

分类号: _____

密级: _____

UDC: _____

编号: _____

工学硕士学位论文

基于多尺度方向特征的行人检测算法

硕士研究生 : 杨志辉

指导教师 : 黄凤岗 教授

学位级别 : 工学硕士

学科、专业 : 计算机应用技术

所在单位 : 计算机科学与技术学院

论文提交日期: 2009 年 1 月

论文答辩日期: 2009 年 2 月

学位授予单位: 哈尔滨工程大学

Classified Index:

U.D.C:

A Dissertation for the Degree of M. Eng

Pedestrian Detection Algorithm Based on Multi-scale orientation feature

Candidate: Yang Zhihui

Supervisor: Prof. Huang Fenggang

Academic Degree Applied for: Master of Engineering

Specialty: Computer Applied Technology

Date of Submission: January, 2009

Date of Oral Examination: February, 2009

University: Harbin Engineering University

哈尔滨工程大学 学位论文原创性声明

本人郑重声明：本论文的所有工作，是在导师的指导下，由作者本人独立完成的。有关观点、方法、数据和文献的引用已在文中指出，并与参考文献相对应。除文中已注明引用的内容外，本论文不包含任何其他个人或集体已经公开发表的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者（签字）：

日期： 年 月 日

哈尔滨工程大学 学位论文授权使用声明

本人完全了解学校保护知识产权的有关规定，即研究生在校攻读学位期间论文工作的知识产权属于哈尔滨工程大学。哈尔滨工程大学有权保留并向国家有关部门或机构送交论文的复印件。本人允许哈尔滨工程大学将论文的部分或全部内容编入有关数据库进行检索，可采用影印、缩印或扫描等复制手段保存和汇编本学位论文，可以公布论文的全部内容。同时本人保证毕业后结合学位论文研究课题再撰写的论文一律注明作者第一署名单位为哈尔滨工程大学。涉密学位论文待解密后适用本声明。

本论文（在授予学位后即可 在授予学位 12 个月后 解密后）由哈尔滨工程大学送交有关部门进行保存、汇编等。

作者（签字）：

导师（签字）：

日期： 年 月 日

年 月 日

摘 要

随着人类社会的发展，社会的不安全因素也随之增多。国际上的每一次恐怖主义袭击事件都为各国的安防部门敲响了警钟。因此，许多国家越来越重视采用视频监控技术对重要部门、敏感地点、公共场所等进行监控。虽然人脸目标检测、车牌目标检测等典型目标检测方法已经日趋成熟，但是复杂环境下高可靠性的移动人体目标检测却仍然面临着很大困难，使之成为了本领域内最为重要和迫切的研究问题。同时，开展此方面的研究对目标的模式表达，以及本领域内的核心问题的研究也具有重要的理论意义。

目前对于特征的描述主要分为对色彩的处理和对轮廓的提取两方面。其中最具有代表性的就是 Viola 提出的 Haar-Like 和 Dalal 采用的 HOG 作为物体的特征描述子，而且在人体检测上都达到了很好的效果。本文作者受到以上两种特征描述子的启发，提出了一种新的特征--多尺度方向特征，这种新特征不仅囊括了上面两种特征各自的优点，还弥补了它们的不足。这种特征是针对特征区域的形状进行统计的，由“全部特征组合起来所形成的特征集”反映的是“图像在不同尺度上的方向特征”。并且分别在 SVM 和 AdaBoost 机制下进行训练，利用训练出来的模型在视频和图片上进行行人检测，通过在公共测试集和本文自己的测试集上进行测试，并将结果与国际上其它领先算法相对比分析，实验证明了：在相同的检测标准下，使用本文所提出的理论框架，无论在运算速度上，还是在检测结果的精度上，本文算法都表现出了明显的优势。

关键词：行人检测；多尺度方向特征；级联 AdaBoost；支持向量机

Abstract

With the development of human society , there are more and more social insecurity factors at the same time. Each terrorist attack on the international gives a wake-up call for every national security departments. As a result, many countries pay more and more attentions to the use of video surveillance technology to the important sectors, sensitive locations, such as to monitor public places. Although the object detection technology such as face detection and license plate detection have become more mature, however the high reliability detection of moving human which is on the complex environment is still faced with great difficulties. Meanwhile, to carry out the research also has important theoretical significance to the objective model expression as well as the core of this area.

At the present, the description of the main features contains tow parts which are the processing of color and the extraction of the outline. The most representative descriptors are Viola's Haar-Like and Dalal's HOG., and achieve good results in human detection. In this thesis, we put forward a new descriptor-Multi-scale orientation feature. The new features not only include the two characteristics of their respective advantages, but also make up for their lack. This feature is described according to the shape of the characteristics region, and the set of all the features describe the different orientation of each scale in a picture. Training these features in SVM and AdaBoost separately and use the training result to detect the human in video and picture. According to the test on the public testing sets and our own testing sets in this thesis, with the analysis of the result comparing to other international algorithms, it can proves that: on the same condition of detection principles, our algorithms have shown a clear advantage both on the computing speed and the accuracy of detection result.

Key words: Pedestrian Detection; Multi-scale orientation feature set; Cascade AdaBoost; Support Vector Machine(SVM)

目 录

| | |
|---------------------------------------|----|
| 第 1 章 绪论 | 1 |
| 1.1 人体检测的研究背景 | 1 |
| 1.2 人体检测的研究现状 | 2 |
| 1.2.1 概述 | 2 |
| 1.2.2 国内外研究现状 | 2 |
| 1.3 论文主要工作 | 4 |
| 1.4 论文组织结构 | 5 |
| 第 2 章 基于机器学习的行人检测算法 | 6 |
| 2.1 基于 Haar-Like 特征的行人检测算法 | 6 |
| 2.1.1 Haar-Like 特征 | 6 |
| 2.1.2 积分图像 | 8 |
| 2.1.3 AdaBoost 分类器 | 9 |
| 2.1.4 分类器级联(Cascade) | 11 |
| 2.2 基于 HOG 特征的行人检测算法 | 12 |
| 2.2.1 什么是 HOG | 12 |
| 2.2.2 HOG 的实现流程 | 14 |
| 2.2.3 Integral HOG 和 Extend HOG | 16 |
| 2.2.4 SVM 分类器 | 16 |
| 2.3 本章小结 | 18 |
| 第 3 章 基于新特征的行人检测 | 19 |
| 3.1 本文采用的人体目标检测算法 | 19 |
| 3.2 MSO 特征集的构造 | 21 |
| 3.2.1 特征单元的构造 | 21 |
| 3.2.2 AdaBoost 特征集的构造 | 24 |
| 3.2.3 SVM 特征集的构造 | 24 |
| 3.2.4 MSO 特征集的特点 | 25 |
| 3.3 分类器模型 | 25 |
| 3.3.1 SVM 分类器 | 26 |

| | |
|--------------------------------|----|
| 3.3.2 AdaBoost 分类器 | 27 |
| 3.4 物体检测 | 31 |
| 3.5 本章小结 | 31 |
| 第 4 章 实验结果与对比分析 | 33 |
| 4.1 数据集的形成 | 33 |
| 4.1.1 数据集 (Data Set) 的描述 | 33 |
| 4.1.2 数据标定的标准 | 33 |
| 4.2 实验结果 | 35 |
| 4.3 和其它的领先技术相比较 | 40 |
| 4.4 本章小结 | 43 |
| 结 论 | 44 |
| 参考文献 | 46 |
| 攻读硕士学位期间发表的论文和取得的科研成果 | 50 |
| 致 谢 | 51 |

第 1 章 绪论

1.1 人体检测的研究背景

目前，人体检测在计算机视觉领域已经得到了广泛的关注，研究人员们意在通过建立一个系统模型来实现对人的检测以及对其外表、行为、活动的理解。人体检测在现实上有着很大的远景，如：视频监控，基于内容的图像、视频处理，以及保护行人安全的车载行人检测系统等等^[1-5]。而且对于在其基础上进行扩展的一些技术也将产生巨大的影响，如：步态识别，对人体动作的语义识别等。在人体检测上本文按照不同的方法可以分成不同的类别，是否需要预处理（背景剪除，直接检测），是否需要外部特征的描述（形状，颜色，动作），是否需要人体模型，机器学习……。但总体上本文可以简单的划分为两类：基于动态视频的和基于单帧静止图片的。基于动态视频的人体检测受到了一定的约束，因此在应用方面会受到一些影响；而基于单帧图片的人体检测的定位上相对繁琐一些，然而一旦单帧图片的人体检测的研究得到突破，那么它将会有非常远大的前景。对于人体形态而言，人体是一个非刚性的、形变大的、多角度的物体，因此研究多视、多态的人体检测将会面对很多困难与挑战，目前越来越多的 CV 学者被吸引到这个课题上来。而我毕业论文的方向则是在单帧图像中对多视、多态的行人进行检测。

由此，本课题以在图像中实现多视角的行人检测为主要研究方向。其在以下方面具有重要的应用价值：

在国防领域，可以用来探测部队的周边环境，防止枪械库、弹药库的重要地点被偷袭或者破坏等；在民用领域，该技术可用于重点建筑（核电站、水电站、水库、银行、重要桥梁等）及重要地点（国家重点保护建筑、博物馆、体育场馆、地铁站、监狱等）的安全保卫工作；

1. 在智能交通领域，可以在十字路口、新建高速路段等进行自动行人检测和安全预警。
2. 随着城市规模的扩大，住宅小区逐渐成为安全防范的重点，结合红外、近红外等视频设备，该技术可增强小区安全监控、减少犯罪，从而增

强居民的安全感。

在近年来的研究中，人脸目标检测。车牌目标检测等典型目标检测方法已经日趋成熟。而复杂环境下高可靠性的移动人体目标检测却仍然面临着很大困难，成为了本领域内最为重要和迫切的研究问题。因此，开展此方面的研究，将对目标的模式表达等本领域内的核心问题的研究具有重要的理论意义。

整个检测系统是在 VC++ 6.0 编译环境下实现的，引入了 Intel 开源的图像处理库 OpenCV，大大节省了程序代码编写上的开销，在 CPU 主频为 3.0G，内存为 1G 的机器上，对 640*480 的图像进行检测，可以达到近似实时的处理速度。

1.2 人体检测的研究现状

1.2.1 概述

在人体（或者其他物体^[1-3]）的检测上，在理论上，国内外已经形成了几种比较经典的框架。但总体上本文可以简单的划分为两类：基于运动区域的和图像/视频特征的。基于运动区域的人体检测受到了一定的约束，因此在应用方面会受到一些影响；而基于图像特征的人体检测的定位相对繁琐一些，然而一旦在单帧图片的人体检测的研究得到突破，那么它将会有非常远大的前景。对于人体形态而言，人体是一个非刚性的、形变大的、多角度的物体，因此多视、多态的人体检测具有很大难度与挑战，目前越来越多的 CV 学者被吸引到这个课题上来。

在人体检测上待于解决的问题是：1.提高检测率；2.降低误检率；3.时间上的高效性。

1.2.2 国内外研究现状

本文的任务是在视频或是在一张图片中检测出行人。本文根据是否使用背景剔除或是分割的方法，可以将以往的相关文献大致的分为两类：基于运动区域（背景剔除的方法获取）的检测；基于图像/视频特征直接检测。

1. 背景剔除技术通常是先在视频中找到前景物体，然后再将其进行分类，通常是基于形状、颜色和其行为以及其它的相关特征。在下面的表 1.1 中列出了关于背景剔除的一些方法。

表 1.1 基于背景剔除的方法

| 论文 | 背景剔除 | 人体特征 |
|-------------------------------|----------|----------|
| Wren 等人 ^[6] | 颜色、相关图像 | 颜色, 轮廓 |
| Beleznai 等人 ^[7] | 颜色、相关图像 | 区域模型 |
| Haga 等人 ^[8] | 颜色、相关图像 | F1-F2-F3 |
| Eng 等人 ^[9] | 颜色、相关图像 | 颜色 |
| Elzein 等人 ^[10] | 运动、帧差 | 小波 |
| Toth 和 Aach ^[11] | 运动、帧差 | 傅立叶形状 |
| Lee 等人 ^[12] | 运动、帧差 | 形状 |
| Zhou 和 Hoang ^[13] | 运动、帧差 | 形状 |
| Yoon 和 Kim ^[14] | 运动+颜色 | 几何, 像素值 |
| Jiang 和 Lijun ^[15] | 深度 | 运动 |
| Li 等人 ^[16] | 深度 | 形状 |
| Han 和 Bhanu ^[17] | 红外 | 红外+颜色 |
| Jiang 等人 ^[18] | 红外 | 红外+颜色 |
| Naneet Dalal ^[20] | 运动+颜色+光流 | HOG |

此外，还有一些研究团体，如戴姆勒克莱斯勒的研究机构，D. M. GAVRILA 等人，将行人检测技术应用到了车载智能设备上^[21]。通过在汽车上的应用来检测车道上的行人，起到对司机预警的作用，关键时刻能够自动采取应急措施避免事故的发生，其中采用了对模型的级联，以及一些视觉标准来不断地缩小搜索范围。同时采用 4 种模型的组合（少量的基于实体的区域生成器，基于形状的检测器，基于纹理的分类器和多个基于实体的校验器），此外还采用了一个新颖的专家混合结构，基于纹理进行各个组件的分类，而且每个的权重受到形状匹配的结果的影响。

2. 对于直接检测，是在图像和视频段中提取特征，并依据这些特征将事物划分为人和非人。这些特征通常包括形状（包括轮廓或者其他的描

述形式), 颜色 (皮肤颜色检测), 动作, 或者是这些的混合。下面表 1.2 列举了近年来关于直接检测的方法。

表 1.2 基于直接检测的方法

| 论文 | 人体模型 | 分类器 |
|---|-----------------|-------------------------|
| Cutler 和 Davis ^[22] | 周期运动 | 运动相似性 |
| Utsumi 和 Tetsutani ^[23] | 几何, 像素值 | 距离 |
| Gavrila 和 Giebel ^[21] | 形状模板 | 倒角距离 |
| Viola 等人 ^[5,18] | 形状+运动 | Adaboost 分类器级联 |
| Sidenbladh ^[24] | 光流 | SVM |
| Dalal 和 Triggs ^[25] | 梯度直方图 | SVM |
| K. Mikolajczyk 等人 ^[26] | 共生+方向特征+概率组装 | 以概率组装的分类器级联 |
| Oncel Tuzel 等人 ^[27] | 黎曼流型 | LogitBoost |
| Cong Hou, Haizhou Ai, and Shihong Lao ^[13] | 主方向+拓展 HOG+非矩形块 | Vector Boosting+树形结构检测器 |
| Qixiang Ye 等人 ^[28] | 运动轮廓匹配 | 运动轮廓+动态规划 |
| Thomas Serre 等人 ^[29] | Gabor | SVM+Boosting |

近年来, 越来越多的研究人群倾向于对物体的直接检测, 通过机器学习的方法训练出相应的人体模型, 在检测时用此来与被检测的物体相匹配。目前主要的方法有 Haar-Like、HOG、黎曼流型、Gabor、Sharplet 等。

