

# 基于简单局部特征学习的物体检测方法



论文作者签名: 陈伟东

指导教师签名: 李海勇

论文评阅人 1: 徐正全教授 武汉大学

评阅人 2: 马利庄教授 上海交通大学

评阅人 3: 魏宝刚教授 浙江大学

评阅人 4: (隐名评阅)

评阅人 5: (隐名评阅)

答辩委员会主席: 汪国昭教授 浙江大学

委员 1: 董金祥教授 浙江大学

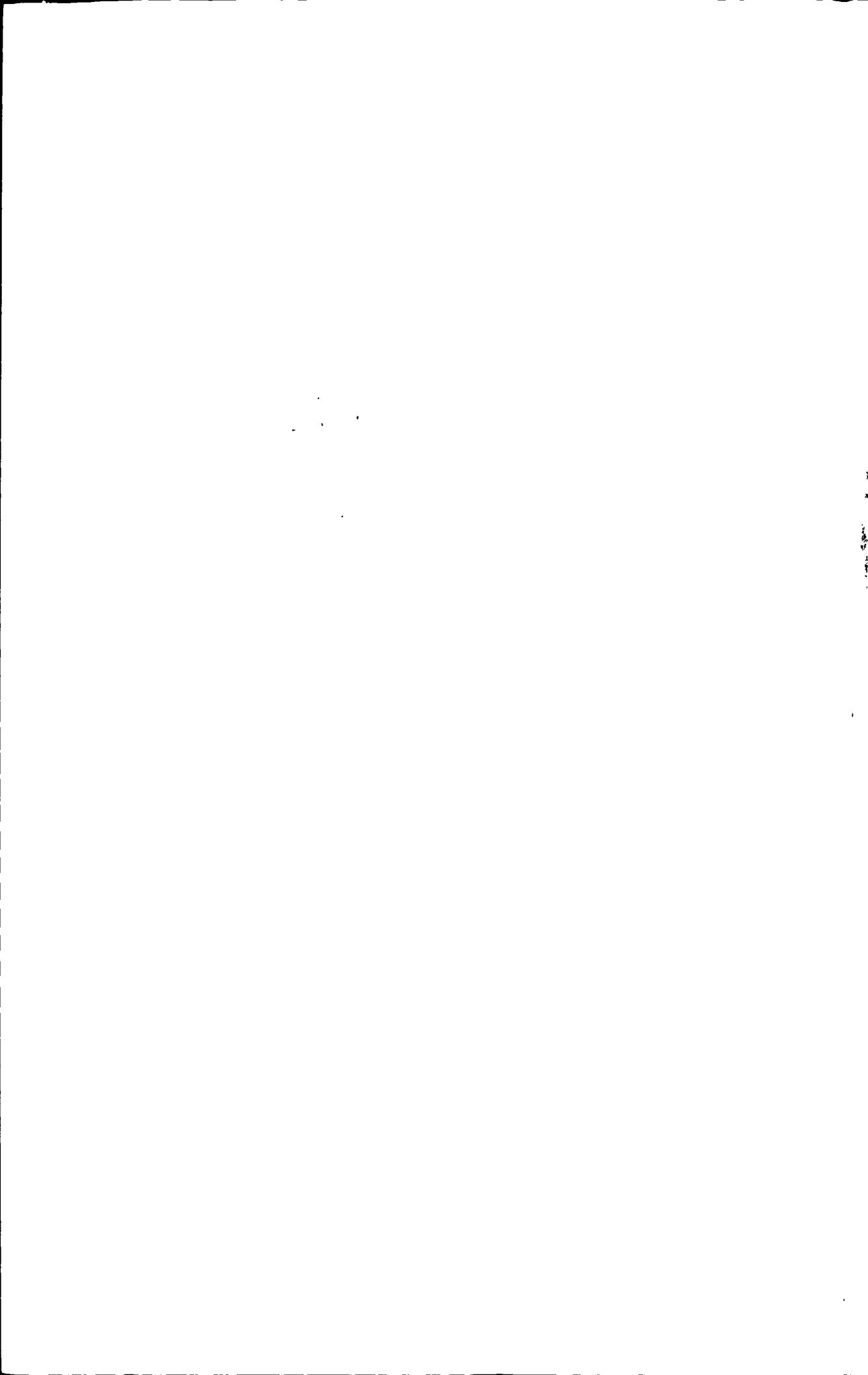
委员 2: 陈刚教授 浙江大学

委员 3: 魏宝刚教授 浙江大学

委员 4: 王丽萍教授 浙江工业大学

委员 5: 童若锋教授 浙江大学

答辩日期: 2010年6月8日



Y1853379



## Learning Simple Local Features for Object Detection



Author's signature: 

Supervisor's signature: 

### External Reviewers:

Prof. Zhengquan XU, Wuhan University

Prof. Lizhuang MA, Shanghai Jiaotong University

Prof. Baogang WEI, Zhejiang University

(Anonymity)

(Anonymity)

### Examining Committee Chairperson:

Prof. Guozhao WANG, Zhejiang University

### Examining Committee Members:

Prof. Jinxiang DONG, Zhejiang University

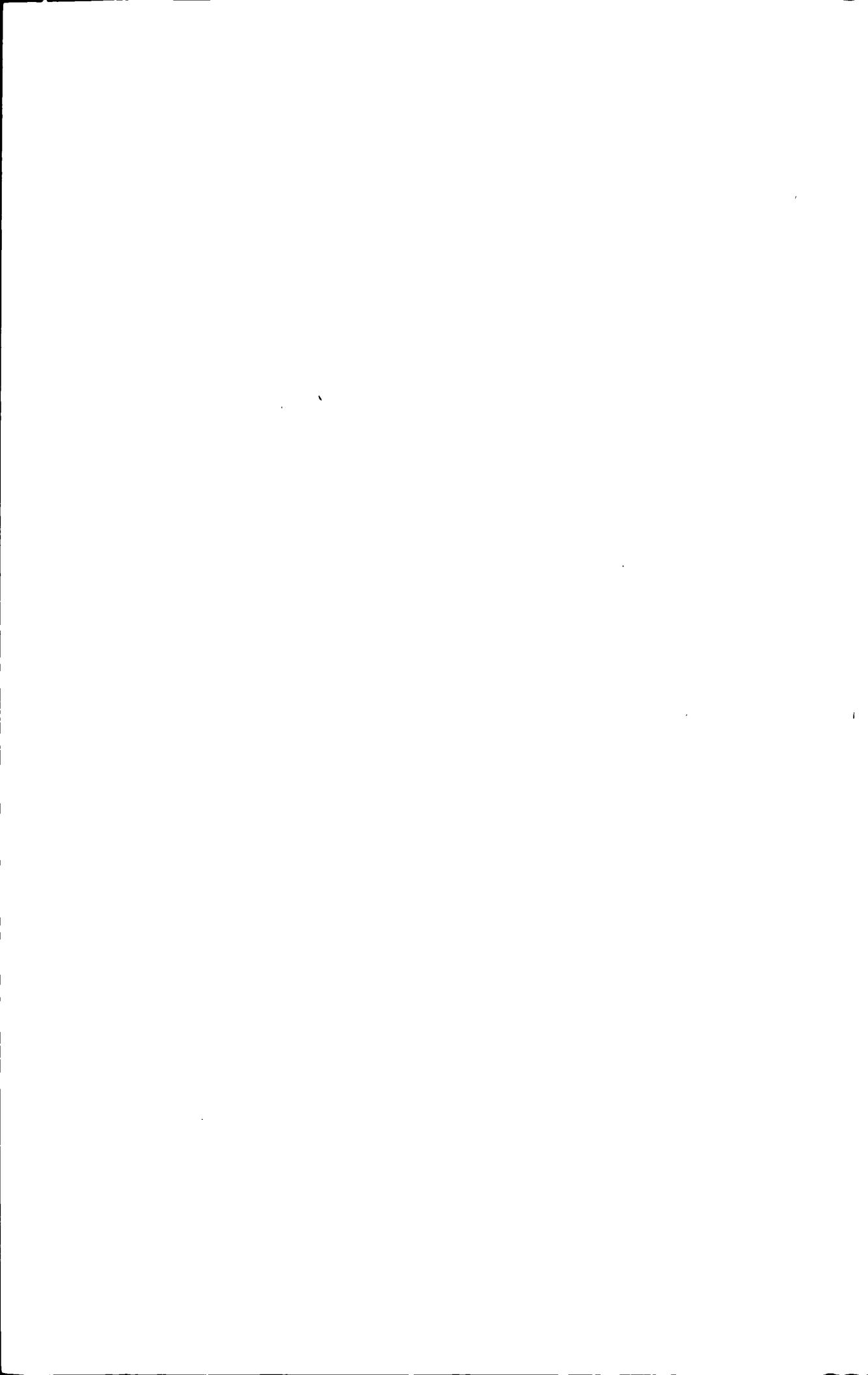
Prof. Gang CHEN, Zhejiang University

Prof. Baogang WEI, Zhejiang University

Prof. Liping WANG, Zhejiang Univ. of Technology

Prof. Ruofeng TONG, Zhejiang University

Date of oral defence: June 8, 2010



## 浙江大学研究生学位论文独创性声明

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

学位论文作者签名： 签字日期： 年 月 日

## 学位论文版权使用授权书

本学位论文作者完全了解 浙江大学 有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权 浙江大学 可以将学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名： 导师签名：

签字日期： 年 月 日 签字日期： 年 月 日



## 摘要

本文研究的物体检测是计算机视觉领域一个非常具有挑战性的课题，即从真实世界的图像中识别某类物体并在混乱的图像背景上定位该物体。这是人眼视觉系统的核心能力之一，但计算机视觉系统目前还远未达到接近该层级的效果。该课题的主要难点在于寻找一种有效的物体表示方式，一方面必须容忍同类物体不同个体之间的类内变化，另一方面又能够区分不同类别物体的类间变化，同时还需要鲁棒的处理混乱背景、光照变化以及部分遮挡等问题。

本文回顾了物体检测方法的研究现状，分析和总结了常用的物体模型和特征，从简单局部特征和基于特征的学习算法两个层面，研究了如何构建准确的、鲁棒的物体模型。主要的研究成果和创新点有如下两个方面：

在特征层面，本文提出了一种基于表形的局部特征——分散矩形特征，和一种基于形状的局部特征——霍夫变换线段（组）。

分散矩形特征是类 Haar 特征的一种变体。和类 Haar 特征一样，它也是一种基于矩形模板的简单特征，但模板内的矩形不需要在水平或垂直方向上相邻对齐，因而不仅能够表示任意的方位信息，而且在几何关系上能够表示错切、分离和重叠的形状信息，使得分散矩形特征能够更加灵活的表示物体部分，具有更好的表示能力。同时，可以利用积分图，在常数时间内计算图像窗口上任意位置任意尺寸的分散矩形特征值，解决了由于缺少相邻对齐约束后带来的大量特征的计算量问题。另外，采用构造方法，严格的证明了任意一个非退化的分散矩形特征与满足一定几何约束关系的多个类 Haar 特征之间的等价性。该等价性说明一个非退化的分散矩形特征所包含的物体部分的信息等效于多个类 Haar 特征所包含的信息的综合，因此该特征更加鲁棒。在 MIT 和 CMU 人脸测试集上的对比实验结果显示，基于分散矩形特征的分类器性能优于类 Haar 特征分类器。

霍夫变换线段（组）是受线条画启发而提出的一种简单的形状特征。该特征不是由其两个端点表示，而是通过其法向与横轴的夹角角度、到坐标原点的垂直距离、线段中心到垂线的距离以及线段长度所组成的四元组表示。该四元组表示形式不仅能够唯一的确定任意线段，而且能够很方便的处理缩放、旋转和平移变换。给定以物体中心为原点的局部坐标系，霍夫变换线段与物体中心之间的几何关系被隐含于四元组中，成为一个

紧凑的隐式形状模型，后者在物体检测研究中已经被证明为一个有效的模型。相连的霍夫变换线段构成霍夫变换线段组，通过引入线段之间的局部几何信息进一步增强特征的区分能力。霍夫变换线段（组）之间的相似度由四元组表示空间中的一个加权欧氏距离衡量。通过调整四元组元素相应的权重，该距离能较好的容忍不可靠的边缘检测所引起的噪音问题，从而能够用来从训练样本中选取有区分能力的霍夫变换线段组，建立该类别物体码表。通过形状匹配实现的物体检测实验表明形状确实是定义物体类别的一个重要特征，也能胜任物体检测任务。

在学习算法层面，本文提出了一种 AdaBoost 算法变体——双阈值 AdaBoost 算法。该变体与原来的算法使用同样的框架，但使用了具有两个阈值的弱假设，其核心思想是选择分类性能更好的弱假设可以使学习到的强假设更加鲁棒和高效。双阈值弱假设在不退化为单阈值弱假设的情况下，通过对样本在特征值空间更为精细的划分，确保分类错误比单阈值弱假设更小。为了快速确定两个阈值的最优值，本文相应的提出了一个双阈值弱学习算法。该弱学习算法将确定阈值问题转换为寻找最大和连续子序列问题，后者可以采用线性的动态规划算法解决。对类 Haar 特征和分散矩形特征的学习结果表明，在同样的训练性能要求下，该变体收敛速度更快，生成的检测器具有更少的层次，使用更少的特征。在 MIT 和 CMU 人脸测试集上的检测实验结果显示，该变体学习的分类器性能总体上优于原来的 AdaBoost 算法。

除以上成果外，本文在人脸检测实验过程中，构建了一个分辨率为  $19 \times 19$  像素的正面人脸训练集，提出了一个针对 MIT 和 CMU 人脸测试集的检测标准。该标准利用测试集所提供的关于人脸五官的真实位置信息，定义了最小人脸矩形和正确检测可能的最大分布区域，可以作为一个客观严格的检测标准。

**关键词** 物体检测；计算机视觉；类内变化；类间变化；局部特征；类 Haar 特征；模板；积分图；霍夫变换；隐式形状模型；边缘检测；弱假设；强假设；最大和连续子序列；动态规划；检测标准；

## Abstract

This thesis is concerned with object detection, one of the most challenging problems in Computer Vision, that is recognizing objects of some category, and localizing them in cluttered real-world images. This capability is one core competency of the human visual system. Yet, computer vision systems are still far from reaching a comparable level of performance. The main difficulty lies in finding an effective object representation that is tolerant of intra-class variations in appearance and geometry, while distinctive to inter-class variations, and meanwhile robust to image clutters, illumination changes, partial occlusion, etc..

The thesis reviews the object detection approaches with an analysis of object models and features of common use. Based on this investigation, the research on how to build an accurate and robust object model is carried out on two levels: simple local feature and learning algorithm.

On the level of local feature, an appearance-base feature named Scattered Rectangle Feature (SRF) and a shape-based feature called Hough Transformed Line Segment (HTLS) are proposed.

SRF is a variant of Haar-like feature (HLF). It is also template-based, while the rectangles in the template are not required to be adjacent and aligned horizontally or vertically. Therefore the rectangles can not only explore more orientation cues, but also encode misaligned, detached and overlapped shape information, resulting in a more free and distinctive feature. Meanwhile, since SRF takes the rectangle template as HLF does, it can make full use of integral image, and be computed in constant time no matter its scale or location. Moreover, the thesis proves by construction that any non-degenerated SRF is equivalent to several HLFs that are constrained by some geometric relationship. The cue of the object part represented by such a SRF is therefore equal to the combined cues of several HLFs, which makes the detector based on SRF more robust. The experiments on the MIT and CMU face test set show that the detector based on SRF outperforms that based on HLF.

HTLS is a simple shape feature, motivated by the fact that a line drawing conveys most of the information. HTLS is represented by a quadruple of the inclination of the normal, the distance from the origin, the shift distance of the center from the foot and the length, rather than the end points. The quadruple does not only uniquely define any HTLS but also handle the rotation, translation and scaling conveniently. Given a local coordinate system on the

object centroid, the quadruple incorporates the geometric relationship between the HTLS and the centroid implicitly, resulting in a compact Implicit Shape Model, which has been proved effective in object detection tasks. To even enhance the distinctness of HTLS, connectiveness is employed to form HTLS groups. A weighted Euclidean distance in HTLS space to measure the similarity between the HTLS (groups) is introduced accordingly, which can handle well partial match and other noises caused by the unreliable edge detection, after a careful pick of the weights. By using the distance, distinctive HTLS (groups) from the training set are collected into a codebook. The experiments on motorbike and cow category show that shape cues are not only important but also competent for object detection.

On the level of learning algorithm, a variant of AdaBoost named 2-threshold AdaBoost is proposed. It shares the same framework with AdaBoost except the 2-threshold weak hypothesis and the 2-threshold weak learner. The motivation is that selecting better weak hypothesis makes the final strong one more robust and efficient. The 2-threshold weak hypothesis guarantees a smaller (or the same if degenerated) classification error, since the extra threshold enables a finer splits over the feature values. The selection of the optimal values for the two thresholds is transformed by the weak learner into a maximum-sum consecutive subsequence problem, which can be solved by Dynamic Programming. Applying the learning algorithm on HLF and SRF, the experiments demonstrate that the variant converges faster, with fewer stages and features in the cascade, given the same training goal and in general the cascade learned by the variant outperforms that by AdaBoost on MIT and CMU face test set.

Besides those results, a face training set of 19\*19 resolution is built for the experiments in the thesis. An objective detection criterion is also introduced for MIT and CMU face test set, which relies on the minimal face rectangle extracted from the ground truth information of the test set.

**Keywords** Object Detection; Computer Vision; Intra-Class Variation; Inter-Class Variation; Local Feature; Haar-Like Feature; Template; Integral Image; Hough Transform; Implicit Shape Model; Edge Detection; Weak Hypothesis; Strong Hypothesis; Maximum-Sum Consecutive Subsequence; Dynamic Programming; Detection Criterion;

## 目 录

摘要 .....	I
Abstract.....	III
第1章 绪论 .....	1
1.1 本文研究的问题 .....	2
1.2 难点和性能指标 .....	4
1.3 物体检测方法现状研究 .....	8
1.3.1 基于三维模型的物体检测方法 .....	8
1.3.2 基于全局特征的物体检测方法 .....	9
1.3.3 基于局部特征的物体检测方法 .....	10
1.3.4 研究现状小结 .....	19
1.4 本文研究的内容 .....	20
1.5 本文的组织结构 .....	21
第2章 基于分散矩形特征的物体检测 .....	23
2.1 局部特征分析 .....	23
2.2 分散矩形特征 .....	25
2.2.1 类 Haar 特征 (Haar-Like Feature, HLF) .....	25
2.2.2 分散矩形特征 (Scattered Rectangle Feature, SRF) .....	27
2.2.3 HLF 和 SRF 对比 .....	28
2.3 实验结果和分析 .....	29
2.3.1 人脸训练集构造 .....	29
2.3.2 实验设置 .....	31
2.3.3 训练结果 .....	31
2.3.4 检测标准 .....	32
2.3.5 检测结果 .....	33
2.4 本章小结 .....	36
第3章 基于双阈值 AdaBoost 算法的物体检测 .....	37
3.1 AdaBoost 算法 .....	37
3.2 双阈值弱假设 .....	40

3.3 双阈值弱学习算法 .....	42
3.3.1 符号表示 .....	43
3.3.2 双阈值弱学习算法 .....	43
3.4 讨论 .....	46
3.5 实验结果与分析 .....	47
3.5.1 实验设置和检测标准 .....	47
3.5.2 训练结果 .....	48
3.5.3 检测结果 .....	49
3.6 本章小节 .....	53
第 4 章 双阈值 AdaBoost 选取的分散矩形特征 .....	55
4.1 分散矩形特征与类 Haar 特征的等价性 .....	55
4.2 实验结果与分析 .....	60
4.2.1 实验设置 .....	61
4.2.2 训练结果 .....	61
4.2.3 检测结果 .....	61
4.3 本章小结 .....	66
第 5 章 基于霍夫变换线段（组）的物体检测 .....	69
5.1 方法框架 .....	70
5.2 霍夫变换线段（Hough Transformed Line Segment, HTLS）和距离 .....	72
5.2.1 霍夫变换线段 .....	72
5.2.2 霍夫变换线段的提取 .....	73
5.2.3 霍夫变换线段距离 .....	75
5.3 霍夫变换线段码表 .....	76
5.4 霍夫变换线段组 .....	76
5.5 实验结果与分析 .....	78
5.5.1 实验设置 .....	78
5.5.2 训练结果 .....	79
5.5.3 检测标准 .....	80
5.5.4 检测结果 .....	80

---

5.6 本章小结 .....	81
第6章 结论 .....	83
6.1 全文工作总结 .....	83
6.2 未来研究展望 .....	84
参考文献 .....	87
攻读博士学位期间主要的研究成果 .....	97
致谢 .....	99

## 图目录

图 1.1 物体在图像中呈现的各种变化 .....	5
图 1.2 典型的 ROC 曲线 .....	7
图 1.3 类 Haar 特征的扩展和基于模板的直方图特征 .....	12
图 1.4 $2 \times 2$ 网格的 SIFT 计算过程示意图 <sup>[61]</sup> .....	14
图 1.5 图像块样例和对应的 Spin Image .....	14
图 2.1 特征抽取过程中的三个通用组件 .....	24
图 2.2 类 Haar 特征的几个例子 .....	25
图 2.3 积分图的生成和使用: (a) 积分图一次性扫描生成示意图; (b) 利用积分图, 矩形 $ABCD$ 内的像素灰度值之和可以计算为 $I(A) + I(C) - (I(B) + I(D))$ .....	26
图 2.4 分散矩形特征的几个例子: (a) 既是分散矩形特征, 也是类 Haar 特征; (b) 错 切的分散矩形特征; (c) 分离的分散矩形特征; (d) 重叠的分散矩形特征 .....	27
图 2.5 训练集人脸裁剪规则示意图 .....	30
图 2.6 垂直向上正面人脸训练集人脸样例图片 .....	30
图 2.7 AdaBoost 算法在学习过程中选取的前两个类 Haar 特征 (上) 和分散矩形特征 (下)。红色 (或者深色) 表示第一个特征, 绿色 (或者浅色) 表示第二个特征 .....	31
图 2.8 最小人脸矩形和最大正确检测分布区域 .....	33
图 2.9 使用类 Haar 特征(a)和分散矩形特征(b)的检测器的检测结果 .....	34
图 2.10 类 Haar 特征 (HLF) 和分散矩形特征 (SRF) 的检测器的 ROC 曲线 .....	35
图 3.1 AdaBoost 算法框架 .....	38
图 3.2 单阈值二值弱假设函数曲线 .....	39
图 3.3 单阈值的弱学习算法 .....	39
图 3.4 一个用于说明弱假设及其学习算法的简单例子。示例包含了 17 个已经按照某 个特征值从小到大排列的样本, 其中圆形表示正样本, 方形表示负样本, 其面 积代表样本权重 .....	41
图 3.5 双阈值二值弱假设函数曲线 .....	41
图 3.6 双阈值的弱学习算法 .....	45
图 3.7 双阈值弱学习算法的学习过程 .....	45

图 3.8 采用人工控制的训练方式, 分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器 (ACH 和 ZCH) 的 ROC 曲线 .....	50
图 3.9 采用性能驱动的训练方式, 分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器 (AC 和 ZC) 的 ROC 曲线 .....	51
图 3.10 单阈值(a)和双阈值(b)AdaBoost 算法训练的类 Haar 特征检测器 (AC 和 ZC) 的检测结果 .....	52
图 4.1 正区域和负区域在水平或者垂直方向上对齐但不相邻的分散矩形特征(a)等价于两个类 Haar 特征(b-c)的重叠组合(d).....	56
图 4.2 正区域和负区域共享一个顶点的分散矩形特征(a)等价于两个垂直和水平方向共享同一个区域的类 Haar 特征(b-c)的重叠组合(d).....	57
图 4.3 错切的分散矩形特征关于对称中心的分解方式 .....	58
图 4.4 分离的分散矩形特征关于对称中心的分解方式 .....	59
图 4.5 重叠的分散矩形特征关于对称中心的分解方式 .....	60
图 4.6 采用人工控制的训练方式, 分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器 (HLF1Cascade 和 HLF2Cascade), 与对分散矩形特征进行训练得到的检测器 (SRF1Cascade 和 SRF2Cascade) 的 ROC 曲线 .....	62
图 4.7 采用人工控制的训练方式, 分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器的 9 层级联器 (HLF1Cascade_9 和 HLF2Cascade_9) 与 10 层级联器 (HLF1Cascade_10 和 HLF2Cascade_10) 的 ROC 曲线 .....	63
图 4.8 单阈值(a)和双阈值(b)AdaBoost 算法学习的类 Haar 特征检测器 (HLF1Cascade 和 HLF2Cascade) 的检测结果 .....	64
图 4.9 单阈值(a)双阈值(b)AdaBoost 算法学习的分散矩形特征检测器 (SRF1Cascade 和 SRF2Cascade) 的检测结果 .....	65
图 5.1 Attneave 的猫 .....	69
图 5.2 基于霍夫变换线段的物体检测方法框架 .....	71
图 5.3 霍夫变换线段 .....	73
图 5.4 Douglas-Peucker 多边线简化算法 .....	74
图 5.5 码表中的霍夫变换线段组特征 .....	79
图 5.6 Twenty-percent 码表的检测结果 .....	82

## 表目录

表 2.1 常见几种局部区域特征的特点 .....	23
表 2.2 类 Haar 特征和分散矩形特征对比总结 .....	28
表 2.3 三张测试图像（见图 2.9）上的误检测数量 .....	35
表 2.4 三张测试图像（见图 2.9）上的正确检测数量 .....	35
表 3.1 AC 和 ZC 的训练结果 .....	49
表 3.2 三张测试图像（见图 3.10）上的误检测数量 .....	51
表 3.3 三张测试图像（见图 3.10）上的正确检测数量 .....	51
表 4.1 分散矩形特征分解方式和类 Haar 特征的等价关系 .....	60
表 4.2 三张测试图像（见图 4.8 和图 4.9）上的误检测数量 .....	66
表 4.3 三张测试图像（见图 4.8 和图 4.9）上的正确检测数量 .....	66
表 5.2 码表的训练结果 .....	79
表 5.3 码表的检测性能 .....	81

## 第1章 绪论

自从 Marr 提出著名的三层计算视觉（Computational Vision）模型<sup>[1]</sup>对人脑视觉系统进行理解以来，计算机视觉经历了几十年的快速发展。然而人体视觉的基本能力，如鲁棒的确定常见物体（人、汽车、椅子等），准确的分类场景（沙滩、树林、办公室等），识别简单的活动模式（跳舞、划船、接吻）等，仍然是现今计算机视觉系统所不能完全具备的。另一方面，随着各种视频和图像设备的普及、互联网技术的持续发展，大量的图像信息充斥在我们生活中（Facebook 每月有 25 亿张图片上传<sup>1</sup>，看完 Youtube 上的视频需要超过 600 年的时间<sup>2</sup>）。各种应用领域，包括军事、游戏、医疗、人机交互、机器人、安全监控、交通等，都迫切需要拥有这些能力的计算机视觉技术。

物体检测作为计算机视觉研究领域一个极富有挑战性的课题，与计算机视觉相关的其他问题，包括图像分割、目标跟踪、图像理解等，密切相关，且扮演着重要角色。图像分割，作为图像处理与计算机视觉基本问题之一，从后者的角度来看，可以视为前景和背景，或者具有某些相同性质的图像区域块的检测问题；目标跟踪可以看成是在空间和时间两个维度上的物体检测问题；而基于内容的图像提取、场景分类等图像理解技术则需要采用物体检测技术，然后在此基础上进行更高层次的内容分析。由此可见，物体检测问题是计算机视觉研究的一个重要问题，对该技术的研究可以全面的推动计算机视觉技术的发展。

也正因为如此，物体检测技术成为近年来计算机视觉研究的一个热点，出现了一系列针对不同设备获取的图像信息，对不同物体进行检测的技术，比如飞机航拍或卫星图像中道路的检测<sup>[2]</sup>，车载摄像机图像序列中的车辆行人检测<sup>[3]</sup>，CT 图像中的肿瘤检测<sup>[4]</sup>等等。尽管如此，选取怎样的特征，构建怎样的模型，来准确的、鲁棒的描述目标物体，使得在外部环境变化条件下物体检测依然能够可靠的区分类内变化和类间变化，一直困扰着研究者。出现以上情况有两方面的原因：1) 目前对人眼视觉系统的研究还停留在人脑前期的视觉处理，如何从前期获取的简单的特征转换到后期对复杂的真实世界理解的过程还并不清楚<sup>[5]</sup>，因而不能为该问题提供生物学方面的指导；2) 物体检测作为少数几个突破 Marr 视觉理论框架的技术<sup>[6]</sup>，已经抛弃了在检测之前的感知组合（如前景、背

<sup>1</sup> [http://wiki.answers.com/Q/How\\_many\\_pictures\\_are\\_uploaded\\_onto\\_facebook\\_every\\_second](http://wiki.answers.com/Q/How_many_pictures_are_uploaded_onto_facebook_every_second)  
<sup>2</sup> [http://wiki.answers.com/Q/How\\_many\\_videos\\_are\\_there\\_on\\_YouTube](http://wiki.answers.com/Q/How_many_videos_are_there_on_YouTube)

景)和三维模型的处理方法<sup>[1]</sup>。特征通常是基于传统的图像处理技术，在缺乏先验知识的情况下从图像中抽取的，带有一定的尝试性。因此，针对该问题的研究，不仅可以促进物体检测技术的进步，而且还有可能对人脑视觉研究以及传统的图像处理和分析技术提供某些借鉴经验和辅助作用。

总而言之，物体检测作为计算机视觉研究中的一个重要课题，不仅具有广泛的应用需求和前景，对于解决计算机视觉领域的其它问题也有重大研究意义。与此同时，物体检测也可以为传统的图像处理和分析技术（如图像增强、图像压缩等）提供辅助作用。此外，模仿人脑视觉系统、带有实验性的物体检测研究对于人脑视觉的后期理解过程也有一定的借鉴意义。

## 1.1 本文研究的问题

从1965年Roberts<sup>[7]</sup>对积木世界进行简单物体检测，到Marr<sup>[1]</sup>提出自底向上的计算机视觉研究基本框架，再到最近将物体检测与图像分割、目标跟踪、场景分类等问题相结合<sup>[8-11]</sup>的发展历程中，物体检测的内涵、方法和外延也起了多方面的变化。特别是最近二十年，随着基于表形方法的流行，针对各种不同的应用需求，大量的特征和模型被相应提出，使得物体检测的研究内容变得非常丰富，与计算机视觉其它问题关系也更为密切。本节从一个物体检测系统的输入、对象、方法和效果这几个角度，阐述本文要研究的问题。

物体检测技术可以应用于静态图像或视频(图像序列)。我们将针对前者的研究称为静态的物体检测；针对后者的研究称为动态的物体检测。静态的物体检测利用的仅仅是图像空间的信息，而动态的物体检测还利用了帧与帧之间的时间信息。显然，静态的物体检测技术也完全可以应用于视频，逐帧的进行静态物体检测。但该做法忽略了视频帧之间的时间连续性，导致检测效率和效果都不如动态的物体检测方法。后者通常将静态的物体检测方法和卡曼滤波(Kalman Filter)、粒子滤波(Particle Filter)等目标跟踪技术相结合，获得物体在空间和时间维度上的状态变化信息。本文主要讨论静态的物体检测，且输入是灰度图像。

根据认知心理学类别层次结构<sup>[12]</sup>，本文研究的物体检测对象主要是基本层次类别(Basic-Level Category)的物体。该类别是大部分人对一种物体通用的视觉知识，一个经

常被引用的例子<sup>[13]</sup>是，当人们被问及一张鸟的图片时，他们很可能回答：鸟；除非这张图片给出的是鸵鸟或者企鹅，这种情况人们可能不会回答鸟，而会给出更具体的命名。其他基本层次类别的物体还包括人脸、人体、摩托车、汽车、猫、牛等。基本层次类别是视觉类别，相同类别的物体具有相同的视觉特征。而比其更高一层的超平常类别 (Superordinate-Level Category) 则是一种语义领域的类别，比如动物、食物等。对该层次类别的物体进行检测通常需要更高层次的抽象和知识，而不仅仅是视觉上的观测结果，因此超出了本文研究的范围。在类别层次结构的另一端，比基本层次类别更具体的是次平常类别 (Subordinate-Level Category)，比如人脸可以再分为男性人脸和女性人脸，甚至某个具体的人脸。该层次物体共享相同的视觉特征，但这些特征与同一基本层次的其他次层次类别物体具有不同的空间分布。因此该层次与专业知识存在紧密联系，本文暂不对该层次的物体检测进行具体研究。但次平常类别和基本层次类别这两个层次密切相关，具有一些共同的研究问题，基本层次类别的物体检测技术可作为次平常类别物体检测的预处理步骤（比如先进行人脸识别，然后再确认是否为某个个体），因此本文会引用和分析一些相关的次平常类别的物体检测研究工作。

根据所使用的特征，目前常见的物体检测方法大致分为两类，基于全局特征的物体检测方法和基于局部特征的物体检测方法。简单的说，特征就是一种用来描述整张图像或者其中某个区域的图像模式 (Image Pattern)。全局特征是对整个图像的描述，如颜色直方图；而局部特征则用来描述一个局部区域，通常还包含该局部区域的相对位置信息，如一个图像块。因为局部特征能够更好的处理遮挡、复杂光照和混乱背景等恶劣情况，因此基于局部特征的方法是目前物体检测研究的一个热点。单个局部特征只能描述物体的一部分，因此必须按照某种关系将局部特征组合，以准确、鲁棒的描述整个物体。根据组合关系的来源，组合方法可以粗略的分为基于知识的方法和基于学习的方法。前者是基于人对物体的先验知识，比如正面的人脸轮廓总是某个大致对称的形状，五官之间大概满足一个怎样的分布，它们各自的形状、大小、颜色关系等等，构建一个描述物体的特征模型。该方法的主要问题是将人的知识转换为定义良好的特征；另一个问题是不能有效的处理各种变化对物体表形和形状产生的影响，如物体姿态的变化。而后者将知识到特征表示的转换过程交给学习算法处理，通过对大量样本的学习自动选取有效的特征建立物体模型，不仅节省了人工成本，而且还有可能发现未被人所掌握的知识。本文将研究基于简单特征学习的物体检测技术。

