该系统针对简单的背景采用自适应的背景前景检测方法进行运动检测，然后利用实时模式识别工具将人群从其他运动目标中区别出来，以便准确的统计人数。Zhao^[17]将人体形状作

为模型, 结合人体高度、摄像机模型、图像曲线等先验知识, 利用贝叶斯框架和马尔可夫链蒙特卡罗(Markov chain Monte Carlo)方法从人群中分割出个体的人头轮廓, 从而实现人群计数。Wu^[18]将人体作为各个部分的集合, 提取轮廓方向特征, 用 boosting 训练, 各个部分的检测结果融合成一个或几个人的模型, 最后用最大后验概率估计方法检测, 可以处理多人遮挡问题。

第二类是先检测后跟踪, 获取运动目标或目标某类特征的运动轨迹, 然后分析计数。Rabaud^[19]用 KLT 跟踪记录目标运动轨迹, 然后对运动轨迹进行分析聚类, 获取人群人数。Zhang^[20]针对拥挤人群, 用组跟踪(Group tracking)解决遮挡和运动分割不全问题, 实验结果可以实时跟踪, 实验正确率达 95%。Chen^[21]用迭代 Mean Shift 方法对目标进行跟踪, 实现人群计数。Kilambi^[22]利用场景的先验知识和摄像机标定技术, 对运动分割后的前景进行透视校正, 用卡尔曼滤波方法分别进行个体和团体跟踪计数。

第三类是基于视觉特征的回归分析方法, 这类方法通过对提取到的特征, 如运动像素面积、边缘、纹理等, 进行回归分析或用学习算法进行分类学习, 实现计数或密度估计。Regazzoni^[23]提出基于分布式数据融合的实时人群估计方法, 先提取面积和边缘特征, 然后利用非线性分段模型训练特征和人数之间的静态关系, 用分布式卡尔曼滤波网络进行动态预测, 结合两者实现人群计数。Ma^[24]基于像素特征实现人群密度估计, 用几何方法进行透视校正, 同时对异常人群密度进行检测。但是只能适用于低密度人群, 而且文献中缺少密度估计的实验结果。Kong^[25]用边缘方向直方图和块面积直方图作为特征, 经透视校正后对特征和人数进行线性拟合, 然后计数。Chan^[26]人群图像中提取面积、轮廓、边缘、纹理等共 21 个特征, 然后用高斯过程回归学习特征和人数之间的数学关系, 从而通过特征, 统计出人群人数。

前两类方法都需要对个体进行识别, 不适用于密集或者大规模人群。本文研究的方法属于第三类范畴。

1.3.3. 人群目标检测的发展趋势

目前关于单个人体的识别已经有很成熟的体系, 而在行人数量比较大时, 由于人群中存在大量的遮挡, 难以准确地对人群中的单个行人进行检测、分割和跟踪的情况下, 如何在不进行单个目标的检测的前提下, 利用前景块的特征

通过统计学习的方法建立判决规则，直接估计行人的数量就成为了人群检测方面当前国内外研究的重点。因此，寻找合适的能表述人群密度的特征，并选用合适的学习算法或回归算法进行分析就成了所要关注的问题。

常用的特征有形状、面积、边缘、纹理、运动轨迹等，常用的学习方法有SVM、线性或非线性回归、高斯回归等。

人群计数的难点在于人群图像常常会存在透视和遮挡(或者说是重叠)现象，一般通过对图像做透视校正或者遮挡处理解决。

1.4 本文的研究内容

本文提出一种人群检测方法，能够在提取小维数特征集的情况下对人群目标人数预测估计值，并且提出了分段回归的算法，解决了在不同人群密度下，由于人体目标出现遮挡交错，而导致回归模型不匹配的问题。同时建立了一个视频监控下人群计数的系统原型，能在实验环境下取的较好效果。

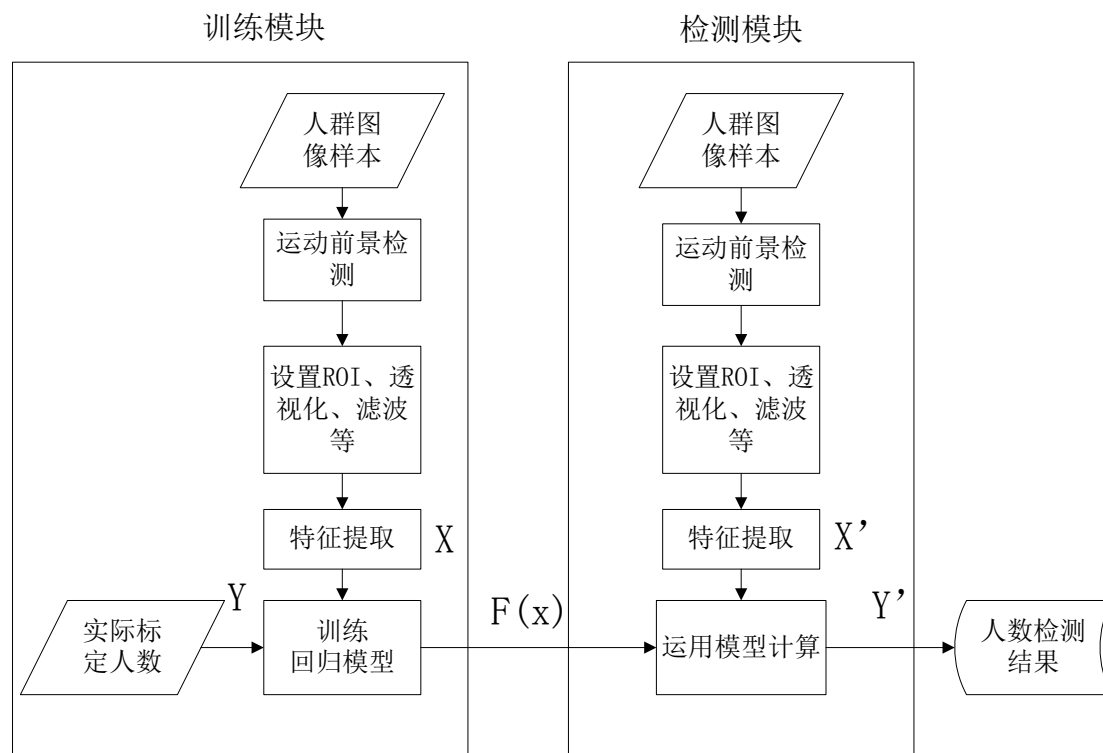


图 1-1 行人流量统计的框图

1.5 本文的组织结构

第一章，绪论。主要论述人群目标检测算法的研究背景和意义，分析国内外的研究现状以及发展趋势，并介绍人群目标检测的主要研究内容和分类，最后说明了本文的主要研究目的和研究内容。

第二章，行人流量统计的研究基础。主要论述行人流量统计算法的基础研究内容，包括运动目标的检测方法，视觉特征概念和分类以及回归分析的基本概念。

第三章，基于视觉特征的行人流量统计。主要介绍视觉特征集的选择，并对比了在此特征向量集合上多元线性回归和基于 RBF 核函数的 SVM 回归方法的实验结果。

第四章，基于分段回归的行人流量统计。提出使用分段回归学习的方法，对不同密度人群分别训练回归模型，从而使人数的统计更加准确。

最后是结束语，总结本文的主要工作，展望了未来工作的方向，以及对如何进一步提高算法准确性等研究热点和难点的探讨。

第二章 行人流量统计的研究基础

行人流量统计作为物流工程领域的重要研究内容，也是计算机视觉领域的热点研究问题之一。目前，如何提高行人流量统计算法的准确性等方面的研究逐步成为行为分析研究的重点。本章将主要介绍基于视觉特征和回归分析的行人流量统计算法的主要构成，并对一些基本方法进行介绍。

2.1 引言

对于一个特定的场景，我们希望计算机能够通过观察场景中人群目标，自动预测场景中的人数。但是大多数监控场景中的人群分布具备很大多样性，并且很难区分个体，通过个体识别来计数。针对这种情况，研究人员提出了基于视觉特征和回归分析的行人流量统计算法，这也是本文提出的算法的基础。基于视觉特征和回归分析的行人流量统计算法通常分为三部分，首先是运动目标检测，在固定场景提取运动前景，然后提取有效的视觉特征，最后通过数据回归分析的方法预测场景中的人数。

2.2 运动目标检测方法

运动目标检测是指在序列图像中检测出变化区域并将运动目标从背景图像中提取出来。通常情况下，目标分类、跟踪和行为理解等后处理过程仅仅考虑图像中对应于运动目标的像素区域，因此运动目标的正确检测与分割对于后期处理非常重要。然而，由于场景的动态变化，如天气、光照、阴影及杂乱背景干扰等的影响，使得运动目标的检测与分割变得相当困难。

根据摄像头是否保持静止，运动检测分为静态背景和运动背景两类。运动目标检测一直是视频序列分析的重点和难点，而大多数视频监控系统是摄像头固定的，因此静态背景下的基于运动信息的运动目标检测算法受到广泛关注。

目前常用的运动目标检测方法主要有光流法、帧间差分法和背景差分法。光流法^[27-31]：

当人的眼睛观察运动物体时，物体的景象在人眼的视网膜上形成一系列连续变化的图像，其连续变化的信息不断通过视网膜，好像一种光的流，故称为光流(Optical flow)。1981年 Horn^[32] 将光流的概念引入到计算机视觉的研究当中，将图像的灰度变化同二维速度场联系起来，Horn 认为客观物体在真实世界中的相对运动时连续变化的，所以运动过程投影到摄像机传感器二维平面上的图像也应该是连续变化的，图像上的光流场就反映了像素点在图像平面上运动的速度矢量。而在图像序列中，物体的运动体现为图像平面上像素值灰度的变化。对此我们就能给出这样的假设:对于客观世界中的同一个点，投影到图像上的一个点，该点的瞬时灰度值是不变的。

基于上述假设，我们可以给出这样的式子：

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (2-1)$$

上式中， $I(x, y, t)$ 表示在 t 时刻处于图像平面上的 (x, y) 位置的像素点的灰度值， $I(x + dx, y + dy, t + dt)$ 表示 $t + dt$ 时刻改点运动到了 $(x + dx, y + dy)$ 位置的灰度值。 dx 、 dy 、 dt 分别表示 x 方向上的偏移量、 y 方向上的偏移量和时间上的偏移量。由于我们假设其灰度值连续变化，所以 dt 趋近于 0 时我们认为其灰度值不变。

上述公式在 dt 趋近于 0 时由泰勒级数展开并忽略二阶无穷小，可以得到：

$$\frac{\partial I(x, y, t)}{\partial t} = \frac{\partial I}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial I}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial I}{\partial t} \quad (2-2)$$

因为被跟踪的像素点在 t 时刻没有发生改变，因此：

$$\frac{\partial I(x, y, t)}{\partial t} = 0 \quad (2-3)$$

所以上式可整理为：

$$-\frac{\partial I}{\partial t} = \frac{\partial I}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial I}{\partial y} \frac{\partial y}{\partial t} \quad (2-4)$$

其中 $\frac{\partial x}{\partial t}$ 可以理解为 t 时刻 x 方向上的变化率，我们用 u 代替， $\frac{\partial y}{\partial t}$ 可以理解为 t 时刻 y 方向上的变化率，我们用 v 代替，则上式最终写为：

$$-\frac{\partial I}{\partial t} = \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v \quad (2-5)$$

在实际应用中，我们可以通过其他约束条件与光流约束方程求解出 u ， v 。设 $w=(u,v)$ ，则像素点 (x,y) 的光流可以表示为 $w(x,y)$ 。在理想情况下，被监控场景中的背景应该是不会变化的，因此，背景像素点的光流应该为 0，而不为 0 的点就应该是检测到的前景运动目标。