1.3 论文主要工作

本文主要工作归纳如下:

本文作者受到 Viola 提出的 Haar-Like 和 Dalal 采用的 HOG 的启发, 提出了一种新的特征, 这不仅囊括了其各自的优点, 还弥补了它们的不足。这种特征是针对特征区域的形状进行统计的, 由全部特征组合起来所形成的特征集反映的是图像在不同尺度上的方向特征。

在获取大量的包含人体图片之后, 对图像中的人体进行标注, 形成一系列的正例人体样本和反例非人体样本。然后使用本文提出的描述子对样本集进行编码, 通过本文自己的一套 Cascade AdaBoost 训练机制对样本集进行训

练, 形成人体有用的特征模板。然后通过生成的模板对单帧的图像进行检测, 最终实现一个检测率高、性能鲁棒的行人检测系统。

1.4 论文组织结构

第 1 章 首先介绍了课题的研究背景, 然后阐述了国内外研究的现状, 最后说明了论文完成的主要工作。

第 2 章 对已有算法的相关知识点做以大致的介绍, 主要是 Viola 提出的基于 Haar-Like 特征的人体检测, 还有 Dalal 采用的 HOG 特征的人体检测。

第 3 章 详细介绍本文的所提出的特征描述子, 并对整个算法流程进行详细的分析。

第 4 章 首先介绍数据集的采集和构造过程, 然后使用不同的方法在公共测试集和我们的测试集上分别进行测试, 根据实验结果将其做一比较分析。

第 2 章 基于机器学习的行人检测算法

2.1 基于 Haar-Like 特征的行人检测算法

2.1.1 Haar-Like 特征

在 1997 年, Oren 提出了用小波提取特征, 再使用支持向量机来做分类的思想。虽然在现在看来, 该算法的效果还远远算不上理想, 但是在当时, 这一算法不仅取得了领先的效果, 更为重要的是: 首次将使用机器学习的思想引入了行人检测领域, 通过对训练样本的学习自动地建立分类模型, 从而摆脱了原先手工建立模型的方法。以此为标志, 人体检测正式进入了机器学习的时代。而且这一算法确立了一个基本的框架, 理解这一方法将帮助我们更好地理解后来出现的更为复杂的算法。下面我们将对这一方法做一概述。

Haar-Like 特征是由 Haar 小波演变而来的^[1-4], Haar 小波的形式如下(图 2.1):

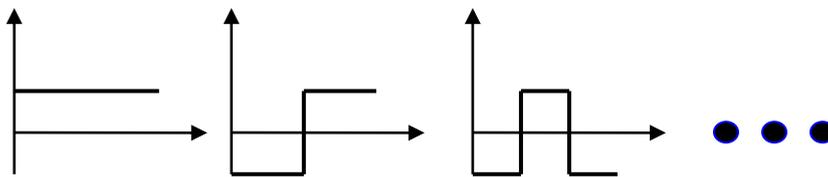


图 2.1 Haar 小波

而在灰度图像上, 可以表现成图 2.2 的形式。本文将这种特征称之为“Haar-Like 特征”。

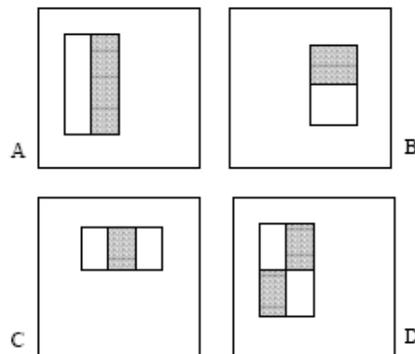


图 2.2 Haar-Like 特征

特征的值等于灰色矩形框中所有像素的颜色值和减去白色矩形框中的所有像素的颜色值。(A)和(B)为两个矩形的特征,(C)为三个矩形的特征,(D)为四个矩形的特征。

在使用 Haar-Like 的目标检测中的图像分类器是基于一些简单的特征值的。关于本文之所以使用特征的方法,而不是直接对像素点进行处理有很多的原因。但最根本的原因是:特征能够表现出一些特殊的信息,这是对于数量有限的训练样本所不能达到的。之所以使用特征方法的另外一个重要的原因是:对于同样的系统,采用基于特征的方法要比采用基于像素的方法要更快些。

在 Papageorgiou 等人的试验中,他们采用了基于 Haar 的函数作为简单的特征(见图 2.2)。对于由 2 个矩形所组成的特征的值,可以通过对两个矩形区域的所有像素分别求和然后再相减来实现。这些相邻的矩形区域都有相同的大小、形状、以及相同的排列方式(见图 2.2)。对于由 3 个矩形组成的特征的值是通过使用中间矩形的 2 倍减去两边矩形的和。最后一种是由 4 个矩形组成的本文可以按照对角线的两个矩形相加然后再求差来得到这种特征的值。

后来,Viola 等人又对上面基本的 Haar 特征进行了拓展,便有了现在的“拓展的 Haar-Like 特征”,见图 2.3。

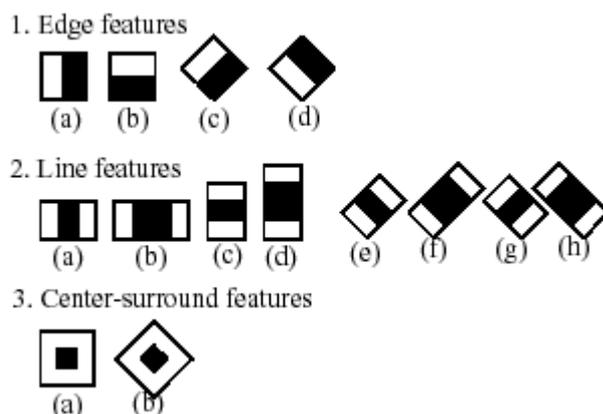


图 2.3 拓展的 Haar-Like 特征

在给定的一个标准的 24*24 分辨率的图片中,那些矩形特征所组成的完备集中的元素的个数非常巨大,超过了 180 000 个。对于数量这么大的特征值,如果每一个矩形区域都要相对独立的计算,那么计算量也是十分巨大的。

为了解决这个问题，Viola 等人提出了下面的“积分图像的”概念。

2.1.2 积分图像

“积分图像”是 Viola 等人为了快速特征估计而提出的一种新的图片表示方法，本文称之为“积分图像”^[5]。有一部分受到 Papageorgiou 等人的启发^[4]。他们的检测系统并不是直接工作在图像的亮度之上的，和其他学者一样，他们也使用了特征集合，这些特征的值就是本文通过标准的 Haar 变换函数获得的。为了在多尺度上快速的计算这些特征，由此便引进了积分图像来对图像表示。而这个积分图像可以通过对图像中的每个像素的一系列的操作来生成。一旦计算出来，任何一个 Harr-Like 特征都可以在任何尺度、任何位置上在一个常量的时间复杂度上计算出来。

现在矩形特征的值可以通过积分图像表达方法得以快速的计算出来。这个积分图像在 x,y 处的值为原图像左上方所有像素值的和，

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (2-1)$$

式中： $ii(x, y)$ ----- 此点的积分图像值

$i(x', y')$ ----- 原始图像中此点相素的灰度值

可以使用下面的这对递推公式来完成上面的计算：

$$s(x, y) = s(x, y-1) + i(x, y) \quad (2-2)$$

$$ii(x, y) = ii(x-1, y) + s(x, y) \quad (2-3)$$

式中： $s(x, y)$ ----- 行累加和

其中： $s(x, y)$ 是行累加和， $s(x, -1) = 0$ ， $ii(x, y) = 0$ 。

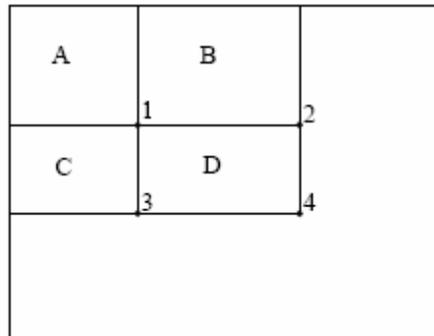


图 2.4 矩形框 D 中的像素和的计算

矩形框 D 中的像素和可以通过 4 个相关矩形计算出来，在积分图像中位置 1 的值就是矩形 A 的所有像素的和，位置 2 的值对应的是 A+B，位置 3 对应的是 A+C，位置 4 对应的是 A+B+C+D。因此矩形 D 的特征的值可以通过“区域 4+区域 1-区域 2-区域 3”的所有像素的像素值累加和来计算。

本文在使用这个积分图像中的任何一块矩形的值时，可以通过 4 块儿相关的矩形进行求得（见图 2.4）。显然，两个矩形的累加和之差可以通过 8 个相关区域来求得。因此，2 矩形的特征的特征值，依据上面的定义，可以通过 6 个矩形来求得，以此类推，3 矩形的特征需要 8 个，4 矩形的特征需要 9 个。

2.1.3 AdaBoost 分类器

当本文给出一组特征集和一组由正样本和负样本组成的图片集，那么本文用任何一个机器学习方法来充当学习的分类函数。在本文的系统中，本文使用 AdaBoost 变种的方法作为小特征集的选取和分类器的训练。它的初级模型中，这个 AdaBoost 的学习算法被用作达到提升一个小学习算法（或称为弱分类器）的分类效果[5、30-35]。

让本文再回想一下前面所提到的，一幅图像中的每一个子窗口所包含的矩形特征超过了 180 000 个，这个数字远比该子窗口的像素数（24x24）大得多。即使本文假设每一个特征值都可以被快速的计算出来，但要处理整个完备集的特征还是要花费很多的时间的。为了实现通过一小部分的特征的合成就可以形成一个高效的分类器，本文就应该求助于 AdaBoost 算法了。本文通过弱学习算法来筛选出“最好的”一个矩形特征作为切分正例样本和反例样本的条件。对于每一个特征，弱学习器都要决定出一个最佳的阈值分类函数，来使训练样本的错误分类数达到最小。对于一个弱分类器函数 $h_j(x)$ 有一个特征 f_j ，一个阈值 θ_j ，还有一个用来指示不等式符号的方向的符号函数 p_j ：

$$h_j(x) = \begin{cases} 1 & \text{if } (p_j f_j \leq \theta_j) \\ 0 & \text{otherwise} \end{cases} \quad (2-4)$$

这里的 x 是一幅图像中的一个 24x24 像素的子窗口。关于 Boosting 过程的摘要概述见图 2.5。

每一次的 boosting 循环都将会从 180,000 个备选特征中选出一个特征。

实际上，没有一个单一的特征能够完成低错误率的分类任务。在 boosting 过程的前期提取出来的特征的错误率是在 0.1 到 0.3 之间，而在之后的循环中所提取出来的特征，因为任务更加困难，所以错误率处于 0.4 到 0.5 之间。

- 给一组样本图片 $(x_1, y_1), \dots, (x_n, y_n)$,
其中, $y_i = 0, 1$ 分别表示为反例样本和正例样本.
- 分别将 $y_i=0$ 和 1 的样本的权重初始化为 $\omega_{1,i} = \frac{1}{2m}, \frac{1}{2l}$,
这里的 m 和 l 分别为反例样本和正例样本的数量.
- $Fort = 1, \dots, T$:
 - 1.归一化权重,

$$\omega_{t,i} \leftarrow \frac{\omega_{t,i}}{\sum_{j=1}^n \omega_{t,j}}$$
 由此, ω_t 服从概率分布.
 - 2.对于每一个特征 j ,
 训练一个分类器 h_j 都要严格的遵守使用一个特征的原则.
 其误差可以通过按照下面的公式估算出来

$$\varepsilon_j = \sum_i \omega_i |h_j(x_i) - y_i|.$$
 - 3.选择分类器中误差 ε_t 最小的.
 - 4.更新权重:

$$\omega_{t+1,i} = \omega_{t,i} \beta_t^{1-e_i}$$
 其中, 如果分类正确则 $e_i = 0$;
 否则, $e_i = 1$, 并且 $\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$.
- 最终的强分类器为:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & otherwise \end{cases}$$
 其中 $\alpha_t = \log \frac{1}{\beta_t}$.

图 2.5 分类学习的 AdaBoost 算法

2.1.4 分类器级联(Cascade)

通过分类器级联的形式，本文不仅可以提高检测的性能，还可以降低检测所需要花费的时间。本文发现这样一种重要的现象：越小的提升（boosted）分类器可以达到更好的效果，它可以在检测几乎所有的正确实例的同时还可以排除掉大量的无关的子窗口（即：本文可以通过调节一个 boosted 分类器的阈值来降低误错率，并使之趋近于 0）。简单的分类器通常被用来排除大量的子窗口，在这之后再利用更加复杂的分类器来完成降低误正率的错误。

从整体的结构来看，整个检测过程实质上就是一棵退化了的决策树，本文称之为“级联”^[2,5]（见图 2.6）。

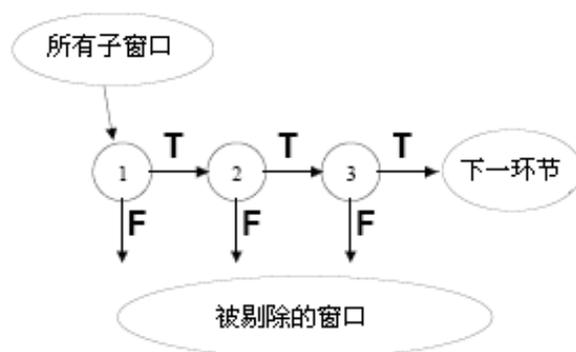


图 2.6 一个级联检测器的示意图

这一系列的分类器将用于所有子窗口的检测。最初的那个分类器可以通过少量的处理来拒绝大量的反例。随后的那些将会继续剔除反例，但相对需要额外的计算。经过几级的处理，子窗口的数量将会明显的减少。下一环节的处理可以是增加的几个级联分类器或者是其他的检测系统。

从第一个分类器出来的被肯定了的的结果将会触发第二个分类器，并通过此分类器来实现更高的检测率。而从第二个分类器出来的被肯定了的的结果又会触发第三个分类器，以此类推。相反，从任何一个结点输出的被否定了的结果都会导致对这个子窗口的检测立即停止。

在级联器中的每一级都是通过利用 AdaBoost 形成的分类器而构造出来的，而后本文可以通过调节相应的阈值来使误错率达到最小。需要指出的是：

默认的 AdaBoost 的阈值是用来降低在训练数据时产生的错误率的。一般而言, 较低的阈值可以导致一个较高的检测率和一个较高的误警率。同样地, 本文渴望级联器尽可能的在早些时候可以将大部分无用的子窗口排除掉。而当一个有意义(包含检测目标)的子窗口通过了级联其中的每一个分类器时, 这种情况相比之下是非常小的概率(因为检测目标在图像上所占的比例很小), 因此级联的 AdaBoost 算法具有较高的学习效率。