根据所获取的物体信息，物体检测效果可以分为三个层次。最基本的检测效果只是说明图像中是否存在某类物体。该层次的物体检测方法通常偏向于使用全局的特征，适用于图像分类（Image Categorization）或者图像提取（Image Retrieval）等不需要知道物体额外信息的应用。第二层次的检测不仅能给出图像中是否存在该物体，还能确认这些物体在图像中位置信息。本文所研究的方法，以及目前绝大多数研究方法，都属于该层次的物体检测技术。第三层次则更进一步，不仅给出了物体的位置信息，而且还能确定物体的大致轮廓或者物体各部分的位置信息<sup>[8, 11, 14]</sup>。属于该层次的物体检测方法可以根据检测结果，将图像按照物体和非物体的边界进行图像分割。

综上所述，本文研究的物体检测问题是：基于局部的简单特征，通过学习建立物体模型或者分类器，然后对静态图像中的特定物体类别进行检测，确认图像中是否存在该类物体，并标记它们在图像中的位置。这也是近年来大多数物体检测方法所研究的问题。

注意，以上关于物体检测问题的描述隐式的将物体检测问题划分为两个阶段。1) 学习阶段。通过对大量样本图像（可能包含物体，即正样本；也可能不包含物体，即负样本，还可能提供其它必要的输入信息）的学习，获取能够表示这类物体的有效特征，构建合适的物体模型或者分类器；2) 检测阶段。对给定图像上所有可能出现物体的区域应用学习阶段生成的物体检测器，根据检测器的输出对这些区域进行标记（物体还是非物体）。该过程通常涉及将检测器中的特征与被检测区域内的特征按照某个测量标准进行比较或者匹配，这些结果通过和检测器结构相一致的方式综合，然后通过和一个阈值的比较，输出一个二值化的结果，表示被检测区域是否为物体。因此，本文研究的核心问题就是选取怎样的特征，构建怎样的模型或者分类器，来准确、鲁棒的表示物体。

## 1.2 难点和性能指标

前一小节描述了本文研究的问题是如何通过基于局部简单特征的学习，建立物体模型或者分类器，然后检测静态图像中是否出现该类物体以及这些物体的位置。本节将限定于该研究内容，讨论物体检测问题的难点。

理想情况下，我们总是希望生成的物体模型或者分类器能够将被检测图像中的所有该类物体标示出来（比如在出现该类物体的区域绘制一个方框），但不会错误的标示没有出现该类物体的区域。

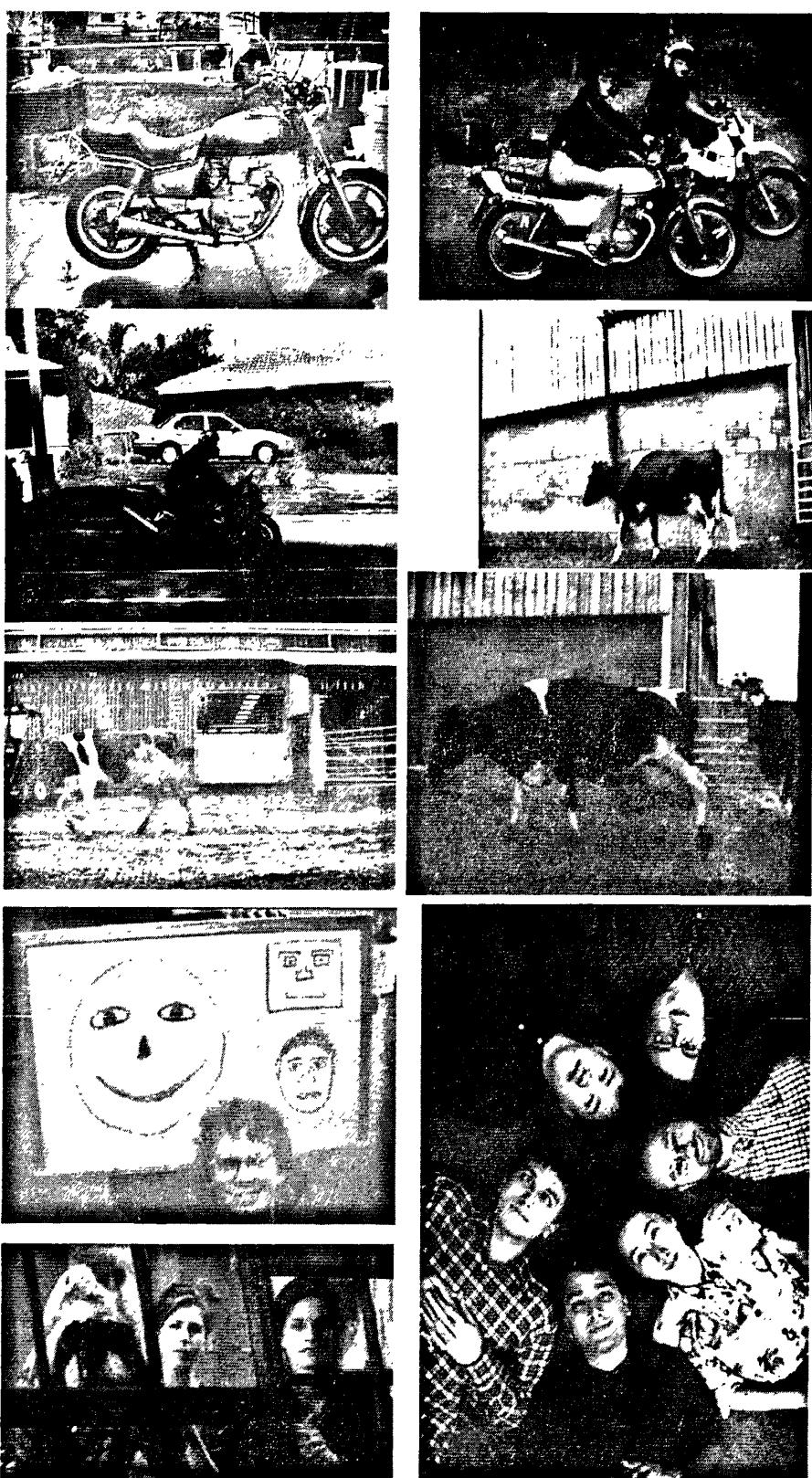


图1.1 物体在图像中呈现的各种变化

然而实际上，物体检测问题被多种因素复杂化了。如图 1.1 所示，因为混乱的图像背景，复杂的光照变化，物体个体之间的类内变化，物体姿态、形状、大小的变化，以及遮挡问题，成像设备自身的性质都会使物体在最终图像中呈现各种各样不同的表现，很容易使得学习过程中所生成的物体模型或分类器无法有效的表示该物体，导致检测过程中做出错误的判断。

检测错误可以分成两类。一类是将没有出现物体的区域标记为物体，这类错误被称为误检 (False Positive, FP)。误检率 (FP Rate, FPR) 被定义为误检数和非物体个数的比值，描述了有多少比例的非物体被误检。另一类错误则相反，将出现物体的区域错误的判断为没有该类物体，这种错误被称为漏检 (False Negative, FN)。漏检率 (FN Rate, FNR) 被定义为漏检数与真实的物体个数的比值，描述了有多少比例的物体被漏检。显然检测率 (Detection Rate or Recall, DR)  $DR=1-FNR$ 。

误检率和检测率是评价物体检测效果很常用的两个指标。显然理想的情况是误检率为零，检测率为百分百，但实际应用中，几乎总会出现漏检和误检。考虑两种极端的情况：

1. 如果对一张图像不做任何处理，那么此时误检率为 0，但是检测率也为 0；
2. 如果将一张图像的所有区域都标记为物体，那么此时检测率为 100%，但同时误检率也为 100%。

一般说来，对检测器的性能进行评价不会仅在一张图像上进行，而是采用一个测试集，其中包括各种背景、光照、姿态、遮挡等因素，表示检测器在实际应用中可能会面临的情况。然后收集检测器在所有测试样例上的检测结果，给出平均的检测率和误检率。因此采用一个通用的测试集一方面能更加准确的测试检测器在现实情况中的性能，另一方面也可以和其它检测器作性能对比。然而最近的一个实验<sup>[15]</sup>表明目前采用的自然图像测试集并不能反映真实世界图像中的变化，一个空模型在测试集上的性能居然超过了当前最好的物体检测系统。关于这方面的思考已经超出了本文的研究范畴，因此，本文还是采用目前能够获得的测试集对文中提出的方法进行评测。

与单张图像的情况类似，在测试集上的检测结果如果以检测率为纵坐标，误检率为横坐标，通过调整检测器的阈值，能得到一条和图 1.2 实线类似的曲线。该曲线总是会通过(0, 0)和(1, 1)这两个阈值取极端值的情况下点。该曲线被称为 ROC (Receiver

Operating Characteristic) 曲线<sup>[16]</sup>, 它一方面表现了在所有正样本上的正确分类情况(好处), 另一方面也说明了在所有负样本上的错误分类情况(代价). 直观上, 我们总是希望这条曲线越靠近左上角(0, 1), 表明该检测器在取合适的阈值时, 可以获得更接近于理想情况的性能.

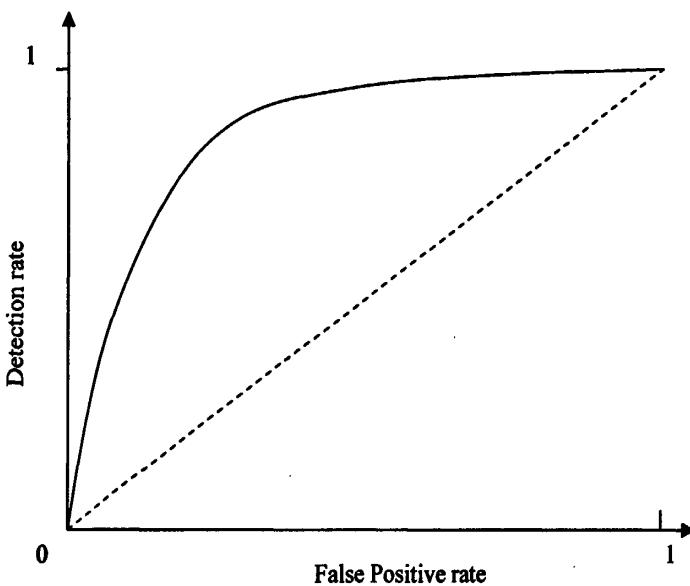


图1.2 典型的 ROC 曲线

除了以上对 ROC 曲线的直观理解外, 还存在各种量化的指标, 比如 AUC(Area Under the ROC Curve), 以 ROC 曲线与横坐标包含的面积作为测度, 越大说明检测器性能越好, 以及 ERR(Equal Error Rate), ROC 曲线上漏检率和误检率相等的时候的错误率, 越小说明检测器性能越好. 但是将一条 ROC 曲线总结为一个数值总是会失去检测率和误检率之间的 tradeoff 模式信息.

实际上, 性能评估的方法不是只有 ROC 曲线. ROC 曲线以及 AUC 或者 ERR, 在某些情况下(如负样例数量远超过正样例)并不能准确的对检测器做出性能评价<sup>[17]</sup>, 甚至会有一定的误导<sup>[18]</sup>, 因此还存在其它的性能评估手段, 比如 Precision-Recall Curve 和 Expected Performance Curve<sup>[18]</sup>等. 但是各种方法都有其局限性, 如何能够准确的测量一个检测器的性能还没有一个定论.

另外, 如果对两个检测器的性能进行对比, 除了采用一样的测试集和性能指标外, 检测标准, 即如何判断检测器所给出的检测是一个正确或错误的检测, 也很重要. 不同

的检测标准情况下的性能差异会很大，导致对比的客观性。其它的因素，比如在检测之前或者之后对测试图像进行各种图像处理也同样会影响到对比实验的结果。

因此，如果测试集、评价指标、检测标准以及其它因素不一致的话，不同检测器之间的对比实验结果就会比较片面，只能在某个程度上表明检测器之间的性能差异。而且根据参考文献<sup>[15]</sup>，即使所有因素一致，也只能表明该检测器在给定测试集上的检测性能，而不是在现实世界的图像上的泛化情况。本文将采用目前通用的测试集，使用检测率、误检率和 ROC 曲线的方法，提出明确的检测标准，对本文介绍的物体检测方法进行评估。

### 1.3 物体检测方法现状研究

物体检测研究经历了四十多年的发展，产生了大量的研究方法。本节从模型和特征的角度，概括的介绍该研究领域的研究现状，描述主要的研究方法。首先简要回顾了基于三维模型和基于全局特征的物体检测方法，然后围绕本文的主题，按照局部特征和通过局部特征的学习构建的模型两个方面对基于局部特征的物体检测方法的研究现状进行详细叙述。

#### 1.3.1 基于三维模型的物体检测方法

早期的物体检测方法通常将物体表示为三维原型模型<sup>[7]</sup>，或者将物体分解，使用超二次曲面（Superquadric）<sup>[19-21]</sup>或体元（Volumetric Primitive）<sup>[22-24]</sup>构建三维体部分。这些方法的一个中心要素是使用以物体为中心的坐标系统，从而使物体检测对视图不变。但是这种方法的难点是在于如何从真实世界图像中可靠的抽取低层的特征，对模型各部分建立视点不变且具有区分能力的表示。因此在这些方法中，物体的结构都比较简单，物体表面的纹理、光照条件和背景都具有一定的限制，使得可以假设图像中的边线与遮挡边界或物体部分的表面不连续相对应。而该假设在实际条件下是不可能满足的，因此这些方法并不能处理实际的物体检测问题。

之后，受 CAD 模型的启发，三维模板开始被用于精确的描述物体<sup>[25-26]</sup>。这些方法建立的物体模型也是视点不变的，但是为一个真实物体构建模型却需要大量的工作。此外，物体的纹理也会严重影响物体检测算法中搜索的复杂度，因此上述方法通常也不能用于具有复杂纹理的物体。相比较于采用三维体部分的模型方法，这些模型采用了点、

线或者点、线的组合来描述物体，使得模型描述歧义性大大提高。因此这些方法在进行物体检测的时候需要对大量的模型假设进行验证。

尽管在九十年代，随着基于表形的物体检测方法的出现，三维物体模型逐渐没落了。但这些三维模型方法中有很多思想被之后的方法所借鉴，包括层次化表示、可变的结构，以及形状对于物体类别定义的重要性等等。而且三维模型在处理视点变化方面有着基于视图的方法无可比拟的优势，始终吸引着一些学者。最近基于三维模型的物体检测研究又有一些新的进展<sup>[27-29]</sup>。

### 1.3.2 基于全局特征的物体检测方法

在基于表形方法早期研究中，全局特征被普遍应用。这些特征是涵盖了整张图像或者整个图像块的信息，如一些简单的统计数据（平均值，颜色直方图）或者对整个图像采用空间降维技术获取有用的信息。空间降维技术，也就是子空间方法，包括主成分分析<sup>[30]</sup>、线形区分分析<sup>[31]</sup>和独立元分析<sup>[32]</sup>等，其主要思想是将原数据投影到一个子空间，使得按照某个先前定义的标准，能最优的表示原来的数据。

Swain 和 Ballard 使用颜色直方图表示物体<sup>[33]</sup>。该直方图可以看作是关于物体颜色分布的一个近似，通过计算和测试图像的颜色直方图的匹配代价进行物体检测。该方法虽然简单，但是至少对于颜色明显的、且用户只对整张图像感兴趣的情况下工作的非常好。Turk 和 Pentland<sup>[30]</sup>，以及之后 Murase 和 Nayar<sup>[34]</sup>分别采用了主成分分析方法，通过对一系列模型图像进行主成分分析，挑选前几个主成分构成投影空间，物体在该空间的投影作为物体的描述器。他们对 PCA 方法的使用上存在区别，前者对不同类别分别建立独立的特征空间，然后根据每个特征空间上的重建错误来确定测试图像对应的类别；而后者对所有的类别只建立一个全局的特征空间，通过该空间内的最近邻查找（Nearest-Neighbor Search）算法确定最接近的训练样本及类别。相对于早期基于三维模型的方法来说，基于表形的方法表现出很好的检测效果，能检测更多种类的物体。

由于全局特征是对整个图像的信息抽象，无法区分前景和背景，因此基于全局特征的物体检测算法只能应用于干净的背景或者物体能够被分割出来的图像上。混乱图像和遮挡问题是基于全局特征的物体检测方法所面临的重大挑战。一种解决方法就是将图像分割为有限的一些区块或段，每个区块或段对应于一个物体或者物体的一部分<sup>[35]</sup>。但是这又产生一个鸡和蛋的问题，因为图像分割本身就是一个很困难的任务，通常需要对图

像内容有更高层次的理解，仅仅靠颜色和纹理信息来获取一个有意义的分割是不可能的。

Leonardis 和 Bischoff<sup>[36]</sup>提出了在特征空间解决遮挡问题的另一个思路。他们不再通过将数据投影到特征图像来计算特征图像的系数，而是采用次采样（Subsampling），使用图像点的一个子集，通过一个鲁棒的假设-测试（Hypothesize-and-Test）机制决定特征空间投影的系数。相互竞争的假设（Competing Hypotheses）则通过最小描述长度原则（Minimum Description Length principle）来选取。但这种方法计算量巨大，对缩放比例敏感。

基于全局特征的物体检测方法还存在一个问题：被检测图像中通常只能存在一个物体。针对该问题的一种解决方法就是在图像每个位置，使用不同的比例，对图像的各个子部分进行穷尽式采样，然后用全局特征描述每个子部分。这种方法也被称为滑动窗口方法，能够检测图像中多个不同比例的物体，但仍然不能处理部分遮挡问题，而且通常要求被检测物体不能具有太大的形状变化。因此，滑动窗口方法在人脸、行人和汽车的检测中最为流行。但是该方法需要异常高效的处理方法对图像中成千上万的子部分进行检测，这严重的限制了滑动窗口方法所能采取的特征。

针对滑动窗口方法性能的问题，可以对图像采用更稀疏的固定网格采样方法<sup>[37]</sup>或者随机采样方法<sup>[38]</sup>，使得图像子部分数大大减少。然而这种方法对于物体检测问题来说，一方面可能会造成物体定位不准确，另一方面可能会漏过对某些出现物体的区域的检测，从而影响了算法的检测率。一种更好的采样方法是有指导的采样方式，如文献<sup>[39-40]</sup>利用边缘信息提出了沿边缘一致的采样方法。该方法对于由形状定义的物体比较有效，但是对于具有混乱背景的图像来说，并不能有效的减少图像的子部分。

尽管有各种针对全局特征方法所存在的问题而提出的解决方法，但全局特征的本质决定了这些方法难以有效的解决视点变化、混乱背景、光照变化等问题；而人眼视觉研究进展<sup>[22]</sup>也表明了局部特征的重要性，越来越多的研究者转向了局部特征。

### 1.3.3 基于局部特征的物体检测方法

局部特征由于其在某些应用中具有明确的语义（比如航拍图像中的线段通常对应于道路等），与某些全局特征（如颜色直方图）相比能够提供准确的定位信息，更好的处理遮挡、复杂光照和混乱背景等问题，因而很快在物体检测研究中得到广泛的应用。

然而相对于全局特征可以表示整个物体的特点来说，局部表示方法在鲁棒性方面要差一些，因此需要通过一定的方式将多个局部特征关联，形成能够准确描述整张图像（整个物体）的模型或者分类器。

理想情况下，我们希望局部特征能够有语义意义的对应于物体某部分。但实际上这是不可能的，因为这需要对场景内容有更高层的理解。而此时我们对有没有目标物体，有多少目标物体等问题都还没有答案，因而更谈不上对物体的各部分进行有语义的表示。这又是一个鸡和蛋的问题，通常的做法是，局部特征从图像的灰度值模式中直接获取。

尽管如此，不同的获取方式产生的特征在可重复性、可区分性、不变性、鲁棒性以及数量、精度等方面存在不同的侧重，需要根据具体的应用需求进行权衡，确定良好特征的获取方式。针对本文研究的物体检测问题，特征的可重复性和可区分性可能是最重要的两个性质。良好的特征应该容忍不同个体之间的类内变化，从而能够在不同的目标个体上重复出现，使得物体检测时能够检测到该物体；但又不会经常出现在与本类物体具有低类间变化的其他类物体上，使得物体检测时能够区分其他类物体。可重复性和可区分性之间存在着竞争关系。一个具有某个灰度值的像素点或者一条线段显然具有很好的可重复性，然而几乎可以确定的是它无助于区分不同类的物体；而某个个体的所有像素或者整个轮廓则因为提供了大量的信息，具有高可区分性，但因为可重复性太低，也不能用于检测同类物体，甚至不能检测姿态或者外部环境光照细微变化后的同一个个体。因此如何从图像的强度信息中获取既包含足够的信息来区分不同类物体，又能在不同个体上重复出现的鲁棒特征是局部特征抽取的一个重要研究方向。

根据局部特征所使用的信息，局部特征可分为两类，基于表形信息的局部特征和基于形状信息的局部特征。前者与像素密切相关，比如图像块（Image Patch）、类 Haar 特征、直方图等；后者使用了图像的边缘信息，通常就指物体的轮廓信息。类似的我们可以将物体检测方法分为基于表形局部特征和基于形状局部特征的方法。与前者不同，后者虽然在九十年代末就被提出<sup>[41-42]</sup>，但应用的环境比较简单，直到最近几年才逐渐受到关注，被应用到比较复杂的图像上。该方法与1.3.1小节所描述的基于三维模型的方法也不同，轮廓信息被用于建立基于视图的二维物体模型。本小节将从特征和模型两个层次分别讨论这两类基于局部特征方法的研究现状。

### 1.3.3.1 基于表形的局部特征

早期基于表形的方法大量使用相对比较简单的特征。这些特征通常具有低维数、低抽象程度的特点，特征值计算也比较容易，因此一般允许通过在图像上密集采样的方式，形成大量的可用特征。

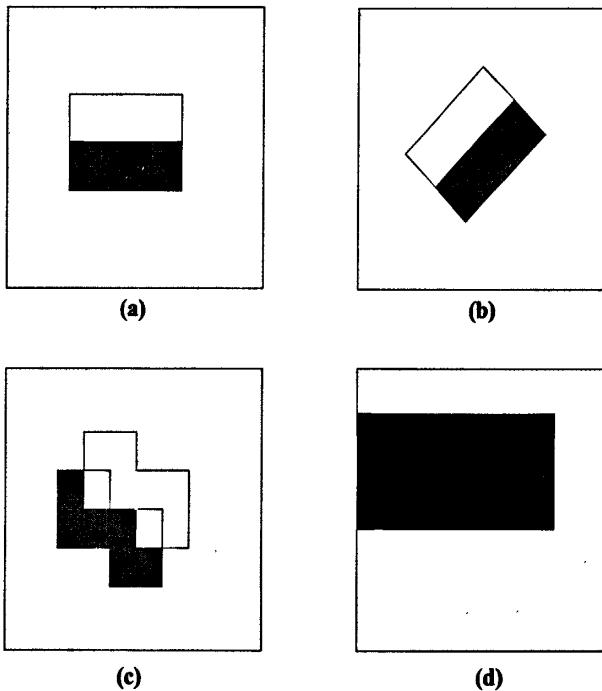


图1.3 类 Haar 特征的扩展和基于模板的直方图特征

其中产生最重要影响的简单特征莫过于类 Haar 特征<sup>[43-44]</sup>，在对人脸<sup>[43-44]</sup>、人体<sup>[43, 45]</sup>、汽车<sup>[46]</sup>等物体的检测研究中都具有广泛的应用。该特征是一个基于模板的特征，如图 1.3(a)所示，每个模板包含多个矩形区域，这些区域大小相同，在水平或者垂直方向相邻对齐。每个区域有符号（正、负），对应特征的特征值就是这些带符号的区域内的像素值之和。由于该性质和 Haar 基本函数很类似，因此被取名为类 Haar 特征。Viola 和 Jones<sup>[44]</sup>提出了积分图的概念，使得任意类 Haar 特征值都能够在常数时间内计算得到，解决了类 Haar 特征在图像上密集采样而产生的数量庞大的问题。之后 Lienhart 等人<sup>[47]</sup>引入了一种倾斜版本的类 Haar 特征，如图 1.3(b)所示，并相应的提出了一种与积分图对应的快速计算倾斜特征值的方法。Viola 等人<sup>[45]</sup>引入一种 S 形的模板（见图 1.3(c)）来对行人进行检测。

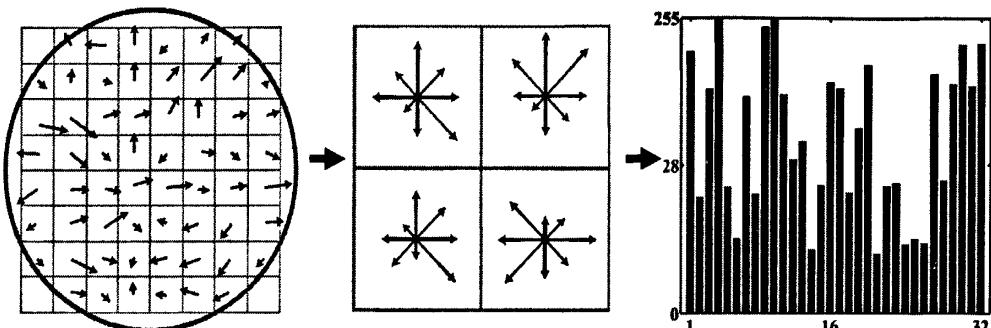
类似的，Wang 等人<sup>[48]</sup>提出了基于模板的直方图特征（Template-Based Histogram Feature）。如图 1.3(d)所示，不同于类 Haar 特征，该模板中只包含一个矩形区域，其位置和大小是预先定义的。根据位置和大小的不同，在  $64 \times 64$  的图像上一共只有 892 不同的直方图特征。为了快速计算特征值，他们使用了积分直方图<sup>[49]</sup>，一个和积分图类似的概念。然后使用 Fisher 线性判别（Fisher Linear Discriminant）将直方图特征值（向量）映射为一个标量，作为判断结果。

基于模板的特征的优势在于计算简单性和有效性。然而由于模板的限制，这种特征通常不能很好的处理具有很大类内变化的物体，因为这意味着需要大量的训练样本去学习该物体的表形变化。此外，由于其特征值和像素值之间只存在很简单的关系，因此难以有效处理剧烈的光照变化。

为了解决以上问题，一些更加抽象的复杂特征相继被提出来。这些特征不再是对像素的简单编码，特征值也不是低维数的向量。实际上，这些特征和像素之间存在着非常复杂的关系，以至于它们通常需要和兴趣点检测方法一起使用，从少量兴趣点位置获取复杂特征，从而很好的解决了计算复杂度问题。这种复杂的特征通常也被称为局部描述器（Local Descriptor）。

兴趣点检测的主要思想是搜索那些足够明显的，在各种图像变换下都能够被可靠抽取的位置。根据应用需求，这些位置可以是边界凹点（Boundary Concavity）或者曲率极值（Curvature Extrema）<sup>[50]</sup>，类角结构（Corner-like Structure）<sup>[51]</sup>，对缩放不变优化的局部操作的极值<sup>[52]</sup>，甚至仿射不变的抽取<sup>[53]</sup>。更多关于兴趣点检测可以参考文献<sup>[54-55]</sup>。通过使用兴趣点检测，物体可以通过少量的局部描述器进行表示。

原始图像块，作为其中可能是最简单的局部描述器，被广泛的应用于基于兴趣点的匹配技术中<sup>[56-58]</sup>。图像块从兴趣点的位置处抽取，包含兴趣点，可以看作是兴趣点邻区的局部信息。图像块的尺寸和应用相关，但通常不能太小。它的优势在于实现简单，人眼能够直接了解已经匹配的区域，因此有利于算法的设计；而缺点则是非常高的维数。该问题可以使用本章1.3.2小节介绍的子空间方法来解决。此外也可以通过引入包括光照正则化等预处理步骤加强原始图像块的表示能力。Weber 等人<sup>[59]</sup>引入高通滤波，提出了一种基于高通滤波图像块的表示方式；Nelson 和 Selinger<sup>[60]</sup>则进一步的通过计算曲率极值，抽取类边缘结构的局部窗口作为基本特征。

图1.4 2×2网格的SIFT计算过程示意图<sup>[61]</sup>

另一方面，Lowe<sup>[61-63]</sup>提出的SIFT (Scale Invariant Feature Transform) 描述器显示了非常高的抽象度。这是一个受人脑视觉皮层中复杂神经元的反射性质启发而设计的局部特征表示。SIFT 使用高斯差 (Difference of Gaussian, DoG) 函数在图像的尺度空间 (Scale Space)<sup>[64]</sup> 中所有位置搜索兴趣点。如图 1.4 所示，每个局部区域被转换为梯度表示，然后又被等分为 16 ( $4 \times 4$ ) 个子区域。对每个子区域使用高斯窗口函数进行梯度加权，然后使用插值方法将梯度方向简化为 8 个主要方向，最后叠加起来形成一个 128 维的包含 8 个方向和 16 个区域的梯度信息的三维直方图。SIFT 对尺度和旋转不变，能够鲁棒的处理光照和视点上的少量变化，但维数很高，计算非常复杂。

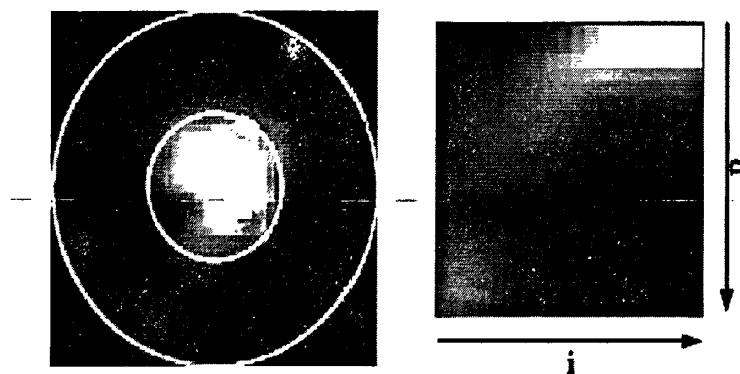


图1.5 图像块样例和对应的Spin Image

Ke 和 Sukthankar<sup>[65]</sup>针对 SIFT 维数高的问题，引入了主成分分析方法，使用 20 个特征向量代替直方图来描述梯度方向，使得特征维数降为 SIFT 的  $1/8$ 。该特征被称为 PCA-SIFT，在匹配实验中显示 PCA-SIFT 只比 SIFT 略差<sup>[66]</sup>。另一个 SIFT 扩展是 GLOH (Gradient Location-Orientation Histogram)<sup>[66]</sup>。与 Shape Context<sup>[67]</sup>的思想类似，该特征的区域被径向 (Radial) 和角向 (Angular) 网格划分 (而不是  $4 \times 4$  常规的网格)，然后

采用 PCA 方法，将 272 维的梯度直方图降低为 128 维的描述器。Lazebnik 等人<sup>[68]</sup>则将区域先转换为一个叫做 Spin Image 的二维直方图，该图的每个点表示灰度值为  $i$  的像素到区域中心距离  $d$  的概率关系，如图 1.5 所示。该特征对平面内旋转不变，通过直方图光滑和正则化步骤，还能够获得仿射光照不变性。

SRUF (Speeded Up Robust Features)<sup>[69]</sup>针对 SIFT 计算复杂的问题，采用 Haar 小波近似 Hessian 矩阵<sup>[70]</sup>，而前者又可以有效的利用积分图<sup>[44]</sup>进行计算，使得该特征能够应用于对计算效率要求比较高的情况。Dalal 和 Triggs<sup>[71]</sup>则更进一步，不在尺度空间建立梯度图像，而是直接将每个子区域（称之为 Cell）的局部梯度直方图在整个区域上进行对比度正则化 (Contrast-Normalization) 后，组合形成一个叫 HOG (Histogram of Oriented Gradient) 的描述器。HOG 计算量很小，因此可以不使用兴趣点而是直接通过在图像上密集采样获取。基于 HOG 的人体检测方法<sup>[71]</sup>显示了目前最佳的检测的人体检测结果。

### 1.3.3.2 基于形状的局部特征

与基于表形的局部特征不同，基于形状的局部特征利用边缘检测器获取图像或者物体的轮廓信息，然后通过点或者线来描述图像或者物体。实际上基于表形的局部特征和基于形状的局部特征的区别并不是非常明显，点线信息也是从图像像素值中获取的，而梯度信息从某种程度上来讲也是一种边线信息。对这两者进行区分的主要方式是该特征是否具有明显的形状属性，而没有包含太多的纹理和颜色信息。

早期的基于三维模型的物体检测方法都是基于形状的，通过从测试图像中检测一些基本的点、线，根据透视投影的性质，和模型进行匹配，从而能够检测一些简单的物体。该方法对物体的复杂度和纹理、图像背景、光照等条件都有严格的限制。尽管如此，这也说明形状是定义一个物体的重要信息。之后在基于表形特征的物体检测方法大行其道的时候，虽然也有一些基于形状的特征被提出<sup>[41-42]</sup>，但通常只被引用于图像配准、或者简单物体的分类等领域，物体也只有简单的形状（锤子、鱼等），没有复杂的纹理，基本上不考虑图像的混乱背景。

基于形状的局部特征在物体检测中没有得到广泛的应用有其自身原因：1) 边缘检测本身是不可靠的。物体轮廓由于光照、视点、背景以及其他噪音的原因可能会被部分的检测或者检测成多个不连续的轮廓，甚至完全被忽略掉，因此自底向上的物体检测方法在原始输入中就存在诸多的不确定性因素。2) 边缘所包含的信息太少。混乱背景（如

草丛)产生的大量杂乱边缘使得一条边缘的意外匹配可能性很高,因此边缘的区分能力相对很弱。这两个因素使得研究者更倾向于使用基于表形的特征。

尽管如此,近年来,越来越多的研究发现对某些物体采用基于表形的特征并不能满足检测的需求,比如自行车、饮料瓶等物体,如果采用表形特征,其中包含的信息更多的是自行车后面的背景或者瓶子上各种标签的信息。这种随机变化(*ad hoc*)的信息显然无助于物体检测。而最新的心理学和人体视觉研究也表明人类在早期视觉中确实使用形状对物体进行分类<sup>[72]</sup>,人类识别动物先是根据形状然后才是纹理,颜色信息使用的最少<sup>[73]</sup>。这些因素都促使研究者重新对基于形状的特征进行审视。同时,边缘检测算法和图像分割算法<sup>[74-77]</sup>的进步也为基于形状特征的物体检测方法研究提供了条件。