光流的研究是利用视频图像序列中像素强度数据的时域变换和相关性来确定各自像素位置的运动，即研究图像灰度在时间上的变化与景象中物体结构极其运动的关系。通过计算光流场可以从图像中近似计算不能直接得到的运动场。

光流法的优点是能够检测独立运动的对象，可以在不预知场景信息的情况下得到完整的运动信息，并且适用于摄像机运动的情况，在处理背景运动和遮挡问题上有很大优势。但是由于该算法要对序列图像中的所有像素进行计算，处理数据庞大，算法复杂耗时，对硬件设备要求较高，很难实现实时检测，并且对图像噪声较敏感，抗噪性差。因此不能被直接应用于实时视频监控场合。可以通过与背景差法结合在特定时刻使用^[33]。

帧间差分法^[34-38]：

帧间差分法，也称为序列差分法，是利用视频序列中的连续两帧或几帧图像的差异来进行目标检测，比较适用于存在多个运动目标和摄像机移动的情况。在连续两帧或多帧图像间计算逐像素的灰度差，并通过设置阈值来确定对应运动前景的像素，进而得到运动前景区域。常用的帧间差分提取方法有两帧差和三帧差。

两帧差提取运动目标的数学表述如下：用 I_n ， I_{n-1} 表述图像序列中任意相邻的两帧图像，则差分图 D_n 可以表示为：

$$D_n(x,y)=|I_n(x,y)-I_{n-1}(x,y)| \quad (2-6)$$

对差分图 D_n 阈值化，即可确定运动前景 M_n ：

$$M_n(x,y)=\begin{cases} 0 & D_n(x,y)\leq T \\ I_n(x,y) & D_n(x,y)>T \end{cases} \quad (2-7)$$

其中 T 是预设的阈值。

两帧差分的方法能够简单快速检测出相邻两帧图像中的运动物体，但是易受干扰导致的光照变化影响，导致检测整张图像都是运动目标的错误结果。

三帧差分法和两帧差分原理相同，只是在具体构造差分图上有些许不同，其能克服两帧差在外界扰动下出现检测失效的情况，过滤非连续像素阶跃变化，对摄像机晃动等干扰具有较强自免疫能力。

帧间差分法的优缺点：由于相邻两帧间的时间间隔非常短，所以，背景部分往往变化不大，差分运算时能相互消去。从而，能避免背景的某些变化而引起的误差，使算法对场景的光线不太敏感，适应光线等变化较频繁的环境，具有较好的稳定性。用前一帧图像作为当前帧的背景模型具有较好的实时性，其背景不积累，且更新速度快、算法简单、计算量小。

算法的不足在于对环境噪声较为敏感，阈值的选择相当关键，选择过低不足以抑制图像中的噪声，过高则忽略了图像中有用的变化。对于比较大的、颜色一致的运动目标，有可能在目标内部产生空洞，无法完整地提取运动目标。

帧间差分法也可以与背景差法结合取得比较好的效果^[39]。

背景差分法^[40-42]：

背景差分法是一种有效的运动目标检测算法，由于算法简单，且监控摄像在多数情况下是固定的，近年来在运动目标检测中大量应用。其基本思想是利用背景的参数模型来近似背景图像的像素值，将当前帧与背景图像进行差分比较实现对运动区域的检测，其中区别较大的像素区域被认为是运动区域，而区别较小的像素区域被认为是背景区域。

其数学表示如下：用 $I(x, y)$ 表示当前帧图像， $B(x, y)$ 为背景图像，则差值图像 $d(x, y)$ 表示为：

$$d(x, y) = I(x, y) - u(x, y) \quad (2-8)$$

对差值图像 $d(x, y)$ 阈值化：

$$D(x, y) = \begin{cases} 1, & |d(x, y)| > T \\ 0, & otherwise \end{cases} \quad (2-9)$$

其中 T 是预设的阈值。可以采用自适应算法进行确定。

背景差分法算法简单，实时性强，检测准确，目前应用较广。它一般能提供最完整的特征数据，原理和算法设计简单，所得结果直接反映了运动目标的位置、大小、形状等信息，能够得到比较准确的运动目标信息。

但是该方法性能依赖于所使用的背景建模技术。由于背景的建立时间和当前检测图像往往有一定时间间隔，当环境光线等发生变化时，当前图像的背景部分往往和开始建立的背景模型发生很大变动，从而使大量的背景被认为是前景目标，影响了算法的检测效果。受光线、天气等外界条件变化的影响较大。常用的背景建模算法有单高斯背景建模算法和混合高斯背景建模算法等。

本文中所做实验采用的即是背景差分法，后文再做详细介绍。

2.3 视觉特征概念和分类

图像特征是表征一个图像最基本的属性，它可以是人类视觉能够识别的自然特征，也可以是人为定义的某些特征^[43]。

特征提取是计算机视觉和图像处理中的一个概念。其内容是使用计算机提取图像信息，决定每个图像的点是否属于一个图像特征。特征提取的结果是把图像上的点分为不同的子集，这些子集往往属于孤立的点、连续的曲线或者连续的区域。至今为止特征并没有通用的定义。其精确定义往往由问题或者应用类型决定。

特征是一个数字图像中“有趣”的部分，它是许多计算机图像分析算法的起点。一个算法是否成功往往由其使用和定义的特征决定。因此特征提取最重要的一个特性是“可重复性”，即同一场景的不同图像所提取的特征应该是相同的。

特征提取是图像处理中的一个初级运算，也就是说它是对一个图像进行的第一个运算处理。它检查每个像素来确定该像素是否代表一个特征。假如它是一个更大的算法的一部分，那么这个算法一般只检查图像的特征区域。由于许多计算机图像算法使用特征提取作为其初级计算步骤，因此有大量特征提取算法被发展，其提取的特征各种各样，它们的计算复杂性和可重复性也非常不同。

特征被检测后可以从图像中被抽取出来。其结果被称为特征描述或者特征向量。

常用的视觉特征分为以下几类：

1) 颜色特征：是一种全局特征，其描述了图像或图像区域所对应的景物的表面性质。一般颜色特征是基于像素点的特征，此时所有属于图像或图像区域

的像素都有各自的贡献。由于颜色对图像或图像区域的方向、大小等变化不敏感，所以颜色特征不能很好地捕捉图像中对象的局部特征。另外，仅使用颜色特征查询时，如果数据库很大，常会将许多不需要的图像也检索出来。

颜色直方图是最常用的表达颜色特征的方法，其优点是不受图像旋转和平移变化的影响，进一步借助归一化还可不受图像尺度变化的影响。颜色直方图的缺点是没有表达出颜色空间分布的信息。

颜色特征常用的特征提取与匹配方法还包括：颜色集、颜色矩、颜色聚合向量、颜色相关图等。

2) 纹理特征：也是一种全局特征，描述了图像或图像区域所对应景物的表面性质。由于纹理只是一种物体表面的特性，并不能完全反映出物体的本质属性，所以仅仅利用纹理特征是无法获得高层次图像内容的。与颜色特征不同，纹理特征不是基于像素点的特征，它需要在包含多个像素点的区域中进行统计计算。在模式匹配中，这种区域性的特征具有较大的优越性，不会由于局部的偏差而无法匹配成功。作为一种统计特征，纹理特征常具有旋转不变性，并且对于噪声有较强的抵抗能力。但是，纹理特征也有其缺点，一个很明显的缺点是当图像的分辨率变化的时候，所计算出来的纹理可能会有较大偏差。另外，由于有可能受到光照、反射情况的影响，从 2-D 图像中反映出来的纹理不一定是 3-D 物体表面真实的纹理。

例如，水中的倒影，光滑的金属面互相反射造成的影响等都会导致纹理的变化。由于这些不是物体本身的特性，因而将纹理信息应用于检索时，有时这些虚假的纹理会对检索造成“误导”。

在检索具有粗细、疏密等方面较大差别的纹理图像时，利用纹理特征是一种有效的方法。但当纹理之间的粗细、疏密等易于分辨的信息之间相差不大的时候，通常的纹理特征很难准确地反映出人的视觉感觉不同的纹理之间的差别。

纹理特征描述方法分类包括：统计方法、几何法、模型法、信号处理法。

纹理特征的提取与匹配方法主要有：灰度共生矩阵、Tamura^[44]纹理特征、自回归纹理模型、小波变换等。

3) 形状特征：各种基于形状特征的检索方法都可以比较有效地利用图像中感兴趣的目标来进行检索，但它们也有一些共同的问题，包括：①目前基于形

状的检索方法还缺乏比较完善的数学模型；②如果目标有变形时检索结果往往不太可靠；③许多形状特征仅描述了目标局部的性质，要全面描述目标常对计算时间和存储量有较高的要求；④许多形状特征所反映的目标形状信息与人的直观感觉不完全一致，或者说，特征空间的相似性与人视觉系统感受到的相似性有差别。另外，从 2-D 图像中表现的 3-D 物体实际上只是物体在空间某一平面的投影，从 2-D 图像中反映出来的形状常不是 3-D 物体真实的形状，由于 viewpoint 的变化，可能会产生各种失真。

常用的特征表示方法包括轮廓特征和区域特征。图像的轮廓特征主要针对物体的外边界，而图像的区域特征则关系到整个形状区域。

常用的形状特征主要有：Harr-Like 特征，HOG 特征，SIFT 特征，MSO 特征等。

4) 空间关系特征：所谓空间关系，是指图像中分割出来的多个目标之间的相互的空间位置或相对方向关系，这些关系也可分为连接/邻接关系、交叠/重叠关系和包含/包容关系等。通常空间位置信息可以分为两类：相对空间位置信息和绝对空间位置信息。前一种关系强调的是目标之间的相对情况，如上下左右关系等，后一种关系强调的是目标之间的距离大小以及方位。显而易见，由绝对空间位置可推出相对空间位置，但表达相对空间位置信息常比较简单。

空间关系特征的使用可加强对图像内容的描述区分能力，但空间关系特征常对图像或目标的旋转、反转、尺度变化等比较敏感。另外，实际应用中，仅仅利用空间信息往往是不够的，不能有效准确地表达场景信息。为了检索，除使用空间关系特征外，还需要其它特征来配合。

提取图像空间关系特征可以有两种方法：一种方法是首先对图像进行自动分割，划分出图像中所包含的对象或颜色区域，然后根据这些区域提取图像特征，并建立索引；另一种方法则简单地将图像均匀地划分为若干规则子块，然后对每个图像子块提取特征，并建立索引。

2.4 回归分析基本概念

回归分析 (Regression analysis)^[45]是确定两种或两种以上变数间相互依赖的定量关系的一种统计分析方法。回归分析的运用十分广泛，按照涉及的自变量

的多少，可分为一元回归分析和多元回归分析；按照自变量和因变量之间的关系类型，可分为线性回归分析和非线性回归分析。如果在回归分析中，只包括一个自变量和一个因变量，且二者的关系可用一条直线近似表示，这种回归分析称为一元线性回归分析。如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为多元线性回归分析。

回归分析的一般步骤如下：

1) 根据预测目标，确定自变量和因变量

明确预测的具体目标，也就确定了因变量。如预测具体目标是下一年度的销售量，那么销售量 Y 就是因变量。通过市场调查和查阅资料，寻找与预测目标的相关影响因素，即自变量 X ，并从中选出主要的影响因素。