2.2 基于 HOG 特征的行人检测算法

2.2.1 什么是 HOG

HOG 是 Histograms of Oriented Gradients 的缩写^[13,19, 20, 30, 34], 由 Dalal 和 Triggs 提出。在 2005 年由 Dalal 提出了基于 hog 特征的算法。其算法框架与上文 Oren 的并无本质不同。唯一的差别只有两点: 其一是由于计算机性能的提高, Dalal 使用了上千维的特征, 而非像 Oren 那样仅仅选出 29 维; 其二是 Dalal 使用了表达能力更强的 hog 特征。但令人惊叹的是, 这个看似简单的算法竟然取得了一流的效果: 在 MIT 行人库上表现近乎完美, 以至于 Dalal 不得不自己创建了一个难度更高的 INTRA 行人库。目前该行人库已成为了行人检测领域的标准测试数据。在 INTRA 行人库上, Dalal 在万分之一的误报率下取得了近 90% 的检测率。即使在今天, 该算法依然是行人检测领域最优秀的算法之一。下面我们就来介绍一下“神奇”的 hog 特征(关于其具体实现见他们发表的文章《Histograms of Oriented Gradients for Human Detection》^[19])。通过他们的试验, 证明了这种算法不仅可以实现对人的检测, 还可以实现对摩托车等物体进行检测。

与 SIFT 特征类似, HOG 特征通过提取局部区域的边缘或梯度的分布, 可以很好地表征局部区域内目标的边缘或梯度结构, 进而表征目标的形状。由于是在局部区域统计求得, HOG 特征对辐射差异和配准误差有较强的鲁棒性。下图 2.7, 反映的是 SIFT 特征的描述子。HOG 特征的构造过程如图 2.8 示:

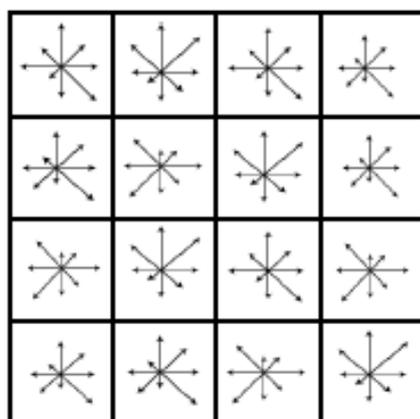


图 2.7 SIFT 描述子

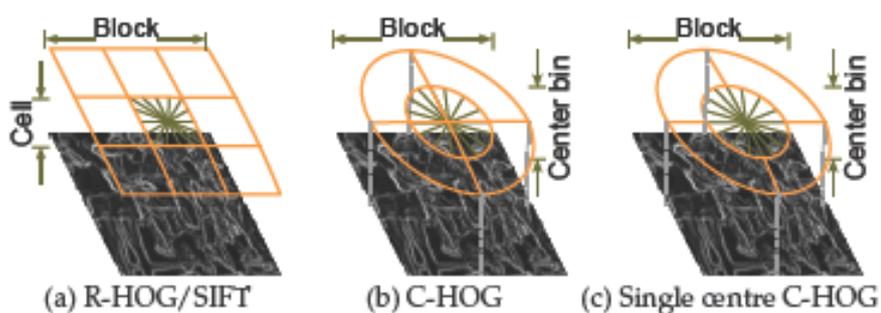


图 2.8 三种不同的特征描述子

(a) 每个矩形的特征描述子 (R-HOG) 是由 3×3 个 cell 组成的 block 来形成的。(b) 每个圆形的 HOG 特征 (C-HOG) 是在形状上将中间的圆形 cell 分成不同角度的区域。(c) 每个共享中间单元的 C-HOG (Single centre C-HOG) 是通过共享中间 cell 来实现的。

图 2.7, 反映的是 SIFT 特征的描述子,其构造过程如下: 首先将子图像划分为小尺寸的单元, 将梯度方向划分为 8 个区间, 计算每个单元对应的用梯度幅值加权的梯度方向直方图, 将其表示为一个 8 维的特征向量。子图像中所有单元的特征向量联结起来, 即构成子图像对应的特征向量。原文中将尺寸为 128×128 的子图像划分为 4×4 组尺寸为 32×32 的单元, 对每个单元计算上述的梯度方向直方图, 可得一组 $4 \times 4 \times 8$ 的特征向量。用整个子图像

的直方图“能量”（如 L1 范数或 L2 范数）对特征向量进行归一化处理，可以进一步去除光照变化的影响。

而对于 HOG 特征而言，作者 Dalal 是将 $64*128$ 的训练样本按照 $8*8$ 个像素的小块（cell）进行划分，这样就形成了 $8*16=128$ 个 cell。然后再将每相邻的 4 个 cell（田字形结构的 4 个 cell）划分为 1 组（block），block 的形成是按照每个 cell 进行滑动生成的，由此原训练样本（ $64*128$ ）便具有 $7*15=105$ 个 block。首先，本文对于每个 cell 都按照 SIFT 的方法，将其中的所有像素的梯度方向进行投影，形成每个 cell 各自的梯度方向直方图。但这里的方向 bins 的数量被设定为 9（如图 2.9），而不是 SIFT 中的 8，即每 20 度一个 bin，0-180 与 180-360 的方向采用对等角相等的方法进行归类划分。然后，再将每个 block 中的 4 个 cell 的梯度直方图的数据串联起来，由于每个 cell 的梯度直方图为一个 9 维的向量，所以每个 Block 便是一个 36 维的向量。再将所有的 block（共 105 个）依次串联起来，便形成了对每个训练图像的 $36*105=3780$ 维的编码。

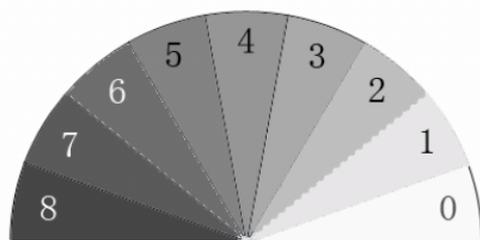


图 2.9 HOG 的梯度方向的划分

2.2.2 HOG 的实现流程

关于 Dalal 和 Triggs 的利用 HOG 进行人体检测的学习过程见图 2.10^[19,20]，特征提取与检测的大致流程如下图（图 2.11）。

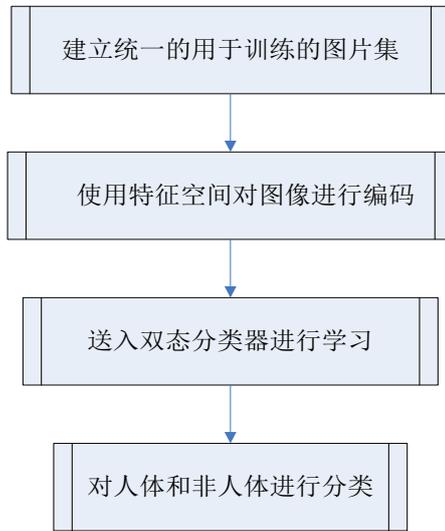


图 2.10 整体的目标检测框架图

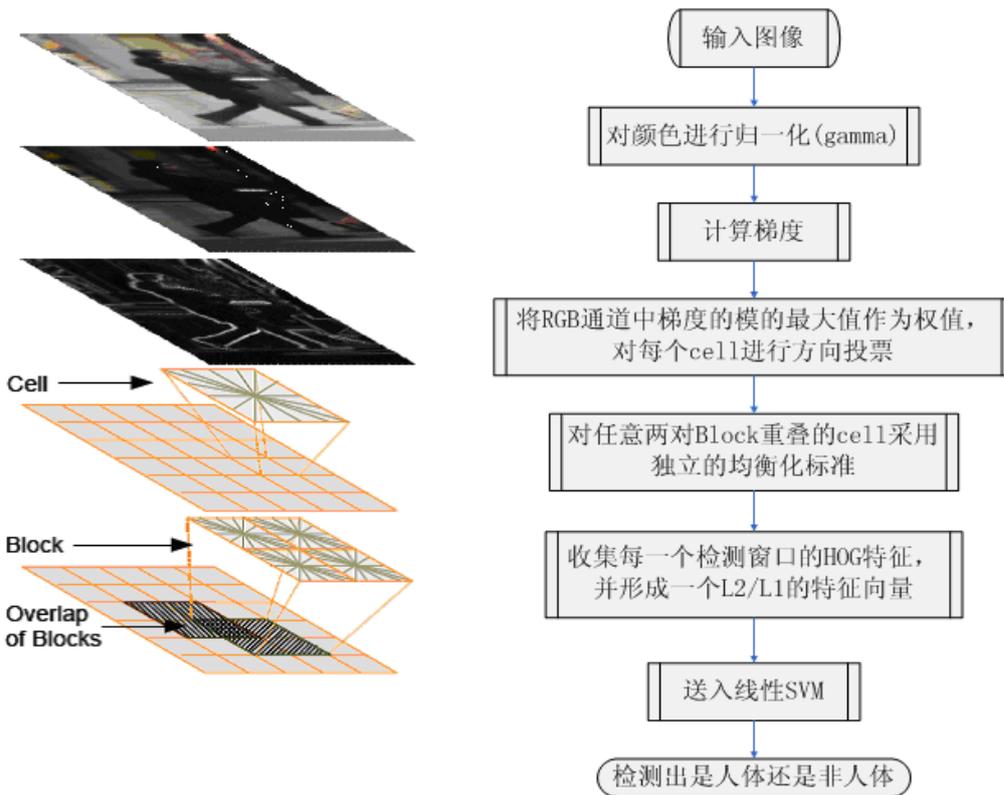


图 2.11 Dalal 和 Triggs 的特征提取与目标检测流程图

首先将输入图片分成若干块，每块再分成若干格子，对每个格子统计其中所有像素的梯度值在各个方向上的分布，得到特征向量，再把一个块内所

有格子的特征向量串联起来便得到了该块的特征向量。通过对固定大小的训练窗口的特征提取，并利用双态分类器进行训练学习。

2.2.3 Integral HOG 和 Extend HOG

Cong Hou 等人在 2007 年发表了一篇文章《Multiview Pedestrian Detection Based on Vector Boosting》，在其中作者在 HOG 的基础上提出了其推广的特征“扩展的 HOG”（Extended Histograms of Oriented Gradients features），并采用了树形结构的检测器作为向量提升算法对多视角的行人进行学习。其中 EHOH 是在梯度方向中对方向角量化，将梯度方向映射到几个主要的方向区间内，从中选取 3 个主要的区间（bin）并对其进行归一化；将两个 Block 作为一个 Block 对，在区域上进行“加”或“减”，从而形成一个新的不规则的特征区间（形状），这便是“EHOH”；采用了参考文献^[8]中的“积分梯度直方图”（Integral HOG）的方法，并结合作者提出的树形检测结构可以达到快速的检测速度，并取得了较好的多视角人体检测效果。

2.2.4 SVM 分类器

Vapnik 提出的支持向量机（Support Vector Machine, SVM）是一种基于结构风险最小化准则的学习方法^[19,24,36,37]，根据所采用核函数的不同，可分为“线性 SVM”和“非线性 SVM”两种，这两种 SVM 在处理不同的数据上，有着各自的优势。

其核心思想是找到最优分类面，此分类面要使两类样本正确分开，而且分类间隔最大（如图 2.12）。

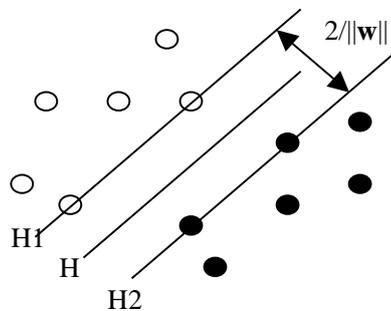


图 2.12 最大间隔分类器

对于线性可分的样本 $\{x_i, y_i\} | i = 1, N$ 。其中 x_i 为样本特征向量, $y_i \in [+1, -1]$, 其分类面方程为 $|w \bullet x_i + b| = 0$ 。将分类面方程归一化, 使两类样本满足 $|w \bullet x_i + b| \geq 1$, 这样分类间隔为 $2/\|W\|$ 。要使分类间隔最大, 即要 $\|W\|$ 最小, 并要求样本满足:

$$y_i [(w \bullet x_i) + b] \geq 1 \quad (2-5)$$

那么, 求解最优超平面问题就可以表示成如下的约束优化问题: 即在式 (2-5) 的约束条件下, 求解最小化函数:

$$\phi(w) = \frac{1}{2} \|w\|^2 = \frac{1}{2} (w \bullet w) \quad (2-6)$$

为了解决约束优化问题, 引入 Lagrange 函数:

$$L = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i (x_i \bullet w + b) + \sum_{i=1}^n \alpha_i \quad (2-7)$$

其中 $\alpha_i > 0$ 为 Lagrange 系数, 现在的问题就是要求 L 的最小值。把式 (2-7) 分别对 W 、 α_i 和 b 求偏微分, 并令其等于 0, 就可以把上述问题转换为一个较简单的对偶问题。即在约束条件下, 对 α_i 求解下列函数的最大值:

$$\begin{cases} Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j x_i \bullet x_j \\ \sum_{i=1}^N \alpha_i y_i = 0, C \geq \alpha_i \geq 0 \end{cases} \quad (2-8)$$

式中: C ----- 惩罚因子。

对于样本本质上是非线性可分的情况, 我们可以采取将特征空间映射到高维空间上去, 是样本达到线性可分的状态。但这样有可能会带来另外一个问题“维数灾难”, 解决这类问题的方法就是使用“核函数”。

关于核函数的构造, 机器学习和模式识别中的很多算法要求输入模式是向量空间中的元素。但是, 输入模式可能是非向量的形式, 可能是任何对象——串、树、图、蛋白质结构、人等等。一种做法: 把对象表示成向量的形式, 传统算法得以应用。问题: 在有些情况下, 很难把关于事物的直观认识抽象成向量形式。比如, 文本分类问题。或者构造的向量维度非常高, 以至于无法进行运算。核函数不要求输入是向量形式, 可以是任何对象。只要这

个函数在对象集上导出的矩阵是正定的。这个核函数就可以用在任何基于点积的算法中，而没有任何问题。因为核函数是特征空间中的点积，所以核函数构造过程是隐含的特征提取过程。特征提取的原则是，特征向量要反映对象的本质信息。所谓本质，就是对象所以属于不同类别的性质。所以构造核函数的原则也是应该使其反映对象的本质。

在 Dalal 所发表的论文中，作者尝试了几种不同的 SVM 分类器方法来对人体数据进行学习，在他的实验过程中，线性 SVM 的检测效果要由于其它类型的 SVM 的结果。

Dalal 使用 SVM 对人体进行训练学习的机理如下：

先将训练样本集（包括正例和反例）进行 HOG 编码，由上可知，每幅 $64*128$ 的训练样本都将编码为 3780 维的特征向量——这便是上面 SVM 公式中的 x_i 。而对于每个样本，本文对其相应的特征数据进行编号，凡是正例样本都标记为+1，反例样本都标记为-1——这便是上面 SVM 公式中的 y_i 。由此，本文将生成的样本数据放到线性 SVM 训练的程序中进行学习分类，经过训练后，SVM 将对每一个训练样本都进行赋权值。进而在后来的人体检测过程中便可以用训练出来的这个模型进行匹配检测。