Belongie 等人<sup>[67]</sup>提出使用点的相对分布位置和方向信息作为区域描述器。这些轮廓点可以通过边缘检测器(如 Canny 边缘检测器<sup>[78]</sup>)和常规的采样方法获得。整个形状就可以通过这些点逐对之间的相对位置和方位信息所表示。但是这种描述方式会随着区域的增大(点的数量增加)维数也快速增长。为了解决该问题,他们提出了 Shape Context 特征,将该高维表示转换为 Log-Polar 空间中的直方图表示,形成一个 60 维的特征。Shape Context 可用于检测手和字符等具有一定复杂度的物体。

Berg 等人<sup>[79]</sup>提出了一个可变形的点模板,通过整数二次规划(Integer Quadratic Programming)算法确定输入点和特征点的对应关系,然后根据该对应关系,估计一个正则化的薄板样条(Regularized Thin Plate Spline)对齐变换,建立两个形状之间密集的对应关系,实现对物体的检测。该方法可以对正面、侧面人脸进行检测。Leordeanu 等人<sup>[80]</sup>则在物体轮廓上均匀的采样,同时获取这些抽象点的法向信息。相互连接的点逐对之间的几何关系被用来去除错误的对应关系。

Shotton 等人<sup>[81]</sup>和 Opelt 等人<sup>[82]</sup>则分别提出以轮廓段作为特征的物体检测方法。前者使用随机的一些矩形对物体轮廓进行切割获取轮廓段,后者采用沿物体轮廓逐步增长的方式获取特征。在这两种方法中,轮廓段特征都包含有与物体中心相对的位置信息,用于去除可能的误匹配。这两种基于轮廓段的方法能够用来检测复杂物体,如牛,马,汽车等。

以上这些方法都是将形状表示为点。虽然能够保持轮廓的细节的信息,但是在处理空间配置的时候计算复杂度比较高。为此用线段表示物体形状的方法被相继提出。Ren<sup>[83]</sup>

使用受限制的三角面片划分方法建立边缘图，然后为不同视图学习相应地线段模板。这些模板只包含一小部分线段，但是定义了所有的线段对之间相对的几何配置关系。通过允许部分匹配，该模板方法可以比较有效的处理人体姿态的大量变化。Ferrari 等人<sup>[84]</sup>在边缘图上链接边缘的断裂处，然后将它们分割成大致成为直线的轮廓段，形成轮廓段网络（Contour Segment Network, CSN）。物体检测问题被转换为在轮廓段网络上寻找相似的连通路径。之后 Ferrari 等人<sup>[85]</sup>提出了一系列的尺度不变的局部形状特征，每个特征包含  $k$  条大致成直线的相近的轮廓段，称为  $k$ AS。这些方法也能够用于人体、马等复杂物体的检测。

### 1.3.3.3 局部特征模型和学习方法

以上我们分别介绍了基于表形和形状的一些常见特征。有关局部特征的更多介绍可以参考文献<sup>[54-55, 66, 86-88]</sup>。局部特征只是对一个物体部分的描述，正如我们不能仅仅看到几个轮子（可能就是几个废弃的轮子）就说该物体是汽车或者摩托车一样，物体检测方法通常也不能仅仅根据一个或者几个局部特征就作出正确的检测判断。多个局部特征需要按照一定的（几何、概率等）关系组合形成物体模型来描述整个物体。通过有效的学习方法，从大量样本中获取局部特征之间的关系，构建准确描述物体的模型对于物体检测方法至关重要。

Rowley 等人<sup>[89]</sup>使用神经网络（Neural Network）训练人脸模型。他们采用了三种隐层单元（Hidden Unit）：4 个检查  $10 \times 10$  像素大小的区域的隐层单元，16 个检查  $5 \times 5$  像素大小的区域的隐层单元，以及 6 个用来检查  $20 \times 5$  像素的水平条带区域的隐层单元。这些隐层单元可以在网络中多次重复使用。神经网络能够发现一些人类无法直接获得的重要特征，但是关于这些特征之间的关系、整个模型如何起作用却同样也不能被人轻易的理解和解释，只能说这些具有条状感受域的隐层单元可以用来检测嘴巴或者两只眼睛，而方形感受域的隐层单元则用来检测单个眼睛、鼻子或者嘴角。

Viola 和 Jones<sup>[44]</sup>提出了一个快速的人脸检测框架。他们在类 Haar 特征上，采用 AdaBoost 算法学习人脸模型。AdaBoost<sup>[90]</sup>是一种将任意给定的准确率仅比随机猜测 ( $1/2$ ) 略高的弱假设提升为准确率很高的强假设的多项式机器学习算法。他们将每个类 Haar 特征作为一个弱假设，通过训练集确定弱假设的分类阈值和分类错误。根据分类错误，通过 AdaBoost 算法选择最优的一些弱假设，线形组合成一个强分类器。因此不同

于神经网络模型，强分类器内的类 Haar 特征权重关系是确定的。由于类 Haar 特征是一个基于模板的特征，因此强分类实际上隐含了其中所有类 Haar 特征个体之间的几何关系信息。分类器将这些特征按照它们之间的几何关系分布在被检测区域各个位置，然后线性加权组合每个特征的输出结果，做出检测判断。基于分类器上述特性，Viola 和 Jones 提出了分类器级联结构（Cascade Structure），根据分类器包含的特征数从少到多构成一个应用层次，前期的应用层次用于快速的排除一些非人脸的区域，而只有通过前期层次的区域才能进入后期层次的检测，大大减少了后期层次分类器越来越复杂的计算量，从而实现了实时的人脸检测速度。因此在该级联结构中分类器是有层次次序的。

支持向量机（Support Vector Machine, SVM）<sup>[91]</sup>是通用的一种统计学习方法，能够在高维空间中，在小样本上学习分类器，使得结构风险最小化。Dalal 和 Triggs<sup>[71]</sup>、Osuna 等人<sup>[92]</sup>、Papageorgiou 和 Poggio<sup>[93]</sup>等将 SVM 学习算法成功应用于人脸、人体、汽车等物体检测系统中。在 SVM 学习算法中，样本被特征值向量（点）所表示。通过采用合适的核函数，将这些样本点投影到高维空间，成为线性可分的问题，最后在该高维空间中确定到最近的不同类样本点（支持向量）距离最大化的分类界面。局部特征之间的几何关系只能包含在特征值向量的构建过程中。

另一种完全忽略局部特征之间的几何关系的模型是 Bag of words (BoW)<sup>[94]</sup>，最先被应用在文本的表示和分类中。在图像的 BoW 模型中，可以通过固定的网格采样、兴趣点检测等（见1.3.2）方法获取局部的图像块特征，经过聚合（Clustering），生成字符集，用于表示其他图像，形成特征值向量（如直方图）。该特征值向量可以通过 SVM、AdaBoost 等算法进行学习，或者通过朴素贝叶斯（Naïve Bayes）分类器<sup>[95]</sup>、层次贝叶斯（Hierarchical Bayesian）模型<sup>[96]</sup>进行概率建模。虽然该模型在场景分类和物体检测中表现了良好的效果，但由于缺少位置信息，物体定位的准确性不够。

与 BoW 模型不同，Weber 等人<sup>[59]</sup>和 Fergus 等人<sup>[57]</sup>提出的星群（Constellation）模型试图通过相互有几何关系约束的部分来表示物体。Fergus 等人采用简单的兴趣点特征表示物体的部分，假设这些部分的位置信息的联合概率分布满足高斯分布。一旦物体的尺度被确定，每个部分的位置就可以通过图像坐标表示，然后使用贝叶斯决策来判断该区域是否为物体。由于学习过程的时间复杂度和物体部分数量成指数关系，模型中物体部分的数量被限制在 6 或者 7，从而限制了模型对复杂物体的表示能力。

Leibe 等人<sup>[8]</sup>提出了一个将局部特征与物体中心建立几何约束的隐式几何模型 (Implicit Shape Model, ISM)。在训练集上学习该模型时，每个局部特征（如图像块）都会绑定一个或者多个投票向量 (Voting Vector)，指向可能的物体中心，每个投票向量都有权重，表示该物体中心位置的置信度，通过对大量样本的学习建立。在检测物体时，投票向量根据其置信度对可能的物体中心进行概率投票。通过在霍夫投票空间中搜索较大的值来确定物体的位置。而对这些物体中心进行投票的特征又可以用来确定物体可能的边界，使得该模型可以同时处理图像分割问题。为了获取特征和物体中心之间的几何关系（投票向量），该方法需要训练集给出正样本的物体中心位置。

类似于 ISM，Opelt 等人<sup>[82]</sup>提出了轮廓段模型 (Boundary Fragment Model, BFM)。该模型中，物体由轮廓段特征表示，每个轮廓段和物体中心也存在着类似于投票向量的几何约束。但不同于 ISM 为特征的几何关系建立概率模型的做法，BFM 采用 Chamfer 距离衡量两个轮廓段的距离，然后对这些轮廓段进行 AdaBoost 学习，选取那些能重复在正样本上出现且和物体中心保持有效的几何约束的特征，生成物体分类器。

与 ISM 或者 BFM 都不同，Ferrari 等人<sup>[85]</sup>提出的  $k$ AS 利用了相连的轮廓线段之间的局部几何关系，并利用该几何关系衡量  $k$ AS 特征之间的相似度。由于  $k$ AS 的通用性，他们采用了不同的  $k$ （相连的线段数量），构建了一个全局的码表 (Codebook)，然后再针对具体物体类别进行训练，比如在该码表之上建立 BoW 模型，或者采用类似于 Dalal 和 Triggs<sup>[71]</sup>的方法，将  $k$ AS 的位置信息隐含于特征向量中，然后使用 SVM 方法学习分类器。

### 1.3.4 研究现状小结

通过以上关于物体检测研究现状的描述可以看出，其最终目的是寻找到一个鲁棒的物体模型，在混乱背景、光照变化、视点变化、部分遮挡等条件下仍然能够表示物体的类内变化；同时这还是一个准确的模型，能够排除其他具有较小的类间变化的物体。从目前的研究情况来看：

1. 三维模型似乎是一个最好的解决方案，但是如何构造三维模型是一个巨大的挑战，特别对于一些纹理和结构复杂的物体。而且目前自底向上的方法不能保证产生良好质量的底层和中层特征（点、线等），因此在混乱背景条件下如何进行可靠的检测也是一个需要解决的问题。

2. 全局的特征由于包含了整个物体的信息，因此相对于局部特征来说，通常更加鲁棒，但是对于遮挡、视点变化等问题则显得不是非常有效。尽管如此，子空间方法、滑动窗口方法等可以应用到基于局部特征的方法中。
3. 局部特征由于采用了分块对物体进行描述的方式，因此对处理遮挡具有一定的优势。不管对于简单局部特征还是复杂局部特征来说，特征的结构非常重要。这里所说的结构不仅仅包含特征内部的结构，还包括特征之间的结构（几何关系），后者被模型和分类器所描述。
4. 基于形状的局部特征相对于基于表形的局部特征来说，比较简单，信息量要小，因此更注重于利用特征的结构。还有一种思路就是与基于表形的局部特征结合<sup>[97-99]</sup>，而这一点实际上在复杂的基于表形的局部特征中已经存在（如果将兴趣点看成是形状特征的话）。
5. 目前特征之间的结构约束有：1) 为所有的特征建立约束；2) 有关联的两个特征之间建立约束；3) 与物体中心建立约束。第一种方式计算量比较大，第二种方式和第三种方式则比较自由，能较好的处理类内变化。

#### 1.4 本文研究的内容

本文研究的主要内容是选取怎样的特征，构建怎样的模型或者分类器，来准确、鲁棒的表示物体，从而提高检测效果。通过对物体检测方法现状的描述分析可知，尽管物体检测研究中提出或者采用了各种不同抽象程度的局部特征来表示物体，到底何种特征能够最有效的描述一般物体还没有定论。通常更抽象的特征显示出更好的区分能力，但是所能描述的物体类别也有更多的限制。仿生物学的物体模型<sup>[5]</sup>表明简单局部特征能够有效的表示一般物体。因此，本文针对简单局部特征展开研究，研究了如何使用学习方法，利用简单局部特征的几何关系，增强模型或分类器对物体的表示能力。

在局部特征研究方面，本文从信息编码的角度，分析了局部特征通用的组件，提出两种通过改变组件而改进局部特征的方法。在基于表形的局部特征中，重点研究了类 Haar 特征。通过修改类 Haar 特征的特征结构，提出了一种类 Haar 特征变体，称为分散矩形特征。该特征和类 Haar 特征同样都是基于矩形模板的简单特征，但是在特征结构上矩形区域能够更加自由对物体进行表示，能够编码更多的方位和形状信息，使得该特

征在区分能力上优于类 Haar 特征。通过构造方法，本文还严格证明了非退化的分散矩形特征与满足一定几何约束的多个类 Haar 特征之间的等价关系，分析了分散矩形特征更为鲁棒的原因。人脸检测实验结果验证了分散矩形特征在表达能力和鲁棒性两方面对类 Haar 特征存在的优势。

在对基于形状的局部特征研究中，本文分析了目前研究中存在的两种表示轮廓的方法。受心理学研究结果的启发，采用了基于线段的轮廓表示方式，提出了霍夫变换线段（组）特征。该特征可以隐式的与物体中心建立几何约束关系，是一种简单紧凑的隐式形状模型<sup>[8]</sup>，而且能够很方便的处理旋转、缩放和平移变换。本文研究了在霍夫变换线段空间中的线段相似度，提出了一个加权的霍夫变换线段距离公式，通过选取合适的权重，可以有效的处理不可靠边缘检测引入的部分匹配问题。为了进一步增强霍夫变换线段的表示能力，本文引入了线段之间的连接性，提出了霍夫变换线段组，有效利用霍夫变换线段的局部几何关系，降低混乱背景上的误检测。本文还研究了如何从大量训练样本的霍夫变换线段组中选取有区分能力的特征构建该物体类别码表的方法。通过对摩托车和牛的形状匹配检测实验表明形状确实是定义物体类别的重要特征。

在学习算法方面，本文着重研究了 AdaBoost 算法。根据选择性能更好的弱假设可以使学习到的强假设分类更加鲁棒和高效的思想，本文提出了双阈值弱假设。该假设所引入的额外的阈值能保证将特征值空间划分的更精细，使分类错误比单阈值弱假设更小，从而提高了该算法的学习能力。相应地，本文研究了用于确定弱假设两个最优阈值的弱学习算法。通过将分类错误最小化问题转换为寻找最大（最小）和连续子序列问题，该算法能够在线形时间内为双阈值弱假设确定最优的阈值。对类 Haar 特征和分散矩形特征的学习结果表明，在同样的训练性能要求下，使用双阈值的 AdaBoost 算法收敛的更快，所生成的检测器具有更少的层次，使用更少的特征；人脸检测实验结果表明，该算法学习的分类器性能总体上要优于原来的 AdaBoost 算法。

## 1.5 本文的组织结构

本文分为六章。第一章是绪论。首先阐述了本文的研究背景和意义，明确本文要研究的问题，介绍了物体检测问题的难点和相关概念；然后回顾了物体检测方法的研究现状，重点介绍了基于局部特征的常见方法；最后提出了本文的研究内容和结构安排。

第二章研究了分散矩形特征。通过适当的修改类 Haar 特征在特征结构上的所受到的限制，使分散矩形特征能够表示更多方位和形状信息，成为一种更具有表示能力的基于表形的局部特征。

第三章研究了双阈值 AdaBoost 算法。提出了双阈值弱假设，通过对特征值空间更精细的划分，确保比对应的单阈值弱假设分类错误更小；提出了一个线性最优的双阈值弱学习算法，使得双阈值 AdaBoost 算法在保持学习速度的同时，具有更好的学习能力。

第四章结合了分散矩形特征和双阈值 AdaBoost 算法。实验显示了分散矩形特征不仅能更好的表示物体，而且能鲁棒的处理过匹配问题；通过构造方法，严格证明了一个非退化的分散矩形特征等价于满足一定几何约束关系的多个类 Haar 特征，并利用该等价性对实验结果做出了解释。

第五章研究了霍夫变换线段（组）。提出了霍夫变换线段四元组表示方式，与物体中心建立隐式的几何约束关系；提出了霍夫变换线段的相似度衡量方法，通过调整四元组各元素权重，处理部分匹配问题；根据连接性构建霍夫变换线段组，使其成为一个能定义多种物体类别的局部形状特征。

第六章是本文的结论，对论文的主要工作和成果的进行了总结，并对进一步的研究进行了展望。

## 第2章 基于分散矩形特征的物体检测

在绪论中，我们综述了不同时期物体检测方法中所使用的各种常用特征，讨论了特征随着研究进展和应用需求变化的演进过程。本章将从信息编码的角度出发分析局部区域特征，从中抽取出这些特征相同的组成部分（组件）。然后基于该分析结果，对类 Haar 特征<sup>[43-44]</sup>的结构进行扩展，形成一种新的特征变体，我们称之为分散矩形特征。由于分散矩形特征和类 Haar 特征之间的区别仅局限于特征结构，因此在本章的实验中，可以不修改 Viola 和 Jones 的检测框架<sup>[44]</sup>而直接替换类 Haar 特征。实验结果显示基于该变体的人脸检测效果优于基于类 Haar 特征的效果。

### 2.1 局部特征分析

局部特征是对物体某个区域的描述，能够有效的处理遮挡、光照变化等问题，因此在物体检测研究中得到了广泛的重视，产生了形形色色各种不同类型的局部区域特征。这些特征可以大致分为基于表形和基于形状的局部特征，前者可分为简单和复杂的区域特征，后者可分为基于点表示和基于线表示的轮廓特征。它们在具体细节上差异很大，表 2.1 从局部区域特征的位置、形状和大小的确定方式、信息的表示方式以及如何建立特征之间的几何关系几个角度总结了绪论中提到的一些特征的特点。

表2.1 常见几种局部区域特征的特点

局部区域特征	位置	形状和大小	表示	特征之间的几何关系
类 Haar 特征	滑动窗口	矩形，滑动窗口	区域像素和	滑动窗口中隐式确定
基于模板的直方图特征	手工	矩形，手工	颜色直方图	滑动窗口中隐式确定
SIFT（以及其变体）	兴趣点检测	圆形，多尺度	梯度直方图	取决于模型
Spin Image	兴趣点检测	圆环（圆形）	灰度-距离直方图	取决于模型
轮廓段	边缘检测	取决于轮廓段	点或者线段	与物体中心的几何约束、两两之间的几何配置等

尽管以上特征的形式和内容变化多样，但是从信息的编码来看，其输入都是一个二维的像素矩阵，其目的都是从该像素矩阵中抽取能够用于建立鲁棒、准确的物体表示的信息。在特征抽取过程中，如图 2.1 所示，原始输入信息首先通过特征过滤器转换为更

可靠的表示，比如梯度或者边缘信息。特征过滤器还包括其他的图像处理方法，如直方图均衡化、高斯平滑等。然后，局部区域特征的位置和形状、大小等信息通过特征结构来确定。可以使用的具体方法包括滑动窗口方法、兴趣点检测、边缘检测，甚至随机采样或者手工指定的方法。经过该步骤后，就收集到了一些原始的局部区域信息，某些情况下（如使用模板的方法确定的局部特征），不同特征之间的空间关系也隐式的包含在特征结构中。特征形式确定该局部区域特征的最终表达，形成鲁棒的、具有区分能力的、可以被后续过程，比如特征匹配或者特征选取，直接使用的局部特征。一般说来，标量特征形式因计算量小，被广泛的用于物体类别检测；而向量特征形式因为信息量更大，可以应用于个体物体的识别。特征形式还可以看作是和物体模型的接口，甚至有时和模型有所重叠，比如如果使用 BoW，那么特征中关于几何关系的信息就完全不必要抽取。

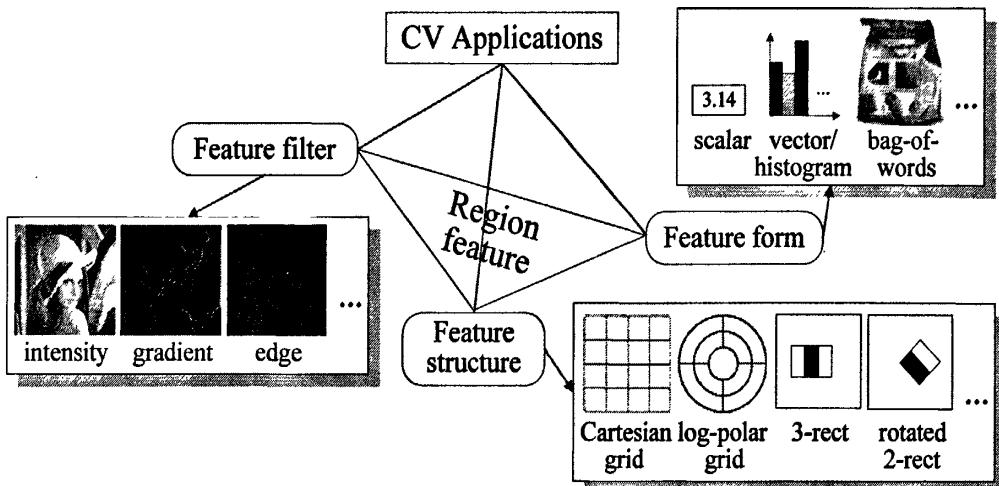


图2.1 特征抽取过程中的三个通用组件

特征抽取过程中，这三个组件的设计其实是紧密相关的，也依赖于具体的应用问题。通用组件的分析只是大致描述了区域特征的主体框架，而没有涉及具体细节。尽管如此，该分析结果指出了两条针对具体应用和需求，改进现有特征设计的直观途径。

1. 修改某个特征的某个（些）组件来提高性能。比如使用混合的特征过滤器，将基于表形的和基于形状的局部特征结合<sup>[100]</sup>，使得该特征能更鲁棒的处理混乱背景。
2. 将多个特征的不同组件组合成一个新的特征。Winder 和 Brown<sup>[55]</sup>显示了类似的一个例子，将包括 SIFT 和 GLOH 的各种特征的不同组件组合，生成一个优于 SIFT 的描述器。

本章采用了第一条途径，通过修改类 Haar 特征的特征结构，提出了一种简单的变体，我们称之为分散矩形特征，具体内容见2.2小节。

## 2.2 分散矩形特征

本节我们首先详细介绍在 Viola 和 Jones 提出的物体检测框架中所使用的类 Haar 特征，对类 Haar 特征的结构进行分析。然后通过去除类 Haar 特征模板的结构限制，形成一种分散的矩形特征。最后我们对这两种特征进行对比，分析新变体的表达能力。

### 2.2.1 类 Haar 特征 (Haar-Like Feature, HLF)

在 Viola 和 Jones 的所提出来的具有里程碑意义的物体检测框架中，他们采用了灰度图像上的类 Haar 特征。根据类 Haar 特征所包含的矩形区域个数不同，该框架中使用了双矩形、三矩形和四矩形三种特征模板，图 2.2 显示了几个类 Haar 特征。

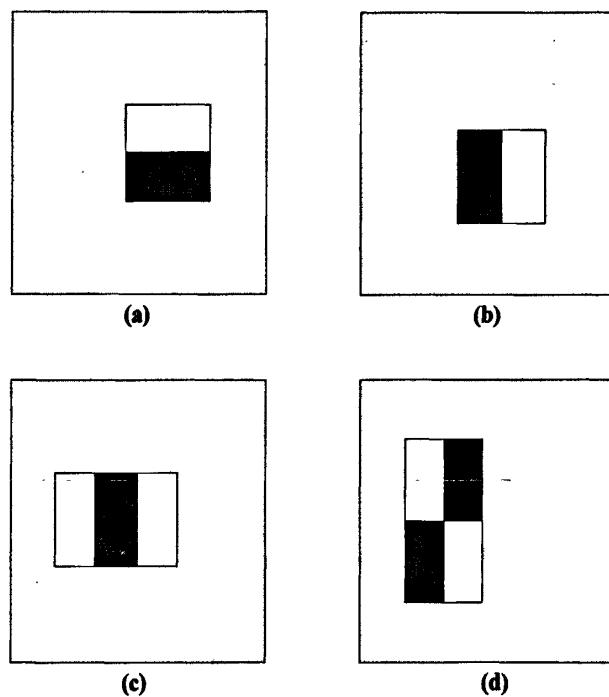


图2.2 类 Haar 特征的例子

类 Haar 特征模板中各子区域具有相同的尺寸和形状，并且在水平或垂直方向上相邻、对齐，其特征值就是灰色区域内的像素灰度值与白色区域内的像素灰度值的平均差值，用数学公式可以表达为

$$f = \sum_{i=1}^n w_i R_i, \quad (2.1)$$

其中  $R_i$  表示特征中第  $i$  个区域的所有像素灰度值之和,  $w_i$  为对应区域的权重,  $n$  取 2, 3 或者 4, 各区域从左到右, 从上到下进行编码。对于图 2.2(a),  $f = R_1 - R_2$ ; 对于图 2.2(c),  $f = R_1 - 2R_2 + R_3$ , 其中  $w_2 = -2$  是为了消除灰色区域面积所带来的影响。类 Haar 特征还可以通过将这些权重对特征的总面积进行标准化, 从而获得尺度不变的特征性质。

由于这些特征模板能够在图像窗口中缩放和平移, 因此通常一个窗口中的类 Haar 特征实例集合将会非常大, 比如在一个  $19 \times 19$  像素的窗口中, 一共有 63,960 个类 Haar 特征, 远远超过图像像素个数。过完备 (Overcomplete) 的特征集合为特征拾取提供了更多可能的选择, 但同时也增加了特征值计算的工作量。为了能够快速计算类 Haar 特征值, Viola 和 Jones 提出了一种积分图 (Integral Image) 的图像中间表示形式。

积分图中的每个像素的值是以原图左上角点和对应像素点为对角顶点的矩形区域内像素灰度值之和:

$$II(x, y) = \sum_{x' \leq x, y' \leq y} I(x', y'), \quad (2.2)$$

其中  $I(x, y)$  表示原图,  $II(x, y)$  表示对应的积分图。

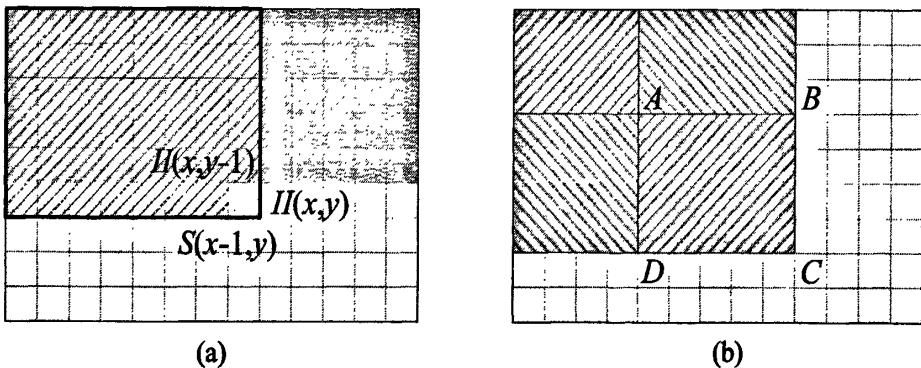


图2.3 积分图的生成和使用: (a) 积分图一次性扫描生成示意图; (b) 利用积分图, 矩形  $ABCD$  内的像素灰度值之和可以计算为  $II(A)+II(C)-(II(B)+II(D))$

在积分图的构建过程中, 我们引入一个行累加器  $S$ , 如图 2.3(a) 所示, 有

$$\begin{cases} S(x, y) = S(x-1, y) + I(x, y) \\ II(x, y) = II(x, y-1) + S(x, y) \end{cases}, \quad (2.3)$$

其中  $S(-1, y) = 0$ ,  $H(x, -1) = 0$ , 则积分图可以通过在原图上从上到下, 从左到右一次性扫描获得。建立积分图之后, 原图上任何位置和尺寸的矩形区域之内的像素灰度值之和就转换为积分图上对应矩形的四个角点之间的一个简单的加减计算。同样, 任何位置和尺寸的类 Haar 特征也均可以在常数时间内计算得到其特征值。

### 2.2.2 分散矩形特征 (Scattered Rectangle Feature, SRF)

从前一节的介绍中知道, 类 Haar 特征的特征结构限制很多, 矩形模板各子区域不仅要具有相同的形状和尺寸, 而且它们之间还需要满足相邻、对齐的规则。我们对该特征结构进行了修改, 产生一个松散的特征结构: 矩形模板各子区域仅需要具有相同的形状和尺寸, 而不受限于相邻、对齐的约束。

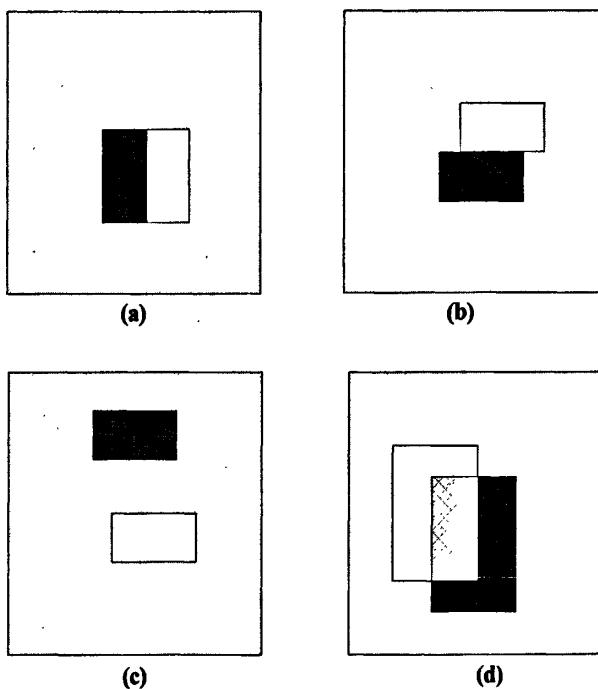


图2.4 分散矩形特征的几个例子: (a) 既是分散矩形特征, 也是类 Haar 特征; (b) 错切的分散矩形特征; (c) 分离的分散矩形特征; (d) 重叠的分散矩形特征

采用该特征结构后, 类 Haar 特征中的多个子矩形可以散乱的分布在图像窗口中, 因此我们将该类 Haar 特征的变体命名为分散矩形特征。由于组成一个分散矩形特征的区域并不需要在水平或垂直方向上相邻或对齐, 导致分散矩形特征数量的组合爆炸。针对该问题, 我们采用的方法是限制特征模板种类, 只采用双矩形模板。此时, 同样在一个  $19 \times 19$  像素的窗口中, 将会有 1,533,300 个分散矩形特征。这已经远远超过了采用三

种模板的类 Haar 特征。图 2.4 显示了一些分散矩形特征的例子。和类 Haar 特征一样，分散矩形特征的特征值也采用公式(2.1)计算，只是此时  $n=2$ 。

### 2.2.3 HLF 和 SRF 对比

尽管类 Haar 特征采用了更多的矩形模板，但是由于各子区域必须在水平或垂直方向上对齐，因此只能表达水平、垂直和对角方向上的信息；同时由于各个子区域必须相邻，不管子区域之间是否覆盖无用或者表达能力弱的区域，这些信息都被特征所包含。因此相邻对齐的规则限制或降低了类 Haar 特征的表达能力。

而另一方面，分散矩形特征尽管只采用了一个矩形模板，但由于不受这两方面的限制，能够编码几乎所有方向上的信息（如果不考虑数字图像的离散性）。在空间位置方面，也能够表达错切、分离和重叠的形状信息。根据分散矩形特征值的计算方式可知，含有重叠区域的分散矩形特征，实际上并不包含重叠区域的信息，而仅仅包含两个 L 形（而不是矩形）区域的像素信息。这些方位和形状信息是单个类 Haar 特征所不能表达的。除特征数量之外，这也弥补了分散矩形特征模板单一的弱点。我们利用 Viola 和 Jones 的物体检测框架，使用类 Haar 特征和分散矩形特征训练的两个人脸分类器（见2.3小节）的结果也经验性的表明，分散矩形特征的表达能力更好。

表2.2 类 Haar 特征和分散矩形特征对比总结

	类 Haar 特征			分散矩形特征
特征模板				
模板限制	子区域形状大小相同，垂直或水平方向上对齐、相邻			子区域形状大小相同
19×19 窗口中特征个数	34,200	21,660	8100	1,533,300
特征值计算 (使用积分图)	$f - e - a + b + 2 \times (c - d)$	$h + a - b - g - 3 \times (f + c - e - d)$	$2 \times (f + b + h + d) - (c + g + i + a + 4 \times e)$	$h + e - f - g - (d + a - b - c)$
计算复杂度*	6	8	10	7

\* 以操作符个数计算，不区分 +、-、× 之间的计算复杂度区别

由于分散矩形特征和类 Haar 特征一样，都使用了矩形模板，具有同样的特征形式，因此分散矩形特征同样可以采用积分图快速计算其特征值。表 2.2 总结了类 Haar 特征和分散矩形特征的特点。此外，因为两者的区别仅限于它们的特征结构上，因此在我们的实验中，分散矩形特征可以简单替换类 Haar 特征，而不需要对 Viola 和 Jones 的物体检测框架做任何改动。

## 2.3 实验结果和分析

我们实现了一个以 Viola 和 Jones 的物体检测框架为基础的人脸检测系统，使用同样的实验设置，分别采用类 Haar 特征和分散矩形特征训练生成两个人脸检测器，在 MIT 和 CMU 的正面垂直人脸测试集上进行实验对比和分析。

### 2.3.1 人脸训练集构造

我们在 Caltech<sup>3</sup>, BioID<sup>4</sup> 和 IMM<sup>5</sup> 人脸库上构建了实验所需的垂直向上正面人脸样本集。通过手工的方式，从这些库中裁剪了 2429 张不同尺寸的人脸。我们在人脸裁剪过程中遵循了如图 2.5 的规则：使用正方形裁剪，使得鼻尖大致在正方形中心；将该正方形等分成九份，鼻子应该基本落在中心区域，左右眼睛基本上应该分别落在左右上角区域，而嘴巴则大部分应该落在最下面的中间区域。如果找不到合适的裁剪正方形，满足该规则，则放弃对该人脸的处理。通常这种情况意味着该人脸有很大的平面内或平面外旋转，而这超出了本实验的测试范围。注意，这个规则是很主观的，只是辅助裁剪过程中人脸位置和大小的标定，使最终获得的人脸能够保持粗略的对齐。这些大小不一的人脸通过缩放对齐在一个  $19 \times 19$  像素分辨率的图像窗口中。然后经过左右对称，形成一个包含 4858 张正面人脸的人脸样本集。

对于训练集中的负样本，则相对来说很容易获得。首先，我们从 Caltech 中的非人脸库图像以及互联网上其他图像来源中，下载了一系列不包含人脸的图像。然后利用一个简单的程序，利用类似于滑动窗口检测的方法，在这些图像上自动收集了大量的非人脸图片。最后我们从这些图片中随机挑选了 5000 张形成初始负样本集，其余的图片用于训练中每个层次的负样本自举（Bootstrap）过程。