2) 建立回归预测模型

依据自变量和因变量的历史统计资料进行计算，在此基础上建立回归分析方程，即回归分析预测模型。

3) 进行相关分析

回归分析是对具有因果关系的影响因素（自变量）和预测对象（因变量）所进行的数理统计分析处理。只有当变量与因变量确实存在某种关系时，建立的回归方程才有意义。因此，作为自变量的因素与作为因变量的预测对象是否有关，相关程度如何，以及判断这种相关程度的把握性多大，就成为进行回归分析必须要解决的问题。进行相关分析，一般要求出相关关系，以相关系数的大小来判断自变量和因变量的相关的程度。

4) 检验回归预测模型，计算预测误差

回归预测模型是否可用于实际预测，取决于对回归预测模型的检验和对预测误差的计算。回归方程只有通过各种检验，且预测误差较小，才能将回归方程作为预测模型进行预测。

5) 计算并确定预测值

利用回归预测模型计算预测值，并对预测值进行综合分析，确定最后的预测值。

2.5 本章小结

本章对行人流量统计的研究基础进行了综述。首先，分别介绍了运动目标检测的几种主要方法及其优缺点；然后，介绍了视觉特征提取的概念和分类；最后，简要介绍了回归分析的概念和一般步骤。

第三章 基于视觉特征的行人流量统计

3.1 基于视觉特征的行人流量统计算法框架

在本文第一章已经介绍了人群目标检测的分类和现在研究较多的行人流量统计算法，本文的实验即是基于视觉特征和回归分析的行人流量统计方法的研究。图 3-1 显示的是 Chan^[46]等人采用人群数据集中的一帧图像，行人流量统计的研究目的简单说就是要统计该帧图像上的人数。



图 3-1 人群图像场景监控实例

基于特征与回归方法的人群流量估计主要分三个部分，首先通过运动目标检测分割出运动的人群，然后从前景图像中提取出需要的视觉特征，最后通过训练学习特征和人数之间的关系，得到特征与人数之间的回归模型。

3.2 平均背景建模

在第二章第二节，介绍了三种常用的运动目标检测的方法，并分析了其优缺点和各自的应用范围。本文重点研究行人流量检测部分，背景建模只是其中的一部分，我们希望该部分的计算消耗较少，所以采用较简单且有效的平均背景建模。

其基本思想是，计算每个像素的平均值作为背景模型，检测当前帧时，只需将当前帧像素值 $I(x, y)$ 减去去背景模型中相对像素的平均值 $u(x, y)$ ，得到差

值 $d(x, y)$ ，将 $d(x, y)$ 与一个阈值 TH 进行比较，那么得到输出图像 $output$ 的值如下：

$$d(x, y) = I(x, y) - u(x, y) \quad (3-1)$$

$$output(x, y) = \begin{cases} 1, & |d(x, y)| > TH \\ 0, & otherwise \end{cases} \quad (3-2)$$

这里 TH 可以采用自适应算法进行确定，需要计算每个像素帧间差的平均值 $u_{diff}(x, y)$ 和标准差 $diff_{std}(x, y)$ 。令 $I_t(x, y)$ 代表 t 时刻的图像中 (x, y) 处像素值， $inter$ 代表两帧之间的间隔， $F_t(x, y)$ 代表 t 时刻的图像中 (x, y) 处像素帧间差，公式如下：

$$F_t(x, y) = |I_t(x, y) - I_{t-inter}(x, y)| \quad (3-3)$$

$$u_{diff}(x, y) = \frac{1}{M - inter} \sum_{t=inter+1}^M F_t(x, y) \quad (3-4)$$

$$diff_{std}(x, y) = \sqrt{\frac{1}{M - inter} \sum_{t=inter+1}^M (F_t(x, y) - u_{diff}(x, y))^2} \quad (3-5)$$

M 代表背景模型帧数，通常要足够大保证 u_{diff} 和 $diff_{std}$ 的精确性。 TH 可以这样确定：

$$TH = u_{diff} + \beta \times diff_{std} \quad (3-6)$$

其中 β 一般设置为 2。

为了提高算法鲁棒性，可以在检测完后对背景模型进行更新，对所有像素，令 $u(x, y)$ ， $u_{diff}(x, y)$ ， $diff_{std}(x, y)$ 更新为 $u'(x, y)$ ， $u'_{diff}(x, y)$ ， $diff'_{std}(x, y)$ ：

$$u'(x, y) = (1 - \alpha) \times u(x, y) + \alpha \times I(x, y) \quad (3-7)$$

$$u'_{diff}(x, y) = (1 - \alpha) \times u_{diff}(x, y) + \alpha \times F(x, y) \quad (3-8)$$

$$diff'_{std}(x, y) = (1 - \alpha) \times diff_{std}(x, y) + \alpha \times |F(x, y) - u'_{diff}(x, y)| \quad (3-9)$$

这里 α 为学习率 (0 ~ 1)， α 越大，对背景变化的适应速度越快。

图 3-2 为在实验数据集上获得的背景图实例：



图 3-2 背景建模实例

3.3 视觉特征提取

在基于视觉特征的行人流量统计算法中，选取什么样的特征至关重要，一般认为良好的特征应该具有好的区分性，可靠性，独立性，并且数量均衡。

本文所采用的特征主要分为图像像素统计特征和纹理特征两个大类，下面就依次详细介绍。

3.3.1. 前景面积特征

前景面积特征简单的说就是计算前景运动目标的像素个数，由于人群规模越大，人数越多，前景面积就显著的增长，因此，前景面积是统计行人流量最简单有效的特征。这里我们采用减背景的方法得到运动前景，得到图 3-3(a)，再经过选择 ROI(Region of Interest)区域和使用形态学方法滤波去噪得到图 3-3(b)。在本文所使用的数据集将 ROI 区域手工设置为道路部分。统计图 3-3(b)上白色部分像素个数即为提取的前景面积特征。



(a)

(b)

图 3-3 前景面积特征提取实例

本文所使用的前景面积特征按照文献[26]中的方法对不同坐标的像素进行了加权处理，以此消除人物靠近摄像机和远离摄像机前景面积大小不同所产生的误差。对景深影响不大的场景可省略此步骤。

图 3-4 是前景面积统计特征和人数之间的关系图，x 轴是前景像素面积统计值，y 轴为人数。如图所示，前景面积特征与人数之间具有明显的线性关系。

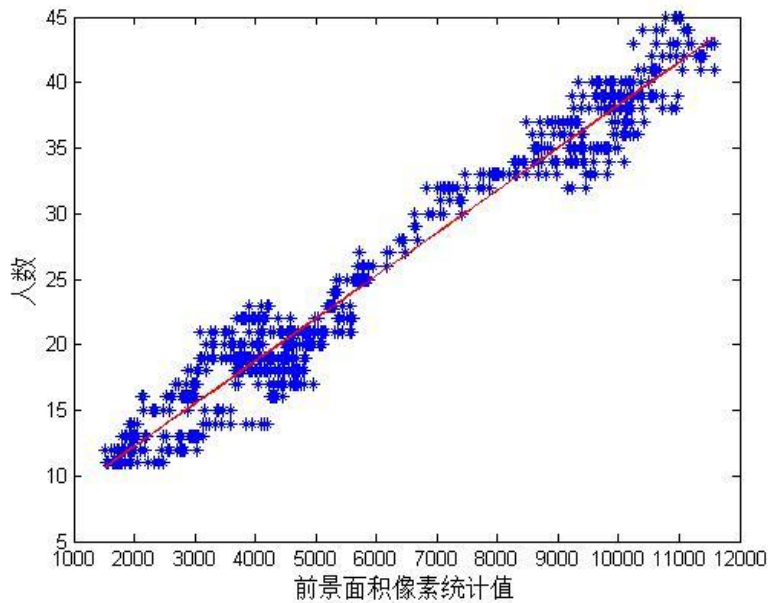


图 3-4 前景面积特征和人数关系图

3.3.2. Canny 边缘特征

图象的边缘是指图象局部区域亮度变化显著的部分，常存在于目标与背景、目标与目标、目标与影子之间。当人群密度越高，人群重叠就越多，人与人之间的轮廓线边缘点也就越丰富。因此边缘特征也可以作为人群检测的特征之一。

图象的边缘部分集中了图象的大部分信息，边缘的确定与提取对于整个图象场景的识别与理解非常重要，同时也是图象分割所依赖的重要特征，边缘检测主要是图象的灰度变化的度量、检测和定位。

图象边缘检测的基本步骤^[47]：

1) 滤波。边缘检测主要基于导数计算，但受噪声影响。但滤波器在降低噪声的同时也导致边缘强度的损失。

2) 增强。增强算法将邻域中灰度有显著变化的点突出显示。一般通过计算梯度幅值完成。

3) 检测。但在有些图象中梯度幅值较大的并不是边缘点。最简单的边缘检测是梯度幅值阈值判定。

4) 定位。精确确定边缘的位置。

自从 1959 提出边缘检测以来, 经过五十多年的发展, 已有许多种不同的边缘检测方法。本文选择使用 Canny 边缘检测算子 (Canny, 1986)^[48] 检测图像边缘。

Canny 边缘检测算子是当前较受欢迎的边缘检测技术, 它由三个主要目标行成^[47]:

- 1) 无附加响应的最优检测: 可以通过最优平滑处理来减少噪声效应;
- 2) 检测边缘位置和实际边缘位置之间距离最小的正确定位: 即正确性, 在正确的位置检测到边缘, 可以通过非极大值抑制 (non-maximum suppression, 相当于峰值检测) 处理来实现;
- 3) 减少单边缘的多重响应而得到单响应: 限制的是单个边缘点对于亮度变化的定位, 这是因为并非只有一条边缘表示为当前检测到的边缘。

Canny 指出高斯算子对图像平滑处理是最优的。高斯算子 $g(x, y, \sigma)$ 可表示为

$$g(x, y, \sigma) = e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (3-10)$$

沿坐标轴对单位向量 $U_x[1, 0]$ 和 $U_y[0, 1]$ 进行微分处理, 可得

$$\nabla g(x, y) = \frac{\partial g(x, y, \sigma)}{\partial x} U_x + \frac{\partial g(x, y, \sigma)}{\partial y} U_y = -\frac{x}{\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} U_x - \frac{y}{\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} U_y \quad (3-11)$$

上式给出高斯模板导数系数的计算方法, 高斯模板把一阶微分和高斯平滑结合在一起进行处理。图像经过平滑处理后, 边缘就成为数据岭 (ridge of data)。想要在一个正确点上标记一个边缘 (并减少多重响应), 我们可以在图像与边缘法线方向上给出一阶导数的算子进行计算, 这个函数的极大值应该是边缘数据的峰值, 在该点原图像的梯度是最清晰的, 因而成为边缘位置。

因此, 我们要找到一个算子 G_n , 它就是高斯函数 g 在法线 n_{\perp} 方向上的一阶导数:

$$G_n = \frac{\partial g}{\partial n_{\perp}} \quad (3-12)$$

式中 n_{\perp} 可通过高斯函数 g 的一阶导数与图像 P 进行卷积计算求出，并适当缩放尺度。真正边缘点的位置就是与图像进行卷积运算的 G_n 的极大值点。当微分值为 0，其极大值为

$$\frac{\partial(G_n * P)}{\partial n_{\perp}} = 0 \quad (3-13)$$

代入上式得

$$\frac{\partial^2(G * P)}{\partial n_{\perp}^2} = 0 \quad (3-14)$$

上式给出一个算子的基底，该算子符合 Canny 边缘检测算子的一个标准，即应该在正确位置检测到边缘。这正是极大值抑制，相当于返回峰值，因而可以细化边缘算子的响应，给出正确位置的边缘点，而没有多重响应，而且对噪声的响应最小。