2.3 本章小结

在本章，我们介绍了当今基于机器学习的行人检测的两种主要的研究方法。基于图像块颜色的方法，致力于在图像上寻找最有效的特征区域来描述行人；基于梯度方向的方法，着力于对图像中相应的区域进行方向统计，以适应行人在动作、姿势上的变化。另外，还有基于多部位、多视角的方法等等。究竟哪个方向才是最有效呢？这个问题很难回答。因为在行人检测领域，还缺乏一个标准的测试集。目前使用最多的是 Inria 行人数据集。在该数据集上，目前最好的方法还是基于特征的方法，如 shapelet, HOG 等等。由此也可见，特征的选取对于行人检测是多么地举足轻重。当然，就如我们曾经说过的，这些方法是从不同的角度出发解决不同的问题，是相互补充的。一个理想的行人检测系统应当是将这些方法有机地结合到一起，才能达到最佳的性能。

第 3 章 基于新特征的行人检测

由上所述，本文可以得出下面的结论：行人的外貌表现得千差万别，而且会产生局部形变，这使得多视角的行人检测仍然是目前研究的重点。在这篇论文中，本文尝试使用一种新的特征描述子来进行人体检测。与目前存在的成果相对比，具有以下两个主要贡献：

提出了一种新的特征描述算子，用来进行行人检测；

在行人检测器构造上，本文尝试了两种分类器的学习方法：SVM 和 AdaBoost。SVM 更加注重结果的正确性，而 AdaBoost 则强调的检测的高效性，从结果上来看，两者可以达到较高的检测率。

本章节中的剩余部分组织如下：

第 1 部分详细介绍本文的人体检测算法；

第 2 部分为 MSO 特征集的构造方法；

第 3 部分进行分类器模型的构建；

最后在第 4 部分使用新特征进行物体检测。

3.1 本文采用的人体目标检测算法

通过在第二章中的介绍，我们已经知道 Dalal 提出的 hog 特征在行人检测上的优异表现。但是 Dalal 提出的算法也存在速度过慢的问题，即使用线性的支持向量机，面对每个图片窗口中三千多维的特征向量也显得力不从心。根据 Dalal 的实验，如果每张图片检测 4000 个窗口，那么还是需要 1 秒的运行时间。显然这是无法满足实时的要求的。之后 Qiang Zhu 将 hog 取代了 haar 特征，利用 boosted cascade 的算法框架将检测速度大大提升。但是这样的速度还是无法达到实时的要求。受到 Bo Wu 将 haar 特征与 edgelet 特征结合起来的启发，我们提出将 haar 特征与 hog 特征结合起来，提出了我们自己的特征描述子 MSO (MSO: Multi-scale orientation feature, 多尺度方向特征)，将其计算量得到了降低。

通过本文提出的算法，本文可以在图像或视频中实现对人体目标检测，

并具有很好的鲁棒性。本文的系统框架大致的分为 2 部分，分别为训练样本和目标检测，流程图分别为图 3.1 和图 3.2。在接下来的部分，本文对 MSO 特征集的构造、分类器模型的设计以及目标检测的方法分别作以分析。

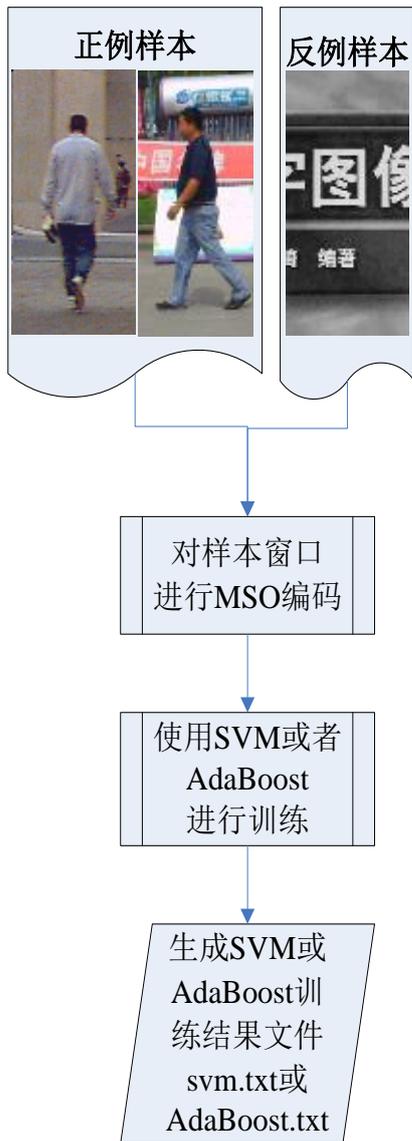


图 3.1 训练样本框架图，
分别构造正面、侧面两个分类器

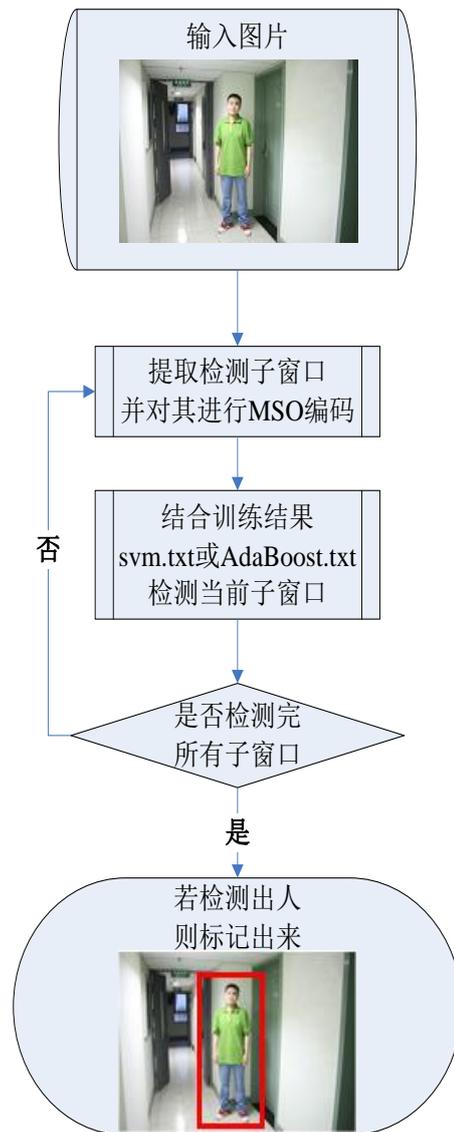


图 3.2 目标检测框架图，
根据两个分类器标记出结果

3.2 MSO 特征集的构造

在目标检测中，对物体的描述的好坏将直接影响其最终的检测性能。正如前面引言部分所述，目前对于特征的描述主要分为对色彩的处理和对轮廓的提取。这两方面最具有代表性的就是 Viola 提出的 Haar-Like 和 Dalal 采用的 HOG 作为物体的特征描述子，而且在人体检测上达到了很好的效果。通过前一章的介绍，本文可知，Haar-Like 训练是基于灰度图像的一种学习模式，相当于对一幅“灰度图像”在“尺度、位置”变化的基础上，在其区域内的“像素灰度差值”的概率统计。而 HOG 训练是基于对 R、G、B 任意通道上最大“梯度方向”的一种学习方法，它反映的是被识别物体轮廓的方向在某一区域内的一种“区域投票”的概率统计。然而，这两种方法都存在着各自的不足，前者对于颜色和光照过于敏感，后者由于是一种区域内的方向投票，因此容易受到纹理的影响，而且运算速度慢也是一个待于解决的问题。

本文作者受到以上两种特征描述子的启发，提出了一种新的特征，这不仅囊括了其各自的优点，还弥补了它们的不足。这种特征以图像的颜色特征作为基础，通过对特征区域像素的颜色值进行累加换算，转换成相应的方向特征。然后再针对特征区域的形状进行统计的，由全部特征组合起来所形成的特征集反映的是图像在不同尺度上的方向特征。

3.2.1 特征单元的构造

在固定尺度上的单个特征的计算如图 3.3 所示，其中每个特征区域都是一个 $m*m$ 的正方形，对于这个正方形特征的特征值的计算，本文可以分为两步：1、计算正方形区域的毛梯度方向（因为本文这里是在大尺度上求解梯度，相对于原始定义的梯度而言更加粗糙，因此这里将其命名为“毛梯度”）；2、将计算出来的梯度方向映射成固定的特征值编码。

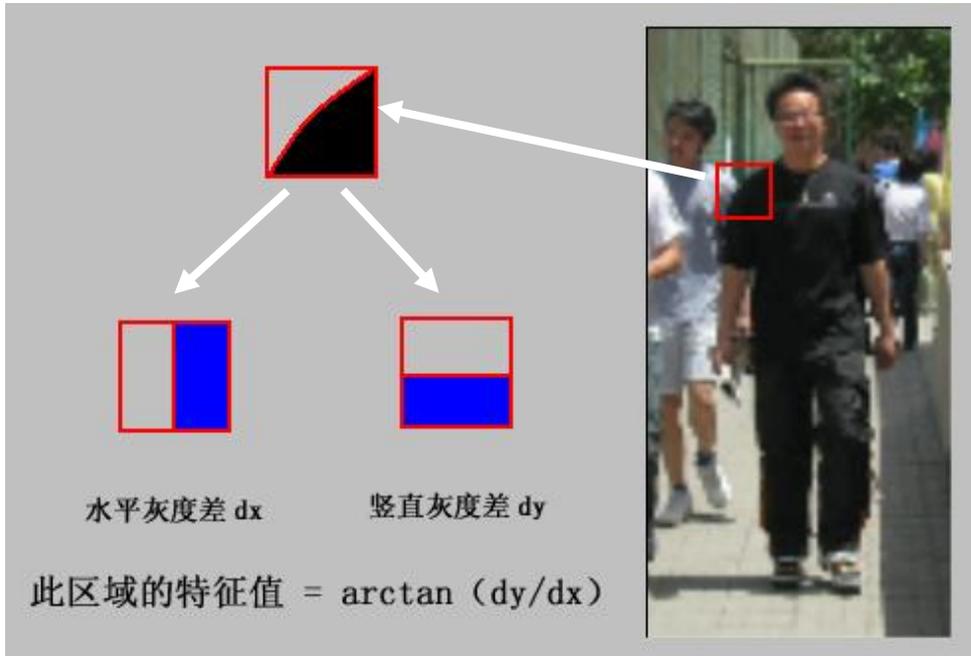


图 3.3 单个特征的计算

首先，本文将正方形分为左右两个等份，计算左右两个小区域的所有像素的灰度值之和，分别记为 SUM_L 和 SUM_R ，得出水平灰度差 dx ；再将正方形分为上下两等份，计算出竖直灰度差 dy 。这样这个正方形在形状上便体现了田字形的特征。

如图 3.4 所示，区域 A 的灰度值累加和 SUM_A 的计算公式如下：

$$SUM_A = \sum_{x_1 < x < x_2, y_1 < y < y_2} PixelValue \quad (3-1)$$

其中， $PixelValue(x, y)$ 表示坐标为 (x, y) 像素点的灰度值，这是一种比较直观的求法，然而本文可以参考积分图的求解过程，这会在很大程度上降低整幅图像的运算复杂度。这样本文便可以求出正方形特征的水平灰度差 dx 和竖直灰度差 dy 。

$$\begin{aligned} dx &= SUM_R - SUM_L \\ dy &= SUM_D - SUM_U \end{aligned} \quad (3-2)$$

其中， SUM_R 、 SUM_L 、 SUM_D 和 SUM_U 分别表示为正方形特

征区域的右半部、左半部、下半部和上半部的灰度值累加和。

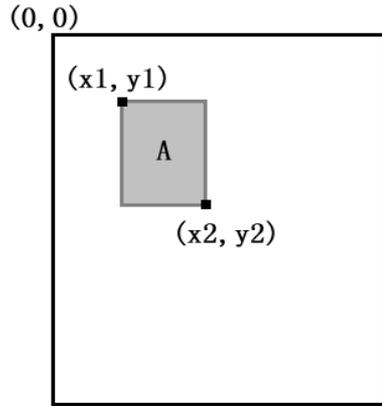


图 3.4 矩形区域的灰度值求和

接着，本文先将 0~360 度等分为 18 个区间（如图 3.5 所示），并保证对顶角区域属于同一区间。这样本文就可以根据所求出来的 dx 和 dy 计算出这个矩形区域内的“大致”梯度方向 Ori_Rect ，

$$Ori_Rect = F(\arctan(dy/dx)) \quad (3-4)$$

其中，F 为将角度转换为数值 0, 1, 2, ..., 8 的函数，如：当 $\arctan(dy/dx)$ 的值在 1~20 之间的时候，则 Ori_Rect 的值为 0；当 $\arctan(dy/dx)$ 的值在 21~40 之间的时候，则 Ori_Rect 的值为 1；依此类推。

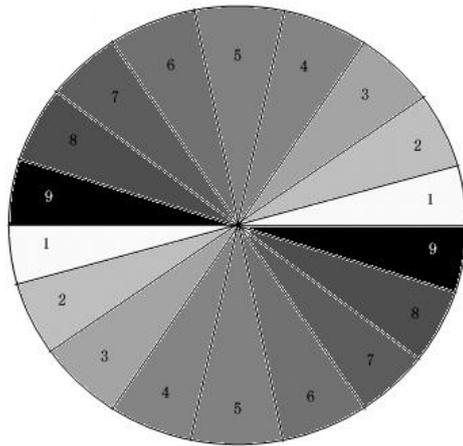


图 3.5 特征空间的划分

当使用 SVM 和 AdaBoost 两种方法分别检验本文的特征描述子时，本文发现这种新的特征描述子对于不同的学习算法都是等价的。并且在 AdaBoost

验证框架里，本文发现如果将特征空间（如图 3.5）等分为 18 份，将会得出更好的效果（而对于 SVM 的机理，本文采用的仍然是 9 个空间的形式）。

当本文将训练样本（或被检测子窗口）内不同尺度的特征组合起来，这边形成了“特征集”，即：将图像编码为一个高维的向量。在本文的两个检测系统中，这个特征集合的构造方法和大小是存在些差异的。

3.2.2 AdaBoost 特征集的构造

本文吸取了 Viola 的 Haar-Like 特征“可变尺度”的特性，本文打破了 HOG 单一尺度特征上的束缚，由此便形成了本文的“多尺度方向特征”。在 AdaBoost 框架下，本文将正方形特征区域的尺度规定为 $6*6, 8*8, \dots, 20*20$ 的大小，如图 3.6。