<sup>3</sup> <http://www.vision.caltech.edu/html-files/archive.html>

<sup>4</sup> <http://www.bioid.com/downloads/facedb/index.php>

<sup>5</sup> [http://www2.imm.dtu.dk/pubdb/views/publication\\_details.php?id=3160](http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3160)

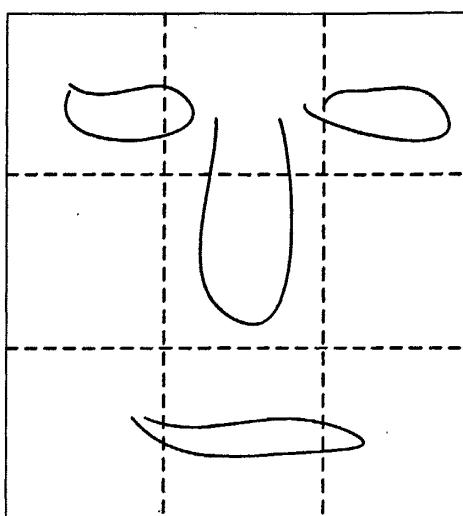


图2.5 训练集人脸裁剪规则示意图



图2.6 垂直向上正面人脸训练集人脸样例图片

图 2.6给出了我们构造的训练集中一些人脸图片。可以看出，该人脸集包含了多个个体在不同光照条件下不同表情的人脸。由于这些人脸是从不同的人脸库中提取，因此

光照条件变化非常大。某些人脸具有较大的平面内旋转和平面外旋转，一些人脸图像中还包含眼镜或者其他遮挡，其中甚至还有几张手工绘制的人脸。

### 2.3.2 实验设置

我们采用与 Viola 和 Jones 同样的训练算法，使用 AdaBoost 选取有区分能力的特征构成强分类器，通过性能指标驱动级联检测器的生成。但是在我们的系统中，并没有采用一个真正的独立的验证集来确认当前级联器的总体性能，对训练过程中每一层强分类器进行交叉验证并调整阈值。这是一个很费时间的操作步骤。在我们的分类器训练过程中，每个强分类器根据其在训练集上的检测率和误警率目标，独立的调整各自的阈值。该训练机制加快了训练过程，但是可能导致最终的级联分类器的过匹配/过优化(Overfitting)。

基于该训练机制，我们分别使用类 Haar 特征和分散矩形特征，训练人脸检测器。在训练过程中，每层强分类器都需要检测出所有的人脸，同时确保误警率不超过 50%。一旦最终的级联检测器的总体误警率达到  $2.5 \times 10^{-6}$ ，训练过程就结束。

### 2.3.3 训练结果

训练结果显示，这两个级联器都包含 17 层强分类器，其中使用类 Haar 特征训练的分类器包含 680 个特征，有大约 49.71% 的特征是三或四矩形模板特征；而使用分散矩形特征的分类器选择了 1422 个特征，其中有大约 86.22% 的特征是各种方向上错切、分离和重叠的分散矩形特征。

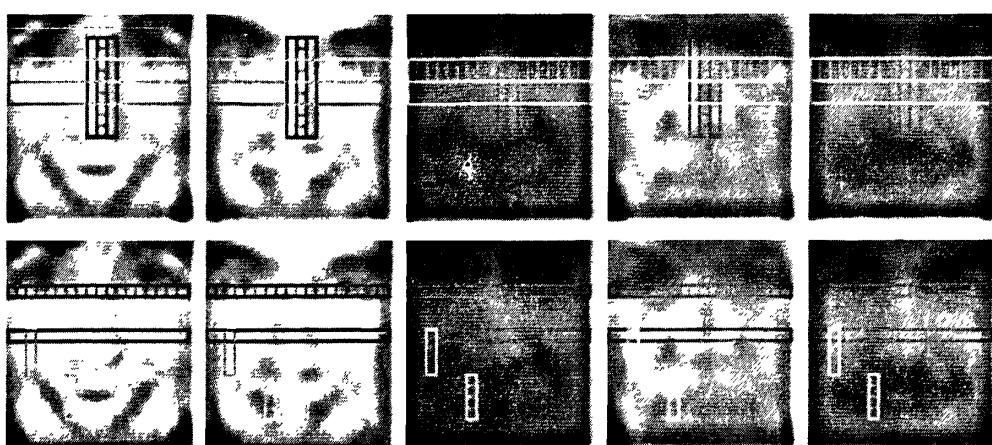


图2.7 AdaBoost 算法在学习过程中选取的前两个类 Haar 特征（上）和分散矩形特征（下）。红色（或者深色）表示第一个特征，绿色（或者浅色）表示第二个特征

图 2.7 显示了这两个分类器的前两个特征在五个典型的训练人脸上的分布情况。可见，第一个类 Haar 特征大致在鼻子上，第二个则在眼睛和脸颊上；第一个分散矩形特征和第二个类 Haar 特征相似，也处于眼睛和脸颊上。只不过因为没有区域相邻的限制，该分散矩形特征可以不用包含眼睛和脸颊之间的区域信息，而只是包括了眼睛（偏暗）和脸颊的高光（偏亮）的部分；而第二个分散矩形特征同样也选择了脸颊区域（偏亮）和鼻子下面的嘴角区域（偏暗）。

注意，第二个类 Haar 特征同时也是一个分散矩形特征，尽管和第一个分散矩形特征相似，但是并没有在基于分散矩形特征的检测器的前两次学习过程中被选中。可能的原因是因为眼睛和脸颊之间的区域信息并不是很具有区分能力，但是由于类 Haar 特征的子区域必须保持相邻对齐，不得不将该信息也包含在特征中，因而降低了该类 Haar 特征的区分能力。由此可见，通过去除相邻对齐的限制条件，确实给分散矩形特征的表达能力带来好处，2.3.5 小节的检测结果也经验性的验证了这一结论。

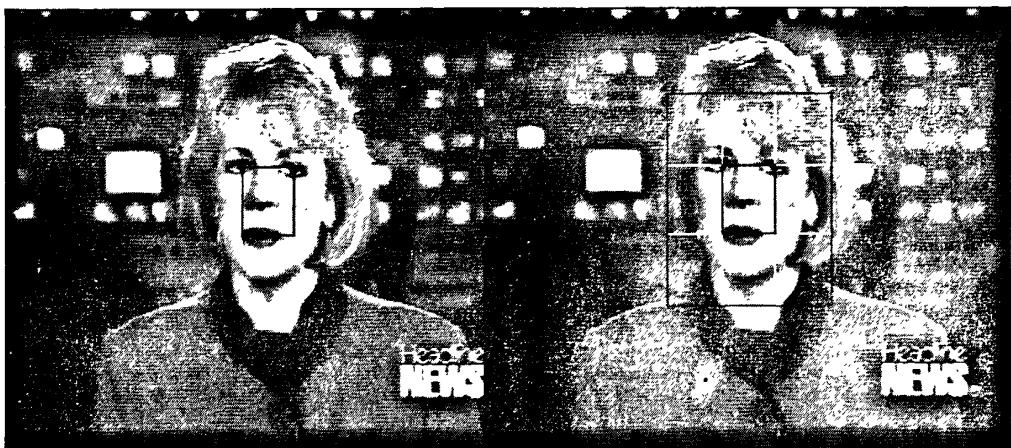
#### 2.3.4 检测标准

我们在 MIT 和 CMU 的垂直人脸测试集<sup>6</sup>上对两个级联检测器进行性能评估。该检测集包含 130 张图像和 511 张人脸，其中有些侧面人脸以及两张上下颠倒的人脸。在我们的实验中，我们并没有去除这些非正面垂直人脸。

该测试集还为每一张测试人脸提供了额外的关于人脸位置的真实信息，具体包括左眼、右眼、鼻子、左嘴角、嘴中心和右嘴角在测试图片中的坐标信息。我们的测试实验利用了这些信息来对检测的结果进行判断。

首先我们定义能覆盖这 6 个点的不经过旋转的最小矩形为最小人脸矩形，如图 2.8(a) 所示。当利用滑动窗口在图像上对区域块进行检测时，当且仅当该窗口包含最小人脸举行，并且窗口面积不超过最小人脸矩形的面积的四倍，我们认为这是一个正确的检测。其他的情况都被认为是一个误检。根据该检测标准，我们可以确定正确的检测可能分布的最大区域，如图 2.8(b) 所示。注意，该区域内的检测不一定就是正确的检测，但是该区域之外的检测一定是误检测。对比该最大区域和真实的人脸大小，可以看出，该定义是一个非常严格的检测标准。

<sup>6</sup> [http://www.vasc.ri.cmu.edu/idb/html/face/frontal\\_images/index.html](http://www.vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html)



(a) 利用 MIT 和 CMU 提供的人脸测试集的真实数据定义的最小人脸矩形

(b) 基于最小人脸矩形的检测标准所确定的正确检测可能分布的最大区域

图2.8 最小人脸矩形和最大正确检测分布区域

检测过程中，由于滑动窗口每次平移的步长通常是几个像素，因此会出现大量检测窗口重叠的情况。我们不对这些检测做任何后处理（合并、删除等），因为在某些情况这有可能会减少误检的次数，导致性能比较不准确。比如，一张脸上出现两个重叠的检测，一个检测窗口大于四倍的该人脸的最小人脸矩形，而另一个检测窗口却被该最小人脸矩形所包含。按照我们的检测标准，此时产生两个误检。但是如果采用一些后处理方法，比如平均这两个检测，那么很有可能导致只有一个误检，甚至是一个正确的检测。同样，我们对孤立的检测也不做任何类似删除这样的后处理。而对于一张脸上出现的多次正确检测，我们在最终的性能评测中只计算一次。从而确保在对人脸检测器检测效果评估中，不引入任何主观判断或者后期处理的偏向，使得实验结果更具有可靠性和可比较性。

### 2.3.5 检测结果

基于上一节所描述的客观检测标准，我们在 MIT 和 CMU 的检测集上对两个在同样的实验设置下训练所得的人脸检测器的性能进行评估。我们采用了类似 Viola 和 Jones 的检测方法。每个检测器以 1.0 像素的平移步长和 1.0 的缩放比例开始从上到下，从左到右，在测试图像上滑动窗口，对每个窗口检测。然后以 1.25 的缩放步长，按照同样的方式对测试图像进行多轮滑动窗口检测，直至滑动窗口的某一维的尺寸超过测试图像对

应的尺寸。按照该方法，每个检测器对测试集中 130 张图像的测试均扫描了 71,446,448 个子窗口。

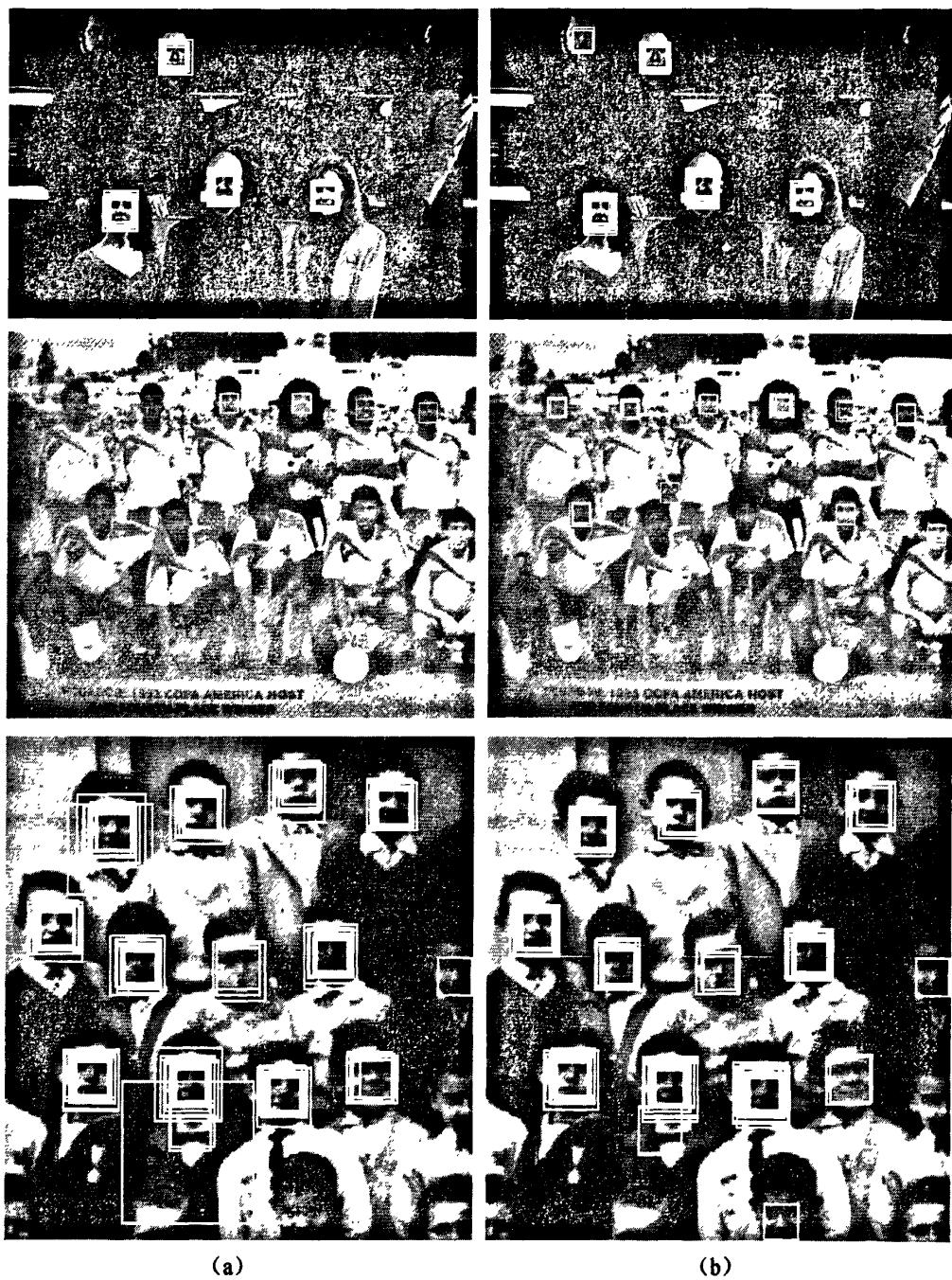


图2.9 使用类 Haar 特征(a)和分散矩形特征(b)的检测器的检测结果

图 2.9 显示了两个检测器在 MIT 和 CMU 人脸测试集上的一些检测结果，其中白色方框表示检测器的输出。除了美化排版所需要的缩放之外，输出结果没有经过任何后处

理。基于我们提出的检测标准，这三张测试图像中所出现的误检测数量和正确检测数量分别见表 2.3 和表 2.4。对比图 2.9 中的实际检测效果和表 2.3、表 2.4 中的数据，再次表明我们提出的检测标准是非常严格的。

表2.3 三张测试图像（见图 2.9）上的误检测数量

测试图像	类 Haar 特征检测器	分散矩形特征检测器
1	13	19
2	0	10
3	22	10

表2.4 三张测试图像（见图 2.9）上的正确检测数量

测试图像	类 Haar 特征检测器	分散矩形特征检测器
1	4	5
2	4	9
3	13	14

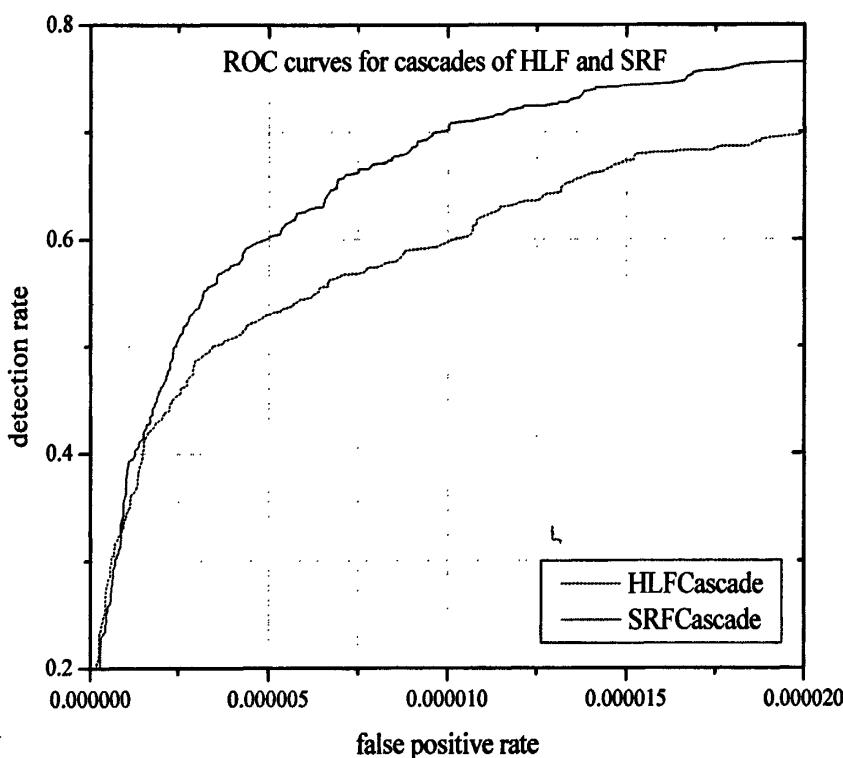


图2.10 类 Haar 特征（HLF）和分散矩形特征（SRF）的检测器的 ROC 曲线

两个检测器的 ROC 曲线如图 2.10 所示。该曲线是通过调整最后一层的分类器的阈值获得的。当该阈值从  $+\infty$  向  $-\infty$  调整时，检测率和误警率会同时提高，从 0.0 到某一个点。对比两条 ROC 曲线可以清楚的看出，基于分散矩形特征的级联检测器性能要优于基于类 Haar 特征的检测器。

注意，我们的检测器并没有展示出像 Viola 和 Jones 文章中的性能，这可能是因为第一，我们的训练集中每张人脸的分辨率只有  $19 \times 19$  像素，而 Viola 和 Jones 使用的是  $24 \times 24$  像素，因此训练的有效信息相对比较少；我们所采集的这些训练图片来自多个人脸库，除缩放对齐外，没有经过任何类似直方图均衡化（Histogram Equalization）等后处理，训练时的输入信息质量并不高，影响 Boosting 算法的泛化性能<sup>[101]</sup>；第二，我们在实验设置中并没有使用真正独立的验证集，使得最终的检测器可能发生过匹配；第三，我们采用了比较严格的检测标准，对检测结果不做任何处理，而 Viola 和 Jones 并没有明确说明他们的检测标准。因此我们实现了他们的算法，在同样的实验设置下，依照同样的客观检测标准，所给出的实验结果具有更好的可比较性。

## 2.4 本章小结

本章首先分析了局部区域特征抽取过程中的三个通用组件，提出了两种改进特征的方法。然后采用修改类 Haar 特征的特征结构的方式，提出了一种新的特征变体，我们称之为分散矩形特征。它同样采用了矩形模板，但是模板中各子区域并不需要在水平或垂直方向上保持相邻对齐。该特征结构上的变化使得分散矩形特征能更自由的表示区域的方位信息和几何信息，从而提高了特征的表达能力。同时该特征变体也能有效的利用积分图进行快速的计算，解决了大量特征所导致的计算量问题。

然后本章使用这两个特征，基于同样的实验设置条件下，通过 AdaBoost 算法分别学习了两个人脸分类器。训练结果验证了在特征结构中去除相邻对齐约束给分散矩形特征在表达能力上所带来的优势；基于同一个严格和客观的检测标准，这两个检测器在 MIT 和 CMU 的人脸测试集上的对比实验结果表明，基于分散矩形特征的检测器性能要优于基于类 Haar 特征的检测器。

在实验过程中，本章还构建了一个人脸训练集，定义了一个客观的基于最小人脸矩形的检测标准。该训练集和检测标准还会被后续章节所采用。

## 第3章 基于双阈值 AdaBoost 算法的物体检测

Boosting 算法是机器学习领域用来提高学习精度的重要算法，其主要思想是将一些粗糙的不是很精确的弱假设（Weak Hypothesis）组合起来，形成一个非常精确的强假设（Strong Hypothesis）。因而在设计复杂分类器时，不需要（通常也不可能）直接构建非常准确的分类规则，而是寻找比随机猜测略好的弱分类器，后者不仅比较容易获得，而且还能有效利用人的先验知识。AdaBoost 作为 Boosting 家族中最具代表性的算法，由于能够在学习过程中，自适应的调整错误率，而无需事先知道弱假设的正确率下限，因而被广泛应用于包含物体检测<sup>[44, 81-82, 102-103]</sup>在内的各种分类问题中。

本章首先介绍 AdaBoost 算法在 Viola 和 Jones<sup>[44]</sup>所提出的快速物体检测框架中的应用，然后简要分析了其他几种算法变体在该框架下所做的改进。在此基础上，本章提出了一种不同于这些变体的思想，即在弱假设中引入另一个阈值，通过增强弱假设自身的准确率来提高强分类器的分类能力，而不是修改训练机制或者寻找更好的特征。由于该算法使用了两个阈值的弱假设和相应的弱假设学习算法，因此被称为双阈值 AdaBoost 算法。我们将通过对类 Haar 特征的学习和人脸检测实验分析该算法的优缺点。

### 3.1 AdaBoost 算法

图 3.1 描述了 Viola 和 Jones<sup>[44]</sup>所采用的 AdaBoost 算法。该算法的基本思想是：首先给定任意一个弱学习算法和训练集  $\{(x_i, y_i) | x_i \in X, y_i \in Y = \{+1, -1\}, i = 1, 2, \dots, M\}$ ， $X$  表示某个域或者实例空间，在本文中就是静态图像空间，而  $Y$  表示分类标签，+1 为目标分类，其对应的图像称为正样本，而 -1 代表非目标分类，其对应的图像称为负样本。初始化训练集的权重，如每个样本被指定权重为  $1/M$ 。然后调用弱学习算法进行  $T$  次迭代。每次迭代都训练产生一个弱假设  $h_t$ ，然后根据训练结果更新训练集的权重，对于分类错误的样本则赋予更大的权重，使得下一次迭代更加关注这些训练样例。迭代过程中产生的弱假设根据各自的预测效果进行加权组合，预测效果好的，相应的权重越大，形成最终的强假设。

一般情况下，任何错误率低于 50% 的弱假设都可以被选择进行线性组合，构成最终的强假设。然而，选择性能更好的弱假设可以使得强假设更加鲁棒和高效。这也是本章研究方法的一个重要动机。

弱假设设计和学习依赖于实际的应用。在绝大多数的物体检测系统中，通常将弱假设和一个特征联系起来，从而使得 AdaBoost 算法中弱假设的学习过程成为特征的选取过程。每次循环迭代，AdaBoost 算法就从大量的可获得的特征中，选取了一个具有区分能力的特征。这些特征最后通过弱假设的线性组合，形成相应的物体检测器。

给定带分类标签的样本  $(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)$ ，其中每个样本  $x_i \in X$ ，标签  $y_i \in \{+1, -1\}$ ， $M = P + N$ ， $P$  表示正样本（标签  $y_i = +1$ ）的数量， $N$  表示负样本（标签  $y_i = -1$ ）的数量

初始化样本的权重为  $w'_i, i = 1, 2, \dots, M$

从  $t=1$  到  $T$  循环

1) 规则化样本权重， $w'_i \leftarrow \frac{w'_i}{\sum_{j=1}^M w'_j}$ ，使得  $w'_i$  成为一个概率分布

2) 使用给定的弱学习算法，基于样本的概率分布  $w'_i$ ，学习相应的弱假设

$$h_t : X \mapsto \{+1, -1\}$$

3) 计算弱假设的分类错误

$$e_t = \frac{1}{2} \sum_{i=1}^M w'_i |h_t(x_i) - y_i|$$

4) 更新样本的权重： $w'_{t+1} = w'_t \beta_t^{1-h_t(x_i)-y_i}$ ，其中  $\beta_t = \frac{e_t}{1-e_t}$

输出最终的强假设

$$h(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^T \alpha_i h_i(x) \geq 0 \\ 0 & \text{otherwise} \end{cases}, \text{ 其中 } \alpha_i = \log\left(\frac{1}{\beta_i}\right)$$

图3.1 AdaBoost 算法框架

在 Viola 和 Jones 的物体检测框架中，他们将类 Haar 特征与弱假设关联，使用 AdaBoost 算法从过完备的类 Haar 特征集中，选取有区分能力的特征。每个弱假设是一个简单的二值分类函数：

$$h(x; f, p, \theta) = \begin{cases} 1 & \text{if } pf(x) < p\theta \\ -1 & \text{otherwise} \end{cases}, \quad (3.1)$$

其中  $x$  为输入图像， $f$  为某个特征， $f(x)$  表示该特征  $f$  在图像  $x$  上的特征值， $\theta$  是该二值函数的阈值，偏置  $p \in \{-1, +1\}$  用于确定不等式的方向，如图 3.2 所示。

弱学习器（算法）用于为特征  $f$  确定合适的阈值  $\theta$  和  $p$ ，使得该假设能够尽好的区分目标物体和非目标物体。该弱学习过程是通过在训练集上的有监督学习实现的。

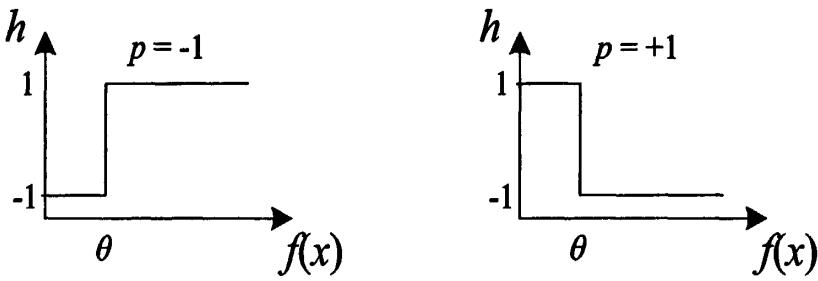


图3.2 单阈值二值弱假设函数曲线

给定带分类标签的样本  $(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)$  及其概率分布  $w_i$ , 和某个特征  $f$ , 该算法确定使得二值弱假设分类错误最小的阈值  $\theta$  和偏置  $p$ . 将样本以及权重按照特征值  $f(x_i)$  排序, 获得一个有序的序列  $(x'_i, y'_i, w'_i)$ . 在该序列中引入额外的一项  $(x'_{M+1}, y'_{M+1}, w'_{M+1})$ , 其中  $f(x'_{M+1}) = +\infty$ . 计算当  $\theta$  取值为  $f(x'_i)$  时的分类错误

$$\begin{cases} e_{+1} = \sum_{y'_j=+1} w'_j & \text{when } p = +1 \\ e_{-1} = \sum_{y'_j=-1} w'_j & \text{when } p = -1 \end{cases}$$

初始化  $e_{+1} = e_{-1}$ ,  $e_{-1} = e_{+1}$ ,  $i_{+1} = i_{-1} = 1$

从  $i=2$  到  $M+1$  循环

累加式的更新分类错误

$$\begin{cases} e_{+1} \leftarrow e_{+1} - y'_{i-1} w'_{i-1} \\ e_{-1} \leftarrow e_{-1} + y'_{i-1} w'_{i-1} \end{cases}$$

如果  $e_{+1} < e_{-1}$ , 则  $e_{+1} = e_{-1}$ ,  $i_{+1} = i$

如果  $e_{-1} < e_{+1}$ , 则  $e_{-1} = e_{+1}$ ,  $i_{-1} = i$

如果  $e_{+1} < e_{-1}$ , 则  $p = +1$ ,  $\theta = f(x'_{i_{+1}})$ ;

否则  $p = -1$ ,  $\theta = f(x'_{i_{-1}})$

图3.3 单阈值的弱学习算法

在给定样本权重下, 弱假设对某个训练集的分类错误被计算为

$$e_t = \frac{1}{2} \sum_{i=1}^M w'_i |h_t(x_i) - y_i| \quad (3.2)$$

由此可知,  $e_t$  与对应的弱假设  $h_t$  所取的阈值  $\theta$  和偏置  $p$  相关, 可以通过最小化错误率来确定阈值和偏置的值。首先对所有训练样本按照特征值从小到大排序, 然后从最小的特征值开始, 一次性遍历所有可能的阈值和偏置取值, 一直到最大的特征值结束。在该过

程中记录最小的分类误差及对应的阈值和偏置，那么一次性扫描完毕后，就确定了该弱分类器及其错误率。由于错误率的累加性，该遍历过程中，对错误率的计算不需要再次扫描，因此该算法是线形的。具体算法见图 3.3。

### 3.2 双阈值弱假设

尽管任何比随即猜测稍好的弱分类器都可以被增强成非常精确的强分类器，更好的弱分类器却能使强分类器更加鲁棒和高效。因此已经有许多工作，围绕 AdaBoost 的最佳弱分类器的选取机制，提出了各种算法变体。AsymBoost<sup>[104]</sup>通过修改权重，从而影响错误率的计算，间接的改变了算法对特征的选取机制，使得算法更偏重于能正确分类正样本的特征；FloatBoost<sup>[105]</sup>采用了一种事后补救的方法，通过搜索，去掉之前已经被加入到强分类器但是却并不是很高效的弱分类器；EntBoost<sup>[103]</sup>则采用了完全不同的错误率计算方法。

不同于这些方法，本节从弱假设自身出发，考虑通过引入另一个阈值来提高弱分类器的分类能力，从而使得强分类更加高效和鲁棒。这种假设我们称之为双阈值弱假设。

因为引入了另一个阈值，双阈值弱假设能够在更细的粒度上划分正样本和负样本，因此也就能获得更好的分类效果。图 3.4 显示了一个简单直观的例子。根据单阈值的弱学习算法（图 3.3），针对该特征的弱分类器具有： $\theta = c$ ， $p = -1$ 。此时该特征值空间被划分成两个区间：下区间  $(-\infty, \theta)$  和上区间  $[\theta, +\infty)$ 。所有特征值落入上区间的样本都被分类成物体。在这个简单的例子中，可见此时的分类错误是所有的在上区间内的方形面积总和。如果采用双阈值弱假设，我们现在可以很容易的通过人工的方式（本章后续章节将会介绍双阈值弱假设的学习方法，通过一个线形算法确定两个阈值）确定这两个阈值  $\theta = a$  和  $\theta' = b$ 。这两个阈值将该特征值空间划分为三个区间：中间区间  $[\theta, \theta')$  和上下两个端点区间  $(-\infty, \theta)$  和  $[\theta', +\infty)$ 。所有特征值落入后两个区间的样本都被分类成物体。实际上后两个区间可以看作一个区间，因为没有什么特征值能落在  $\pm\infty$  之外。在这个简单的例子中（图 3.4），该双阈值弱假设的分类错误就是两个端点区间内的方形和中间区间内的圆形面积之和，这要比之前单阈值假设的错误要小。其原因就在于双阈值能够将特征值空间划分的更细，因此也就能分类的更加准确。

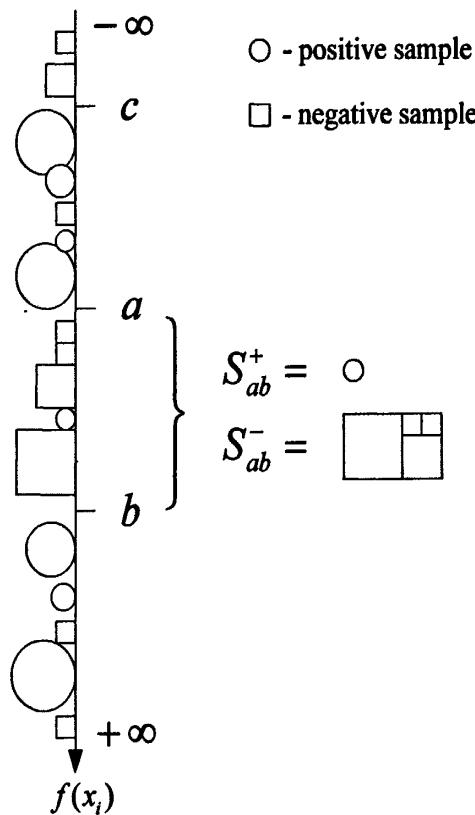


图3.4 一个用于说明弱假设及其学习算法的简单例子。示例包含了 17 个已经按照某个特征值从小到大排列的样本，其中圆形表示正样本，方形表示负样本，其面积代表样本权重

基于这个直观的例子，我们可以很自然的扩展单阈值的弱假设公式(3.1)。通过引入第二个阈值，双阈值弱假设被定义为

$$h(x; f, p, \theta, \theta') = \begin{cases} 1 & \text{if } pf(x) < p\theta \text{ or } pf(x) \geq p\theta', \text{ when } p = +1; \text{ or} \\ & \text{if } pf(x) < p\theta \text{ and } pf(x) \geq p\theta', \text{ when } p = -1 \\ -1 & \text{otherwise} \end{cases} \quad (3.3)$$

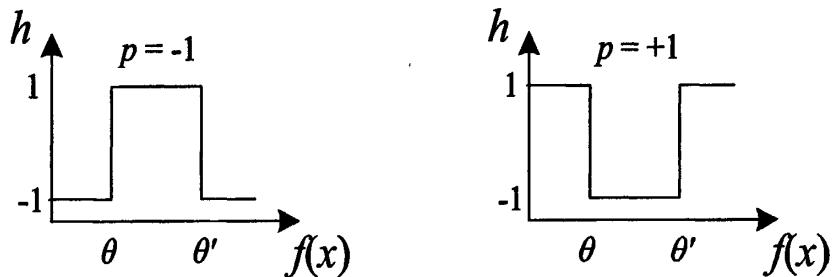


图3.5 双阈值二值弱假设函数曲线

图 3.5 显示了该二值函数的意义。注意，不失一般性，这里我们假设  $\theta < \theta'$ 。另外需要注意的是在双阈值弱假设的定义中我们并没用引入额外的偏置  $p'$ ，这是因为公式(3.3)中已经隐式的假设  $p' = -p$ 。否则如果  $p' = p$ ，那么公式(3.3)就转换为

$$h(x; f, p, \theta, \theta') = \begin{cases} 1 & \text{if } pf(x) < p\theta \text{ or } pf(x) < p\theta', \text{ when } p = +1; \text{ or} \\ & \text{if } pf(x) < p\theta \text{ and } pf(x) < p\theta', \text{ when } p = -1 \\ -1 & \text{otherwise} \end{cases}$$