Canny 边缘算子的实现基本步骤

- 1) 图像读取和灰度化
- 2) 图像的高斯滤波
 - a) 根据一维高斯核进行两次滤波，生成一维高斯滤波系数，分别进行 x 向和 y 向的一维加权滤波，保存滤波后的数据。
 - b) 根据二维高斯核进行滤波，生成二维高斯滤波系数，采用高斯核进行高斯滤波
- 3) 图像增强——计算图像梯度及其方向
- 4) 非极大值抑制
- 5) 链接边缘点的滞后阈值处理

图 3-5 为从原始图像提取的 Canny 边缘，本文统计 ROI 区域的边缘像素数目，作为边缘特征。

图 3-6 是 Canny 边缘特征和人数之间的关系图，x 轴是 ROI 区域的边缘特征值，y 轴为人数。如图所示，Canny 边缘特征与人数之间具有明显的线性关系。



图 3-5 Canny 边缘特征提取实例

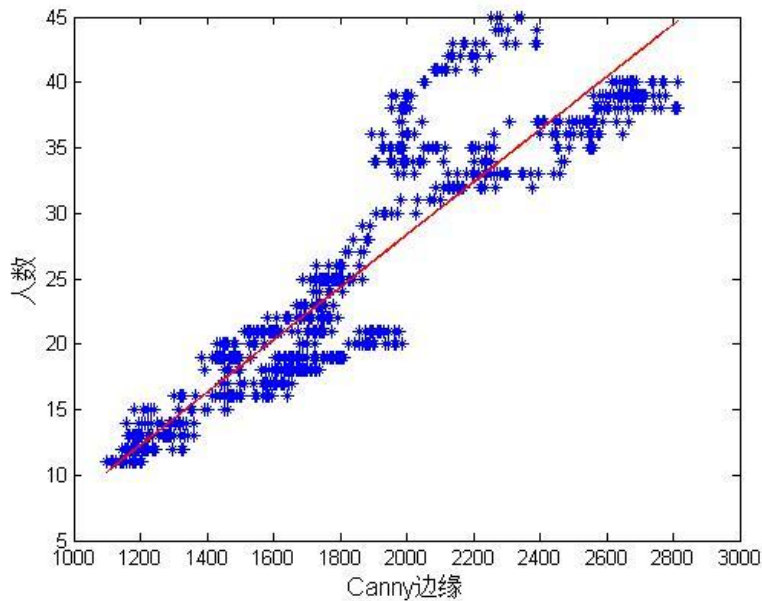


图 3-6 Canny 边缘和人数关系图

3.3.3. 轮廓特征

轮廓特征主要包括三类，轮廓个数，轮廓周长和轮廓周长面积比。本文采用遍历轮廓点的方法，实现对提取的二值前景图像区域的一次轮廓跟踪，得到所有轮廓信息。在轮廓跟踪过程中，首先扫描所有轮廓起点，然后跟踪其轮廓，得到一个封闭的轮廓信心，接着转向下一个轮廓，直到所有的轮廓跟踪完毕为止。能实现一次扫描，得到所有的轮廓信息。

图 3-7 为从经过背景差分法得到的二值前景图像提取的轮廓特征。如图所示，轮廓周长与面积比的特征描述了人群团块的复杂性，同等面积的情况下，轮廓周长越长，团块组成越复杂，可能容纳的人数相应越大。这也是影响人数的一个重要因素，因此，本文将轮廓个数和轮廓面积比作为一个有效特征。



图 3-7 图像轮廓特征提取实例

3.3.4. 基于灰度共生矩阵的纹理特征

灰度共生矩阵(Gray Level Dependence Matrix, GLDM)是 Haralick^[49]提出的基于二阶联合条件概率密度函数 $f(i, j | d, \theta)$ 的统计方法。

灰度共生矩阵是一种通过研究灰度的空间相关特效来描述纹理的常用方法, 每一个 $f(i, j | d, \theta)$ 是一对灰度值 (i, j) 发生在沿角度 θ 、相距 d 的两个像素上的概率, 这些估计值形成一个可写成矩阵形式的二维直方图。若将图像的灰度级定为 N 级, 那么共生矩阵为 $N * N$ 矩阵, 可表示为 $M_{(d, \theta)}(i, j)$, 其中位于 (i, j) 的元素的值 M 表示一个灰度值为 i 而另一个灰度值为 j 的两个相距为 (d, θ) 的像素对出现的次数。

研究发现, 不同密度的人群图像对应纹理模式不同, 高密度的人群图像表现为细纹理, 而低密度的人群图像表现为粗纹理, 因此可将人群图像看做纹理图像, 用相关的纹理特征来对场景图像进行统计分析, 找出其和人数之间得到关系。

纹理粗糙度可由直方图主对角线的相关分布来度量。对粗纹理的区域, 像素趋于具有相同的灰度, 因此对应灰度共生矩阵中的 m_{ij} 值较集中于主对角线附近, 此时 $(i - j)$ 值也较小。相反, 对于细纹理的区域, 其 m_{ij} 值较为分散。因此, 可以使用灰度共生矩阵的各种统计量作为纹理特征的度量。

本文采用基于 $\theta = 0^\circ$ 及 $d = 1$, 分别建立对应图像的灰度共生矩阵, 提取基于灰度共生矩阵的 4 类特征: 对比度、均匀度、能量和熵来建立 4 维的特征向量描述图像粗糙程度。

各特征表示如下:

(1) 对比度(contrast)

$$s_c(d, \theta) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-j)^2 f(i, j | d, \theta) \quad (3-15)$$

(2) 能量(Energy)

$$s_g(d, \theta) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} f(i, j | d, \theta)^2 \quad (3-16)$$

(3) 熵(Entropy)

$$s_p(d, \theta) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} f(i, j | d, \theta) \log f(i, j | d, \theta) \quad (3-17)$$

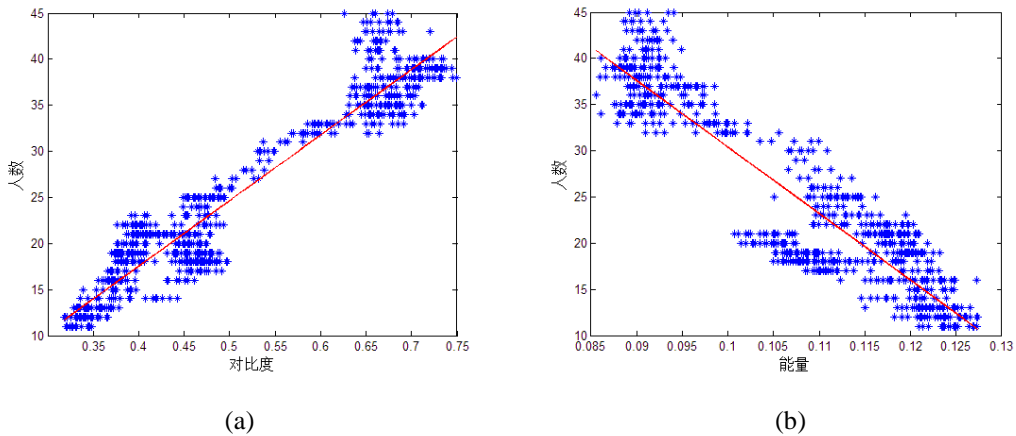
(4) 相关度(correlation)

$$s_{co} = \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} i * j * f(i, j | d, \theta) - \mu_x * \mu_y}{q_x * q_y} \quad (3-18)$$

其中 μ_x 、 μ_y 、 q_x 、 q_y 分别为 m_x 、 m_y 的均值和标准差, 而 m_x 是灰度共生矩阵中一行元素之和, m_y 是灰度共生矩阵中一列元素之和。

对于粗纹理, m_{ij} 值较集中于主对角线附近, 此时 $(i-j)$ 值也较小, 对比度、熵也越小, 能量和相关度越大, 相反, 对于细纹理相应的熵、对比度较大, 能量和相关度越小。

根据实验结果显示, 基于灰度共生矩阵统计值的纹理特征与实际人数基本符合线性关系。



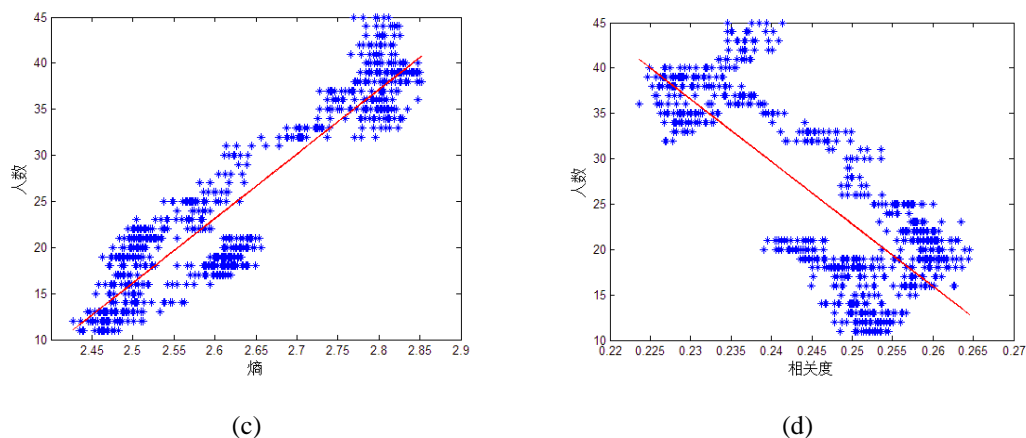


图 3-8 纹理特征和人物关系图
(a)对比度 (b)能量 (c)熵 (d)相关度

3.4 多元线性回归

变量的相关关系中最简单的是线性相关关系，设随机变量 x 与变量 y 之间存在线性相关关系，则由实验数据得到的点 $(x_i, y_i) (i=1, 2, \dots, n)$ 将散布在某一直线周围，因此可以认为 x 关于 y 回归函数的类型为线性函数，即 $y(x) = a + bx$ ，下面用最小二乘法估计估计参数 a, b ，设 y_i 服从正态分布 $N(a + bx_i, \sigma^2) (i=1, 2, \dots, n)$ ，分别求 $S = \sum_{i=1}^n (y_i - a - bx_i)^2$ 对 a, b 的偏导数^[45]，并令它们等于零，得方程组

$$\begin{cases} na + \left(\sum_{i=1}^n x_i\right)b = \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i\right)a + \left(\sum_{i=1}^n x_i^2\right)b = \sum_{i=1}^n x_i y_i \end{cases} \quad (3-19)$$

解得

$$\begin{cases} a = \bar{y} - b\bar{x} \\ b = \frac{l_{xy}}{l_{xx}} \end{cases} \quad (3-20)$$

其中：

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3-21)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (3-22)$$

$$l_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \quad (3-23)$$

$$l_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3-24)$$

线性方程 $y = a + bx$ 称为 x 关于 y 的线性回归方程， a ， b 称为回归系数。

在回归分析中，如果有两个或两个以上的自变量，就称为多元回归。事实上，一种现象常常是与多个因素相联系的，由多个自变量的最优组合共同来预测或估计因变量，比只用一个自变量进行预测或估计更有效，更符合实际。因此多元线性回归比一元线性回归的实用意义更大。

3.5 基于 RBF 核函数的 SVM 回归预测方法

支持向量机是在统计学习理论上发展起来的，1995 年 Vapnik^[50]提出了完整的统计学理论，并在此基础上发展了一种新的通用学习方法 SVM。其基本思想^[51]是通过一个非线性映射，把原空间上的非线性问题转化为另一个高维特征空间上的线性问题。但是在低维输入空间向高维特征空间映射过程中，由于空间维数急速增长，这使得在大多数情况下难以直接在特征空间直接计算最佳分类平面。SVM 通过定义核函数，巧妙的将这一问题转化到输入空间进行计算。