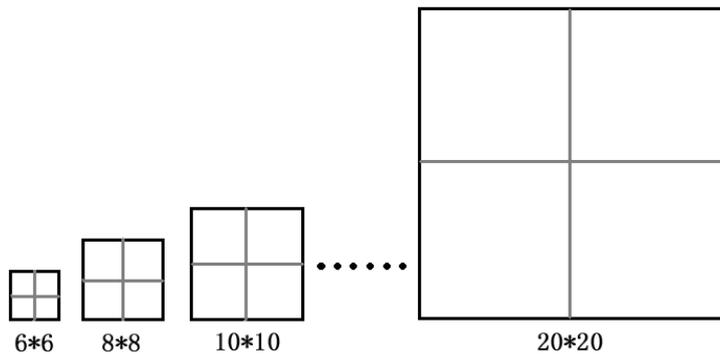


图 3.6 AdaBoost 下的特征窗口的划分

对于某个特定尺度上的田字形特征来说，按照自左向右，由上到下的顺序平移特征窗口，水平和垂直方向上的平移的步长 $iHPace$ 和 $iVPace$ 都设定为 2，这样便产生当前尺度上的所有特征。例如：当田字形特征尺度为 $6*6$ 时，在大小为 $64*128$ 的训练样本上，可以得到 $((64-6)/2+1)*((128-6)/2+1)=1860$ 个特征值。当所有尺度上的田字形特征都计算完后，将这些特征依次排列，便形成一个的 11180 维的高维特征向量。

3.2.3 SVM 特征集的构造

由于 SVM 的分类的本质是按照检测样本与支撑向量相比较，而进行近邻

的划分。也就是在检测时需要用到每个支撑向量的所有的特征进行运算，而且要进行大量的乘法运算，因此在检测前应该保证每个样本的特征数量尽可能的少，所以这就需要在训练样本之前要对特征进行选取。

在 SVM 的特征集的构造中，本文将特征窗口的大小分为 $6*6$ ， $10*10$ ， $14*14$ 的大小的特征单元块，并针对各自的特征单元块，通过 SVM 的筛选机制选出一定数量的特征，将这些特征组合一起便形成了对人体的特征描述。在实验中，本文选取了 400 个特征来作为 SVM 的训练特征集。

3.2.4 MSO 特征集的特点

本文的这个 MSO 特征集涵盖了不同尺度上的特征，加之特征窗口的固定步长的平移，因此能够有效地解决因为人体局部偏移而造成的区域特征的丢失，相对于以往的方法大大的降低了逻辑上的复杂度（避免了因需要线性插值等步骤而带来的额外开销）。也就是说，田字形特征窗口越大，其对于局部偏移的问题解决得就越好。然而由于实际当中背景的千差万别，对于反映边界特征的 MSO 特征来说，当特征窗口过大时，必然会受到过多的不规则背景区域的干扰，这对于后期的特征选择训练来说是没有什么价值的，但窗口太小也不利于对噪声的处理。因此本文将特征窗口的尺度规定为从 6 到 20 等差为 2 的数值序列，这样便可以正确并且有效地反映目标物体的固有特征。

除了上面所提到的可以去除目标物体由于扰动而造成的误差的特性外，MSO 特征另外一个显著地优势就在于“特征组合”，它不仅仅可以实现对同一尺度上的特征进行组合，还可以实现多尺度上的特征的组合。这是 HOG 等一些特征描述子无法实现的。关于 MSO 所选特征的具体表现形式见下面的特征选择的结果图。

3.3 分类器模型

在形成所有训练样本的特征文件后，应选取特定的分类器对编码数据进行学习分类，为此本文分别选择了 SVM 和 AdaBoost 两类分类器。

3.3.1 SVM 分类器

正如上面所述，SVM 是一种基于结构风险最小化准则的学习方法。在本文的算法中，其基本的训练流程为：

关于 SVM 的特征的选择，本文是采取限定特征阈值的范围方法。即：对某一个特征对应的方向选定一个区间，保证为每一个特征所选的区间必须为当前特征所有可能区间集中划分性能最好的那个（如图 3.7），这样将某一尺度上的所有特征都集中起来，选择在各自最好区间的条件下，选出划分性能最好的前 n 个特征最为当前尺度上的所选特征组合——特征选择是按照“ $\max(\frac{\text{正确划分比率}}{\text{错误划分比率}})$ ”的原则来进行筛选的。这样，将所有尺度上所筛选出来的特征组合起来便形成了 SVM 的特征集。

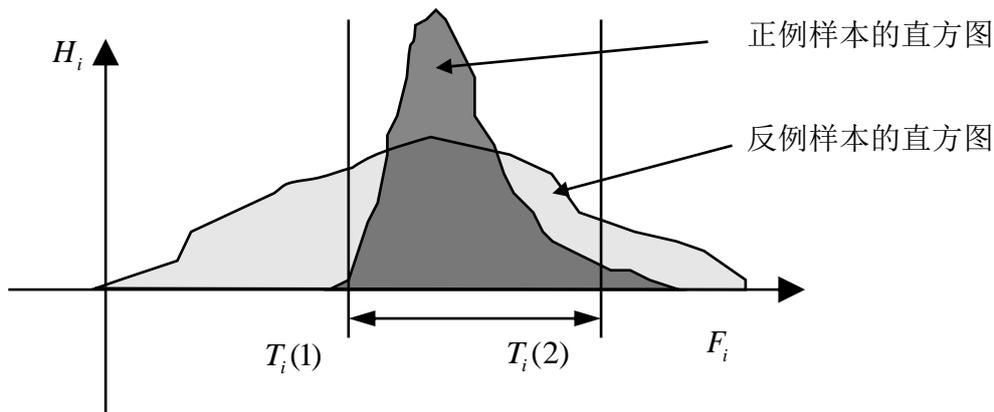


图 3.7 双阈值的划分

横坐标表示第 i 个特征，纵坐标表示第 i 个特征的某个方向在所对应的方向直方图中占的比率， $T_i(1)$ 和 $T_i(2)$ 表示第 i 个特征所对应的特征区间。

对于 SVM 的训练，本文选择的是非线性的 SVM 学习方法。本文把每一个训练样本按照所选特征进行编码，然后放到非线性 SVM 里进行训练，最后的到训练的模型（模型中的部分结果见下图 3.8）。这样本文就可以根据训练出来的模型对测试图片进行行人检测。

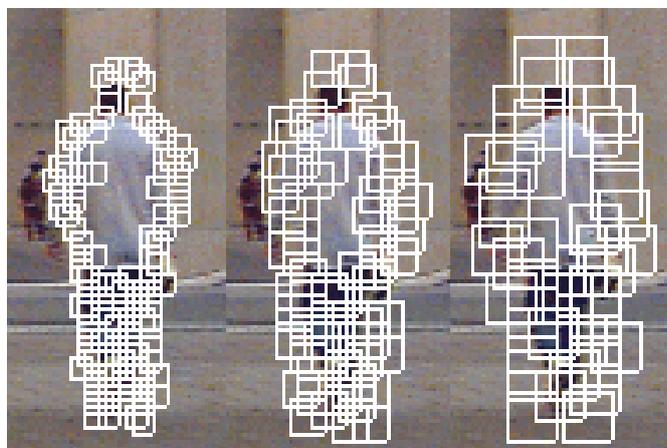


图 3.8 SVM 特征选择

图 3.8 分别在 $6*6$ ， $10*10$ ， $14*14$ 三个尺度上的所选出来的特征，SVM 也正是在有这些特征所组成的特征集上进行训练样本的。

3.3.2 AdaBoost 分类器

事实上，在行人检测领域中，不同的方法往往都会面临着相同的难题：无论使用的是 Haar-Like、HOG、edgelet，还是 shapelet 等算法中，潜在的特征成千上万，我们应该如何选择相应的特征呢？此外，从实用角度上看，检测的速度必须越快越好。一张图片中的大部分区域都不包含行人，如何利用这一点来加速检测呢？本小节将要介绍的正是对这两个问题的一个经典解决方案，我们称之为“级联 AdaBoost”。该方法最初由 Viola 提出，将当时的人脸检测速度提高 10 倍以上，成为之后人脸检测领域的一大经典算法框架，围绕该算法框架展开了大量的研究。在行人检测领域，该算法框架也同样十分适用。在之前提到的各种算法中，有近一半的算法采用或借鉴了该框架。而在这一节中我们将讨论对该算法做出的一些改进，以使之能够更加适应本文提出的 MSO 特征，从而在实现较高的检测精度的同时实现了更快的检测速度。

AdaBoost 的优缺点分析：

由于“人体检测”只是要求将目标划分为真和假即可，这完全符合离散 AdaBoost 的特点。并且这种分类器的特点是采取了弱分类器投票的方式来构

造强分类器，程序实现简单，因此在本文的系统中采用离散的 AdaBoost 作为分类器是非常恰当的。

基本的训练流程：

在实验过程中，由于特征值的离散性，本文并没有采取 Viola 的 AdaBoost 中弱分类器的构造方法，而是采用了一种特征区间的划分方法。这种方法的依据是：对于任何事物的特征描述，会产生大量的特征，然而并不是特征集中的所有特征在物体的划分上都是有用的，本文所要求的是找出能够凸显人体个性的那些特征。当训练样本集非常巨大时，对于人体的代表特征来说，出现在某一特定区域的密度会大一些，而这类特征在大量的反例（非人体）样本上反映的则是近似符合均匀分布，见图 3.9。本文顺着这个思路便可以提取出那些有用的人体特征。

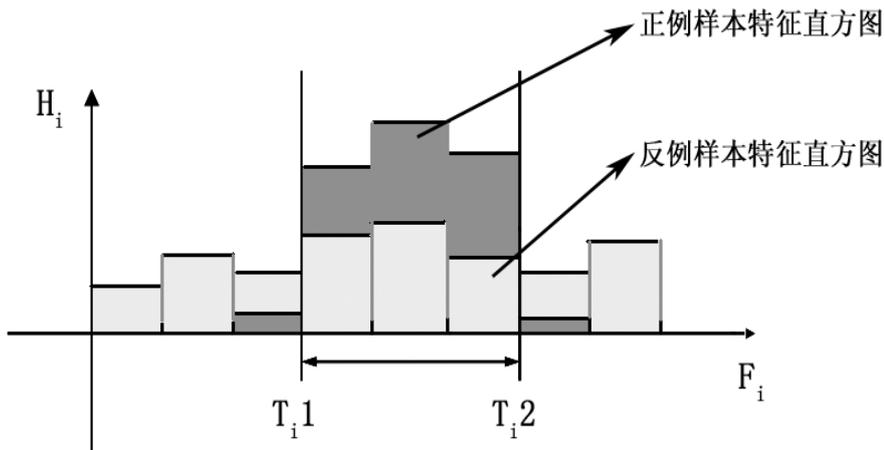


图 3.9 训练样本在某一特征上的分布

对于某个特征 i 而言，纵坐标 H_i 表示所有训练样本在某一特征值上所占的比重；横坐标 F_i 表示将特征区域分成的 9 等份或者 18 等份； T_{i1} 和 T_{i2} 分别记录特征选择时的左右阈值；淡灰色直方图反映的是反例样本的分布情况，深灰色直方图反映的是正例样本的分布情况。

为了提高检测速度，本文采用了级联（Cascade）AdaBoost 的方法，本文通过在观察训练的过程中，惊奇的发现在第一级 AdaBoost 结束时只筛选出了一个特征，并且仅通过这一特征可以排除 90% 以上的反例窗口。级联 AdaBoost 的算法流程见图 3.10。