此时两个阈值将特征值空间分别划分为两个上下区间，而不能形成中间区域。因此实际上就只有一个阈值在对特征值空间起划分作用。而公式(3.3)也能通过将某个阈值设定为  $\pm\infty$  来表达这种情况，因此我们不需要引入额外的偏置。

注意，如果公式(3.3)的某个阈值取值为  $\pm\infty$  时，公式(3.3)也就退化为公式(3.1)。当使用最小化弱假设的错误率来确定阈值和偏置的弱学习算法时，这意味着双阈值弱假设总能获得与相应的单阈值假设至少同样的分类性能。当然在大部分情况下，我们期望为双阈值弱假设选取合适的阈值从而获得更低的错误率。

### 3.3 双阈值弱学习算法

在前一节中，我们借助一个简单的例子，定义和讨论了双阈值弱假设。现在的问题就是如何确定这两个阈值与偏置。

一个很直观的考虑是在按特征值排序后的序列上，使用两个嵌套的循环对每对可能的阈值及偏置进行错误率检查，最后确定最佳值。该算法确保能找到最优的阈值和偏置，但是算法复杂度为  $O(n^2)$ ，其中  $n$  为训练集的大小，通常这是一个非常大的数值（在我们的实验中， $n > 9000$ ）。相比较于单阈值学习算法的  $O(n)$  复杂度，这意味着整个 AdaBoost 学习过程耗时是原来的  $n$  倍，训练需要持续数月才能完成，这对研究和开发来说都是不能容忍也是支付不起的。

另一个直接的实现就是基于单阈值假设的学习算法（图 3.3）。在学习获得第一个阈值之后，再次应用该学习算法，从第一个阈值处开始环形遍历该序列，从而获得第二个阈值。该算法调用了两次已有的线性算法，因此仍然是线性的。但是在大多数情况下，该算法都会退化成为单阈值学习算法。如图 3.4 所示，第一次遍历该序列确定第一个阈值  $\theta = c$ ；从  $\theta$  处开始第二次环形遍历，确定第二个阈值  $\theta' = +\infty$ ，所有特征值落入中间

区间的样本都被分类成物体。这完全与单阈值假设分类结果一样。该算法存在的主要问题是第一次遍历所确定的阈值并不一定是最优的两个阈值当中的一个，而第二次遍历是基于第一次遍历的结果的，因此导致该算法不能确定最优的阈值。

本节将提出一个能在线性时间内确定两个最优阈值的学习算法。为了描述和分析该算法，我们需要介绍一些符号表示，来改写弱假设的错误率（见公式(3.2)）。

### 3.3.1 符号表示

我们定义在特征值区间  $[a, b)$  上的正样本和负样本的区间权重和分别为

$$\begin{cases} S_{ab}^+ = \sum_{\substack{a \leq f(x_i) < b \\ y_i=+1}} w_i \\ S_{ab}^- = \sum_{\substack{a \leq f(x_i) < b \\ y_i=-1}} w_i \end{cases}, \quad (3.4)$$

如图 3.4 所示， $\pm\infty$  可以看作是该特征在给定训练集上的特征值的上下界。对于  $a > b$ ，我们定义  $S_{ab}^\pm = -S_{ba}^\pm$ ，从而使得区间权重和满足传递关系，即对于任何给定的  $c \in \mathbb{R}$ ，有

$$S_{ac}^\pm + S_{cb}^\pm = S_{ab}^\pm. \quad (3.5)$$

注意，对于给定的特征和训练集（包含权重）， $S_{-\infty+\infty}^\pm$  是常数。此外我们定义  $S_{ab}$  为带符号区间权重和

$$S_{ab} = S_{ab}^+ - S_{ab}^-. \quad (3.6)$$

### 3.3.2 双阈值弱学习算法

利用区间权重和，我们可以将双阈值弱假设的错误率表示为

$$\begin{aligned} e_t &= \frac{1}{2} \sum_{i=1}^M w_{t,i} |h_t(x_i) - y_i| \\ &= \min(S_{-\infty\theta}^+ + S_{\theta\theta'}^- + S_{\theta'\infty}^+, S_{-\infty\theta}^- + S_{\theta\theta'}^+ + S_{\theta'\infty}^-) \\ &= \min(S_{-\infty+\infty}^+ + S_{\theta\theta'}^- - S_{\theta\theta'}^+, S_{-\infty+\infty}^- + S_{\theta\theta'}^+ - S_{\theta\theta'}^-) \\ &= \min(S_{-\infty+\infty}^+ - S_{\theta\theta'}^-, S_{-\infty+\infty}^- + S_{\theta\theta'}) \end{aligned} \quad (3.7)$$

根据公式(3.7)，双阈值学习算法的目标就转换为查找最大和最小的带符号区间权重和  $S_{\theta\theta'}$ ，后者是经典的最大和连续子序列问题（Maximum-Sum Consecutive Subsequence Problem），可以通过线性的动态规划算法解决<sup>[106]</sup>。

但是公式(3.7)的推导过程并没有考虑到按照特征值大小排序的序列实际上是一个环序列，即双阈值将特征空间划分为两个端点区间后，所有特征值落入这两个区间的样本具有同样分类结果。为了使用动态规划算法，需要在合适的位置将该环序列展开，我们首先证明以下引理：

在环序列上使得错误率最小的阈值  $\theta$  必须满足：分别处于  $\theta$  上区间和下区间中距离  $\theta$  最近的两个特征值所对应的样本具有不同的分类标签。

也就是说阈值只可能在不同的样本类别对应的特征值之间取值。可以使用反证法证明：不失一般性，假设在一个环序列上已经确定了最优的  $n$  个（在本节中  $n=2$ ）阈值  $\{\theta_1, \theta_2, \dots, \theta_n, \theta_{n+1} | \theta_1 < \theta_2 < \dots < \theta_n, \theta_{n+1} = \theta_1\}$ 。这些阈值将环序列划分成  $n$  个区间  $\{r_1, r_2, \dots, r_n, r_{n+1} | r_{n+1} = r_1\}$ ， $r_i$  由  $\theta_i$  和  $\theta_{i+1}$  界定，所有特征值落入区间  $r_i$  的样本被分类为  $y(r_i)$ ；显然，相邻区间具有不同的分类标签，即  $y(r_i) \neq y(r_{i+1})$ （否则若  $y(r_i) = y(r_{i+1})$ ，则可以去除  $\theta_i$  而保持分类错误不变，最终可以获得满足该假设的  $n$  个阈值）。假设  $r_j$  和  $r_{j+1}$  中离  $\theta_j$  最近的特征值所对应的两个样本  $x'_j$  和  $x'_{j+1}$  具有同样的分类标签  $y$ ，因为  $r_j$  和  $r_{j+1}$  具有不同的分类标签，因此其中有一个样本分类错误。不失一般性，假设被分类错误的样本为  $x'_j$ ，那么可以稍微将  $\theta_j$  往  $\theta_{j+1}$  方向移动一下，如  $\theta'_j = f(x'_j)$ ，使得  $x'_j$  落入正确的区间。在此过程中，区间  $r_j$  和  $r_{j+1}$  上分类错误变小了，而其他区间都没有受到影响，因此用  $\theta'_j$  替换  $\theta_j$  后，新的  $n$  个阈值  $\{\theta_1, \theta_2, \dots, \theta'_j, \dots, \theta_n, \theta_{n+1} | \theta_1 < \theta_2 < \dots < \theta_n, \theta_{n+1} = \theta_1\}$  的总体分类错误更小，这与给定的最优阈值条件相矛盾，定理得证。

我们将符合该引理的特征值称为候选阈值。根据该引理，我们就可以将环序列在任意的候选阈值处展开，然后应用动态规划算法获取最优的阈值和偏置，图 3.6 显示了本文采用的一种双阈值弱学习算法。

注意图 3.6 中我们引入了一个合并操作。该操作通过一次线性扫描，将所有非候选阈值对应的样本合并，将权重转换为带符号的权值，并处理了环形序列首尾之间的问题。动态规划算法在合并后的列表上检查所有可能的阈值候选区间，最终通过公式(3.7)确定使得错误率最小的阈值和偏置。因此该算法的时间复杂度是  $O(n)$ ，获得的阈值和偏置也是最佳的。

给定带分类标签的样本  $(x_1, y_1), (x_2, y_2), \dots, (x_M, y_M)$  及其概率分布  $w_i$ , 和某个特征  $f$ , 该算法确定使得双阈值二值弱假设分类错误最小的阈值  $\theta$ ,  $\theta'$  和偏置  $p$

- 1) 将样本和权重按照特征值  $f(x_i)$  排序, 形成一个有序的序列  $(x'_i, y'_i, w'_i)$
- 2) 扫描该有序序列, 将其中具有相同分类标签的连续子序列, 即  $(x'_j, y'_j, w'_j), (x'_{j+1}, y'_{j+1}, w'_{j+1}), \dots, (x'_{k-1}, y'_{k-1}, w'_{k-1}), (x'_k, y'_k, w'_k)$  , 其中  $y'_j = y'_{j+1} = \dots = y'_{k-1} = y'_k$ ,  $j < k$ , 合并为一项  $\left( f(x'_j), y'_j, \sum_{i=j}^k w'_i \right)$ , 最终生成一个大小为  $\bar{M}$  的合并后的序列  $(f(\bar{x}_i), \bar{w}_i)$
- 3) 如果  $\text{sgn}(\bar{w}_{\bar{M}}) = \text{sgn}(\bar{w}_1)$ , 则  $\bar{w}_1 \leftarrow \bar{w}_1 + \bar{w}_{\bar{M}}, \bar{w}_{\bar{M}} = 0$
- 4) 在合并后的序列上应用动态规划, 确定最大和最小和连续子序列  
根据公式 (3.7), 确定最佳的阈值和偏置

图3.6 双阈值的弱学习算法

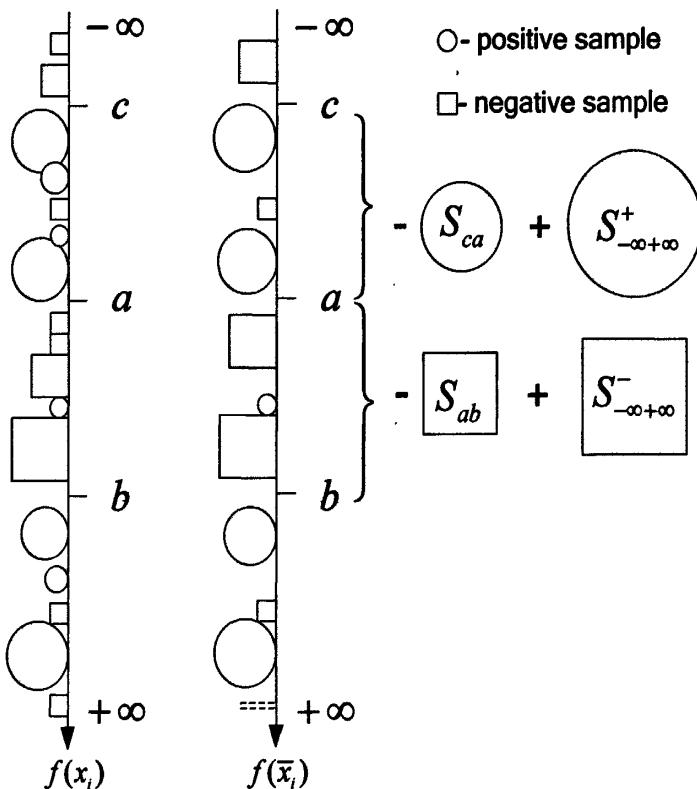


图3.7 双阈值弱学习算法的学习过程

将该算法应用到图 3.4 的简单例子上, 如图 3.7 所示, 首先对该序列进行按标签的连续性合并操作, 形成一个特征值和带符号的权值序列。注意在该合并操作过程中, 需要对特征空间两头的样本进行特殊处理, 如果两者属于同一个类别, 则也需要合并, 并将

合并结果放在起始处。然后在新生成的序列上应用动态规划算法，获取具有最大和最小和的子序列： $S_{ca}$  和  $-|S_{ab}|$ 。最后根据公式(3.7)确定  $S_{-\infty+\infty}^- - |S_{ab}| < S_{-\infty+\infty}^+ - S_{ca}$ ，因此选择最佳阈值  $\theta = a, \theta' = b$  和偏执  $p = +1$ 。

### 3.4 讨论

综合以上两节介绍的内容，我们可以将双阈值弱假设和双阈值弱假设学习算法应用到 AdaBoost 算法框架（图 3.1）中，形成一个双阈值 AdaBoost 变体。除了使最终的强分类器更加鲁棒之外，该变体相较于单阈值 AdaBoost 来说，在实际的检测问题中还有如下几个方面的好处：

1. 在训练时，因为双阈值弱分类器的分类误差比对应的单阈值弱分类器通常要小（至少相同），因此在同样的训练性能要求下，该变体能够收敛的更快，使用更少的弱分类器和层次（强分类器）；
2. 更少的弱分类器也就意味着更少的特征。这对于某些并没有大量候选特征可供选择的应用特别有帮助，双阈值能够更好的利用这些特征的区分能力；
3. 同时更少的弱分类器也意味着检测速度更快。对于 Viola 和 Jones 的检测框中所使用的双矩形模板的类 Haar 特征（见2.2.1小节）来说，计算一个弱分类器的分类结果一共需要 7 个操作，其中 6 个操作（见表 2.2）用来计算类 Haar 特征的特征值，剩余 1 个操作用来比较该特征值和对应弱假设的阈值。而对于双阈值弱假设来说，一共需要 8 个操作来计算分类结果，因为多了一个阈值比较操作。因此一个采用双阈值弱假设的分类器所使用的弱分类器数量只要是对应的单阈值分类器的弱分类器数量的 7/8 倍或者更少，其检测速度也就更快。如果是使用三矩形或者四矩形模板的类 Haar 特征，那么该条件就更容易达到。实际上我们的实验就展示了这种情况。
4. 使用该变体进行特征选取，还可以发现一些被单阈值 AdaBoost 所忽略的特征。如图 3.4 所给出的简单例子中，如果使用单阈值弱假设，那么因为错误率较高，该特征很有可能被认为无效；而如果使用双阈值弱假设，那么情况就可能相反。因此使用该变体，将会有更多有效的特征可以从中选取，也就有可能产生一个分类错误更小的强分类器。

此外，该变体也很容易跟其他变体的思想相结合，比如 FloatBoost，从而进一步提高学习效果。

该变体的有效性也可以通过边界（margin）<sup>[107]</sup>来解释。Schapire 等人从边界的角度对泛化误差做了分析，样本  $(x, y)$  的边界  $\text{margin}(x, y)$  定义为

$$\text{margin}(x, y) = y \sum_i \alpha_i h_i(x) \quad (3.8)$$

其中  $\alpha_i = \frac{1}{2} \ln \left( \frac{1 - e_i}{e_i} \right)$ ,  $\text{margin}(x, y) \in [-1, 1]$ 。边界可以看成是对预测函数的预测结果的可信度。较大的正边界表示可信度高的正确预测，而较大的负边界则表示可信度高的错误预测。利用边界，Schapire 等人证明了对于任何  $\theta > 0$ ，泛化误差以很高的概率不会超过上限

$$\hat{\Pr}(\text{margin}(x, y) \leq \theta) + \hat{O}\left(\sqrt{\frac{d}{M\theta^2}}\right) \quad (3.9)$$

其中  $\hat{\Pr}(\cdot)$  表示在训练集上的经验概率， $d$  是学习算法的 VC 维。

因为每个双阈值弱假设对样本的边界几乎总是要比对应的单阈值假设所产生的边界大，因此尽管双阈值弱假设的 VC 维数高，但是公式(3.9)中的第一项变小；或者说为了维持公式(3.9)中第一项不变，就必须选择更大的  $\theta$ ，这就抵消了第二项中因为 VC 维增大的结果。然而，该变体确实对噪音更加敏感，如果在有噪音的数据上最大化最小边界将会导致过匹配问题。

### 3.5 实验结果与分析

我们基于 Viola 和 Jones 的物体检测框架，实现了前文描述的 AdaBoost 算法变体，采用类 Haar 特征，训练了多个人脸检测器，并和原来的 AdaBoost 算法产生的结果进行了对比。实验结果经验性的验证了我们在3.4小节的讨论。

#### 3.5.1 实验设置和检测标准

本章的所有实验均采用了和前一章实验同样的人脸训练集。该人脸训练集包含 4858 张正面人脸的正样本图片，和 5000 张非人脸的负样本图片。具体构建过程和人脸样例

图片见2.3.1小节。这些训练图片被统一缩放到  $19 \times 19$  像素分辨率的图像窗口中，所有图片均没有进行任何类似高斯模糊、直方图均衡化等后处理。

本章所有人脸级联器的训练过程中，我们均采用了和2.3.2小节描述的同样设置。在训练过程中，每层强分类器都需要检测到所有的人脸，同时确保误警率不超过 50%，一旦最终的级联检测器的总体误警率达到  $2.5 \times 10^{-6}$ ，训练过程就结束。此外，为了验证3.4小节中关于过匹配等问题的讨论，我们还引入了一个手工控制的训练模式。在该训练模式中，可以手工指定训练过程中每个层次使用的特征数量，一旦最终的级联检测器层数超过某一个指定的数值，训练过程就结束。

同样，本章也采用了 MIT 和 CMU 的垂直向上正面人脸测试集，来评估训练的各个级联器的检测性能。我们也采用了前一章中提出的基于最小人脸矩形的检测标准（见2.3.4小节），当利用滑动窗口在图像上对区域块进行检测时，当且仅当该窗口包含最小人脸举行，并且窗口面积不超过最小人脸矩形的面积的四倍，我们认为这是一个正确的检测。其他的情况都被认为是一个误检。本章也不对重叠或者孤立的检测结果做合并或者删除等后处理。

### 3.5.2 训练结果

我们首先采用手工控制的训练模式，分别使用单阈值 AdaBoost 和双阈值 AdaBoost 算法进行类 Haar 特征选取，训练生成两个人脸级联器，我们用符号 ACH 和 ZCH 分别指代这两个级联器。

ACH 和 ZCH 两个级联器都具有 10 个层次，每个层次使用的特征数量分别为 7, 9, 12, 15, 18, 22, 26, 30, 35, 40。因此两个级联器都使用了 214 个特征。通过对比，我们发现其中 120 个特征是两个级联器公有的（当然，具体的阈值和偏置以及分类错误是不同的）。也就是说，在训练过程中，ZCH 发现了 94 个（约 43.92%）被 ACH 忽略的有效特征。此外我们发现，ZCH 中只有 3 个（约 1.40%）特征对应的假设退化为单阈值假设（即两个阈值中的其中一个取值为  $\infty$ ）。

之后，我们采用性能驱动的方式，使用单阈值 AdaBoost 和双阈值 AdaBoost 算法进行自动训练，生成两个人脸级联器，我们用符号 AC 和 ZC 分别指代这两个级联器。训练的结果如表 3.1 所示。

表3.1 AC 和 ZC 的训练结果

	AC	ZC
# Layers	17	16
# Features	680	320
	=7+10+14+22+27+33 +38+39+36+40+53+53 +54+61+63+63+67	=6+10+12+13+17 +17+18+16+19+24+28 +31+32+33+34
Overall fp rate	$2.22158 \times 10^{-6}$	$2.14264 \times 10^{-6}$

从表 3.1 中可以看出，ZC 和 AC 虽然使用了差不多的层数（只差一层），但是由于 ZC 使用的双阈值弱假设几乎总是比对应的单阈值假设的错误率小，因此几乎每一层都能以更少的特征达到同样的训练性能指标，导致最终的检测器中，ZC 使用的特征不到 AC 特征总数的一半，而总体上训练达到的误检率却还要低一些。

同样我们也对比了 ZC 和 AC 这两个级联器中的所有特征，发现它们共享了 279 个类 Haar 特征。ZC 中还有 41 个（约 12.81%）特征是 AC 所忽略的。该数据要比之前手动训练的结果低，一方面是因为 ZC 本身所使用的特征就少，另一方面是因为 AC 相对来说选取了大量的特征，使得其中与 ZC 中的某个特征一致的可能性提高。若是以单阈值 AdaBoost 算法所选择的特征总数来论，那么使用手工控制的训练模式所生成的两个人脸级联器之间的特征共享率为 56.07%，而使用性能指标驱动的训练模式所生成的两个人脸级联器之间的特征共享率为 41.03%。因此双阈值 AdaBoost 算法确实能够发现一些被单阈值 AdaBoost 算法忽略的特征。此外，ZC 中仅有 2 个（约 0.62%）特征退化为单阈值特征。

### 3.5.3 检测结果

我们在 MIT 和 CMU 的垂直向上正面人脸测试集上，使用滑动窗口的方法，对这四个级联器进行性能评估。每个检测器以 1.0 像素的平移步长和 1.0 的缩放比例开始从上到下，从左到右，在测试图像上滑动窗口，对每个窗口检测。然后以 1.25 的缩放步长，按照同样的方式对测试图像进行多轮滑动窗口检测，直至滑动窗口的某一维的尺寸超过测试图像对应的尺寸。按照该方法，每个检测器对测试集中 130 张图像的测试均扫描了 71,446,448 个子窗口。

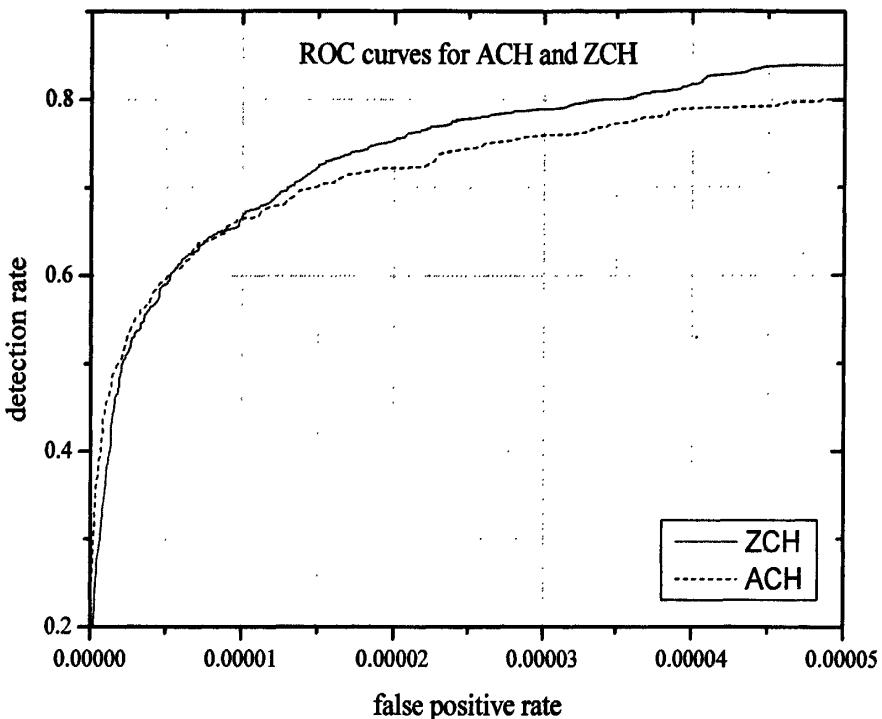


图3.8 采用人工控制的训练方式, 分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器 (ACH 和 ZCH) 的 ROC 曲线

我们依照2.3.5小节描述的 ROC 曲线绘制方法, 分别绘制了 ACH 和 ZCH, AC 和 ZC 的对比曲线, 如图 3.8 和图 3.9 所示。可以看出, 使用手工控制方式训练的两个人脸检测器, 虽然在级联层次上和使用的特征数量上完全相同, 但是使用双阈值 AdaBoost 算法训练的检测器的性能要明显优于单阈值 AdaBoost 算法训练的检测器; 而对于使用性能驱动的训练方式所生成的两个人脸检测器来说, 尽管使用双阈值 AdaBoost 算法在训练过程中所选取的特征数不到单阈值 AdaBoost 算法所选取的特征数的一半, 级联层次还少一层, 但是最终性能相差不大。

从两个训练模式生成的检测器的性能对比中我们可以看出, 使用双阈值 AdaBoost 算法在训练过程中比较容易产生过匹配问题, 这也是为什么 ZCH 性能明显优于 ACH, 而 ZC 却并没有优于甚至还稍稍弱于 AC 的一个重要原因。导致过匹配的主要因素是, 第一, 双阈值弱假设的 VC 维数高, 在训练过程中对训练集样本的边界几乎总是要比对应的单阈值假设所产生的边界大, 因此在我们构建的不经过任何处理的人脸训练集上, 很容易因为噪音问题, 造成最大化最小边界的负作用; 第二, 我们在训练过程中, 没有独立验证集, 对训练结果进行逐层的验证, 来减小过匹配的可能性。

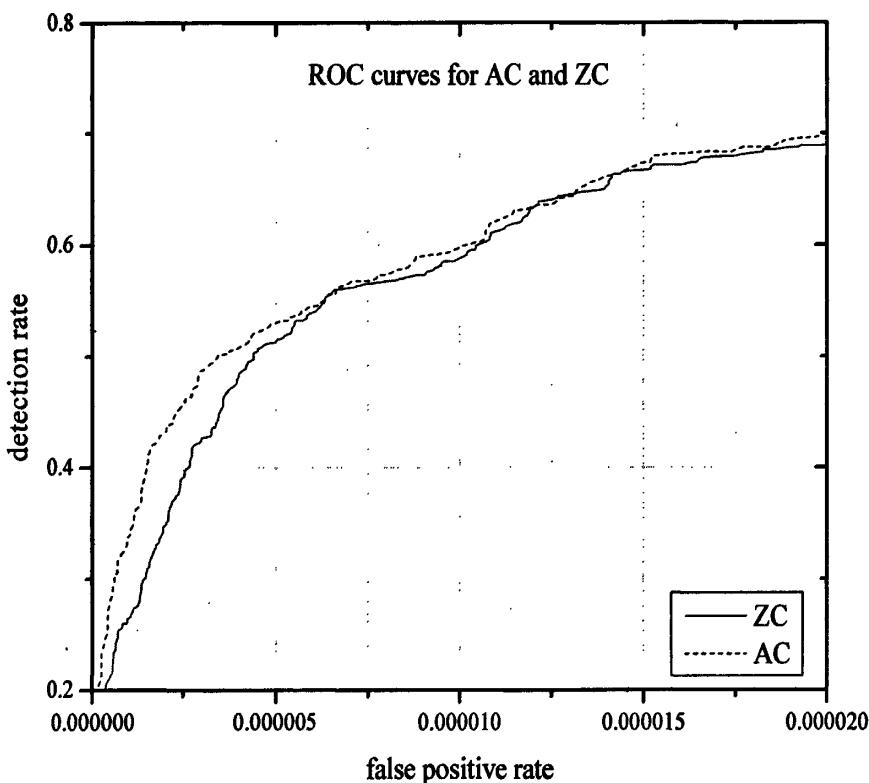


图3.9 采用性能驱动的训练方式, 分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器 (AC 和 ZC) 的 ROC 曲线

表3.2 三张测试图像 (见图 3.10) 上的误检测数量

测试图像	单阈值 AdaBoost 算法训练的类 Haar 特征检测器	双阈值 AdaBoost 算法训练的类 Haar 特征检测器
1	7	25
2	49	75
3	495	521

表3.3 三张测试图像 (见图 3.10) 上的正确检测数量

测试图像	单阈值 AdaBoost 算法训练的类 Haar 特征检测器	双阈值 AdaBoost 算法训练的类 Haar 特征检测器
1	7	9
2	9	9
3	57	57

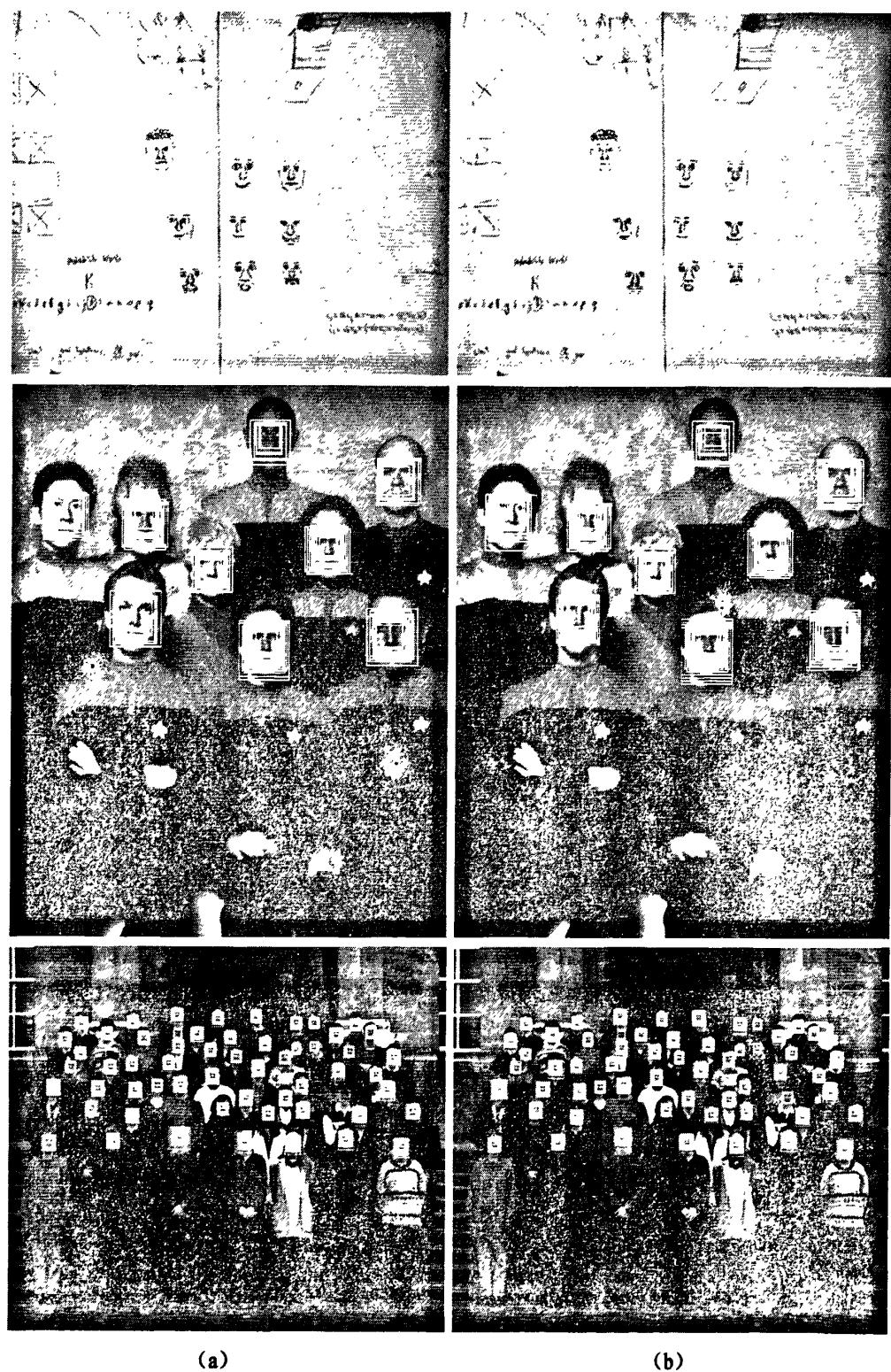


图3.10 单阈值(a)和双阈值(b)AdaBoost 算法训练的类 Haar 特征检测器 (AC 和 ZC) 的检测结果

图 3.10 显示了 AC 和 ZC 两个检测器在 MIT 和 CMU 人脸测试集上的一些检测结果，其中白色方框表示检测器的输出。除了美化排版所需要的缩放之外，输出结果没有经过任何后处理。基于最小人脸矩形的检测标准，这三张测试图像中所出现的误检测数量和正确检测数量分别见表 3.2 和表 3.3。对比图 3.10 中的实际检测效果和表 3.2、表 3.3 中的数据，再次表明基于最小人脸矩形的检测标准是非常严格的。而且从这三张测试图像上的误检测数量来看，该严格检测标准实际上对双阈值 AdaBoost 算法训练的类 Haar 特征检测器带有严重的偏向（即如果稍微放松该标准，那么该检测器的错误率将会大大降低，而单阈值 AdaBoost 算法训练的类 Haar 特征检测器的错误率基本上保持不变），这也是双阈值 AdaBoost 算法训练的类 Haar 特征检测器并没有优于单阈值 AdaBoost 算法训练的类 Haar 特征检测器的原因之一。

此外实验还给出了 AC 和 ZC 在测试集上的测试时间。在相同的计算机上，单阈值 AdaBoost 算法训练的类 Haar 特征检测器检测测试集中的 130 张图片共花费时间 1,167,843 毫秒，而双阈值 AdaBoost 算法训练的类 Haar 特征检测器则只花了 977,269 毫秒。注意这两个时间均包括了将这 130 张图片写到磁盘上所花费的时间。

### 3.6 本章小节

本章详细描述了 AdaBoost 机器学习算法，以及它如何在 Viola 和 Jones 的物体检测框架中实现特征选取和强分类器构建。不同于以往对 AdaBoost 算法的改进工作，本章从 AdaBoost 使用的弱假设自身出发，通过引入另外一个阈值，提高弱假设对特征值空间的类别划分精度，确保产生分类错误更小的弱分类器，从而使得经过 AdaBoost 算法增强后的分类器更加鲁棒和高效。

为了快速确定弱假设的两个最优阈值，本章分析了弱假设阈值可能的取值位置（候选阈值），提出了一个线性最优的弱学习算法。通过将弱假设对应的特征值空间中的样本环序列在候选阈值处展开，该算法将确定最优阈值问题转换为最大（最小）和连续子序列问题，后者可以使用动态规划算法，通过一次性遍历展开后的序列，确定该弱假设的最优阈值和偏执。

结合双阈值弱假设和双阈值弱学习算法，我们讨论了双阈值 AdaBoost 算法在使用的特征数量、检测性能等方面的好处。基于前章所提出的客观检测标准，在 MIT 和 CMU

人脸测试集上的实验结果经验性的验证了我们对该算法变体在特征选取方面的优势的讨论。

## 第4章 双阈值 AdaBoost 选取的分散矩形特征

我们在第二章介绍了分散矩形特征，根据训练结果经验性的分析了分散矩形特征在物体表示方面优于类 Haar 特征的原因，通过人脸检测的实验，也表明了基于分散矩形特征的分类器性能要优于基于类 Haar 特征的分类器。