目前在 SVM 的学习中，对 SVM 和核函数的选择仍然没有形成一个统一的模式，SVM 算法参数选择大多还是凭借经验、实验对比、大范围搜索或交叉验证功能搜索最优的参数组合。常见的核函数有线性函数、多项式函数、径向基（RBF）函数、多层感知器函数等。

SVM 函数回归问题可以表述为：给定训练样本集 $T = \{x_i, y_i, i = 1, 2, \dots, l\}$ ，

$x \in R^n$ 且 $y \in R$ ，在特征空间 F 中构造一个线性回归函数：

$$f(x) = w^T \varphi(x) + b \quad (3-25)$$

W 是特征空间 F 中的向量， $\Phi(x)$ 把输入 x 映射到特征空间 F 中。 W 和 b 可以通过解如下约束优化问题获得：

$$\min_{w,b} T = \frac{1}{2} w^T w + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (3-26)$$

$$s.t. \begin{cases} y_i - w^T \varphi(x_i) - b \leq \varepsilon + \xi_i \\ w^T \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i^*, \xi_i \geq 0 \end{cases} \quad (3-27)$$

由于径向基(RBF)核函数在实际问题中表现出了良好的学习能力，因此拟选用其进行试验，其数学表达式为

$$K(x, x_i) = \exp[-\gamma \|x - x_i\|^2] \quad (3-28)$$

RBF 核函数可以将样本映射到一个高维空间，因此可以较好的处理非线性特征。RBF 核与多项式核相比具有参数少的优点，而参数的个数直接影响到模型选择的复杂性。

表 3-1 不同参数 c 对应的回归模型预测误差 ($g=1.0$ $p=0.25$)

C	0.01	0.1	1.0	10.0	100.0
Err	5.0600	2.7126	2.4734	2.5327	2.7381
MSE	7.1196	3.4184	2.9611	3.0530	3.3400

表 3-2 不同参数 g 对应的回归模型预测误差 ($c=1.0$ $p=0.25$)

g	Err	MSE
0.1	2.4880	2.9616
0.2	2.4688	2.9357
0.3	2.4342	2.9024
0.4	2.4138	2.8802
0.5	2.4021	2.8669
0.6	2.4180	2.8793
0.7	2.4312	2.9044
0.8	2.4389	2.9155
0.9	2.4527	2.9330

本文实验是使用 LIBSVM^[52]工具包实现的。 c 表示惩罚因子, g 表示核参数。关于参数 c 和 g 的选择, 我们在数据集上做了实验, 实验结果如表 3-1 和表 3-2 所示:

从表 3-1 和表 3-2 列举实验数据可以看出在 $c=1.0$, $g=0.5$ 的情况下, 得到的结果最优。即 Err 和 MSE 值均较小。

3.6 基于视觉特征的行人流量统计算法实验

本节总结了本章所提出的基于视觉特征的行人流量统计算法, 并且在数据集上进行实验。实验数据表明本章提出的行人流量统计算法在实验数据集上具有较好的效果。

3.6.1. 实验步骤

在本章的前几节, 本文已经详细介绍了实验所用的视觉特征集合和回归分析方法, 这里不再赘述。

本文采用由图像像素统计特征和纹理特征组成的特征向量集合, 并对比了线性回归和基于 RBF 核函数的 SVM 回归方法的实验结果, 该算法步骤如下:

1) 训练部分

a) 背景建模初始化.

读入训练视频, 对前 $n \times \text{inter}$ 帧进行背景建模, n 为预设值, 实验定为 100, inter 为取样间隔, 设为 5, 即每隔 5 帧取一帧;

b) 循环读入视频帧.

当前帧与得到的背景图像相减, 得到前景图像, 并进行 ROI 区域选择和形态滤波; 对每帧图像提取视觉特征信息;

c) 用所提取的视觉特征向量与标定的真实人数训练回归模型。

2) 测试部分

a) 导入已有背景.

b) 循环读入视频帧

当前帧与得到的背景图像相减, 得到前景图像, 并进行 ROI 区域选择和滤波; 对每帧图像提取视觉特征信息;

c) 使用训练好的回归模型进行回归预测

3.6.2. 测试集合及算法评估

本文采用的数据集是文献[26]作者 Chan 在网上提供的人群图像序列数据集。该数据集提供了一个小时视频图像序列，采集于 UCSD 的一条人行道上的一个静止摄像头。前 2000 帧提供了人工的标定数据用于实验。这 2000 帧用于实验数据，并分为训练和测试两部分。第 601 帧到 1400 帧用于训练统计模型，其余的 1200 帧用于测试和验证。在整个实验过程采用 ROI 区域计数。整个数据集含有人数 49,885 人。人数范围从 10 到 45。本文是在 Intel Core(TM) i7CPU (2.93GHz) 4GB RAM 的机器上完成实验的。



图 3-9 实验数据集实例

我们把回归模型计算得到的数值四舍五入取整作为人数测量值。实验结果用两个指标度量：平均绝对误差（absolute mean error）和均方差（mean squared error）。平均绝对误差用 Err 表示，均方差用 MSE 表示。假设人数真实数据是 x_1, x_2, \dots, x_n ，测量数据是 y_1, y_2, \dots, y_n ，则

$$Err = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (3-29)$$

$$MSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n-1}} \quad (3-30)$$

3.6.3. 实验结果及分析

图 3-8 为在实验数据集上使用 RBF 核函数 SVM 回归方法预测的人数和标定的真实人数的结果对比。从曲线我们可以看出，基于视觉特征的回归方法基本有效，能够估算出大致的人数。

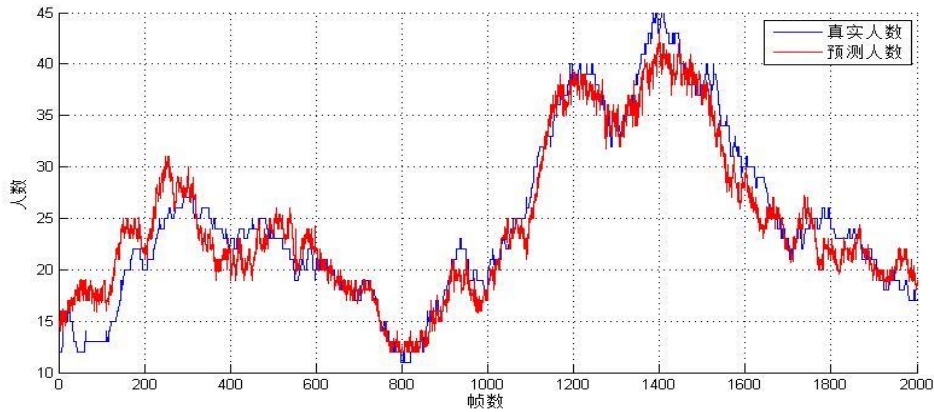


图 3-10 SVM 回归预测人数与真实人数对比图

图 3-9 是测试集 1200 帧的绝对测量误差曲线图。从该曲线图可以看出大部分帧的流量估计误差较小，并且持续时间较短。

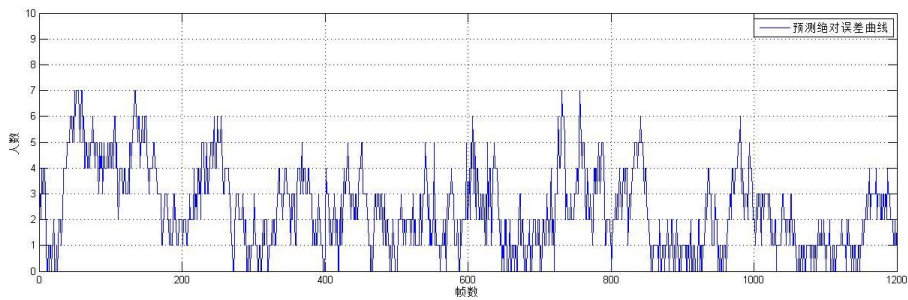


图 3-11 测试集绝对误差变化图

表 3-3 是在实验数据集上进行的多元线性回归和使用 RBF 核函数的 SVM 回归实验结果对比。由表中数据可以看出使用 RBF 核函数的 SVM 回归的实验结果优于多元线性回归。

同时，表 3-3 列出了文献[53]所使用的特征集在相同数据集上的实验结果，文献[53]采用了多组特征组合进行实验对比，并且采用了 K-近邻和多层感知回归联合和文献[26]中使用的高斯回归分别实验。这里为了避免由于个人实现差异带来的不必要的误差，直接引用文献[53]给出的实验结果。由表中数据可以看出，本文使用的特征集和回归方法得出的实验结果，均方差优于文献[53]的结果。

图 3-10 是在实验数据集上进行的多元线性回归和使用 RBF 核函数的 SVM 回归实验累积误差对比。由图中曲线可以看出，基于视觉特征的回归方法可以

把误差人数基本控制在五人以内，并且使用 RBF 核函数的 SVM 回归明显优于多元线性回归方法。

表 3-3 不同方法实验结果对比

特征	回归方法	Err	MSE
本文特征集	多元线性回归	2.6800	3.3455
本文特征集	RBF 核函数 SVM 回归	2.4050	2.8908
文献[53]特征集	K-近邻和多层感知 回归联合	-	4.157
文献[53]特征集	高斯过程回归	-	3.774