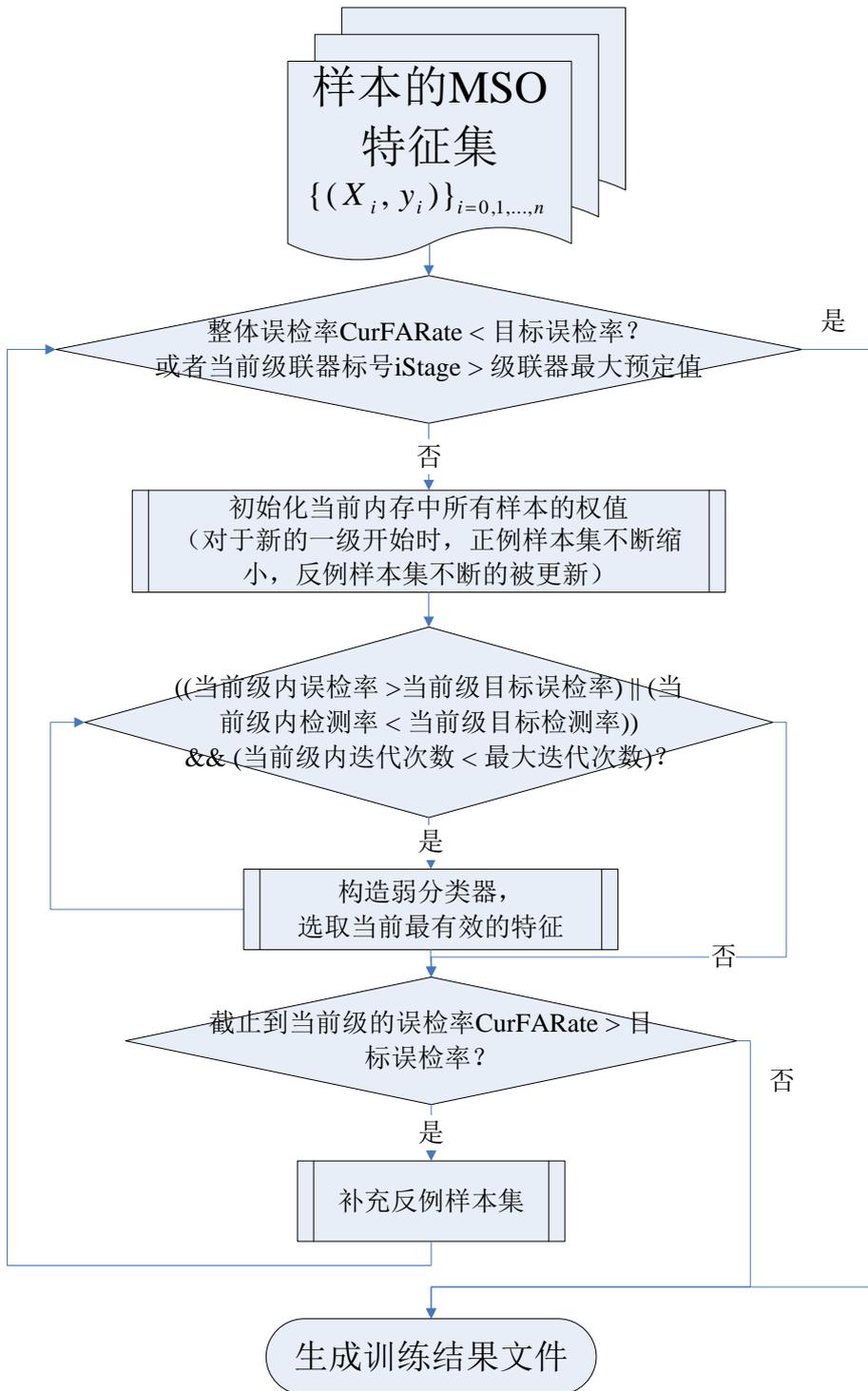


图 3.10 AdaBoost 训练流程图

图 3.11 为使用本文的级联 AdaBoost 后, 分别在前 1、2、3、22 级所提

取出来的有效人体特征。通过最后一个图本文已经可以明显的看到一个人体大体轮廓，这也恰恰验证了本文前面所提出的“MSO 反映的是在不同尺度上的粗糙梯度”的观点。

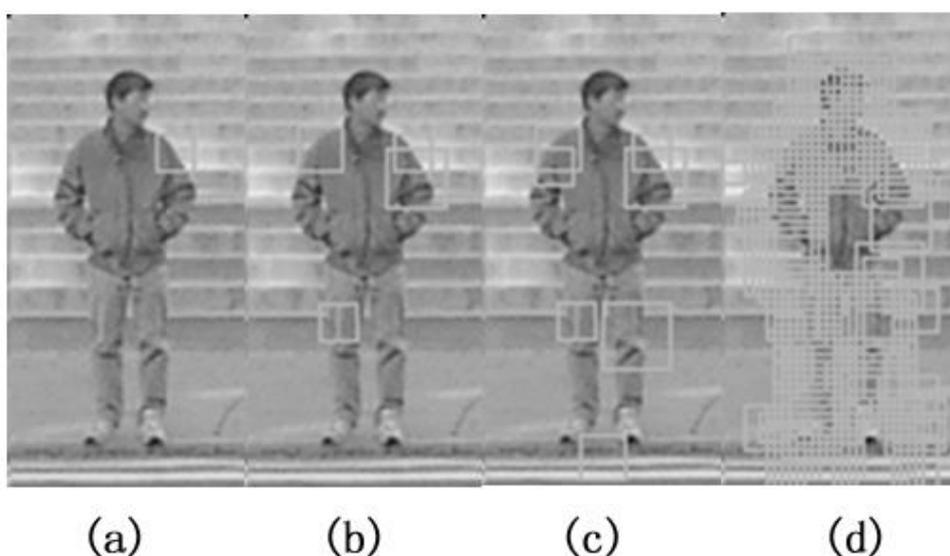


图 3.11 Cascade AdaBoost 选取特征结果图

(a) 为第 1 级，只选择了一个特征；(b) 为前 2 级，一共使用了 4 个特征；(c) 为前 3 级，一共使用了 7 个特征；(d) 是前 22 级所筛选出来的有用特征。

在系统中，本文所采用的 Cascade AdaBoost 算法是按照 Viola 一文中的流程进行编写的，并且对弱分类器的构造采用了本文提出的“特征区间划分”的方法，并对代码进行了一定的优化。通过本文的系统，对 3000 张尺寸为 64×128 的正例样本，4500 张尺寸为 64×128 的反例样本，以及 2000 张 640×480 的反例大图(作为 Adaboost 训练时补充反例使用)进行训练时，在 CPU: 3.0G, 内存: 1G 的机器上，只要 1 个多小时的时间就可以完成，这相对其它已存方法的训练时间要少得多。相比之下，本文在使用 OpenCV 自带的训练器对尺度为 32×32 大小人脸样本集进行训练时，使用了 1000 张正例样本，2000 张反例样本，在如上的配置下，却需要耗时 13 个小时。然而，在本文的框架下，只需要不到 1 个小时就可以完成，训练的结果也是比较让人满意的。

3.4 物体检测

当本文通过级联分类器训练样本生成“模型文件 AdaBoost.txt”后，本文便可以加载这个模型进行检测了。

在程序中，本文采用了“积分图”的理念，并为每个检测子窗口的灰度方差和每个田字形区域特征的 dx 、 dy 分别设定一个阈值（目的是快速排除被测图片中的“过平坦区域”），这些都可以使检测速度得以大幅度的提升，基本上可以达到实时。最终的检测器是在整幅图片的不同尺度和不同位置上进行检测的，检测窗口尺度的缩放是通过迭代乘以一个缩放因子实现的，并通过在“积分图”上截取子窗口来实现加速，本文设定的尺度缩放因子 $scale$ 为 1.25。在同一尺度窗口下，通过平移检测子窗口来遍历整幅图像，水平方向和垂直方向的平移步长为 $\delta * scale$ ， δ 本文设定为 1.5。按照级联 AdaBoost 的原理，只有被检测子窗口通过全部预定的级联分类器（见[2,5]），才能判定当前窗口区域为人体区域，任何一级达不到标准都不能判定为正例。

在本文得到检测结果时，会发现：正确的人体周围会有好几个检测结果，在非人体区域也可能出现错误的检测结果。这就需要本文采取一些方法将那些层叠的检测框合并，同时去掉检测错误的结果。关于检测窗口的融合与误检窗口的去除，本文主要遵循以下几条原则：

(1) 在检测器保证鲁棒的前提下，如果图像上的人体非常符合正例样本集的特征，那么在其相邻区域、相近尺度上都有一些检测结果的层叠；

(2) 对于非人体的图像窗口，则不会出现像正例窗口那样高的检测结果数；

(3) 在同一个人体的检测结果的不同尺度中，越是符合分类器要求的（即与分类器中的模型越相近的），在这个尺度上检测出来的结果才越接近真实的情况。

3.5 本章小结

在本章中，本文使用了我们提出的 MSO 特征，并将 Cascade AdaBoost 和 SVM 两种机制应用到了行人检测中。我们随后对两种机制分别进行了详细的分析，从实验结果来看，MSO+Cascade AdaBoost 既能避免原始的 HOG

算法计算复杂度高，不适合实时检测的缺点，又能弥补 Haar-Like 特征过于简单，精度不大的缺陷；而 MSO+SVM 在运算速度快于 HOG 的前提下，具有更高的检测效果。

除此之外，我们根据 MSO 特征描述子的特点，设计了一种与之相适应的弱分类器构造方法，这种新的 Cascade AdaBoost 方法既简化了计算，又提高了性能，达到了更高的检测精度。

第 4 章 实验结果与对比分析

4.1 数据集的形成

4.1.1 数据集 (Data Set) 的描述

本文一共准备了 1592 个正面人体样本 (其中包括 MIT CBCL 行人数据库样本 924 张和本文自己标注的 668 张, 然后再将每个样本左右翻转, 相当于将样本数量扩大了两倍, 共 3184 张正例训练样本。其中 MIT 的训练样本可在 <http://cbcl.mit.edu/software-datasets/PedestrianData.html> 上下载); 同时本文自己又准备了 3050 张 (经左右翻转后) 侧面人体样本。

关于本文自己标定的训练数据可以到中科院研究生院的网址上下载 (<http://cear.gucas.ac.cn/SDL-HomePage/Dataset.htm>)。

4.1.2 数据标定的标准

由于本文也采用了 MIT 的人体库进行训练, 所以在本文自己的人体库中的样本也应该按照 MIT 的格式进行修整。

为了形成本文自己的训练人体样本库。首先, 本文使用数码相机拍摄了 2000 多张不同背景、不同光照、不同着装、不同姿态的行人图片, 这样便大大的丰富了本文的训练样本。其次, 使用本文自己编写的图片编辑器 (见图 4.1) 来对原始图片 (800*600 大小) 中的行人进行手工标定。由于在原始图片中的行人的尺度表现的千差万别 (有大有小), 因此本文还要将手工标定出来的数据进行转换, 都转换成与 MIT 人体库中的样本的格式相同, 包括图片的大小和人体在样本图片中所在的位置。

图片标定时满足图片左侧中的黑色矩形框为标定的结果, 右侧的图片为由此生成的训练样本图片。

关于“图片的标定”与“样本的生成的标准”如下:



图 4.1 图片标定工具

图片的标定时，对人体的标定框要求满足“上齐头，下齐脚，左右齐肘”的原则；

当对样本提取时，首先要滤除尺寸比较小的样本，在本文的标准中，如果标定的样本的图片小于 $32*64$ 个像素则将其丢弃，不进行样本的提取；对于满足这个标准的标定结果，本文将其按照“图像规格 $64*128$ 个像素，上下边距 14 个像素，左右边距 14 个像素”的规则进行样本归一化。

由于每个行人的高矮、胖瘦等形态都不尽相同，因此无法保证对样本提取时都满足“上下边距 14 个像素，左右边距 14 个像素”，所以这里本着“上下边距”优先的原则，来对样本图片进行提取，这样在图片中虽然会表现出左右边距不等的情况，但这并不影响本文对样本图片集的训练。

由此，便形成了对正面人体的标定与提取工作。对于侧面行人，本文也采用同样的方法，只不过是将标定的标准做以修改。侧面行人的标定标准为“上齐头，下齐脚，左右齐脚”，对提取样本时，按照“图像规格 $64*128$ 个像素，上下边距 14 个像素，左右边距 14 个像素”的规则进行样本归一化，

并且满足“上下边距优先”的原则。

4.2 实验结果

根据实验证明，采用本文所提出的方法，可以对图片中的行人得到很精确的定位，并具有很快的处理速度。图 4.2 中的图片为在本文的“正面行人测试集”中的部分检测结果。图 4.3 为在本文的“正面行人测试集”中具有漏检情况的结果。图 4.4 为在本文的“正面行人测试集”中具有错检情况的结果。

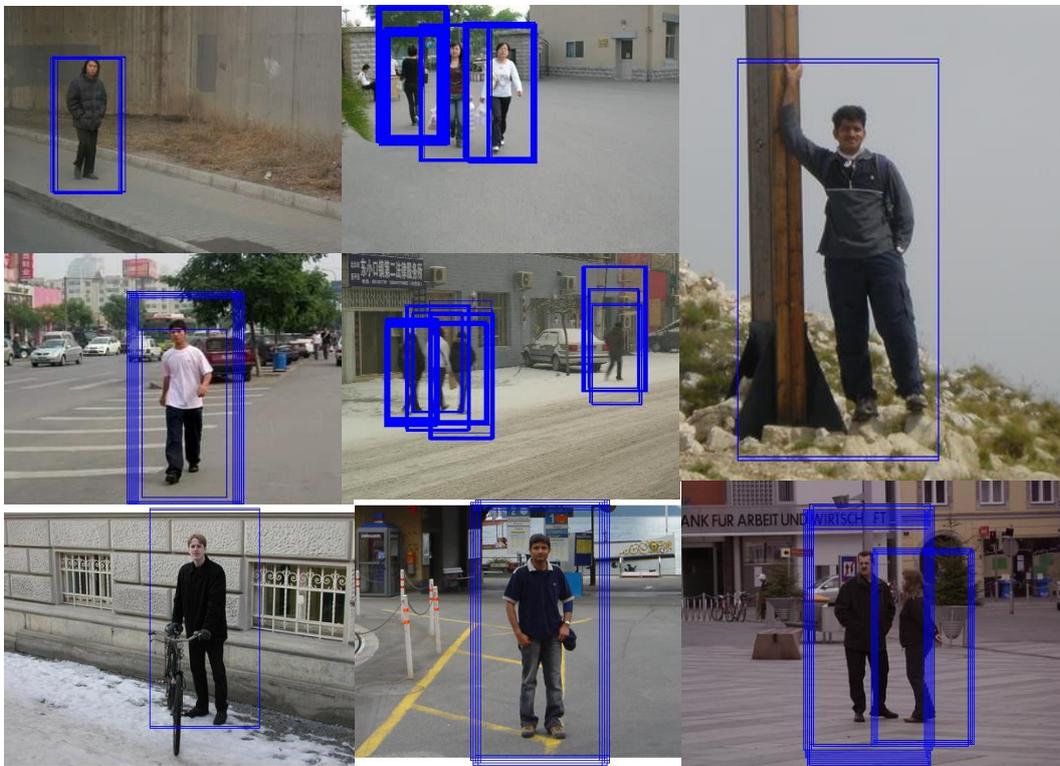


图 4.2 “正面行人测试集”中的部分检测结果



图 4.3 “正面行人测试集”中具有漏检情况的结果

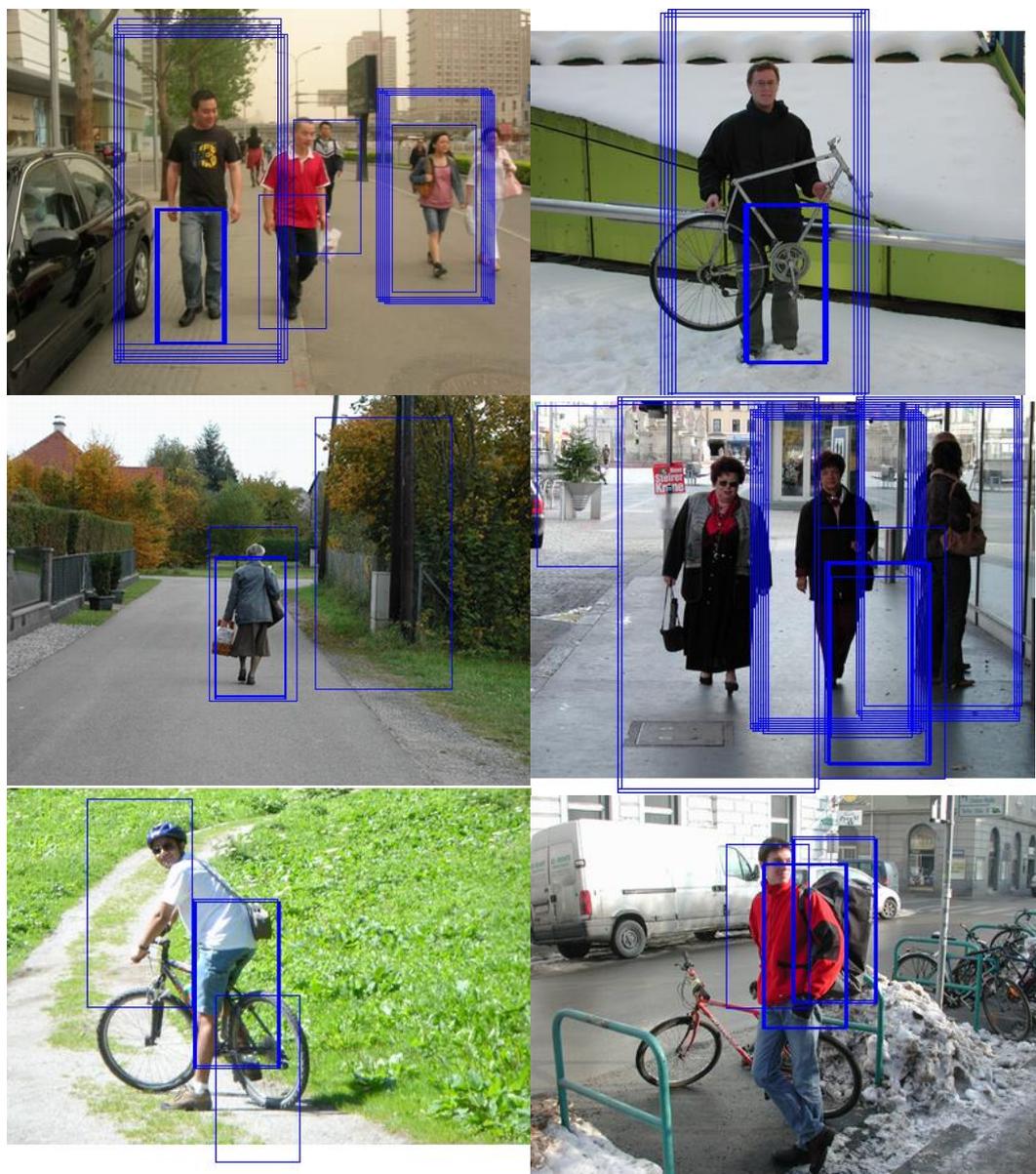


图 4.4 “正面行人测试集”中具有错检情况的结果

本文同时也针对“侧面人体”进行了训练，关于侧面的检测效果见下图。图 4.5 为在“侧面行人测试集”中的部分检测结果。图 4.6 为在“侧面行人测试集”中具有漏检情况的结果。图 4.7 为在“侧面行人测试集”中具有错检情况的结果。在这里，本文将左侧和右侧的行人都归结为“侧面行人”，即：训练样本中既包括“左侧行人”又包括“右侧行人”。