在第三章中，我们介绍了双阈值 AdaBoost 算法，讨论了该变体较单阈值 AdaBoost 算法在特征选取和检测性能方面的好处，以及可能产生的过匹配问题。基于类 Haar 特征，在人脸检测实验中，通过对比单阈值和双阈值 AdaBoost 学习的分类器性能，验证了我们对该算法优势的讨论。

这两章的研究内容分别从特征表示和特征选取（学习）两个层面对 Viola 和 Jones 的物体检测框架进行了改进。自然的，我们就考虑将分散矩形特征和双阈值 AdaBoost 算法这两个变体同时替换原来的类 Haar 特征和单阈值 AdaBoost 算法进行进一步的实验研究。由于这两个改进所做的修改都局限于各层面内部，特征表示和特征选取之间的接口并没有变化，因此实验过程中，不需要过多修改原来的物体检测框架，这也是两个变体在实现上的优势。

实验结果一方面再次验证了前两章的讨论，另一方面也表明，分散矩形特征较类 Haar 特征具有更好的抗过匹配的能力。这是第二章实验所没有显示的，引起了我们对分散矩形特征的重新思考。我们发现非退化的分散矩形特征实际上等价于满足一定几何关系的多个类 Haar 特征。因此该分散矩形特征对应的弱假设对图像区域做出的检测，实际上是多个类 Haar 特征对应的弱假设对该图像区域判断结果的综合，因而更加鲁棒。

这是因实验结果而后引发的重新思考，但是我们在对本章内容进行组织时，采用的顺序是，先给出任意一个非退化分散矩形特征与多个类 Haar 特征等价的严格证明，然后再描述本章的实验过程和结果。

### 4.1 分散矩形特征与类 Haar 特征的等价性

分散矩形特征和类 Haar 特征之间最重要的区别在于类 Haar 特征的子区域不仅要求形状大小一样，而且在水平或者垂直方向必须相邻对齐；而分散矩形特征只需要保持子区域的形状大小一致，这也是分散矩形特征的名字由来。由于没有了子区域相邻对齐的

限制，导致分散矩形特征在单位区域中的数量远远超过类 Haar 特征，因此我们限制分散矩形特征只能采用双矩形模板。本小节将讨论双矩形类 Haar 特征和分散矩形特征之间的等价关系。

由公式(2.1)可知，对于双矩形模板特征来说，特征值就等于其中一个矩形区域内的所有像素值之和减去另一个矩形区域内的所有像素值之和。我们称前者为正区域（正矩形），后者为负区域（负矩形）。要证明分散矩形特征与类 Haar 特征的等价性就是要证明任意一个分散矩形特征的正负区域能够被多个类 Haar 特征的正负区域覆盖。我们将通过构造的方法来证明后者。

很明显，如果分散矩形特征的两个区域恰好在水平或者垂直方向相邻对齐，那么它自身就退化为一个类 Haar 特征，因此也就等价于该类 Haar 特征。除此特殊情况外，显然其他的分散矩形特征都不可能被单一的类 Haar 特征所覆盖。

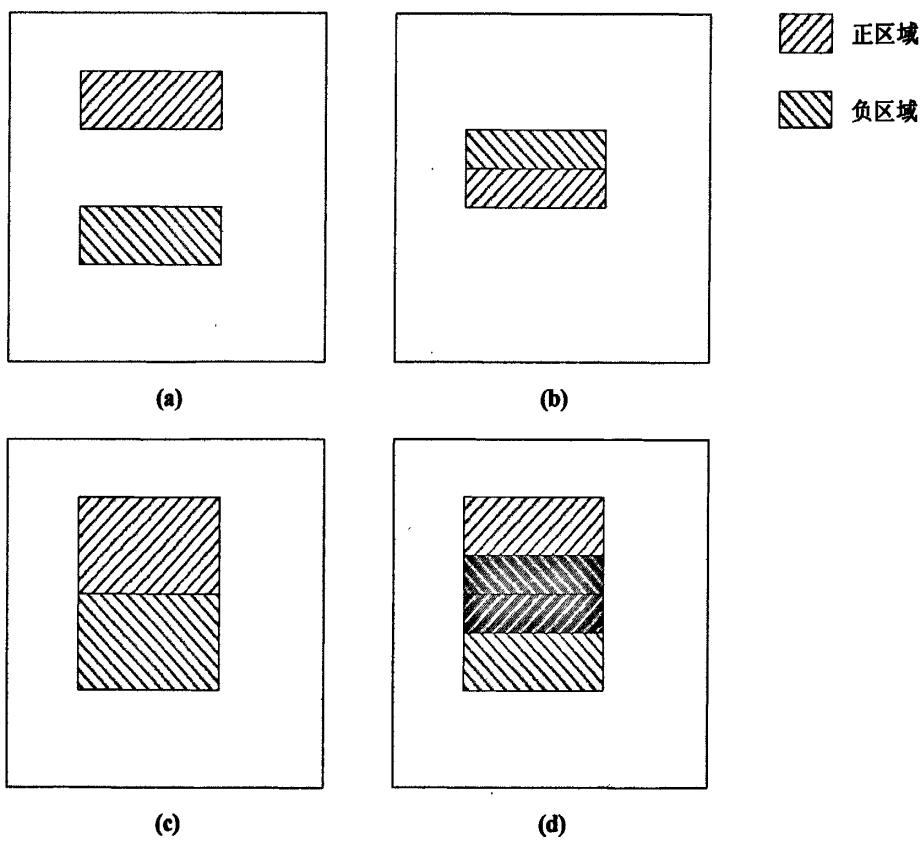


图4.1 正区域和负区域在水平或者垂直方向上对齐但不相邻的分散矩形特征(a)等价于两个类 Haar 特征(b-c)的重叠组合(d)

对于正区域和负区域在水平或者垂直方向上对齐但不相邻的分散矩形特征（图 4.1(a)），很直观的，可以通过一个覆盖正区域和负区域之间的空白区域的类 Haar 特征（见图 4.1(b)）和一个覆盖这两个区域以及中间空白区域的类 Haar 特征（见图 4.1(c)）重叠进行等价构造，如图 4.1(d)所示。

对于正区域和负区域共享一个顶点的分散矩形特征来说，如图 4.2(a)所示，正负区域呈现一个以该顶点为中心的中心对称分布。而类 Haar 特征总是表现为水平和垂直方向以正负区域共享的边线为轴线的镜面对称结构。根据中心对称可以分解为一个水平和一个垂直方面的镜面对称的思路，我们可以在该顶点上建立坐标，通过使用一个水平镜面对称的类 Haar 特征（见图 4.2(b)）和一个垂直镜面对称的类 Haar 特征（见图 4.2(c)），在重叠抵消两者之间共享的区域（见图 4.2(d)）后，构造与目标等价的中心对称的分散矩形特征。

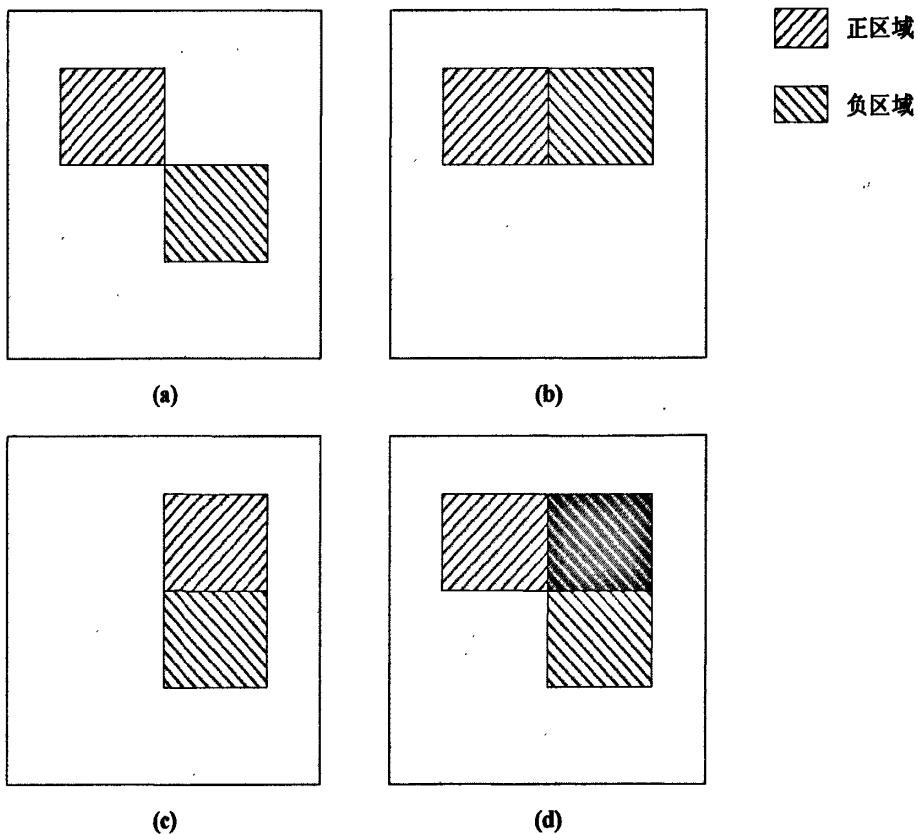


图4.2 正区域和负区域共享一个顶点的分散矩形特征(a)等价于两个垂直和水平方向共享同一个区域的类 Haar 特征(b-c)的重叠组合(d)

注意，以上这两种分散矩形特征都只需要通过两个类 Haar 特征进行等价构造，因此也是最佳的构造方案。我们称这两种分散矩形特征为基本分散矩形特征。在有了对这两个最基本的分散矩形特征的等价类 Haar 特征构造方式之后，其余的分散矩形特征都可以分解为这两个方式。其主要思想是所有的分散矩形特征都呈现中心对称结构，中心对称总是可以通过一次水平对称和垂直对称来获得，而双矩形模板的类 Haar 特征总是呈现水平对称或垂直对称的结构。因此我们采用的分解方式总是先确定分散矩形特征的中心，然后按照正负区域的分布划分为多个基本分散矩形特征，分别对这些基本分散矩形特征采用以上的等价类 Haar 特征构造方式进行构造，从而证明分散矩形特征与类 Haar 特征的等价性。

我们现在分别对错切的分散矩形特征（两个区域对齐于一条边线，但不相邻）、分离的分散矩形特征（两个区域完全分离）和重叠的分散矩形特征（两个区域共享一部分区域）进行分解。

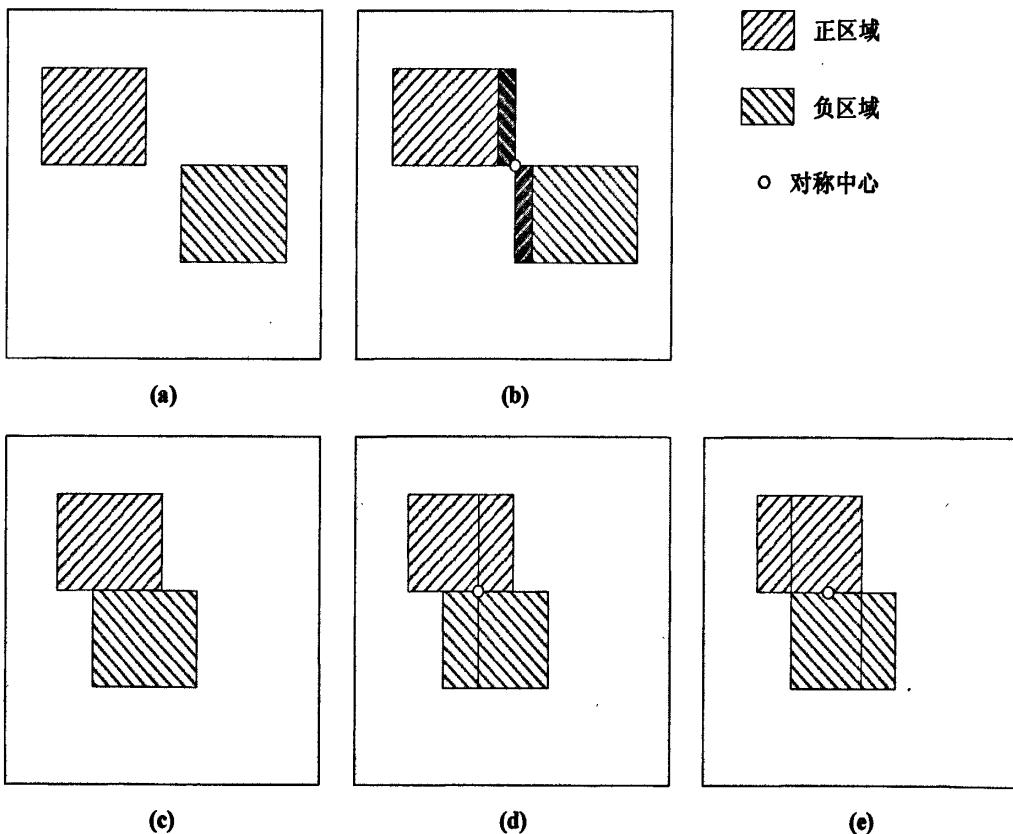


图4.3 错切的分散矩形特征关于对称中心的分解方式

根据两个区域是否相邻，错切的分散矩形特征可以分为两种情况，如图 4.3(a)和(c)所示。它们都可以通过关于对称中心的分解方式，分解为两个共享对称中心的基本分散矩形特征，分别见图 4.3(b)和(d)，因此可以通过四个类 Haar 特征等价构造。当然这种分解方式并不唯一，图 4.3(e)也是图 4.3(c)的一种分解方式，产生一个类 Haar 特征和一个和图 4.3(a)同类型的错切分散矩形特征，不过此时需要五个类 Haar 特征等价构造。同样的，图 4.3(a)以及前面所给出的分散矩形特征的例子也都存在其他可能的分解方式，本文只是给出了其中一种方式。

对于如图 4.4(a)和(c)所示的分离的分散矩形特征，我们分别给出了如图 4.4(c)和(d)的分解方式。一个分离的分散矩形特征可以被分解为两个共享对称中心的基本分散矩形特征和一个错切的分散矩形特征，因此能够通过八个类 Haar 特征等价构造。

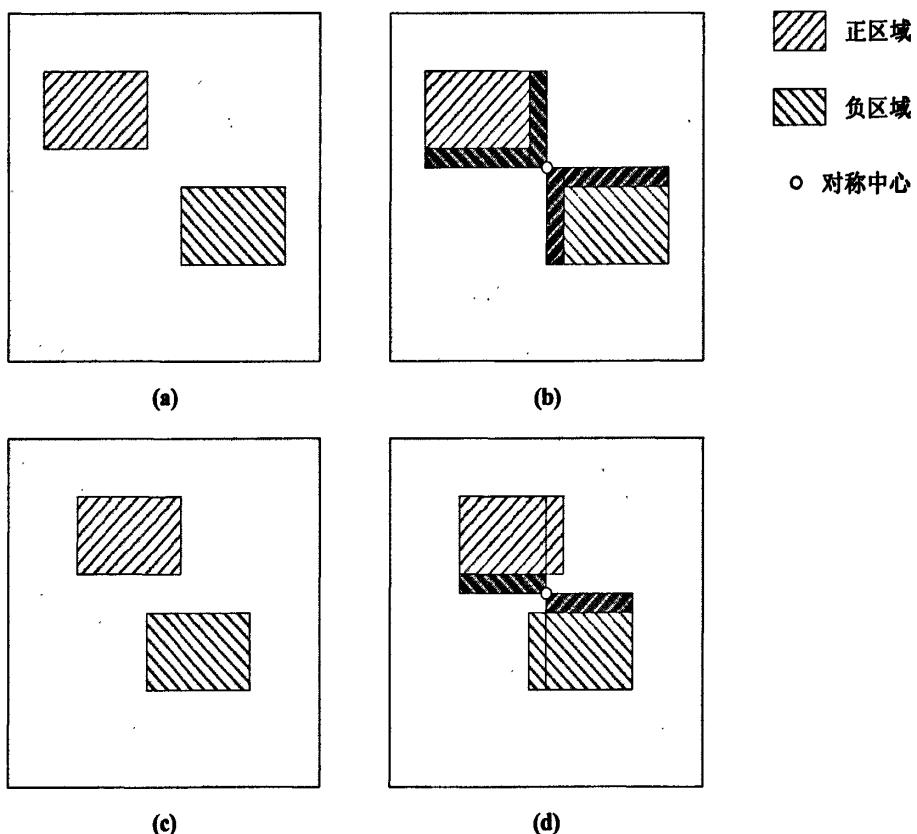


图4.4 分离的分散矩形特征关于对称中心的分解方式

对于重叠的分散矩形特征，如果正负区域对齐，由于中间重叠部分并不会影响特征值结果，因此该特征就退化为图 4.1(a)基本分散矩形特征的情况；否则，如图 4.5 所示，

可以被分解为一个水平方向对齐但不相邻的基本分散矩形特征和一个分离的分散矩形特征，因此能够通过十个类 Haar 特征等价构造。

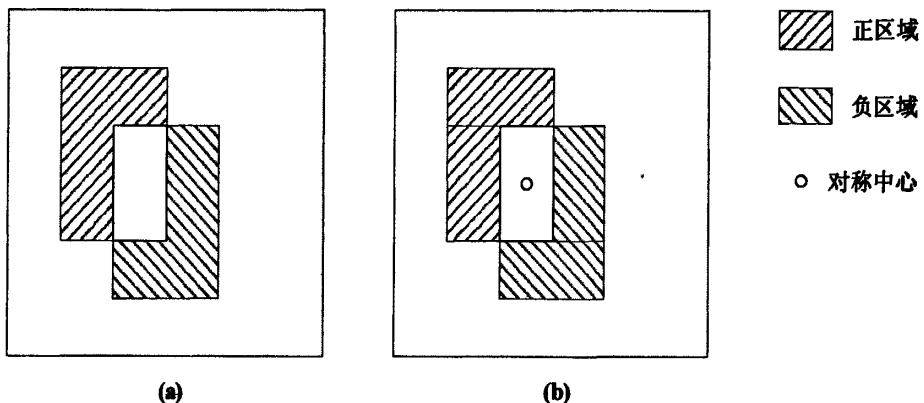


图4.5 重叠的分散矩形特征关于对称中心的分解方式

到此，我们将所有可能的分散矩形特征类型通过分解为基本分散矩形特征，证明了任意一个非退化的分散矩形特征都等价于多个类 Haar 特征按照一定的几何关系的组合。所有分散矩形特征的分解方式和等价关系总结见表 4.1。

表4.1 分散矩形特征分解方式和类 Haar 特征的等价关系

编号	分散矩形特征类型	分解	等价的类 Haar 特征个数
A	退化		1
B	对齐但不相邻		2
C	共享一个顶点		2
D	错切	$2 \times C$	4
E	分离	$2 \times C + D$	8
F	(不对齐) 重叠	$B + E$	10

## 4.2 实验结果与分析

基于前两章介绍的内容，我们利用 Viola 和 Jones 的物体检测框架，将类 Haar 特征和分散矩形特征，与单阈值和双阈值 AdaBoost 算法两两组合，分别训练了四个人脸级联器进行实验对比和分析。

### 4.2.1 实验设置

同前两章一样，本章的实验采用了同样的  $19 \times 19$  像素分辨率的人脸训练集。在训练过程中，为了去除因为级联器结构不同给性能带来的影响，我们对这四个人脸级联器采用了人工控制的训练方式，使得最终训练的所有级联器都具有 20 层，每层使用的特征数量分别为 6, 15, 30, 30, 50, 50, 50, 100, 120, 140, 160, 180, 200, 200, 200, 200, 200, 200, 200, 200, 200，一共使用 2,531 个特征。级联器每一层都需要确保 99% 的人脸检测率，而误检率不超过 70%。

### 4.2.2 训练结果

我们分别将第一组采用类 Haar 特征训练的两个人脸检测器命名为 HLF1Cascade 和 HLF2Cascade，前者表示采用了单阈值 AdaBoost 算法进行特征选取，后者则表示采用了双阈值 AdaBoost 算法。同样的，我们将第二组采用分散矩形特征的两个人脸检测器命名为 SRF1Cascade 和 SRF2Cascade。

通过对比，我们发现 HLF1Cascade 和 HLF2Cascade 共享了 1,772 个（约 70.01%）相同的特征，而 SRF1Cascade 和 SRF2Cascade 则共享了 1,800 个（约 71.12%）个相同的特征。当然这些相同的特征所对应的弱假设的阈值并不相同。由于我们采用了手工控制的训练方式，去除了级联器结构不同所带来的影响，因此这两组级联器之间性能的区别就主要归结于被双阈值 AdaBoost 算法所选取的而被单阈值 AdaBoost 算法所忽略的特征（约占 30%），以及共享的特征的弱假设的不同阈值。

### 4.2.3 检测结果

同样，我们在 MIT 和 CMU 的垂直向上正面人脸测试集上，采用基于最小人脸矩形的检测标准，使用滑动窗口的方法，对这四个级联器进行性能评估。每个检测器以 1.0 像素的平移步长和 1.0 的缩放比例开始从上到下，从左到右，在测试图像上滑动窗口，对每个窗口检测。然后以 1.25 的缩放步长，按照同样的方式对测试图像进行多轮滑动窗口检测，直至滑动窗口的某一维的尺寸超过测试图像对应的尺寸。按照该方法，每个检测器对测试集中 130 张图像的测试均扫描了 71,446,448 个子窗口。

图 4.6 给出了这四个级联器的 ROC 曲线。其中可以看出，对于第二组（采用分散矩形特征）级联器来说，双阈值 AdaBoost 算法相较于单阈值 AdaBoost 算法来说，确实对

性能有提高作用。而第一组（采用类 Haar 特征）的情况则恰恰相反，双阈值 AdaBoost 算法相较于单阈值 AdaBoost 算法来说，反而使得级联器的性能有所下降。

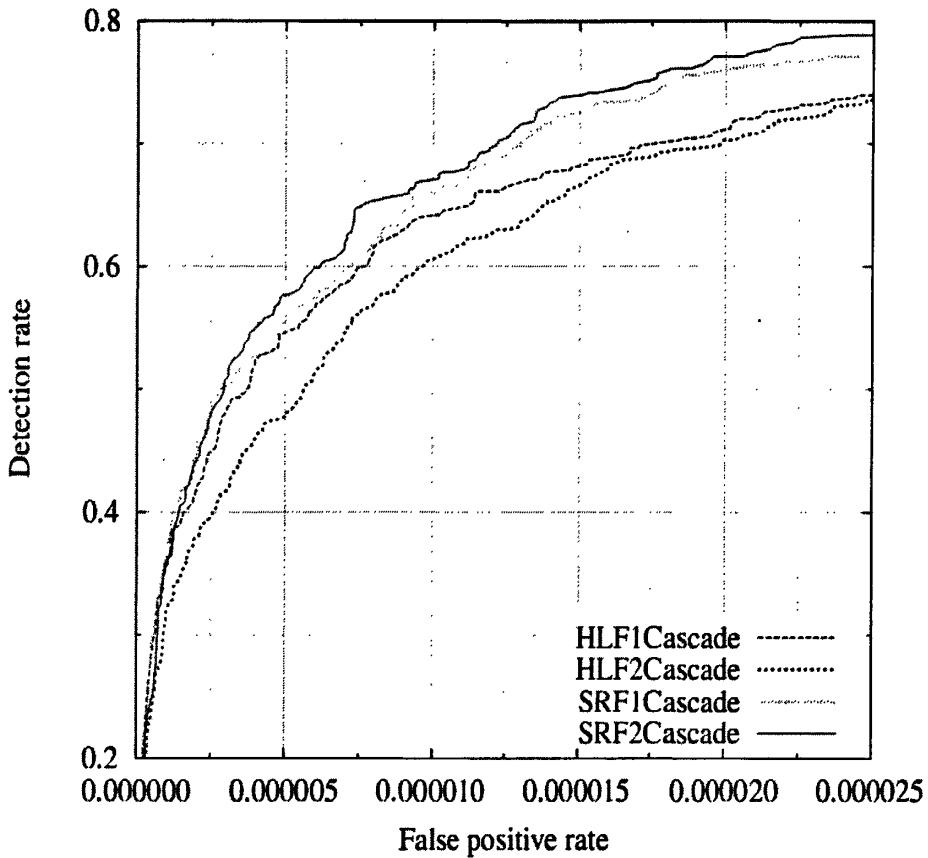


图4.6 采用人工控制的训练方式，分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器（HLF1Cascade 和 HLF2Cascade），与对分散矩形特征进行训练得到的检测器（SRF1Cascade 和 SRF2Cascade）的 ROC 曲线

为了解释该情况，我们对 HLF1Cascade 和 HLF2Cascade 这两个级联器进行了更加深入的对比实验：每次将两个级联器的最后一层去掉，然后查看两者的性能。如图 4.7 所示，我们发现 HLF2Cascade 级联器在第 9 层和第 10 层之间性能有很大的下降，并且第 10 层之后的层次没有给级联器的性能提供很大的贡献；而 HLF1Cascade 级联器却随着层次的增加性能也逐渐提高，并最终超过了 HLF2Cascade 的性能。

通过以上的对比分析，一方面说明双阈值的学习算法确实容易过匹配，另一方面也说明分散矩形特征具有处理过匹配问题的能力，从而使得在分散矩形特征上使用双阈值 AdaBoost 算法训练的级联器并没有出现在类 Haar 特征上使用双阈值 AdaBoost 算法训练

的级联器的性能下降问题。我们在本章的4.1小节中，通过构造的方法，给出了分散矩形特征与多个类 Haar 特征的等价性证明。这很可能是分散矩形特征抗过匹配能力的来源之一。因为一个分散矩形特征的弱假设所作出的分类判断实际上就等价于多个符合一定集合关系的类 Haar 特征弱假设所作出的分类判断的综合，因此也就降低了某个类 Haar 特征因为噪音等原因而作出的错误判断对整体分类结果产生的影响。

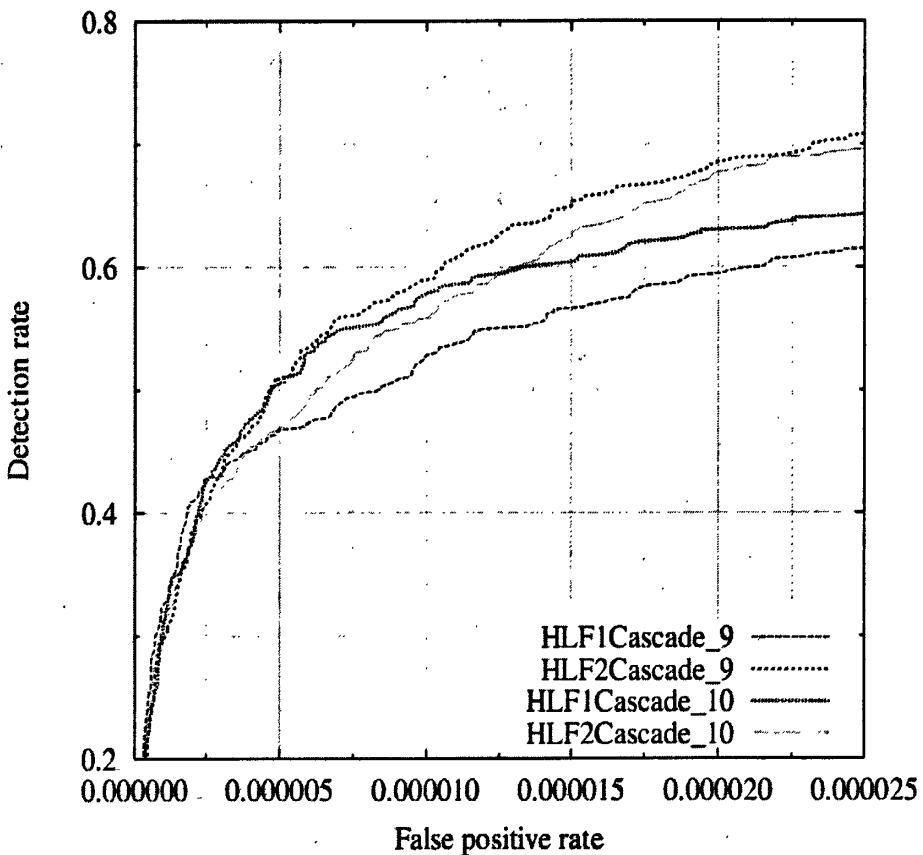


图4.7 采用人工控制的训练方式，分别使用单阈值和双阈值 AdaBoost 算法对类 Haar 特征进行训练得到的检测器的 9 层级联器（HLF1Cascade\_9 和 HLF2Cascade\_9）与 10 层级联器（HLF1Cascade\_10 和 HLF2Cascade\_10）的 ROC 曲线

图 4.8和图 4.9分别显示了这两组级联器在 MIT 和 CMU 人脸测试集上的一些测试结果，其中白色方框表示检测器的输出。除了美化排版所需要的缩放之外，输出结果没有经过任何后处理。基于最小人脸矩形的检测标准，这四个人脸检测器这三张测试图像中所出现的误检测数量和正确检测数量分别见表 4.2和表 4.3。

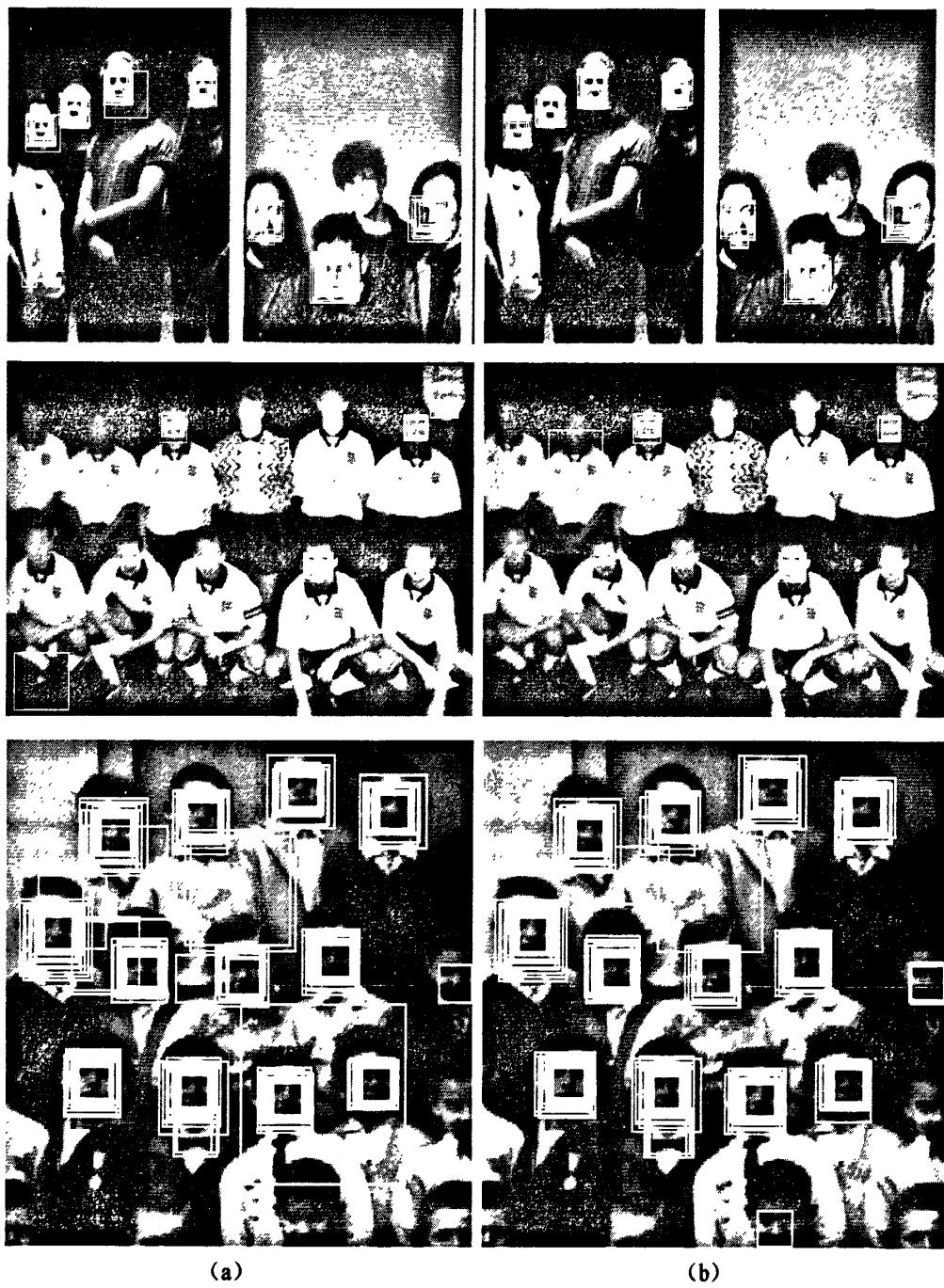


图4.8 单阈值(a)和双阈值(b)AdaBoost算法学习的类Haar特征检测器(HLF1Cascade和HLF2Cascade)的检测结果

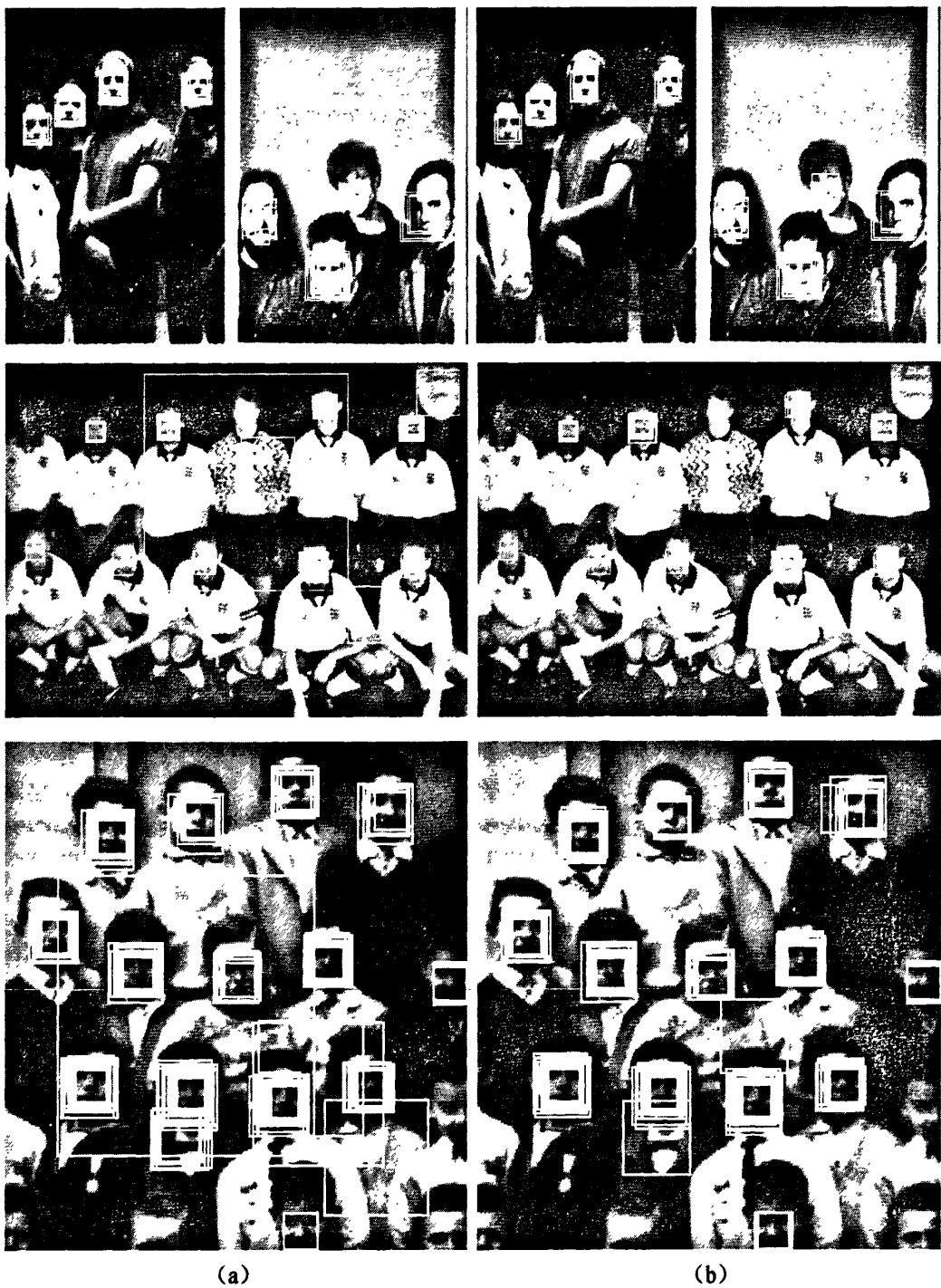


图4.9 单阈值(a)双阈值(b)AdaBoost 算法学习的分散矩形特征检测器(SRF1Cascade 和 SRF2Cascade)  
的检测结果

表4.2 三张测试图像（见图 4.8和图 4.9）上的误检测数量

测试图像	单阈值 AdaBoost 算法训练的类 Haar 特征检测器	双阈值 AdaBoost 算法训练的类 Haar 特征检测器	单阈值 AdaBoost 算法训练的分散矩形特征检测器	双阈值 AdaBoost 算法训练的分散矩形特征检测器
1	8	7	24	17
2	7	11	27	19
3	64	71	29	15