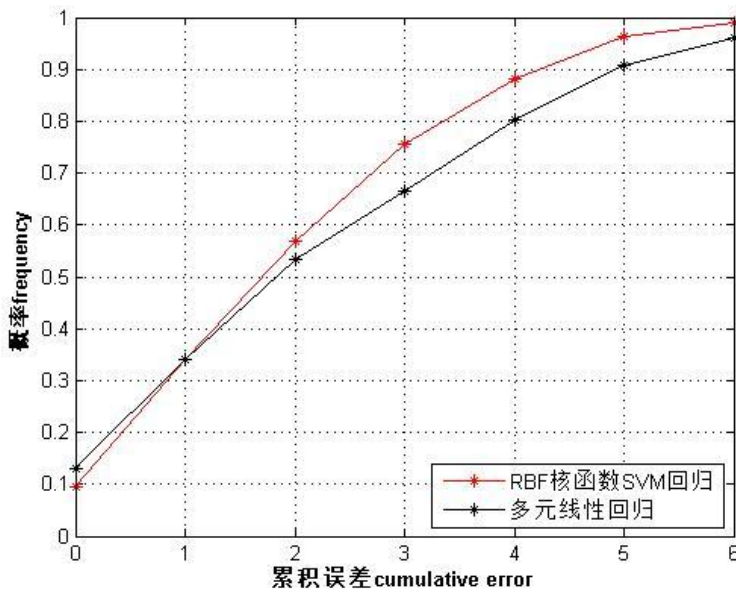


图 3-12 累积误差对比图

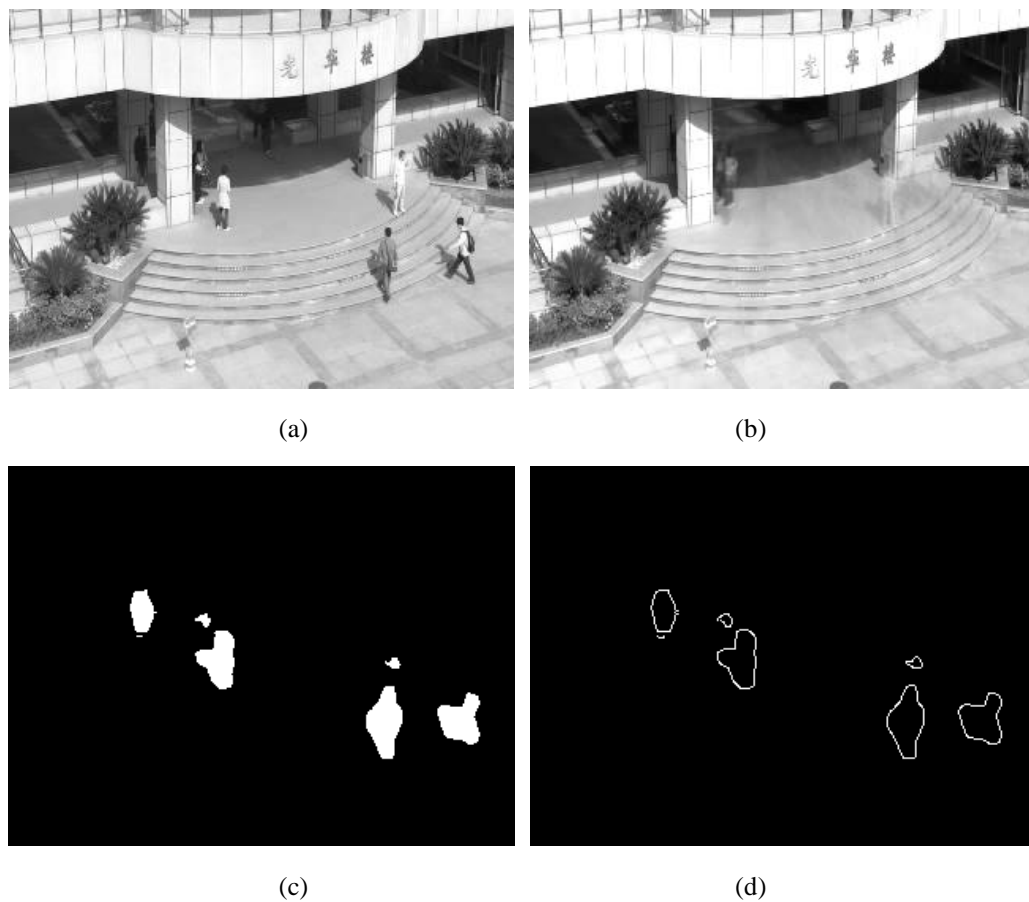
3.6.4. 其他数据集实验

为了验证本文所提出的视觉特征和回归方法有效，本文在复旦大学行人数据集^[54]上做了实验。该数据集拍摄于复旦大学光华楼前，一共 1500 帧，分成 5 段视频进行实验。其中第一段视频用来训练数据，其余作为测试集。该数据集的人数范围为 0 到 15。数据集场景实例如图 3-11。



图 3-13 文献[54]数据集实例

图 3-14 为文献[54]数据集特征提取实例，其中，图 3-14(a)为原始图像，图 3-14(b)为背景建模得到的背景图像，图 3-14(c)为使用背景差分法和形态滤波得到的前景图像，图 3-14(d)为在二值前景图像上得到的轮廓特征，图 3-14(e)为在原始图像提取的 Canny 边缘特征，图 3-14(f)为人工标定的 ROI 区域蒙版。



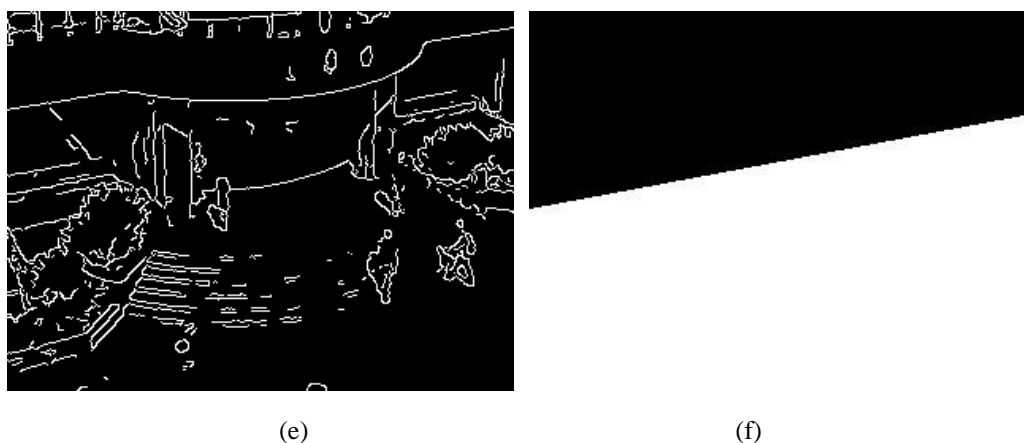


图 3-14 文献[54]数据集特征提取实例

(a)原始图像 (b)背景建模 (c)前景面积特征 (d)轮廓特征 (e)Canny 边缘特征 (f)ROI-MASK

表 3-4 为本文特征集合和回归方法在数据集[54]的实验结果，从表内数据可以看出，使用本文所提特征集合回归得到的 Err 和 MSE 均较小，基本能估算出场景内每一帧的人数，并且在人群较稀疏，范围较小的场景中。使用多元线性回归的结果要由于使用 RBF 核函数的 SVM 回归结果。

表 3-4 本文特征集合和回归方法在数据集[54]的实验结果

回归方法	多元线性回归		RBF 核函数的 SVM 回归	
数据集编号	Err	MSE	Err	MSE
1	0.7767	1.0988	0.6133	0.9608
2	0.8967	1.2089	1.3467	1.7214
3	1.5900	1.9620	1.7100	2.1351
4	1.2033	1.5848	2.2800	2.8461
5	0.8767	1.1318	2.8500	3.1384

3.7 本章小结

本章对基于视觉特征的行人流量统计算法进行了论述和研究。首先，介绍了实验中选取的视觉特征和回归方法，然后对该特征集合分别使用多元线性回归和 RBF 核函数的 SVM 回归方法在数据集上进行实验，并和文献[53]的实验结果做对比，证明选取的特征集合和回归方法有效。此外，本文还在小人数数据集上做了相应实验，证明选取的特征集合和回归方法也适应稀疏人群。

第四章 基于分段回归的行人流量统计

4.1 分段回归的基本理论

由于客观现象之间的联系较为复杂，通常不是简单的线性关系，因而使用线性回归的方法就受到很大限制，而选择非线性方法，使用什么样的曲线拟合，又是一个比较难判断的问题。在有些实际问题中，很难甚至无法找到合适的拟合曲线。如果做残差分析，会发现残差是系统的，而不是随机的，此时，曲线往往在一些区间内与另一些区间内有很大差别，因此，有时想用一种函数来拟合整个数据区域是不适宜的。在这种情况下，一般可以按曲线特征将其分成几个区间，然后在每个区间内分别作回归分析，及所谓分段回归。

如图 4-1 所示

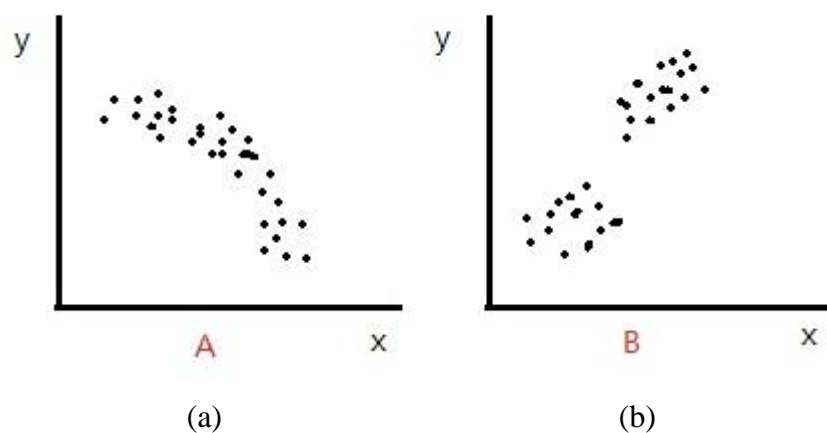


图 4-1 分段回归样本实例

(a)连续函数 (b)不连续函数

显然图 4-1(a)和图 4-2(b)均不呈直线关系，图 4-1(a)为一曲线，但曲线形式尚不能确定，图 4-1(b)则是一个不连续函数，使用传统的方法就比较困难，但是我们可以发现，在图 4-1(a)以 A 为分界线，图 4-1(b)以 B 为分界线，把两个图的数据分别分成两段，前后均成线性关系，因此，我们若能区别不同阶段分别用线性回归方法来进行拟合和进行回归分析，也许能取得较好的结果，而且

方法也比较简单。

所以分段回归方法实际上是设法把较复杂的曲线分割成多段简单函数可拟合的处理方法。

然而在分段回归中，准确的确立拐点是保证拟合效果的关键，从数理上要求是使相邻的两个函数在拐点处连续光滑，即两个函数在拐点具有相同的函数值和相同的 q 阶导数。然而在不完全清楚变量间作用机理的情况下，仅凭数据本身的表现，拐点在实践中不易把握。

4.2 基于分段回归的行人流量统计可行性分析

从第三章所选数据集使用线性回归的实验结果图 3-8 可以看出，相比真实数据，预测数据更趋向于中间值，在人群规模小于 30 的时候，预测数据大于真实数据居多，而人群规模大于 30 的时候，预测数据大多小于真实数据。由此我们推断，在人数规模不同时，人数和特征之间的相关性是在变化的，因此，我们考虑使用分段性回归来进行实验。

为了验证我们的猜想，我们在数据集^[46]使用第三章所提特征集合，用标定数据分段的方法做了验证实验，实验结果如下。

表 4-1 分段线性回归和线性回归实验结果对比图

回归方法	Err	MSE
线性回归	2.6800	3.3455
分段线性回归	2.5483	3.1496

实验证明分段结果优于直接对整体进行线性回归，如何分段就成为我们需要思考的问题。本文使用了两种分段方法，下面将分别介绍。

4.3 分段线性回归学习方法

分段线性回归 (Piecewise Linear Regression) 就是一个比较好的分段回归分析方法。这里我们采用 Giancarlo^[55]提出了一种分段线性回归的学习方法。能够解决连续和非连续的函数回归问题，通过结合空间聚类，分类学习和线性回

归的方法，使用分段映射作为解决方案。

算法步骤如下：

1) 局部回归

建立包含样本 (x_k, y_k) 局部数据集 C_k ，使每一对 $(x, y) \in S$ 和其 $c-1$ 个近邻在同一个局部样本集里。

对每一个局部样本集 C_k 进行线性回归，获得线性回归的权向量 v_k 。

在本实验中， x_k 表示特征向量， y_k 表示训练标定人数。使用欧式距离查找每一对 $(x, y) \in S$ 的 $c-1$ 个近邻， c 设为 10。

2) 聚类

选择一个合适的聚类方法，将权向量集合 v_k 聚类，获得 s 组 U_i 。

这里我们选用的是 K-均值算法^[56]聚类。K-均值算法属于动态聚类法，适合于分类数目已知的情况。该方法基于使聚类性能指标最小化。所使用的聚类准则函数是聚类集中每一个样本点到该类中心的距离平方之和，并使它最小化。其简洁和效率使得其成为所有聚类方法中最广泛使用的。

计算步骤为先随机选取 K 个对象作为初始的聚类中心。然后计算每个对象与各个种子聚类中心之间的距离，把每个对象分配给距离它最近的聚类中心。聚类中心以及分配给它们的对象就代表一个聚类。一旦全部对象都被分配了，每个聚类的聚类中心会根据聚类中现有的对象被重新计算。这个过程将不断重复直到满足某个终止条件。终止条件可以是以下任何一个：

- a) 没有（或最小数目）对象被重新分配给不同的聚类。
- b) 没有（或最小数目）聚类中心再发生变化。
- c) 误差平方和局部最小。

这里，我们将分类数目定为 3。

3) 分类

建立一个新的集合 S' 包含 m 个数据对 (x_k, i_k) ，使聚类 U_{i_k} 包含 v_k 。训练一个多类分类方法获得 \hat{X}_i 的分类矩阵 A_i 。