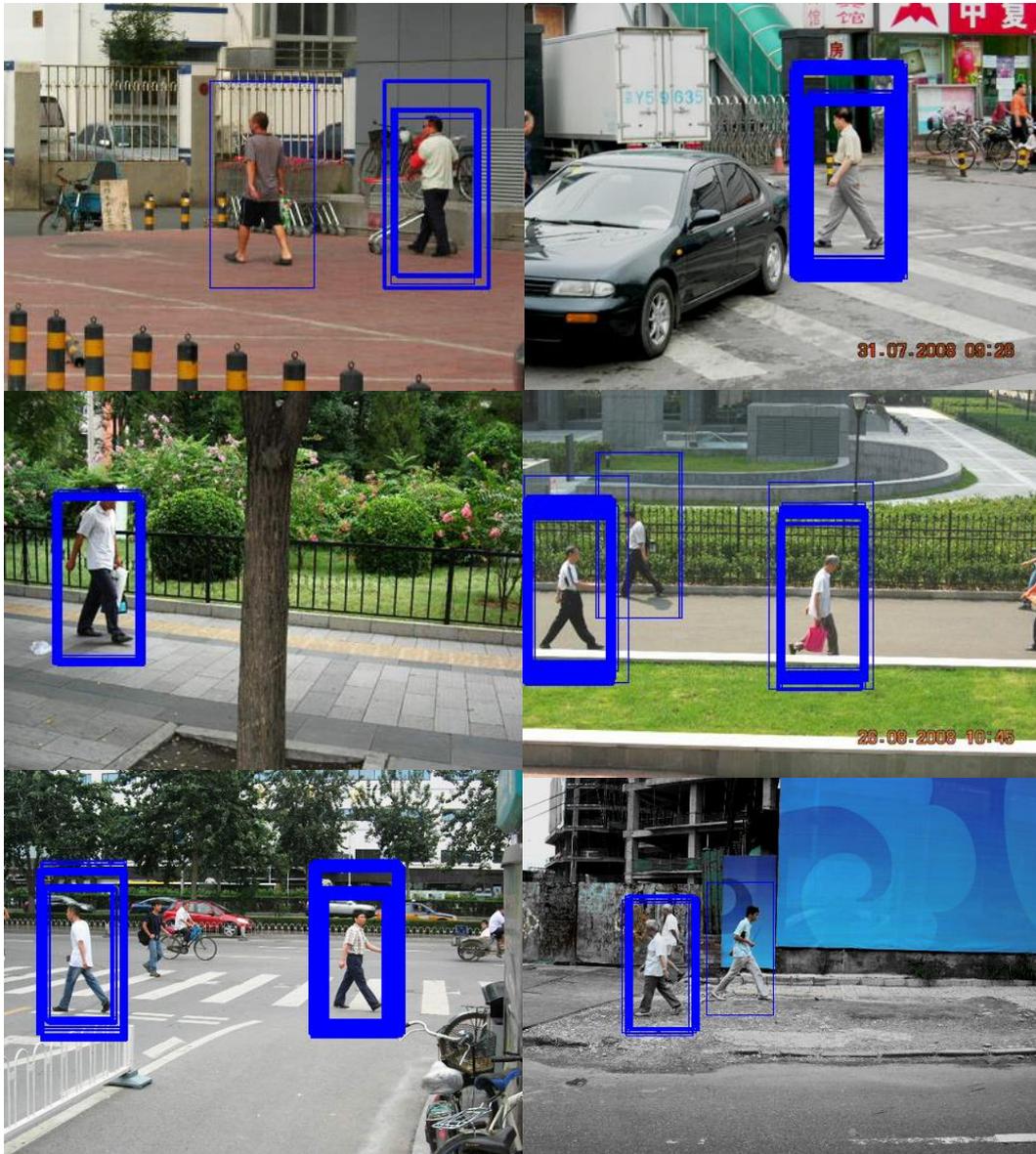


图 4.5 “侧面行人测试集”中的部分检测结果

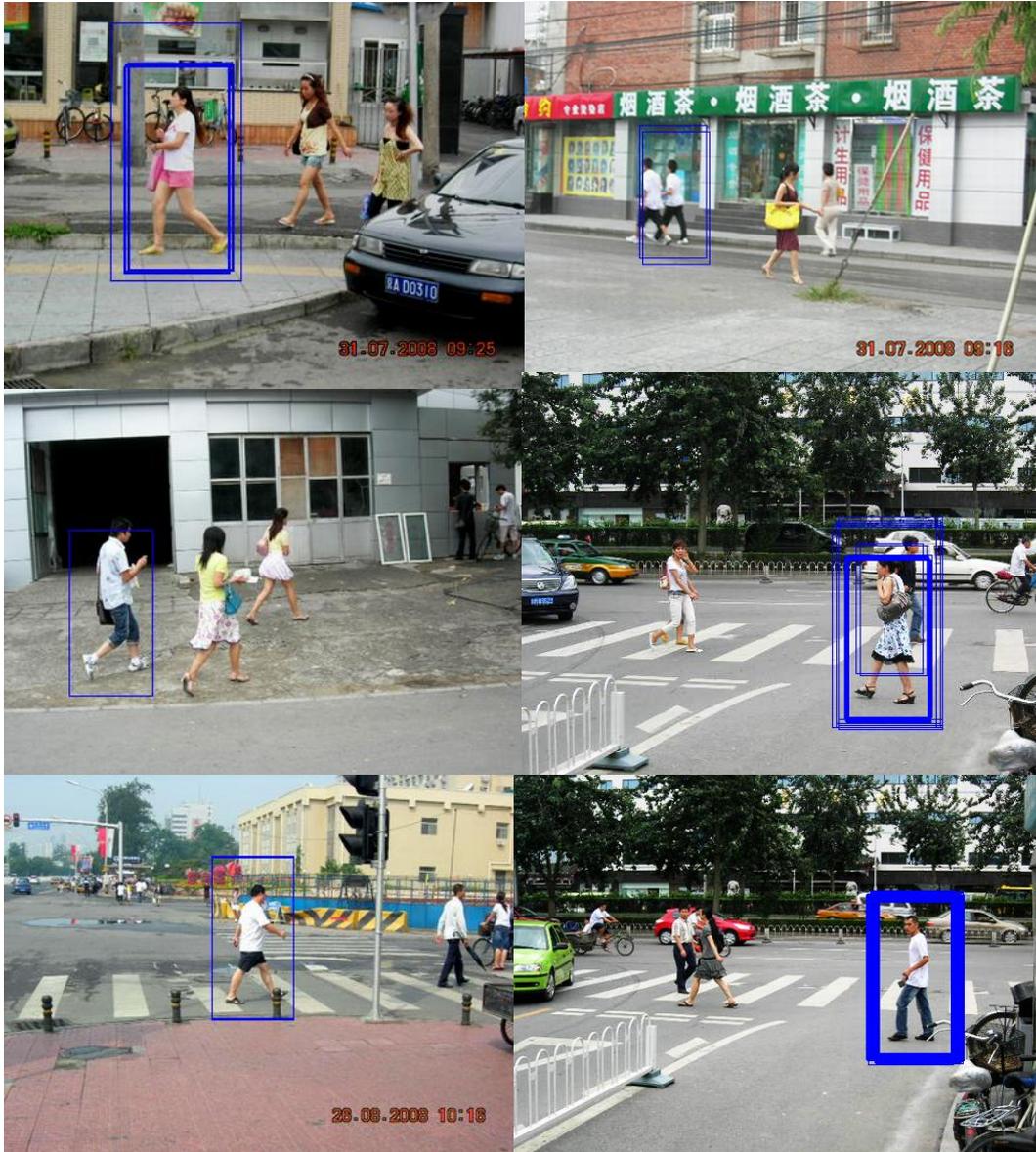


图 4.6 “侧面行人测试集”中具有漏检情况的结果

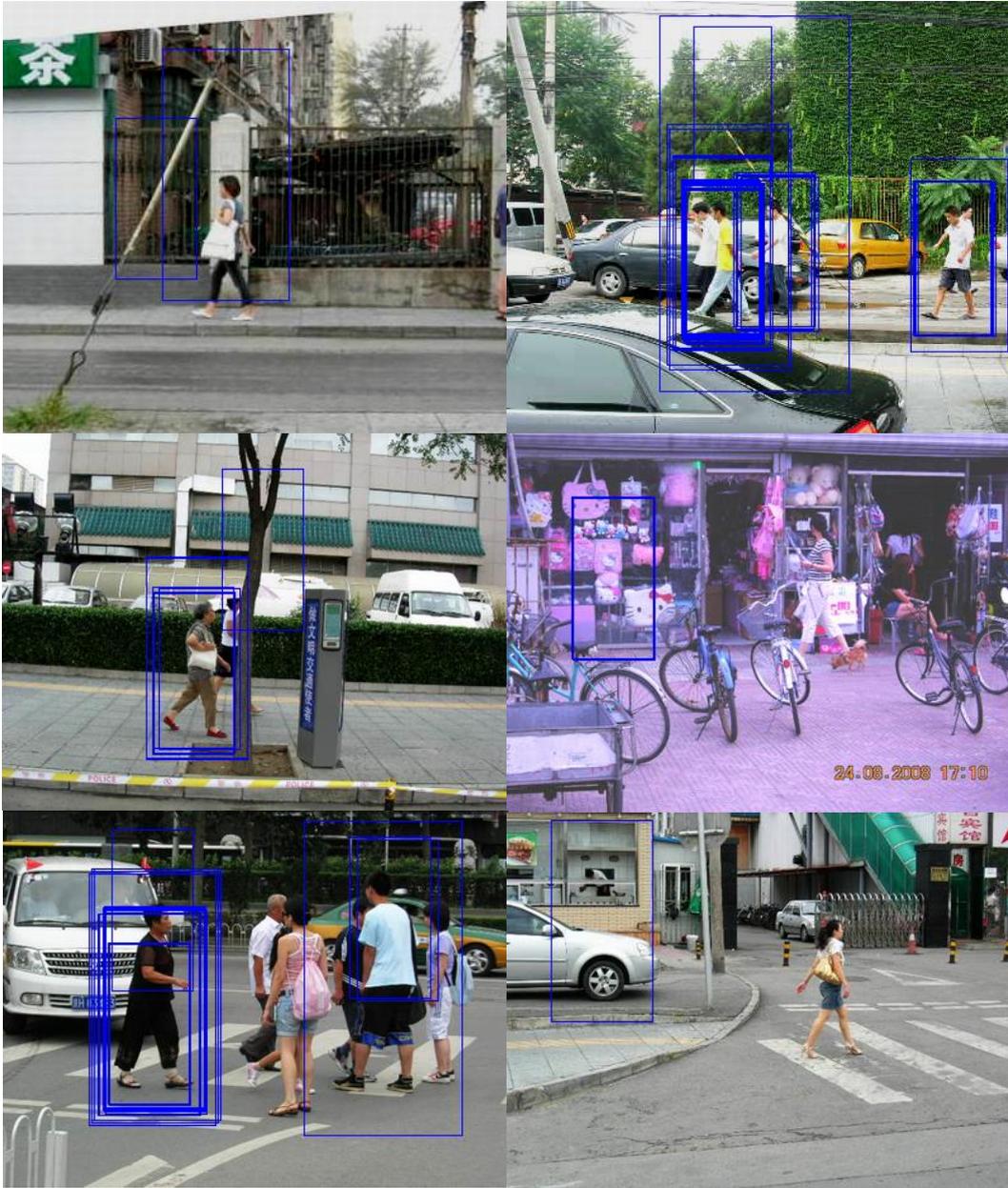


图 4.7 “侧面行人测试集”中具有错检情况的结果

4.3 和其它的领先技术相比较

为了评测本文的方法的性能，本文为此利用 MSO+SVM，MSO+Cascade

AdaBoost, HOG+SVM, Haar-Like+AdaBoost 在两组测试集上进行测试——“INTRA 测试集”和“本文的测试集”。每个测试集上都挑选出 58 幅图片作为测试使用。其中，“INTRA 测试集”为目前国际上默认的公共测试集，可以到它的网站（<http://pascal.inrialpes.fr/data/human/INRIAPerson.tar> 或者 http://www.emt.tugraz.at/~pinz/data/GRAZ_01/）进行下载。而本文的测试集则是包含了不同光照、不同场景、不同姿态的行人数据，关于本文的测试样本集，可以到中科院研究生院的 SDL 实验室的网站上下载（<http://cear.gucas.ac.cn/SDL-HomePage/Dataset.htm>）。