表4.3 三张测试图像（见图 4.8和图 4.9）上的正确检测数量

测试图像	单阈值 AdaBoost 算法训练的类 Haar 特征检测器	双阈值 AdaBoost 算法训练的类 Haar 特征检测器	单阈值 AdaBoost 算法训练的分散矩形特征检测器	双阈值 AdaBoost 算法训练的分散矩形特征检测器
1	7	7	7	8
2	3	3	9	9
3	13	14	14	14

### 4.3 本章小结

本章结合了前两章在特征表示和特征选取两个方面所提出的改进，在同样的实验设置下，通过手工控制训练的方式，使得不同改进方法的组合所生成的人脸级联器具有完全相同的结构；然后通过基于最小人脸矩形的客观检测标准，对检测结果进行了对比和分析。

实验结果一方面再次验证了我们在前两章中所得出的经验性结论，另一方面也说明尽管双阈值 AdaBoost 容易产生过匹配（尤其在有噪音的训练集上）问题，但是在分散矩形特征上，该问题却得到了比较好的抑制，基于分散矩形特征的分类器在使用双阈值学习算法后性能得到明显提升。分散矩形特征所展现的抗过匹配能力是第二章实验所没有反映出来的。

本章从分散矩形特征和类 Haar 特征等价的角度给该实验结果作出了尝试性解释；利用分散矩形特征中心对称与双矩形类 Haar 特征镜面对称的性质，严格证明了任意一个非退化分散矩形特征都可以被分解为多个满足一定几何约束关系的多个类 Haar 特征。因此分散矩形特征所对应的弱分类器的分类结果实际上等价于多个相互关联的类 Haar

特征对应的弱假设分类结果的综合，从而能更鲁棒的处理由于噪音数据等原因产生的过匹配问题。

在等价关系证明过程中，我们首先对分散矩形特征进行更为细致的分类，然后对每类分散矩形特征逐一采用构造分解方式的方法来确定与之等价的类 Haar 特征组合。如表 4.1 所示，对于 A、B 和 C 三类分散矩形特征所对应等价构造方法，我们给出了所需最少的类 Haar 特征个数；而对于 D、E 和 F 后三类分散矩形特征，则只给出了一种构造方法，并没有证明该构造方法是否最优。但可以肯定的是它们都不可能只通过两个类 Haar 特征来等价构造，因此本章的尝试性解释依然成立。



## 第5章 基于霍夫变换线段（组）的物体检测

本章介绍一种简单的形状特征。通过绪论可知，尽管早期基于三维模型的检测方法使用的都是形状信息，但是绝大多数现有的物体检测方法都采用基于表形的特征<sup>[8, 30, 44, 48, 57, 89, 95]</sup>。一方面是因为早期基于三维模型的检测方法效果不理想，而基于表形特征的物体检测方法则在真实世界的图像上大获成功，使得大量注意力被吸引到对表形特征的研究上去了，而忽视了形状信息对定义物体类别的作用；另一方面是因为底层的边缘检测不可靠，形状特征所包含的信息具有很大的不确定性，因此形状特征区分能力较弱，不能有效的描述复杂的场景或物体，特别是在混乱背景上，边缘检测所抽取的杂乱的形状信息很容易引起形状特征的误匹配。因此早期形状信息的应用<sup>[41-42]</sup>对物体和场景有较严格的限制。然而，最近的心理学<sup>[72]</sup>和人体视觉研究<sup>[73]</sup>都表明形状对人类区分物体类别具有重要作用，很多物体类别（自行车、马、杯子）其实是被它们的形状而不是颜色或者纹理所刻画的<sup>[108]</sup>。形状特征在物体检测中的应用重新引起了研究者们的注意，而边缘检测算法和图像分割算法<sup>[74-77]</sup>的进步也为这方面的研究提供了条件。

目前所有基于形状的物体检测方法中所使用的形状都是物体的轮廓。根据对轮廓的表示形式，可以将这些方法归为两类。一类方法用一系列在轮廓上密集采样的点来表示物体形状<sup>[79, 81-82]</sup>，而这些点的空间配置关系被应用于物体分类。这种表示方法能够保留轮廓的细节信息，但通常受制于评估这些点的空间配置关系所带来的计算复杂度问题。另一类方法则使用直线段的方式表示形状<sup>[83-85]</sup>，线段之间的关系被应用于物体分类。该方式可视为对轮廓的一种稀疏点表示，在失去细节信息的代价下，提高了效率。因此采用线段表示形式的轮廓特征，信息量相对更少，区分能也相对更弱。



图5.1 Attneave 的猫

尽管如此，Attneave 的猫<sup>[109]</sup>（如图 5.1）表明，当轮廓被连接轮廓曲线曲率极值点的线段替代后，所生成的线条画仍然能够表达绝大部分的物体信息。本章所提出的方法正是基于该观察，试图采用一系列与物体轮廓近似的线段来定义物体类别。首先我们给出本章方法的框架，然后再针对框架中相关模块展开叙述。

## 5.1 方法框架

图 5.2 描述了本章所提出的方法的整体框架。与 Opelt 等人<sup>[82, 97]</sup>提出的 BFM 类似，该框架需要如下两个数据集学习物体模型：

1. 训练图像集。每张图像包含物体，物体被一个包围盒所标示，物体中心的位置信息已知；
2. 验证图像集。由正样本（即图像中只包含物体）和负样本（即图像中只包含非物体）组成，所有的正样本的物体中心位置已知。

训练集和验证集中的物体必须呈现相似的视图和尺度。

在训练过程中，训练集和验证集首先需要通过一个霍夫变换线段抽取器，转换为一系列的霍夫变换线段图。来自训练集的所有霍夫变换线段将作为物体模型（本文又称为码表）的候选特征。为了使码表保持一个合适的大小，同时又具有良好的区分能力，候选特征需要在验证集上进行评测。具有最小的验证代价的候选特征才会被收集到码表中。为了进一步降低冗余度，码表中相似的特征通过聚合算法合并。

在测试阶段，采用滑动窗口的方法，通过将该码表和被测试窗口进行形状匹配实现物体检测。首先，将霍夫变换线段抽取器应用到该未曾见过的测试图片上，将其转换为霍夫变化线段图。然后通过缩放码表的方式，对测试图中不同位置不同尺寸的窗口进行检测。一旦在滑动窗口和码表中具有足够数量的霍夫变换线段匹配对时，就将该窗口标记为物体。

在对候选特征进行验证、对码表进行聚合以及形状匹配检测过程中，都需要一个衡量霍夫变换线段之间的距离（或相似度）的方法。由于霍夫变换线段抽取器的不可靠性，霍夫变换线段图可能具有各种噪音，因此设计一个对这些噪音具有一定容忍能力的距离衡量方法非常重要。本章根据霍夫变换线段的表示形式，提出了一个在对应空间中简单有效的加权距离。

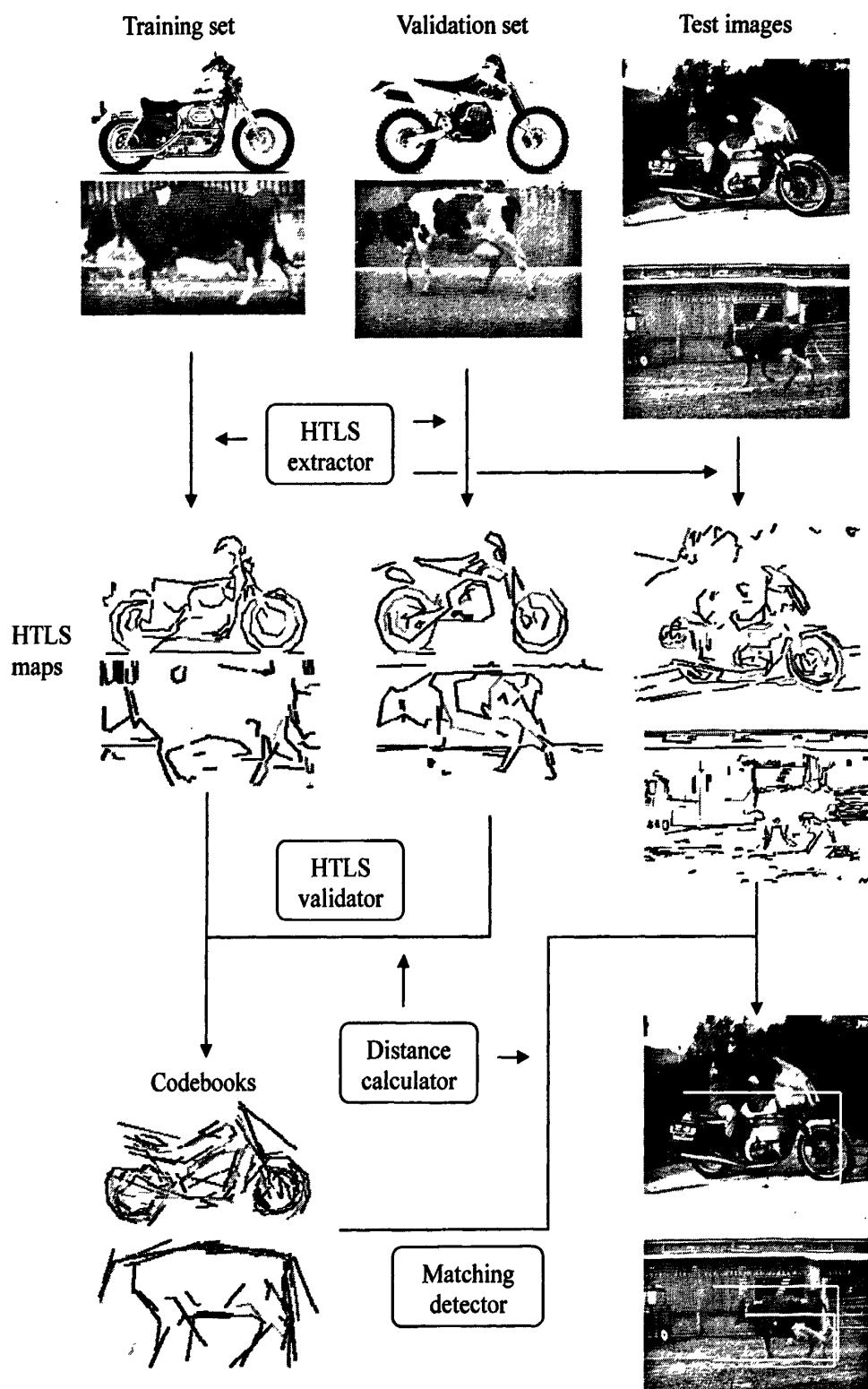


图5.2 基于霍夫变换线段的物体检测方法框架

## 5.2 霍夫变换线段 (Hough Transformed Line Segment, HTLS) 和距离

### 5.2.1 霍夫变换线段

霍夫变换<sup>[110]</sup>是图像处理中识别几何形状的一种常用方法。X-Y平面内的一条直线通过霍夫变换可以转换为  $\rho-\theta$  平面的一个点

$$\rho = x \cos(\theta) + y \sin(\theta). \quad (5.1)$$

受霍夫变换思想的启发, 如图 5.3(a)所示, 我们将一条线段表示为四元组  $(\theta, \rho, d, l)$ , 其中  $\theta$  为该线段法线和横轴的夹角,  $\rho$  为该线段到坐标原点的距离,  $d$  为该线段中心到垂线的距离,  $l$  为该线段的长度。注意, 为了确保四元组表示唯一性,  $d$  是一个矢量。我们定义  $d$  的正方向单位向量为

$$\vec{d}_u = (-\sin(\theta), \cos(\theta)), \quad (5.2)$$

即该线段的单位法向量

$$\bar{\rho}_u = (\cos(\theta), \sin(\theta)) \quad (5.3)$$

逆时针旋转  $\pi/2$  后的向量。我们称该四元组表示为霍夫变换线段。当  $\theta \in [0, \pi)$ , 我们称该表示为标准的霍夫变换线段。

霍夫变换线段能够很方便的处理刚性变换和缩放, 如图 5.3(b-d)所示。易得

$$\begin{aligned} (\theta, \rho, d, l) &\xrightarrow{r} (\theta + r, \rho, d, l) \\ (\theta, \rho, d, l) &\xrightarrow{s} (\theta, s\rho, sd, sl) \\ (\theta, \rho, d, l) &\xrightarrow{t} (\theta, \rho + \bar{t} \cdot \bar{\rho}_u, d + \bar{t} \cdot \bar{d}_u, l) \end{aligned}, \quad (5.4)$$

其中  $r$  表示逆时针旋转角度,  $s$  是缩放因子,  $\bar{t}$  是在 X-Y 平面上的平移量。因此, 一条霍夫变换线段在经过旋转、缩放, 然后平移操作后, 所得到的新的霍夫变换线段为

$$(\theta, \rho, d, l) \xrightarrow{r, s, t} (\theta + r, s\rho + \bar{t} \cdot \bar{\rho}_u, sd + \bar{t} \cdot \bar{d}_u, sl). \quad (5.5)$$

注意, 以上操作顺序中, 除旋转和缩放可以调换顺序, 其他操作顺序的改变都会引起公式(5.5)结果错误。另外, 对于该四元组表示, 显然有

$$(\theta, \rho, d, l) = (\theta + k\pi, (-1)^k \rho, (-1)^k d, l). \quad (5.6)$$

因此在经过任意的旋转、缩放和平移操作后，霍夫变换线段都可以很方便的转换为标准形式。

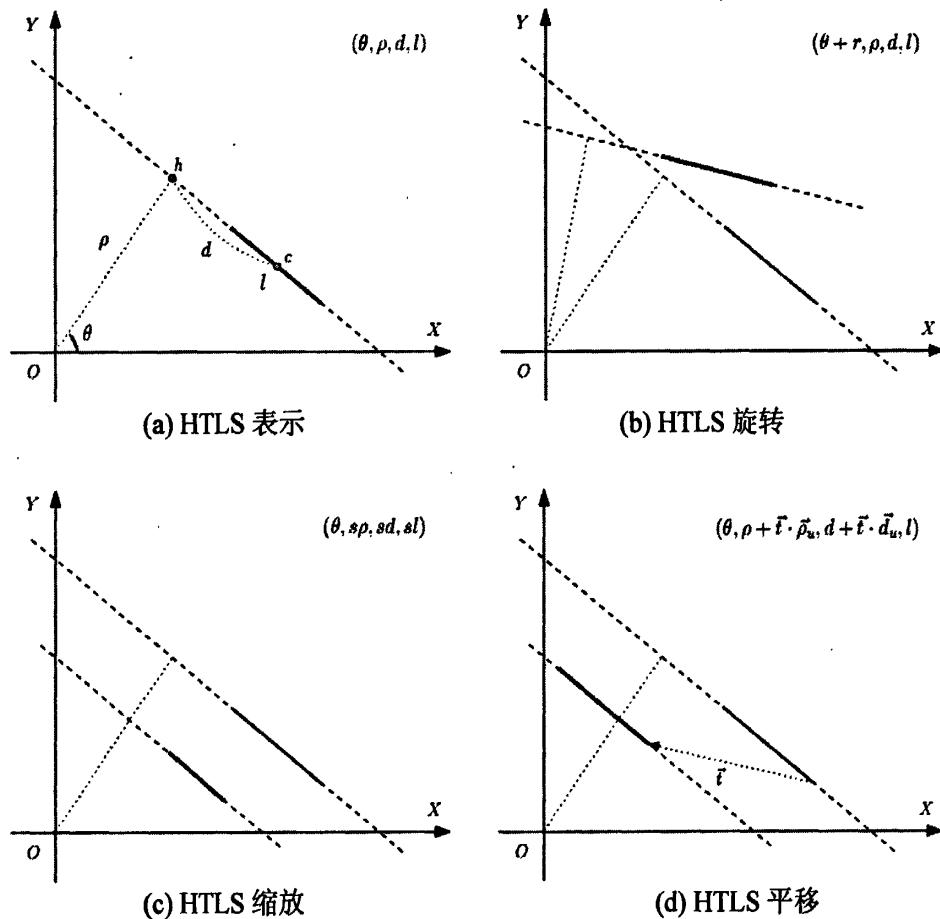


图5.3 霍夫变换线段

### 5.2.2 霍夫变换线段的提取

将图像转化为霍夫变换线段图需要经过三个步骤，1) 使用边缘检测器，将图像转换为边缘图；2) 从边缘图中抽取线段，构建线段图；3) 建立局部坐标，将线段图转换为霍夫变换线段图。对于第一个步骤，本文采用了最常用的 Canny 边缘检测器<sup>[78]</sup>。对于第二个步骤，本文考虑了采用了 Douglas 和 Peucker<sup>[111]</sup>提出的多边线简化（Polyline reduction）方法。给定一条多边线（如图 5.4(a)所示），该方法流程如下：

1. 选取两个端点 ( $v_1$  和  $v_7$ ) 构成第一条多边线，作为对原始多边线的近似；
2. 在原始多边线顶点中寻找离当前近似多边线最远的端点 ( $v_4$ ) 和距离；

3. 如果该距离小于某个阈值 ( $\varepsilon$ )，则认为当前的近似多边线已经足够好，算法停止；否则将该端点作为一个新的顶点加入近似多边线；
4. 迭代执行 2-3 步骤。

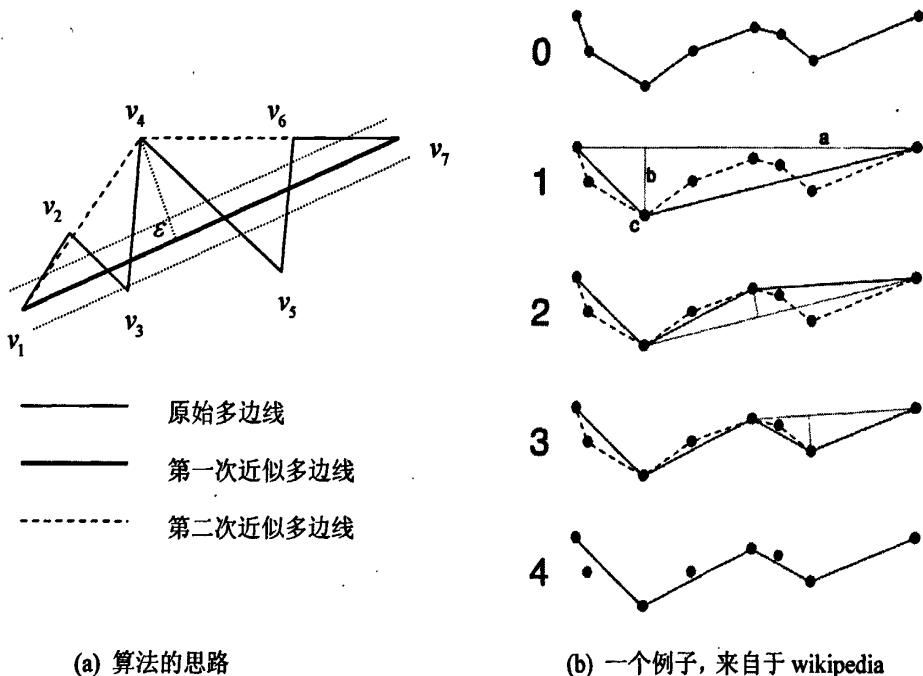


图 5.4 Douglas-Peucker 多边线简化算法

图 5.4(b)<sup>7</sup>给出了该算法的一个简单例子。相对于基于霍夫变换的直线检测算法，采用 DP 算法能够基本上保留物体轮廓段之间的几何关系。该特性对本文引入霍夫变换线段组（见5.4）非常有利，但 DP 算法不能处理环状的或者自交叉的多边线。此问题可以通过将边缘图抽象为一个无向图  $G = \{V, E\}$  而解决，其中  $V$  为相邻边缘像素大于 2 的像素， $E$  为 DP 算法可以处理的边缘多边线。对边缘图所有的多边线处理完毕后，再根据  $G$  的拓扑结构构建完成线段图。

在第三个步骤中，需要为不同的类型的线段图选取不同的局部坐标系。对于训练集中的图像来说，由于物体中心已知，因此可以将局部坐标系定位在物体中心建立霍夫变换线段图。此时，物体模型的所有候选霍夫变换线段特征都隐式的与物体中心建立了几

<sup>7</sup> [http://en.wikipedia.org/wiki/File:Douglas\\_Peucker.png](http://en.wikipedia.org/wiki/File:Douglas_Peucker.png)

何约束关系。对比隐式形状模型<sup>[8]</sup>可知，一条霍夫变换线段就是一个简单的紧凑的隐式形状模型，不需要存储额外的投票向量来表示与物体中心的几何关系。

由于已知验证集中的正样本图像的物体中心位置，因此可以使用和训练集一样的处理方式建立霍夫变换线段图。而对于其中的负样本或者测试图像来说，情况有所不同，它们要么没有物体中心要么物体中心位置未知。考虑到霍夫变换线段能够很方便的处理旋转、缩放和平移变换，我们可以任意选择一个局部坐标系（实际上我们选择在图像的中心位置建立该坐标系）生成霍夫变换线段图。之后，在训练或者测试时，我们可以按需要在负样本上平移被验证的候选霍夫变换线段，或者在测试图像上变换码表。

### 5.2.3 霍夫变换线段距离

针对霍夫变换线段，我们提出了一个在该四元组空间内的加权的欧氏距离，用来计算两条霍夫变换线段之间的相似度

$$dis(f, h) = w_\theta(\theta_f - \theta_h)^2 + w_\rho(\rho_f - \rho_h)^2 + w_d(d_f - d_h)^2 + w_l(l_f - l_h)^2, \quad (5.7)$$

其中  $w_{[i]}$  表示四元组对应元素的权重。我们将该距离称为霍夫变换线段距离。结合公式(5.5)，定义带参霍夫变换线段距离

$$\begin{aligned} dis(f, h; r, s, \bar{t}) = & w_\theta(\theta_f + r - \theta_h)^2 + w_\rho(s\rho_f + \bar{t} \cdot \bar{\rho}_{f_s} - \rho_h)^2 \\ & + w_d(sd_f + \bar{t} \cdot \bar{d}_{f_s} - d_h)^2 + w_l(sl_f - l_h)^2. \end{aligned} \quad (5.8)$$

霍夫变换线段距离非常简单，可以通过选择不同的权重控制四个元素对相似度的影响。由于边缘检测的不可靠性，按照前一节方法构建的霍夫变换线段图中，一些霍夫变换线段可能只被部分的检测，或者检测成多条不连续的霍夫变换线段，此时， $\theta$ 和 $\rho$ 应该保持不变（不考虑数字图像的离散性），而 $d$ 和 $l$ 会有不确定的变化。因此，我们很自然的就偏向于给 $\theta$ 和 $\rho$ 更大的权重，使得该距离能处理好部分匹配的情况。

尽管如此，给霍夫变换线段距离的四个权重（以及一个匹配阈值）确定合适的值，还是非常困难，因为除了 $\theta$ 和 $\rho$ 应该有更大的权重外，它们的取值范围几乎没有限制。针对该问题，我们参考了在图像匹配应用中表现出色的线段 Hausdorff 距离<sup>[112]</sup>，设计了一个将大量自动生成的线段匹配随机产生的参考线段的实验，选取一组合适的参数值使得霍夫变换线段距离表现出与线段 Hausdorff 距离类似的性能。

### 5.3 霍夫变换线段码表

霍夫变换线段码表中的特征是从训练集中提取的，在验证集上验证后的，被认为能够表示物体类别的霍夫变换线段。与轮廓段模型（BFM）<sup>[82]</sup>类似，霍夫变换线段特征要求：

1. 频繁的匹配验证集正样本而不是负样本中的霍夫变换线段；
2. 很好的定位正样本的物体中心位置。

相应的我们将每条候选霍夫变换线段  $f$  的验证代价定义为  $C(f) = c_s(f)c_g(f)$ ，其中  $c_s(f)$  为形状代价，定义为

$$c_s(f) = \frac{\sum_{i=1}^{L^+} dis^+(f, P_{v_i}) / L^+}{\sum_{i=1}^{L^-} dis^-(f, N_{v_i}) / L^-}, \quad (5.9)$$

其中， $L^+$  表示在验证集正样本  $P_{v_i}$  的个数， $L^-$  表示在验证集负样本  $N_{v_i}$  的个数。 $dis^+(f, P_{v_i})$  表示  $f$  通过平移操作后与  $P_{v_i}$  中最为相似的霍夫变换直线的距离，即

$$dis^+(f, P_{v_i}) = \min_{\bar{t}_i, j \in P_{v_i}} dis(f, j; \bar{t}_i) \quad (5.10)$$

$dis^-(f, N_{v_i})$  表示  $f$  通过随机的几次平移操作与  $N_{v_i}$  中最为相似的霍夫变换直线的平均距离，即

$$dis^-(f, N_{v_i}) = \frac{1}{r} \sum_{n=1}^r \min_{j \in N_{v_i}} dis(f, j; \bar{t}_n). \quad (5.11)$$

$c_g(f)$  为几何代价，定义为  $f$  在  $P_{v_i}$  上的平移量（见公式(5.10)中的  $\bar{t}$ ）的平均值。

由于霍夫变换线段很容易与混乱背景产生错误的匹配，因此在计算公式(5.10)时，我们记录 10 个平移操作和对应的形状代价，然后再根据它们的验证代价，选取最优的霍夫变换线段特征。

### 5.4 霍夫变换线段组

在实验中我们发现，基于单条霍夫变换线段特征的码表区分能力很弱，在混乱的背景上容易产生大量的误检测。虽然可以通过降低匹配阈值，提高区分能力，但是此时霍夫变换线段距离对噪声变得非常敏感，不能有效的处理部分匹配等问题。

类似于 Ferrari 等人<sup>[85]</sup>的 kAS，我们引入连接属性构建更具有信息量的霍夫变换线段组特征，用于物体模型。我们定义一个霍夫变换线段组为  $g_f = \{g_f^1, g_f^2, \dots, g_f^m\}$ ，其中  $m$  为该霍夫变换线段组的大小， $g_f^i$  为该霍夫变换线段组中的各条霍夫变换线段 ( $i=1, 2, \dots, m$ )，分别与  $g_f^{i-1}$  和  $g_f^{i+1}$  相连 ( $i=2, \dots, m-1$ )。 $g_f^1$  和  $g_f^m$  也可能相连，此时该霍夫变换线段组构成一个霍夫变换线段环。根据该定义可知，在霍夫变换线段组内，每个霍夫变换线段除了与物体中心的几何约束关系外，还有与它相邻两边（或者一边）组内成员之间的局部几何信息。这些信息可以用来排除在混乱背景上错误的偶然匹配。

霍夫变换线段组的引入并不需要改变整个系统框架，如图 5.2 所示，只需要修改霍夫变换线段图抽取器和霍夫变换线段距离两个模块。

在霍夫变换线段图抽取（见 5.2.2 小节）过程中，由于采用了 Douglas-Peucker 多边线简化算法，所获得的霍夫变换线段图保持了与原始轮廓边缘一致的连接属性，因此很容易从中提取出霍夫变换线段组。需要注意的是，相对于单霍夫变换线段来说，霍夫变换线段组会随着组的大小  $m$  的增大呈指数增长，因此我们将  $m$  限定为一个比较小的数值。在我们的实验中，只为  $m=1$ ,  $m=2$  和  $m=3$  的霍夫变换线段组建立了一个码表。

与此同时，还需要扩展霍夫变换线段距离，使得能够计算霍夫变换线段组之间的距离（相似度）。我们采用了一个简单的按照各霍夫变换线段长度加权的霍夫变换线段组距离

$$dis_g(g_f, g_h) = \sum_{i=1}^m \frac{L(g_f^i)}{L(g_f)} dis(g_f^i, g_h^i), \quad (5.12)$$

其中  $g_f$  和  $g_h$  为两个具有同样大小的霍夫变换线段组， $L(\cdot)$  计算霍夫变换线段或者霍夫变换线段组的线段总长度。对于两个大小不同的霍夫变换线段组来说，我们定义它们的距离为无穷大。

注意公式(5.12)是一个不对称的距离，而公式(5.7)是对称的。在验证或者检测的时候，需要将候选的霍夫变换线段组或者码表中的霍夫变换线段组特征作为  $g_f$ ，而将验证集或者测试图像上产生的霍夫变换线段组作为  $g_h$ ，进行距离计算，使得霍夫变换线段组之间每对霍夫变换线段的距离使用同样的权重。

另外一个需要注意的问题是，根据公式(5.12)，霍夫变换线段组内是有次序的。 $g_f = \{g_f^1, g_f^2, \dots, g_f^m\}$  和  $g'_f = \{g_f^m, g_f^{m-1}, \dots, g_f^1\}$  是不一样的两个霍夫变换线段组，当  $m > 1$ 。根据霍夫变换线段组内连接性可知，每个霍夫变换线段组也只有两个次序。因此可以通过在训练集上抽取候选霍夫变换线段组的时候生成两份次序不同的霍夫变换线段组，保持算法框架不变。

## 5.5 实验结果与分析

我们使用本章介绍的方法，以摩托车和牛为例，分别构建了两个码表。其中摩托车是人造物体，不管表形还是形状，在不同个体之间都存在较大的差异；牛是有关节的动物，个体的姿态会呈现各种变化。因此对这两类代表性的物体进行物体检测非常具有挑战性。

### 5.5.1 实验设置

首先需要为每个类别准备两个图像集，分别用于训练和验证（见图 5.2）。我们使用了 Motorbike Side Training Set 和 Cow Side Training Set，两者都可以在 Leibe 的主页<sup>8</sup>上获取。这两个图像集中都提供了物体掩码图像，因此通过简单的图像处理方法就可以获取满足我们实验要求（见 5.1 小节）的训练集和验证集。此外这两个图像集也被频繁的应用于最近的基于表形的方法研究中，我们按照这些研究的实验进行相同的设置。

Motorbike Side Training Set 包含 153 张图像。我们将其随机的分成三等分，从中再随机的挑选两份，一份作为训练集，一份作为验证集。因此对于摩托车类别来说，训练集和验证集都包含 51 张图片（个体）。图像中的摩托车个体宽度大致在 200 像素，高度根据车体类型有所变化，基本上在 100 到 130 像素之间。Cow Side Training Set 包含 112 张图像。我们将其随机分为四等分，同样也随机抽取两份作为训练集和验证集。因此对于牛类别来说，训练集和验证集都包含 28 张图片（个体）。图像中的牛个体宽度大致在 180-200 像素，高度基本上在 100 到 120 像素之间。

在我们的实验中，霍夫变换线段组的大小被限制为 3，因此在训练过程中，只产生包含 1 条，2 条或者 3 条的霍夫变换线段组特征。霍夫变换线段距离中，针对霍夫变换线段四元组  $(\theta, \rho, d, l)$  的权重分别被设置为 4000, 0.5, 0.1 和 0.01，匹配阈值被设为 200。

<sup>8</sup> <http://www.vision.ee.ethz.ch/bleibe/data/datasets.html>

### 5.5.2 训练结果

我们为每个类训练了两个码表。其中一个称为 Ten-percent 码表，该码表从训练集上产生的所有候选霍夫变换线段组中收集了其中的 10%; 另一个称为 Twenty-percent 码表，该码表在训练时收集了候选霍夫变换线段组的 20%。码表中霍夫变换线段距离小于匹配阈值的特征被聚合。聚合后的码表再重新回匹配到训练集中的图像，只有那些有大于 20 次匹配的霍夫变换线段组才会保留在最终的物体码表中。

训练结果见表 5.1。两个类别的码表大小差异很大，说明人造物体比有关节的动物的轮廓线段要多。摩托车的码表在收集率提高两倍后，码表的大小并没有相应的增长到原来的两倍，这说明新收集的特征与原来的码表特征之间还有很多相似的霍夫变换线段组被聚合；而对于有关节的动物牛来说码表相应的增长到差不多两倍。为了更清楚的显示码表的内容，我们将这四个码表中所有的霍夫变换线段组特征转换为线段图，见图 5.5。