在本实验中，我们选择 SVM 分类方法。关于 SVM 分类的内容，会在本章

下一节详细介绍。

4) 回归

最后, 在 $(x, y) \in S$ 数据集对每个 $i=1, 2, \dots, s$ 进行线性回归分析, 获得每一层的权向量 ω_i 。

4.4 基于 SVM 的分段回归方法

SVM 是在统计学习理论上开发出来的一种非常有效的机器学习方法, 它是在结构风险最小原则下, 实现复杂度和推广性的最佳折衷, 从而达到了很好的分类性能, 并且有效地解决了学习过程中的小样本问题, 对训练样本量的需求较小。

在第三章我们简单介绍了 SVM 函数的回归问题, 在本章我们采用 SVM 分类器来进行实验。

传统的统计学常常会导致过学习现象, 因此, Vapnik^[50]等人在结构风险最小化原则的基础上创造了 SVM 的学习方法。

SVM 分类的基本思想是: 通过非线性变换 $\Phi(i)$ 将输入空间映射到一个高维特征空间中, 将输入空间中线性不可分的数据集映射后变换为线性可分, 然后在特征空间中求取最大的间隔超平面 $f(x) = w^T \Phi(x) + b$, 其中 w , b 分别是这个超平面的权值和和阈值。

设训练样本集为 $\{(x_1, c_1), (x_2, c_2), \dots, (x_n, c_n)\}$, $i = 1, 2, \dots, n$, $x \in R^d$, $c \in \{+1, -1\}$, 其中 c_i 为 1 或 -1, 分别表示数据点属于正例、反例。 x_i 是表示一个样本的 n 维向量, 为了防止变量随机影响分类结果, 每个元素都被缩放到区间 $[0, 1]$ 或 $[-1, 1]$ 。设分隔超平面的数学表示为 $w \cdot x - b = 0$, w 是垂直于分隔超平面的向量, b 是间隔距离的位移。与最优分隔超平面平行, 且距离支持向量最近的超平面, 可由方程表示为:

$$\begin{cases} w \cdot x - b = +1 \\ w \cdot x - b = -1 \end{cases} \quad (4-1)$$

若训练数据样本线性可分, 则存在两个超平面, 其间没有样本点且距离最

大，通过几何计算可知两个超平面之间距离为 $2/\|w\|$ ，最小化 $\|w\|^2$ 才可以得到最大的距离。此外，为了保证超平面间隔之间没有样本数据点，要确保所有的样本满足如下条件：

$$w \cdot x - b \geq 1 \text{ 或者 } w \cdot x - b \leq -1 \quad (4-2)$$

$$\text{即 } c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n \quad (4-3)$$

经过转换后，寻求最优分隔超平面的问题就变为在 $c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n$ 的约束条件下最小化 $\|w\|$ 的问题，也就是说转化为一个二次规划最优化的问题：

$$\min \|w\|^2 / 2, \text{ s.t. } c_i(w \cdot x_i - b) - 1 \geq 0, 1 \leq i \leq n \quad (4-4)$$

这样，我们就可以得到最优分类面或最大间隔分类面，其中，在分类面 H1 和 H2 上的训练样本点称为支持向量(Support Vector, SV)。

若样本数据点是线性不可分的，则可以在公式 4-4 的条件中增加一个松弛项 $\xi_i, i=1,2,\dots,n$ ，同时在求解中加入松弛变量，要求 $\|w\|^2 / 2 + c \sum_{i=1}^n \xi_i$ 的最小值，

这样公式 4-4 可以转化为：

$$\min \|w\|^2 / 2 + c \sum_{i=1}^n \xi_i, \text{ s.t. } c_i(w \cdot x_i - b) - 1 + \xi_i \geq 0, 1 \leq i \leq n \quad (4-5)$$

折衷考虑最小误差样本和最大分类间隔，则可以得到最优分类面，其中 $c > 0$ 是一个常数，对控制样本分类正确度惩罚的程度。

求解上面的二次规划，我们可以得到最优分类超平面函数为：

$$f(x) = \text{sgn}\{w^* \cdot x - b^*\} = \text{sgn}\left\{\sum_{i=1}^n \partial_i y_i (x_i \cdot x) - b^*\right\} \quad (4-6)$$

该公式中仅包含待分类样本与训练样本中的支持向量机的内积运算，因此，特征空间中的最优线性分类问题则转化为求解空间中的内积运算。

对于非线性问题，可以通过核函数 $K(x, y) = \Phi(x) \cdot \Phi(y)$ 将非线性变化转化为某个高维空间中的线性问题。由于在原空间和转换的高维空间中只用到了内积运算，因此对应的分类超平面的函数变为：

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n \partial_i y_i K(x_i \cdot x) - b^*\right\} \quad (4-7)$$

总之，SVM 专门针对在有限样本信息情况下，在模型的复杂性和学习能力

之间寻求最优解，从而获得最好的学习能力。

通过使用不同的核函数，SVM 可以实现很多不同的学习功能。可作为 SVM 核函数的函数很多，常用的有三种：

(1) 线性核函数：

$$k(x, y) = x \cdot y \quad (4-8)$$

(2) 多项式核函数：

$$k(x, y) = (x \cdot y + \theta)^d, d = 1, 2, \dots \quad (4-9)$$

(3) 径向基(RBF)核函数，为最常见的径向基函数：

$$k(x, y) = \exp\left\{-\frac{|x - y|^2}{\sigma^2}\right\} \quad (4-10)$$

在行人流量统计的训练过程中，首先对所有的训练样本进行标号，人数 0~20 标号为 1，人数 20~30 标号为 2，30 人以上标号为 3。

然后对所有的训练样本提取第三章所描述特征向量集合，每个训练样本的特征记为 x_i ，这样 (x_i, y_i) ， $i = 1, 2, \dots, l$ 构成了整个的训练样本集，将这个训练样本集送入合适的模型求解，求出所有的支持向量及其相应的权重，从而可以求得最终的分类决策边界。

对于基于 SVM 的分段方法核函数的选择和参数的选择，我们在测试集做了如下实验：

表 4-2 不同核函数和惩罚因子 c 值对应的 accuracy 值 (g=0.5 p=0.25)

c \ 核函数	1.0	10.0	100.0	1000.0
线性	72.9	73.1	74.65	71.35
多项式	78.55	83.1	86.55	82.75
RBF 函数	73.65	78.55	81.75	71.35

从表 4-2 列举数据可以看出，使用多项式核的 SVM 分类方法的分类效果最好，准确度最高。

4.5 基于分段回归的行人流量统计算法实验

本节总结了本章所提出的基于分段回归的行人流量检测算法，并且在数据集上进行实验，分析实验结果，并且将本文的算法与文献[57][58]所使用的特征集加高斯回归实验结果进行对比，实验数据表明基于分段回归的行人流量检测算法在数据集上具有较好的效果。

4.5.1. 实验步骤

基于分段回归的行人流量检测算法的基本步骤如下：

1) 训练部分

a) 背景建模初始化.

读入训练视频，对前 $n \times \text{inter}$ 帧进行背景建模， n 为预设值，实验定为 100， inter 为取样间隔，设为 5，即每隔 5 帧取一帧；

b) 循环读入视频帧.

当前帧与得到的背景图像相减，得到前景图像，并进行 ROI 区域选择和滤波；对每帧图像提取视觉特征信息，按标定的真实人数对特征集标定类号；

c) 将所提取的视觉特征向量与标定的类号训练分类模型。

d) 对每一类特征分别训练回归模型。

2) 测试部分

a) 导入已有背景.

b) 循环读入视频帧.

当前帧与得到的背景图像相减，得到前景图像，并进行 ROI 区域选择和滤波；对每帧图像提取视觉特征信息，根据特征使用分类模型分类,对各类使用对应训练好的回归模型进行回归预测

4.5.2. 测试集合及算法评估

本章采用的数据集同样是文献[26]作者 Antoni. B. Chan 在网上提供的人群图像序列数据集。该数据集提供了一个小时视频图像序列，采集于 UCSD 的一条人行道上的一静止摄像头。前 2000 帧提供了人工的标定数据用于实验。这

2000 帧用于实验数据，并分为训练和测试两部分。第 601 帧到 1400 帧用于训练统计模型，其余的 1200 帧用于测试和验证。在整个实验过程中，Antoni 采用 ROI 区域计数。整个数据集含有人数 49,885 人。本文是在 Intel Core(TM) i7CPU (2.93GHz) 4GB RAM 的机器上完成实验的。

数据集实例如图 3-9 所示。

我们把回归模型计算得到的数值四舍五入取整作为人数测量值。实验结果用两个指标度量：平均绝对误差（absolute mean error）和均方差（mean squared error）。平均绝对误差用 Err 表示，均方差用 MSE 表示。

4.5.3. 实验结果及分析

表 4-3 为使用分段回归和整体回归的实验对比，并且将第三章和本章实验结果与文献[57][58]的实验结果对比。从表内数据可以看出分段回归结果要优于同类别的整体回归。

表 4-3 使用分段回归的实验对比

特征	回归方法	Err	MSE
本文特征集	SVM 分段+SVM 回归	2.1942	3.0969
本文特征集	RBF 核函数 SVM 回归	2.4050	2.8908
本文特征集	分段线性回归	2.6792	3.3412
本文特征集	多元线性回归	2.6800	3.3455
文献[57]特征集	高斯过程回归	3.151	>6.149
文献[58]特征集	高斯过程回归	3.103	>7.272

但是效果不是非常明显，这和分段的选择较为相关，且实验数据集人数的疏密并不是十分明显。实验结果不理想的原因还包括前景目标的提取准确度较低，以本文所使用数据集为例，有些行人目标灰度与背景相差不大的在前景分割这部分容易被忽略，这个问题可以通过选择更好的背景建模算法来改善。总的来说，使用分段的策略还是有一定的改进效果。

4.6 本章小结

本章对所提出的基于分段回归的行人流量统计算法进行研究。首先，介绍

了分段回归基本理论和在行人流量统计方面应用的可行性分析。然后，论述了两种分段回归方法，详细说明了如何利用分段回归进行行人流量的统计学习。最后，给出了该算法的检测结果实验，结果证明，本文所使用的行人流量统计算法有效，且分段策略在同样条件下的特征回归模型上能使结果准确性有一定的提高。

总结与展望

行人流量是物流工程领域的重要研究内容。基于视觉特征的行人流量统计方法通常会面临人群密度不均，遮挡严重，无法识别个体等问题，本文通过对前景目标提取的视觉特征来进行回归分析得到人数统计结果。

本文首先简述了行人流量统计的研究背景和意义、国内外研究现状和目前已存在的研究方法，分析了目前行人流量统计检测算法中存在的难题，介绍了本文的两个主要研究模块：视觉特征集合和回归分析方法。

在详细介绍了现有的行人流量统计的研究方法的优劣后，本文提出了一个简单有效的特征集合，并采用了基于 RBF 核函数的 SVM 的回归方法与多元线性回归方法对比。考虑到监控场景中人群规模大小相差较大时，由于回归模型不匹配的问题，会导致预测人数趋于中间值，降低预测的准确度，本文提出了基于分段回归的行人流量统计算法。基于上述的两种算法，在人群图像序列数据集上进行了选择不同参数的对比实验用来验证算法，并且将实验结果与其他算法进行了比较。实验结果表明，基于视觉特征和基于 RBF 核函数的回归方法可以取得估计值比较准确的实验效果，并且使用正确的分段策略能够一定程度上提高实验结果准确性。