本文为以上 4 种算法分别在两组测试集上绘制了各自的 DET 曲线图（如图 4.8 和图 4.9）。

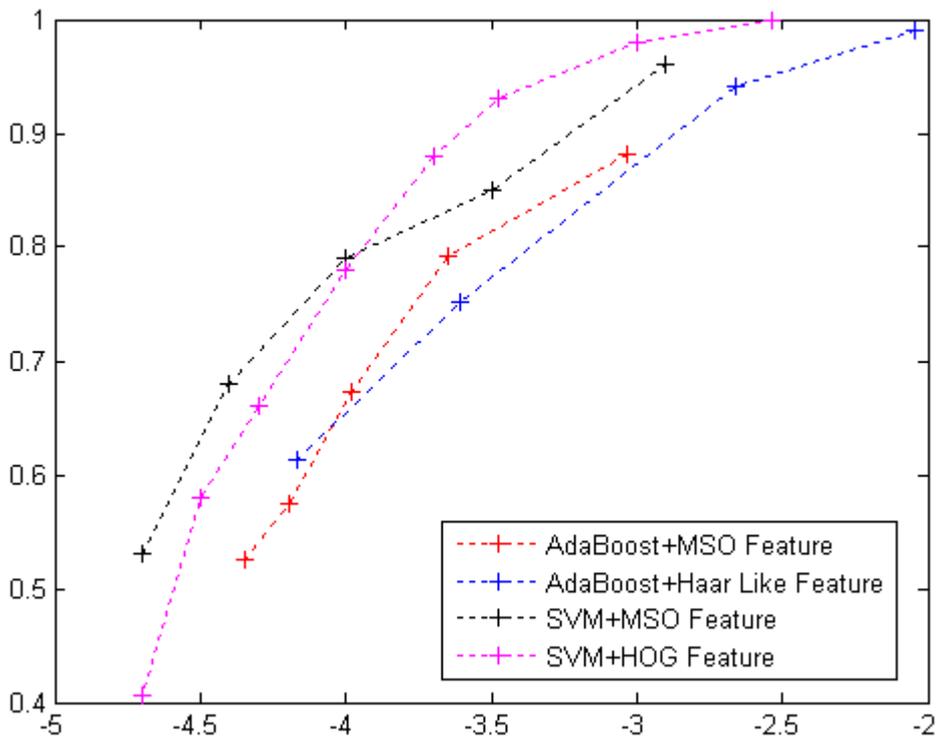


图 4.8 在 INTRA 测试集上的 DET 曲线图

其中横坐标表示的是窗口错误率，每一个坐标点表示 10 的相应幂次方；纵坐标表示的是检测率。

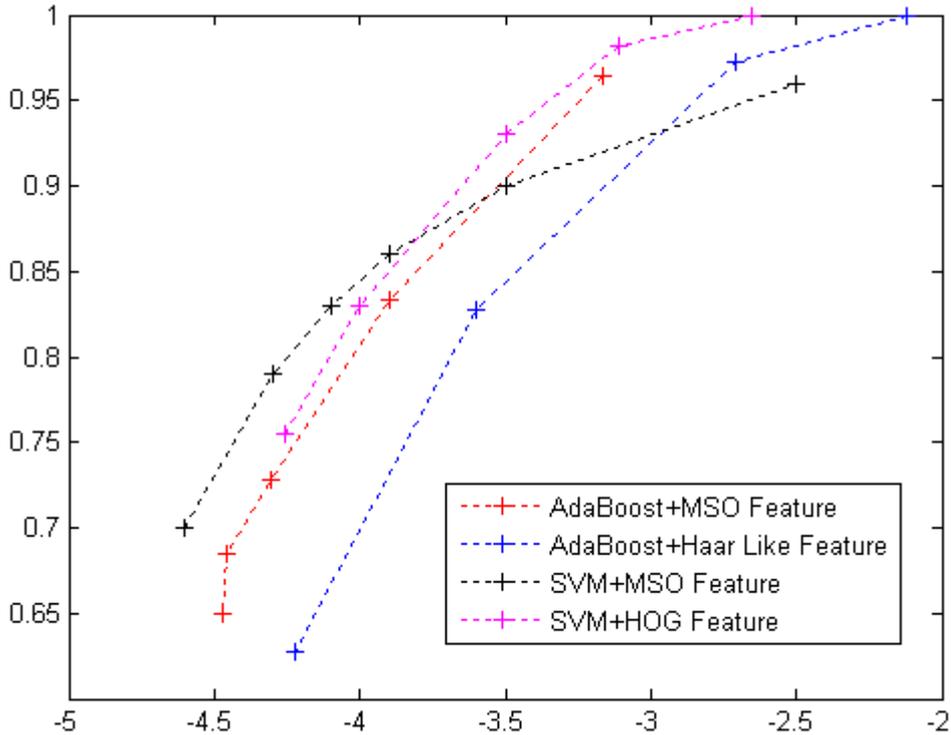


图 4.9 在本文的测试集上的 DET 曲线图

其中横坐标表示的是窗口错误率，每一个坐标点表示 10 的相应幂次方；纵坐标表示的是检测率。

从两个曲线图上本文可以看出，本文的算法和目前比较领先的算法在检测性能上是不相上下的，在不同的要求下，都有着各自的优势。然而，这些算法中，只有本文的“MSO+Cascade AdaBoost”算法和“Haar-Like+Cascade Adaboost”算法具有近于实时的检测效果，而其他的两种算法在时间上都是很耗时的。但是，本文可以从 DET 曲线图上可以发现，本文的算法要优于“Haar-Like+Cascade Adaboost”算法，这便是本文的算法的优越之处。当本文在大小为 640*480 的同一幅图像上进行测试时，MSO+SVM 耗时 1 分多钟，MSO+Cascade AdaBoost 为每秒钟 10 多帧，HOG+SVM 原作者的程序处理 320x240 大小的图片也要耗时 1 秒钟（本文自己模拟的更慢些），Haar-Like+Cascade Adaboost 耗时与 MSO+Cascade AdaBoost 算法相似，可见 MSO+Cascade AdaBoost 在处理速度上有着明显的优势，在满足要求的条件

下,基本上可以达到实时的效果。

4.4 本章小结

在本章中,首先介绍了实验数据集的采集与指定的标准,然后给出了本文算法在 INRIA 数据集上的一些具体的实验结果,通过在 INRIA 测试集和我们自己的测试集上的运行结果的对比,我们可以看出本文所提出的 MSO 特征以及新的机器学习方法的已经达到了国际上的检测水平。

结 论

本文作者将 MSO 特征描述子与相应的机器学习机制相结合，通过 C++ 程序语言实现行人检测算法。在实际测试并与其它领先技术的对比中，可知 MSO 特征是非常有效的。同时本文还使用 MSO 特征对其它物体（如：人脸）进行了训练检测，也得到了很好的效果。

从机器学习的角度来看，基于“组件”的特征描述子（如 Haar-Like 特征），虽然具有很好的“可选性”，但很容易受到物体的不同表现形式所限制；而基于直方图的特征描述子（比如 HOG、SIFT 等），这些特征对于形变都有很强的鲁棒性，然而却没有体现出组合的特性。本文提出的这种 MSO 特征恰好在“可选性”与“不变性”做到了平衡，它既比基于块特征更灵活，又比基于局部直方图的特征更有可选性。但关于 MSO 特征集并不完备性，训练样本数量的不足，检测窗口的精确定位，以及检测速度的提高等问题，都有待于解决，因此本文在今后的研究工作中将对此加以改善。另外，本文的检测算法始终都是对静态图片进行处理的，如果本文将其应用到视频中，加上物体的运动信息，无论是在检测结果上还是运算速度上，性能都应该会得到进一步的提升。

本文作者在对基于学习的行人检测问题进行了较为深入的研究后，总结出本文具有以下 4 贡献：

1. 本文作者受到 Haar-Like 和 HOG 的启发，提出了一种新的特征（多尺度方向特征，Multi-scale orientation feature - MSO），这不仅囊括了其各自的优点，还弥补了它们的不足。这种特征是针对特征区域的形状进行统计的，由全部特征组合起来所形成的特征集反映的是图像在不同尺度上的梯度特征。使用这种特征可以有效地反映整个物体的形状特征。
2. 基于本文的 MSO 特征，提出了与其相适应的弱分类器构造机制---“区域双阈值”划分方法。在 Cascade AdaBoost 训练中，通过这种弱分类器的构造方法，可以有效的提取有用的物体特征。

3. 构建了一个用于训练和测试的图像和视频数据库。该库是在多个实际景点下（包括雨天、阴影、复杂纹理等多种情况），使用数码相机进行拍摄和录制的图片和视频，不仅可以用来测试物体检测的结果，而且还可为背景建模、目标跟踪、视频稳像、轨迹分析等算法的验证提供有效的实验数据，为进一步的研究奠定了良好的基础。
4. 实现了一个近于实时的检测系统。该系统集成了本文所提出的理念，可以在复杂的场景环境下，在较短的时间内表现出很好的效果。

参考文献

- [1] C. Huang, et al. High-performance rotation invariant multiview face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2007, 29(4): 671-686P
- [2] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. Kauai, HI, 2001. Institute of Electrical and Electronics Engineers Computer Society, 511-518P
- [3] J. Jiao, et al. A configurable method for multi-style license plate recognition. *Pattern Recognition*. 2009, 42(3): 358-369P
- [4] C. Papageorgiou and T. Poggio. Trainable pedestrian detection. *IEEE International Conference on Image Processing*. 1999, 4: 35-39P
- [5] D. M. Gavrila and S. Munder. Multi-cue pedestrian detection and tracking from a moving vehicle. *International Journal of Computer Vision*. 2007, 73(1): 41-59P
- [6] C. R. Wren, et al. Pfnder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997, 19(7): 780-785P
- [7] C. Beleznai, et al. Human detection in groups using a fast mean shift procedure. Singapore, 2004. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 349-352P
- [8] T. Haga, et al. Human detection in outdoor scene using spatio-temporal motion analysis. Cambridge, United Kingdom, 2004. Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States, 331-334P
- [9] H.-L. Eng, et al. A bayesian framework for robust human detection and occlusion handling using human shape model. Cambridge, United

- Kingdom,2004.Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States, 257-260P
- [10] H. Elzein, et al. A motion and shape-based pedestrian detection algorithm. IEEE Intelligent Vehicles Symposium, 2003.500-504P
- [11] D. Toth and T. Aach. Detection and recognition of moving objects using statistical motion detection and fourier descriptors. International Conference on Image Analysis and Processing, 2003. 430-435P
- [12] D.-J. Lee, et al. Shape-based human detection for threat assessment. Orlando, FL, United States, 2004. International Society for Optical Engineering, Bellingham, WA 98227-0010, United States, 81-91P
- [13] C. Hou, et al. Multiview pedestrian detection based on vector boosting. Tokyo, Japan,2007. Springer Verlag, Heidelberg, D-69121, Germany, 210-219P
- [14] S. M. Yoon and H. Kim. Real-time multiple people detection using skin color, motion and appearance information. Okayama, Japan, 2004. Institute of Electrical and Electronics Engineers Inc., New York, NY 10016-5997, United States, 331-334P
- [15] L. Jiang, et al. Perceptual-based fusion of IR and visual images for human detection. Hong Kong, China, Hong Kong,2004.Institute of Electrical and Electronics Engineers Inc., New York, NY 10016-5997, United States, 514-517P
- [16] L. Li, et al. Object-Oriented Scale-Adaptive Filtering for human detection from stereo images. Singapore, 2004.Institute of Electrical and Electronics Engineers Inc., New York, NY 10016-5997, United States, 135-140P
- [17] J. Han and B. Bhanu. Detecting moving humans using color and infrared video. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2003:228- 233P
- [18] P. Viola, et al. Detecting pedestrians using patterns of motion and appearance. Nice, France, 2003. Institute of Electrical and Electronics

- Engineers Inc., 734-741P
- [19] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. San Diego, CA, United States, 2005. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 886-893P
- [20] N. Dalal, et al. Human detection using oriented histograms of flow and appearance. Graz, Austria, 2006. Springer Verlag, Heidelberg, D-69121, Germany, 428-441P
- [21] P. Viola and M. Jones. Robust Real-time Object Detection . VANCOUVER, CANADA. 2001
- [22] R. Cutler and L. S. Davis. 2. Robust real-time periodic motion detection, analysis, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000, 22(8): 781-796P
- [23] A. Utsumi and N. Tetsutani. Human Detection using Geometrical Pixel Value Structures. Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. 39P
- [24] H. Sidenbladh. Detecting human motion with support vector machines. Cambridge, United Kingdom, 2004. Institute of Electrical and Electronics Engineers Inc., Piscataway, NJ 08855-1331, United States, 188-191P
- [25] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. San Diego, CA, United States, 2005. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 886-893P
- [27] O. Tuzel, et al. Human detection via classification on riemannian manifolds. Minneapolis, MN, United States, 2007. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States, 427-431P
- [28] Q. Ye, et al. Multi-posture human detection in video frames by motion contour matching. Tokyo, Japan, 2007. Springer Verlag, Heidelberg, D-69121, Germany, 896-904P

- [29] T. Serre, et al. Robust object recognition with cortex-like mechanisms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2007, 29(3): 411-426P
- [30] J. Begard, et al. Real-time human detection in urban scenes: Local descriptors and classifiers selection with AdaBoost-like algorithms. Anchorage, AK, United States, 2008. Institute of Electrical and Electronics Engineers Computer Society, Piscataway, NJ 08855-1331, United States,456-471P
- [31] D. Geronimo, et al. Haar wavelets and edge orientation histograms for on-board pedestrian detection. Girona, Spain,2007.Springer Verlag, Heidelberg, D-69121, Germany, 418-425P
- [32] 郭烈. 基于 Adaboost 算法的行人检测方法.计算机工程. 2008, 34(3): 202-204 页
- [33] 赵春晖. 一种基于三角特征的行人检测算法. 计算机工程与应用. 2008, 44(7): 202-205 页
- [34] 朱文佳. 基于机器学习的行人检测关键技术研究.上海交通大学硕士. 2008: 1-56 页
- [35] 朱谊强. 基于 Adaboost 算法的实时行人检测系统.计算机测量与控制. 2006: 1462-1465 页
- [36] D. Goel and T. Chen.Pedestrian detection using global-local motion patterns. Tokyo, Japan, 2007. Springer Verlag, Heidelberg, D-69121, Germany, 220-229P
- [37] S. Kang, et al. Real-time pedestrian detection using support vector machines. International Journal of Pattern Recognition and Artificial Intelligence. 2003, 17(3): 405-416P

攻读硕士学位期间发表的论文和取得的科研成果

致 谢

这篇论文的研究受到了“中国科学院百人计划”和“国家自然科学基金”的赞助。

感谢导师黄凤岗教授。感谢您两年来对我在学习上的悉心教导和生活上亲切的关怀，您对科学教研事业的热情永远值得我学习。黄老师治学严谨，学识渊博。从论文的选题、资料的收集到论文的撰写、结构的安排、修改直至最后的定稿，处处浸透了导师的心血。黄老师深厚的理论功底、兢兢业业的工作精神和耐心细致的从教态度是值得我终生学习的。在此特别向导师黄凤岗教授表示衷心的感谢和崇高的敬意。

感谢焦建彬老师（北京中科院研究生院教授）和叶齐祥老师（北京中科院研究生院副教授）。感谢两位老师一年来在课题研究中对我的指导、帮助和对学习生活的关心。在课题研究和论文完成过程中，两位老师给予了充分的信任，并给出指导建议和解决思路。两位老师那种忘我的工作作风，严谨的治学态度和平易近人的处世方式也非常值得我学习。感谢与我一同合作的同学们，在你们的协助下，我顺利的完成课题的设计和论文的撰写。

感谢 603 教研室的李智慧老师和实验室的同学们，以及在中科院研究生院一同学习的朋友们，感谢您们在学习和生活上给我的关心与帮助，在此表示同样真挚的谢意！

感谢我的父母、家人和女友，家是出门在外的孩子永远的港湾。只有在学习工作中不断奋发向上，努力争取更好的成绩，才能回报您们的爱。

最后，祝福母校哈尔滨工程大学明天更美好！