本文所提出算法虽然取得了一定的成功，但是仍然一些不足。例如，分段回归算法虽然解决了人群规模导致的回归模型不匹配的问题，但带来了算法中的分段选择问题。在将来的工作中，我们拟对算法的这个问题进行深入地研究。

参考文献

- [1] 吕金刚, 杨建全, 文代明等, 智能视频监控技术的应用与发展[J]. 通信电源技术, 2006(05).
- [2] R. T. Collins, A. J. Lipton and T. Kanade. A system for video surveillance and monitoring[R]. *VSAM final report. Carnegie Mellon University: Technical Report CMU*. 2000.
- [3] Smart Surveillance System(S3)[OL].
http://www-03.ibm.com/products/retail/cn/innovation/solution_smart.shtml
- [4] I. Haritaoglu, D. Harwood and L. S. Davis. W4: A real time system for detecting and tracking people [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, Nara, Japan, 1998.
- [5] T. Matsuyama. Cooperative Distributed Vision: Dynamic Integration of Visual Perception, Action, and Communication[Z]. *KI'99*, 1999:75-88.
- [6] 庞伟, “3111”工程揭竿而起, 安防商机百花齐放[J]. 中国公共安全(综合版). 2008 Vol(4).
- [7] CBSR 生物识别与安全技术中心: 智能视频监控展示[OL].
<http://www.cbsr.ia.ac.cn/china/Surveillance%20Demos%20CH.asp>
- [8] A. N. Marana, S. A. Velastin and L. F. Costa, et al. Estimation of crowd density using image processing[C]. *IEEE Conference on Image Processing for Security Applications*, 1997: 11: 1-8
- [9] A. N. Marana, S. A. Velastin and L. F. Costa, et al. Automatic Estimation of Crowd Density using Texture[J]. *Safety Science*, 1998, Vol 28(3):165-175
- [10] A. N. Marana, L. F. Costa, S. A. Velastin, et al. On the efficacy of texture analysis for crowd monitoring[C]. *SIBGRAPHI'98: Proceedings of the International Symposium on Computer Graphics, Image Processing, and Vision*, 1998:354-361
- [11] A. N. Marana, L. F. Costa and R. A. Lotufo, et al. Estimating crowd density with minkoski fractal dimension[C]. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, 6: 3521-3524

- [12] S. Cho, T. W. S. Chow and C. Leung, A neural-based crowd estimation by hybrid global learning algorithm[J]. *IEEE transactions on systems, man, and cybernetics-part B: cybernetics*, 1999, Vol29(4):535-541
- [13] H. Rahmalan, M. S. Nixon, J. N. Carter. On crowd density estimation for surveillance[C]. *The Institution of Engineering and Technology Conference on Crime and Security*, 2006:540-545
- [14] X. Wu, G. Liang and K. K. Lee, et al, Crowd density estimation using texture analysis and learning[C]. *IEEE International Conference on Robotics and Biomimetics*, 2006: 214-219
- [15] W. Ma, L. Huang and C. Liu, Crowd estimation using multi-scale local texture analysis and confidence-based soft classification[C]. *Second International Symposium on Intelligent Information Technology Application*, 2008: 142-146
- [16] C. Sacchi, G. Gera, L. Marcenaro, and C. S. Regazzoni, Advanced image-processing tools for counting people in tourist site-monitoring applications[J]. *Signal Processing*, 2001, Vol81(5):1017-1040
- [17] T. Zhao, R. Nevatia. Bayesian human segmentation in crowded situations[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003: 459-466
- [18] B. Wu, R. Nevatia. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors[C]. *IEEE International Conference on Computer Vision*, 2005: 90-97
- [19] V. Rabaud, S. Belongie. Counting crowded moving objects [C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006:705-711
- [20] E. Zhang, F. Chen. A fast and robust people counting method in video surveillance[C]. *Computational Intelligence and Security*, 2007: 339-343
- [21] L. Chen, J. Tao and Y. Tan, et al. People counting using iterative mean-shift fitting with symmetry measure[C]. *International Conference on Information, Communication & Signal Processing*, 2007:1-4
- [22] P. Kilambi, E. Ribnick and A. J. Joshi, et al. Estimating pedestrian counts in groups[J]. *Computer Vision and Image Understanding*, 2008, Vol110(1):43-59

- [23]C. S. Regazzoni, A. Tesei, Distributed data fusion for real-time crowding estimation[J]. *Signal Processing*, 1996, Vol 53(1) :47-63
- [24]R. Ma, L. Li, W. Huang and Q. Tian. On pixel count based crowd density estimation for visual surveillance[C]. *IEEE Conference on Cybernetics and Intelligent Systems*, 2004,1:170-173
- [25]D. Kong, D. Gray and H. Tao. Counting pedestrians in crowds using viewpoint invariant training[C]. *British Machine Vision Conference*, 2005
- [26]A. B. Chan, Z. J. Liang and N. Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008:1-7
- [27]R. Rajagopalan, M.T. Orchard and R.D. Brandt. Motion field modeling for video sequences[J]. *IEEE Transactions on Image Proceedings*. 1997, Vol6(11):1503-1516.
- [28]Y. Altunbasak, R. M. Mersereau and A. J. Patti. A fast parametric motion estimation algorithm with illumination and lens distortion correction[J]. *IEEE Transactions on Image Proceeding*. 2003, Vol12(4):395-408.
- [29]E. Trucco, T. Tommasini and V. Roberto. Near-recursive optical flow from weighted image difference[J]. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics* , 2005, Vol35(1):124-129.
- [30]A. Criminisi, G. Cross and A. Blake, et al. Bilayer segmentation of live video[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2006, 1:53-60.
- [31]马颂德, 张正友, 计算机视觉—计算理论与算法基础[M], 北京:科学出版社.1998.
- [32]B.K.P. Horn, B.G. Schunck. Determining Optical Flow[J]. *Artificial Intelligence*, 1981, Vol17(1-3):185-203.
- [33]A. J. Lipton, H. Fujiyoshi, R. S. Patil. Moving target classification and tracking from real-time video[C]. *IEEE Workshop on Applications of Computer Vision*, 1998:8-14.
- [34]C. Anderson, P. Bert and G. V. Wal. Change detection and tracking using pyramids transformation techniques[C]. *SPIE Conference on Intelligent Robots*

- and Computer Vision*, 1985, 579:72-78.
- [35]N. Paragios, C. Tziritas. Detection and Location of Moving Object Using deterministic relaxation algorithm[C]. *International Conference on Pattern Recognition*.1996, 1:201-205.
- [36]H. Song, F. Shi. A real-time algorithm for moving objects detection in video images[C]. *Fifth World Congress on Intelligent Control and Automation*.2004, 5:4108-4111.
- [37]N. Thakoor, J. Gao. Automatic video object shape extraction and its classification with camera in motion[C]. *IEEE International Conference on Image Proceeding*.2005, 3:437-440.
- [38]Z. Wei, X. Ji and P. Wang. Realtime moving object detection for video monitoring systems[J]. *Journal of System Engineering and Electronics*, 2006, Vol.17(4):731-736
- [39]李刚,邱尚斌,林凌等. 基于背景差法和帧间差法的运动目标检测方法[J].*仪器仪表学报*, 2006, Vol27(8):961-96
- [40]A. Minnet, A. Mittal, and N. Paragios, et al. Background modeling and subtraction of dynamic scenes[C]. *IEEE International Conference on Computer Vision*. 2003, 2:1305-1312.
- [41]E. Hayman, J.-O. Eklundh. Statistical background subtraction for a mobile observer[C]. *IEEE International Conference on Computer Vision*, 2003, 1:67-74.
- [42]L. Wang, T. Tan and H. Ning, et al. Silhouette analysis-based gait recognition for human identification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, Vol25(12):1505-1418.
- [43]刘直芳, 数字图像处理与分析[M], 北京: 清华大学出版社, 2011.
- [44]H. Tamura, S. Moil, T. Yamawaki. Texture features corresponding to visual perception[J]. *IEEE Transactions on Systems Man and Cybernetics*,1978, Vol8(6):460-473.
- [45]王明慈,沈恒范. 概率论与数理统计[M].高等教育出版社, 2001.
- [46] “Privacy preserving crowd monitoring: Counting people without people models or tracking.” [OL]. <http://www.svcl.ucsd.edu/projects/peoplecnt>.

- [47]李实英, 杨高波. 特征提取与图形处理(第二版)[M]. 电子工业出版社, 2010.
- [48]J.Canny. A computational approach to edge detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, Vol28 (6): 679- 6981
- [49]R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification[J]. *IEEE Transaction on Systems, Man, and Cybertinetics*, 1973,Vol 3(6): 610-621.
- [50]V. Vapnik, The Nature of Statistical Learning Theory[M]. New York, USA: Springer-Verlag, 1995.
- [51]张学工, 关于统计学习理论与支持向量机[J], 自动化学报, 2000, Vol26: 32-42.
- [52]C. C. Chang and C. J. Lin. LIBSVM: a library for support vector machines[OL].Soft available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [53]J.Zhang, B.Tan and F.Sha, et al. Predicting pedestrian counts in crowded scenes with rich and high-dimensional features[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2011, Vol12(4):1037-1046.
- [54]B. Tan, J. Zhang and L. Wang. Semi-Supervised Elastic Net for Pedestrian Counting[J]. *Pattern Recognition*, 2011.
- [55]F.T. Giancarlo, M. Marco, A New Learning Method for Piecewise Linear Regression[C] . *International conference In Artificial neural networks*. 2002, Vol 2415: 444 - 449.
- [56]蔡元龙, 模式识别[M], 西北电讯工程学院出版社, 1986.
- [57]A. C. Davies, J. H. Yin and S. A. Velastin. Crowd monitoring using image processing[J]. *Electronics & Communication Engineering Journal*. 1995, Vol7(1):37-47.
- [58]D.Kong, D. Gray and H. Tao. Counting pedestrians in crowds using viewpoint invariant training[C]. *British Machine Vision Conference*, 2005

致 谢

在中国科学院研究生院攻读硕士学位的三年学习生活使我受益匪浅。在毕业论文完成之际，由衷地感谢这三年来曾经给予我无数帮助的老师、同学、朋友和家人。

首先，我要感谢我的导师焦建彬教授给予了我珍贵的求学机会，感谢他在我攻读硕士学位期间对我的每一点进步的指导和鼓励。三年来，恩师对我的学习和工作给予了悉心的指导，给我的研究提供了良好的环境和机会，让我受益匪浅。他严谨的治学之风和对事业的孜孜追求深深地影响着我，他诲人不倦的精神和对我的谆谆教导让我感动，他的言传身教将使我受益终生。

其次，非常感谢叶齐祥老师在我的科研学习中的理论和实验方面给予了耐心的指导，并在论文的撰写和审稿中给予了很多帮助。他渊博的专业知识，自强不息的学习精神、扎实的动手能力和对我的谆谆教导，让我受益匪浅。

感谢参加开题及中期评阅的各位老师和专家们，他们丰富的经验和无私的工作对论文方向和研究进度的把握和指点给整个研究工作带来了巨大的帮助。

我还要衷心感谢我的同学们对我的热心帮助以及一起学习生活中结下的深厚友谊，特别是好友李策、武博、高文、汪舰航、沙月。三年的时间里，我们互相关心、互相帮助，共同面对成长中的喜悦和困惑，共同走过的日子是我最难忘的回忆。

最后要感谢我的父母，感谢他们多年来一直给我最无私的关怀，无微不至的关心和爱护。感谢我的父母为我所做的一切，愿他们能够为我而骄傲。

最后，感谢参加论文评审和答辩的各位老师。

纪颖夏

2012年4月