表5.2 码表的训练结果

类别	收集率 (%)	码表大小
Motorbike	10	90
Motorbike	20	130
Cow	10	24
Cow	20	50

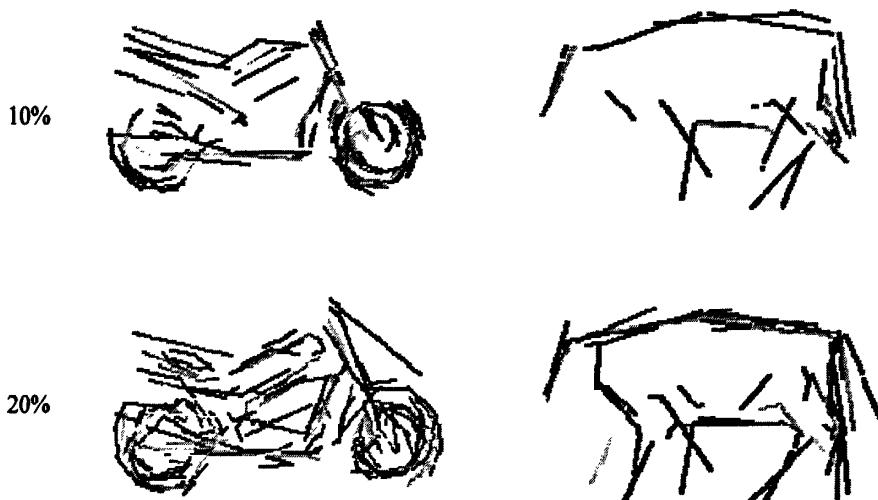


图5.5 码表中的霍夫变换线段组特征

### 5.5.3 检测标准

我们采用 TUD Motorbikes 和 Leeds Cows (同样可以在 Leibe 的主页上获取) 分别作为两个类别的测试集。前者包含 115 张图像，都是现实场景图片，包含多种类型、尺寸、姿态的摩托车，背景和光照变化大，遮挡比较普遍，因此是一个很有挑战性的测试集。本文实验中，剔除了其中遮挡非常严重的一些图片，最终的摩托车测试集大小为 95 张图像。Leeds Cows 包含 14 个视频，显示了共 18 头牛在不同背景和光照条件下从左到右经过的镜头。我们随机的从这些视频中抽取了一些图像帧，构成一个包含 91 张图像的测试集。每张图像只包含一头牛，但是尺寸不一，姿态变化非常明显，因此也是一个很有挑战性的测试集。

我们采用了 VOC 的检测标准，一个正确的检测需要满足

$$\frac{Area(D \cap B)}{Area(D \cup B)} > \frac{1}{2}, \quad (5.13)$$

其中  $D$  表示检测窗口， $B$  表示物体的包围盒。

我们可以直接利用 TUD Motorbikes 提供了关于物体在图像空间中的包围盒信息，对于 Leeds Cows，我们手工的建立测试集中的各帧包围盒信息。

### 5.5.4 检测结果

因为霍夫变换线段能够很容易处理缩放和平移，我们采用缩放比例为 1.25 的滑动窗口的方法对四个码表的性能进行评测。通过简单的形状匹配对每个滑动窗口进行检测，一旦某个滑动窗口中的霍夫变换线段组和经过缩放平移后的码表匹配的次数超过某个阈值，那么就将该区域标记为物体。我们将该阈值和码表大小之间建立关系： $t_d = s_c \times m_r$ ，其中  $s_c$  表示码表的大小， $m_r$  表示匹配比例。不同匹配比例的检测结果见表 5.2。从中可知，两个 Twenty-percent 码表在性能上都要优于 Ten-percent 码表；使用约 50% 的匹配比例后，不管对于摩托车还是牛，两个 Twenty-percent 码表能够提供一个令人满意的性能（检测率>90%，误检率<0.048%）。图 5.6 显示了 Twenty-percent 码表的几个检测结果，其中白色方框表示正确的检测（按照 VOC 检测标准），黑色方框则为错误的检测。实验中我们发现，尽管（摩托车的）测试图像中普遍存在部分遮挡现象，但是霍夫变换线段（组）码表仍然能够获得 90% 以上的检测率，说明该形状特征能够有效的处理遮挡问题。

然而，我们也发现，该特征容易在混乱背景上产生误匹配，实验中大部分的误检集中在类似栅栏、树木、草坪等容易产生大量边缘的背景上。

表5.3 码表的检测性能

类别	码表大小	匹配比例	误检率	检测率
Motorbike	90	0.40	0.000367880	0.952703
		0.42	0.000256016	0.919708
		0.44	0.000181178	0.877049
		0.46	0.000137852	0.840336
Motorbike	130	0.49	0.000219932	0.916129
		0.50	0.000190543	0.905405
		0.51	0.000166843	0.897260
		0.52	0.000118496	0.875000
Cow	24	0.60	0.000728256	0.976744
		0.64	0.000445884	0.947826
		0.68	0.000255947	0.865385
		0.72	0.000130002	0.737374
Cow	50	0.50	0.000474497	0.992481
		0.52	0.000366953	0.944000
		0.54	0.000259414	0.929825
		0.56	0.000181810	0.890909

与 Leibe 等人<sup>[8]</sup>的工作相比，他们也使用了该训练集和测试集（本文牛的测试集是其子集），但采用了基于表形的特征。相对于他们的检测结果而言，本文方法的高检测率是在相对较高的误检率情况下获得的。如何将形状特征变得更有区分能力，从而能够处理混乱背景仍然是我们需要不断努力的一个研究方向。

## 5.6 本章小结

本章提出了一种基于形状的局部特征，称为霍夫变换线段（组）。它具有两个良好的性质：1) 如果将局部坐标建立在物体中心，那么得到的霍夫变换线段自身就是一个紧凑的隐式形状模型，不需要存储额外的投票向量；2) 霍夫变换线段能够很方便的处理缩放、旋转和平移，因此并不需要一定在物体中心（比如测试的时候物体中心位置未知）建立局部坐标来获取霍夫变换线段。相应的，一个简单的衡量霍夫变换线段（组）之间的距离（相似度）的方法被引入。该距离是一个在霍夫变换线段四元组空间内加权版本的欧式距离，通过合理的调整权重，能够有效的处理不可靠边缘检测问题。在摩托车（人造物体）和牛（有关节动物）两类代表性物体上进行的实验结果表明基于形状的局部特征的物体检测方法具有很大的潜力。

本章工作只是展示了霍夫变换线段（组）一些初步的性能，基于霍夫变换线段（组）的方法还有很大的改进空间。一种可能的改进方法是提高霍夫变换线段组的大小，使得更有信息量的特征被收集到码表中，提高码表的区分能力。另一种方法是采用基于概率的 Pb 边缘检测器<sup>[74]</sup>代替 Canny 边缘检测器，从输入上减少混乱背景的信息。当然这种方法也可能导致丢失一部分有用的物体轮廓信息。还可以参考 Opelt 等人<sup>[97]</sup>的工作，通过和基于表形的特征相结合进一步增强码表的区分能力，解决混乱图像易于产生误匹配的问题。

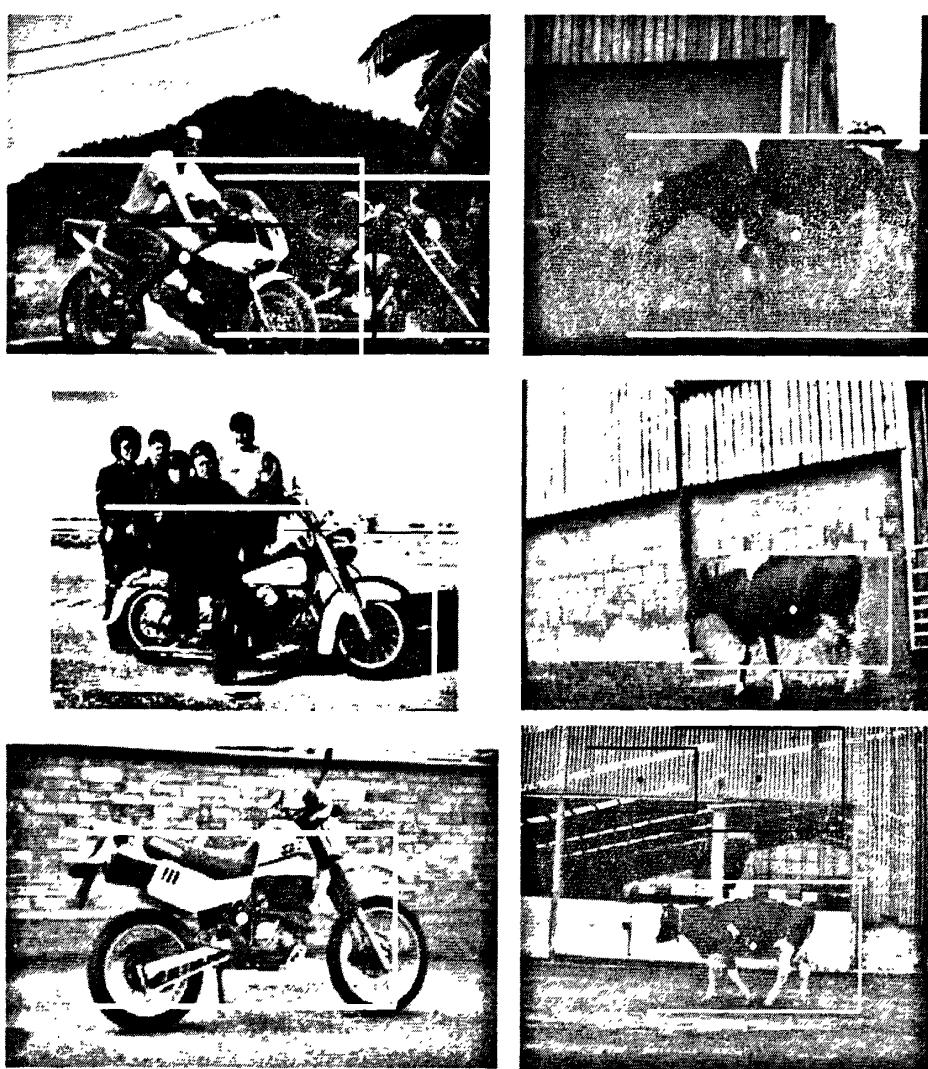


图5.6 Twenty-percent 码表的检测结果

## 第6章 结论

### 6.1 全文工作总结

物体检测是计算机视觉研究领域一个重要课题，近年来采用对局部特征学习，建立物体模型的研究方法引起了广泛的关注。本文重点分析了国内外基于局部特征的物体检测方法的研究现状，从简单局部特征和基于特征的学习算法两个层面，研究了如何构建准确的、鲁棒的物体模型，并取得了以下几方面的成果：

1. 基于表形的分散矩形特征。分散矩形特征和类 Haar 特征的区别在于它们的特征结构。前者由于在特征结构上不需要保持两个矩形在水平或垂直方向上相邻对齐，使得分散矩形特征能更自由的表示区域的方位信息和几何信息，从而提高了特征的表达能力。同时该特征变体也能有效的利用积分图进行快速的计算，解决了大量特征所导致的计算量问题。此外，任意一个非退化的分散矩形特征都可以按照特定的方式分解为满足一定几何关系约束的多个类 Haar 特征，即该分散矩形特征所包含的物体部分的信息等效于多个类 Haar 特征所包含的信息的综合，因此该特征更加鲁棒。基于同样的实验设置条件下，通过 AdaBoost 算法分别学习了多个人脸分类器。训练结果验证了在特征结构中去除相邻对齐约束给分散矩形特征在表达能力上所带来的优势；基于同一个严格和客观的检测标准，在 MIT 和 CMU 的人脸测试集上的对比实验结果表明，基于分散矩形特征的检测器性能优于基于类 Haar 特征的检测器。
2. 基于形状的霍夫变换线段（组）。霍夫变换线段（组）的提出动机是因为线条画能表达绝大部分的物体信息。霍夫变换线段通过其法向与横轴的夹角角度、到坐标原点的垂直距离、线段中心到垂线的距离以及线段长度所组成的四元组表示。该表示具有两个良好的性质：1) 如果将局部坐标建立在物体中心，那么得到的霍夫变换线段自身就是一个紧凑的隐式形状模型，不需要存储额外的投票向量；2) 霍夫变换线段能够很方便的处理缩放、旋转和平移，因此并不需要一定在物体中心（比如测试的时候物体中心位置未知）建立局部坐标来获取霍夫变换线段。相应的，一个简单的衡量霍夫变换线段（组）之间的距离（相似度）的方法被引入。该距离是一个在霍夫变换线段四元组空间内加权版本的欧

式距离，通过合理的调整权重，能够有效的处理不可靠边缘检测问题。在摩托车（人造物体）和牛（有关节动物）两类代表性物体上进行的实验结果表明基于霍夫变换线段（组）的物体检测方法能够处理部分遮挡问题，具有很好的发展潜力。

3. 双阈值 AdaBoost 算法。该算法通过 AdaBoost 算法框架结合了双阈值弱假设和双阈值弱学习算法。前者基于选择分类性能更好的弱假设可以使学习到的强假设更加鲁棒和高效的思想，通过引入另外一个阈值，提高弱假设对特征值空间的类别划分精度，确保产生分类错误更小的弱分类器，从而使得经过 AdaBoost 算法增强后的分类器更加鲁棒和高效。为了快速确定弱假设的两个最优阈值，后者根据候选阈值的性质，将确定最优阈值问题转换为寻找最大（最小）和连续子序列问题，从而在线性时间内完成弱学习过程。对类 Haar 特征和分散矩形特征的学习结果表明，在同样的训练性能要求下，该算法比原来的 AdaBoost 算法收敛速度更快，生成的检测器具有更少的层次，使用更少的特征。在 MIT 和 CMU 人脸测试集上的检测实验显示，基于该变体学习的分类器性能总体上优于原来的 AdaBoost 算法。

此外，本文在人脸检测实验过程中，构建了一个分辨率为  $19 \times 19$  像素的正面人脸训练集，提出了一个针对 MIT 和 CMU 人脸测试集的客观的检测标准。

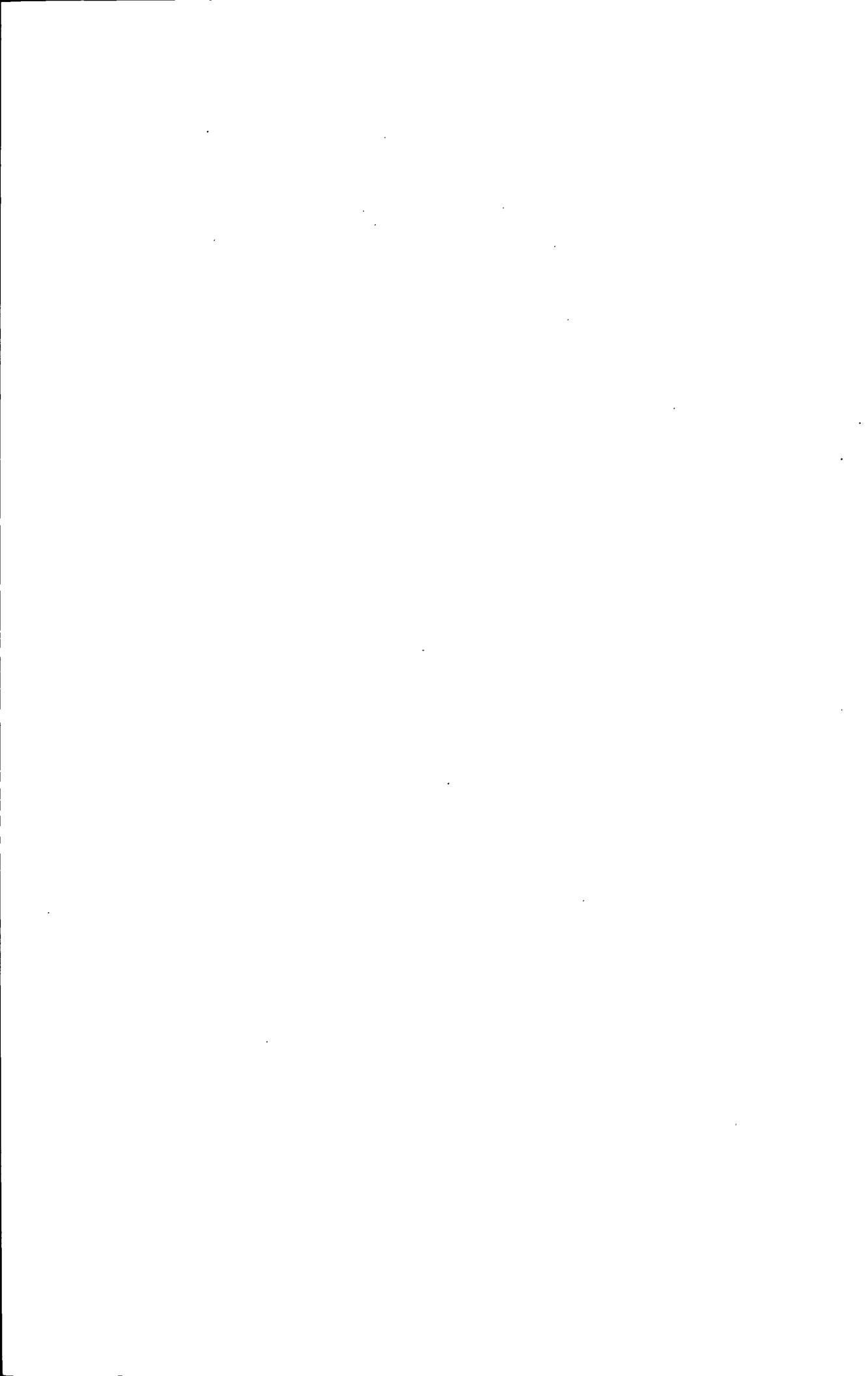
## 6.2 未来研究展望

未来可能的研究方向包括以下几个方面：

1. 进一步增强霍夫变换线段的表示能力。可以参考 Opelt 等人<sup>[97]</sup>和 Leibe 等人<sup>[8]</sup>的工作，1) 增大霍夫变换线段组的大小，使得一个霍夫变换线段组包含更多的信息，提高其区分能力；2) 考虑在码表上应用 AdaBoost 算法，构建更加精确的分类器；3) 对码表中霍夫变换线段组建立概率模型，通过概率投票，而不是简单的计数，进行形状匹配；4) 与其他基于表形的局部特征结合，解决霍夫变换线段在混乱背景上容易出现误检测的问题。
2. 提高双阈值 AdaBoost 抗过匹配的能力。一个可以尝试的方法是和 FloatBoost<sup>[105]</sup>算法结合。采用 FloatBoost 事后补救的思想，在学习过程中，

每次增加弱分类器时都进行回退搜索，去掉之前已经被加入到强分类器但是却并不是很高效的弱分类器。通过交叉验证和事后补救的方法降低双阈值 AdaBoost 产生过匹配的可能性。

3. 扩展物体检测研究的范围。比如在研究对象上，一方面可以将本文提出的方法应用到其他类别（如汽车、人体等）物体检测上，验证其有效性或者发现潜在的其他问题；另一方面可以则研究个体目标的识别或者场景分类问题。在输入上，可以考虑和目标跟踪算法结合处理视频，其中动作<sup>[102]</sup>和事件检测<sup>[113]</sup>是极具有挑战性的课题。



## 参考文献

- [1] D. Marr. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. 1982.
- [2] K. Treash, K. Amaralunga. Automatic road detection in grayscale aerial images. *Journal of Computing in Civil Engineering*, 2000, 14(1): 60-69.
- [3] Z.H. Sun, G. Bebis, R. Miller. On-road vehicle detection: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(5): 694-711.
- [4] J.S. Hong, T. Kaneko, R. Sekiguchi, K.H. Park. Automatic liver tumor detection from CT. *Ieice Transactions on Information and Systems*, 2001, E84d(6): 741-748.
- [5] T. Serre, M. Riesenhuber. Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex. *Massachusetts Institute of Technology, CBCL Paper 239*, 2004, AlMemo volume 17.
- [6] A. Glennerster. Marr's vision: Twenty-five years on. *Current Biology*, 2007, 17(11): R397-R399.
- [7] L.G. Roberts. Machine perception of three-dimensional solids. *Optical and Electrooptical Information Processing*, 1965: 159-197.
- [8] B. Leibe, A. Leonardis, B. Schiele. Robust Object Detection with Interleaved Categorization and Segmentation. *International Journal of Computer Vision*, 2008, 77(1-3): 259-289.
- [9] B. Leibe, K. Schindler, N. Cornelis, L.V. Gool. Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(10): 1683-1698.
- [10] A.B. Torralba. Contextual priming for object detection. *International Journal in Computer Vision*, 2003, 53(2): 169-191.
- [11] V. Ferrari, T. Tuytelaars, L.V. Gool. Simultaneous object recognition and segmentation by image exploration. *Proceedings of the European Conference on Computer Vision*, 2004: 40-54.
- [12] E. Rosch, C.B. Mervis, W.D. Gray, D.M. Johnson, P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 1976, 8: 382-439.
- [13] J.W. Tanaka, M. Taylor. Object categories and expertise: Is the basic level in the eye of the beholder. *Cognitive Psychology*, 1992, 23: 457-482.
- [14] L. Wang, J. Shi, G. Song, I.-F. Shen. Object Detection Combining Recognition and

- Segmentation. Eighth Asian Conference on Computer Vision 2007.
- [15] N. Pinto, D.D. Cox, J.J. DiCarlo. Why is Real-World Visual Object Recognition Hard? *PLoS Computational Biology*, 2008, 4(1): 151-156.
- [16] K.A. Spackman. Signal detection theory: Valuable tools for evaluating inductive learning. *Proceedings of the Sixth International Workshop on Machine Learning*, 1989: 160-163.
- [17] J. Davis, M. Goadrich. The Relationship Between Precision-Recall and ROC Curves. *Proceedings of the 23rd international conference on Machine learning*, 2006: 233-240.
- [18] S. Bengio, J. Mariéthoz, M. Keller. The Expected Performance Curve. *Proceedings of the 22nd International Conference on Machine Learning*, 2005: 9-16.
- [19] D. Terzopoulos, D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(7): 703-714.
- [20] A. Gupta, R. Bajcsy. Volumetric segmentation of range images of 3d objects using superquadric models. *CVGIP: Image Understanding*, 1993, 58(3): 302-326.
- [21] F. Ferrie, J. Lagarde, P. Whaite. Darboux frames, snakes, and super-quadrics: Geometry from the bottom up. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, 15(8): 771-784.
- [22] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 1987, 2(94): 115-147.
- [23] R. Brooks. Model-based 3-D interpretations of 2-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1983, 5(2): 140-150.
- [24] D. Borges, R. Fisher. Class-based recognition of 3d objects represented by volumetric primitives. *Image and Vision Computing*, 1997, 15(8): 655-664.
- [25] T. Silberberg, D. AHarwood, L.S. Davis. Object recognition using oriented model points. *Compututer Vision, Graphics, and Image Processing*, 1986, 35(1): 47-71.
- [26] D. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 1987, 31: 355-395.
- [27] A. Kushal, C. Schmid, J. Ponce. Flexible object models for category-level 3D object recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007: 1370-1377.
- [28] J. Liebelt, C. Schmid, K. Schertler. Viewpoint-independent object class detection using 3D feature maps. *IEEE Conference on Computer Vision and Pattern*

- Recognition, 2008: 2118-2125.
- [29] P. Yan, S.M. Khan, M. Shah. 3D model based object class detection in an arbitrary view. IEEE 11th International Conference on Computer Vision, 2007: 1673-1678.
- [30] M.A. Turk, A.P. Pentland. Eigen faces for face recognition. Proceedings of the Conference on Computer Vision and Pattern Recognition, 1991: 586-591.
- [31] K. Etemad, R. Chellappa. Discriminant analysis for recognition of human face images. J. Opt. Soc. Am., 1997, 14: 1724-1733.
- [32] M. Bartlett, T. Sejnowski. Independent components of face images: a representation for face recognition. Fourth Annual Joint Symposiumon Neural Computation, 1997: 3-10.
- [33] M. Swain, D. Ballard. Color indexing. International Journal in Computer Vision, 1991, 7(1): 11-32.
- [34] H. Murase, S. Nayar. Visual learning and recognition of 3D objects from appearance. International Journal on Computer Vision, 1995, 14(1): 5-24.
- [35] S.B. C. Carson, S. Greenspan and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(8): 1026-1038.
- [36] A. Leonardis, H. Bischoff. Dealing with occlusions in the eigenspace approach. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1996: 453-458.
- [37] J. Winn, A. Criminisi, T. Minka. Object categorization by learned universal visual dictionary. Proceedings of the International Conference on Computer Vision, 2005, 2: 1800-1807.
- [38] F. Jurie, B. Triggs. Creating effcient codebooks for visual recognition. Proceedings of the International Conference on Computer Vision, 2005: 604-610.
- [39] K. Mikolajczyk, A. Zisserman, C. Schmid. Shape recognition with edge based features. Proceedings of the British Machine Vision Conference, 2003: 779-788.
- [40] K. Mikolajczyk, B. Leibe, B. Schiele. Multiple object class detection with a generative model. Proceedings of the Conference on Computer Vision and Pattern Recognition, 2006: 26-36.
- [41] S. Sclaroff. Deformable Prototypes for Encoding Shape Categories in Image Databases. Pattern Recognition, 1997, 30(4).
- [42] A.S. K.Siddiqi, S. Dickinson, and S. Zucker. Shock Graphs and Shape Matching.

- International Journal of Computer Vision, 1999, 30: 1-24.
- [43] C. Papageorgiou, M. Oren, T. Poggio. A general framework for object detection. ICCV, 1998.
- [44] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, 1: 511-518.
- [45] P. Viola, M.J. Jones, D. Snow. Detecting pedestrians using patterns of motion and appearance. Ninth IEEE International Conference on Computer Vision, 2003: 734-741.
- [46] R.N. Hota, K. Jonna, P.R. Krishna. On-road vehicle detection by cascaded classifiers. Proceedings of the Third Annual ACM Bangalore Conference, 2010.
- [47] R. Lienhart, A. Kuranov, V. Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. 25th Pattern Recognition Symposium, 2003: 297-304.
- [48] H. Wang, P. Li, T. Zhang. Histogram Feature-Based Fisher Linear Discriminant for Face Detection. Neural Computing and Applications, 2007, 17(1): 49-58.
- [49] H. Wang, P. Li, T. Zhang. Proposal of novel histogram features for face detection. 3rd International Conference on Advances in Pattern Recognition, 2005, 2687: 334-343.
- [50] J.S. Y .Lamdan, and H. Wolfson. Object Recognition by Affine Invariant Matching. IEEE Conference on Computer Vision and Pattern Recognition, 1988: 335-334.
- [51] C. Harris, M. Stephens. A combined corner and edge detector. Alvey Vision Conference, 1988: 147-151.
- [52] T. Lindeberg. Feature Detection with Automatic Scale Selection. International Journal of Computer Vision, 1998, 30(2): 79-116.
- [53] T. Tuytelaars, L.v. Gool. Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions. British Machine Vision Conference, 2000: 412-422.
- [54] T. Tuytelaars, K. Mikolajczyk. Local invariant feature detectors: a survey. Foundations and trends in computer graphics and vision, 2008, 3(3): 177-280.
- [55] S.A.J. Winder, M. Brown. Learning local image descriptors. IEEE Conference on Computer Vision and Pattern Recognition, 2007: 17-24.
- [56] M.V.-N. S. Ullman, and E. Sali. Visual features of intermediate complexity and their use in classification. Nature Neuroscience, 2002, 5(7): 682-687.
- [57] R. Fergus, P. Perona, A. Zisserman. Object class recognition by unsupervised

- scale-invariant learning. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003: 264-271.
- [58] B. Leibe, B. Schiele. Analyzing appearance and contour based methods for object categorization. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003: 409-415.
- [59] M. Weber, M. Welling, P. Perona. Towards Automatic Discovery of Object Categories. IEEE Conference on Computer Vision and Pattern Recognition, 2000, 2: 101-108.
- [60] R. Nelson, A. Selinger. A Cubist Approach to Object Recognition. Sixth International Conference on Computer Vision, 1998: 614-621.
- [61] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [62] D. Lowe. Object Recognition from Local Scale Invariant Features. Seventh International Conference on Computer Vision, 1999, 2: 1150-1157.
- [63] D. Lowe. Local Feature View Clustering for 3D Object Recognition. IEEE Conference on Computer Vision and Pattern Recognition, 2001: 682-688.
- [64] A. Witkin. Scale space filtering. Proc. International Joint Conference on Artificial Intelligence, 1983: 1019-1021.
- [65] Y. Ke, R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, 2: 506-513.
- [66] K. Mikolajczyk, C. Schmid. A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615-1630.
- [67] S. Belongie, J. Malik, J. Puzicha. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 509-522.
- [68] S. Lazebnik, C. Schmid, J. Ponce. A sparse texture representation using affine-invariant regions. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003: 319-324.
- [69] H. Bay, T. Tuytelaars, L. Van Gool. SURF: Speeded up robust features. Computer Vision - Eccv 2006 , Pt 1, Proceedings, 2006, 3951: 404-417.
- [70] P.R. Beaudet. Rotationally invariant image operators. Proceedings of the International Joint Conference on Pattern Recognition, 1978: 579-583.
- [71] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. Proc. IEEE

- Conf. on Computer Vision and Pattern Recognition, 2005: 886-893.
- [72] P.C. Quinn, P.D. Eimas, M.J. Tarr. Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *Journal of experimental child psychology*, 2001, 79(1): 78-94.
- [73] J.H. Elder, L. Velisavljevic. Cue dynamics underlying rapid detection of animals in natural scenes. *Journal of Vision*, 2009, 9(7): 1-20.
- [74] D.R. Martin, C.C. Fowlkes, J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(5): 530-549.
- [75] P. Dollár, Z.W. Tu, S.J. Belongie. Supervised learning of edges and object boundaries. *CVPR*, 2006: 1964-1971.
- [76] S. Zheng, Z. Tu, A.L. Yuille. Detecting Object Boundaries Using Low-, Mid-, and High-level Information. *IEEE Computer Vision and Pattern Recognition*, 2007.
- [77] X.F. Ren. Multi-scale Improves Boundary Detection in Natural Images. *Computer Vision - Eccv 2008, Pt Iii, Proceedings*, 2008, 5304: 533-545.
- [78] J. Canny. A Computational Approach to Edge-Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(6): 679-698.
- [79] A. Berg, T. Berg, J. Malik. Shape Matching and Object Recognition Using Low Distortion Correspondence. *CVPR*, 2005, 1: 26-33.
- [80] M. Leordeanu, M. Hebert, R. Sukthankar. Beyond local appearance: Category recognition from pairwise interactions of simple features. *IEEE Conference on Computer Vision and Pattern Recognition*, 2007: 928-935.
- [81] J. Shotton, A. Blake, R. Cipolla. Contour-based learning for object detection. *Tenth IEEE International Conference on Computer Vision*, 2005: 503-510.
- [82] A. Opelt, A. Pinz, A. Zisserman. A boundary-fragment-model for object detection. *Computer Vision - ECCV 2006*, 2006, 3952: 575-588.
- [83] X.F. Ren. Learning and Matching Line Aspects for Articulated Objects. *Proc. IEEE Conf. Comput. Vision and Pattern Recogn*, 2007.
- [84] V. Ferrari, T. Tuytelaars, L. Van Gool. Object detection by contour segment networks. *Computer Vision - Eccv 2006, Pt 3, Proceedings*, 2006, 3953: 14-28.
- [85] V. Ferrari, L. Fevrier, F. Jurie, C. Schmid. Groups of adjacent contour segments for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(1): 36-51.

- [86] H.L. Jin, Q.S. Liu, X.O. Tang, H.Q. Lu. Learning local descriptors for face detection. IEEE International Conference on Multimedia and Expo (ICME), 2005: 929-932.
- [87] P. Moreels, P. Perona. Evaluation of features detectors and descriptors based on 3D objects. International Journal of Computer Vision, 2007, 73(3): 263-284.
- [88] E. Seemann, B. Leibe, K. Mikolajczyk, B. Schiele. An Evaluation of Local Shape-Based Features for Pedestrian Detection. British Machine Vision Conference, 2005.
- [89] H.A. Rowley, S. Baluja, T. Kanade. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 23-38.
- [90] Y. Freund, R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 1997, 55(1): 119-139.
- [91] V. Vapnik. The Nature of Statistical Learning Theory. 1995.
- [92] E. Osuna, R. Freund, F. Girosi. Training support vector machines: an application to face detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997: 130-136.
- [93] C. Papageorgiou, T. Poggio. A trainable system for object detection. International Journal of Computer Vision, 2000, 38(1): 15-33.
- [94] L. Fei-Fei, R. Fergus, A. Torralba. Recognizing and Learning Object Categories. CVPR 2007 short course, 2007.
- [95] G. Csurka, C. Dance, L.X. Fan, J. Willamowski, C. Bray. Visual categorization with bags of keypoints. Proc. of ECCV International Workshop on Statistical Learning in Computer Vision, 2004: 1-22.
- [96] J. Sivic, B.C. Russell, A.A. Efros, A. Zisserman, W.T. Freeman. Discovering objects and their location in images. Tenth IEEE International Conference on Computer Vision, 2005: 370-377.
- [97] A. Opelt, A. Pinz, A. Zisserman. Learning an alphabet of shape and appearance for multi-class object detection. International Journal of Computer Vision, 2008, 80(1): 16-44.
- [98] J. Shotton, J. Winn, C. Rother, A. Criminisi. TextonBoost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. Computer Vision - Eccv 2006 , Pt 1, Proceedings, 2006, 3951: 1-15.
- [99] G. Gill, M. Levine. Incorporating Shape Features in an Appearance-Based Object

- Detection System. Computer Analysis of Images and Patterns, Proceedings, 2009, 5702: 269-276.
- [100] J. Thureson, S. Carlsson. Appearance Based Qualitative Image Description for Object Class Recognition. ECCV, 2004: 518-529.
- [101] A.J. Grove, D. Schuurmans. Boosting in the limit: Maximizing the margin of learned ensembles. Fifteenth National Conference on Artificial Intelligence, 1998: 692-699.
- [102] I. Laptev, P. Perez. Retrieving actions in movies. IEEE 11th International Conference on Computer Vision, 2007: 2165-2172.
- [103] D.D. Le, S. Satoh. Ent-Boost: Boosting using entropy measures for robust object detection. Pattern Recognition Letters, 2007, 28(9): 1083-1090.
- [104] P. Viola, M. Jones. Fast and robust classification using asymmetric AdaBoost and a detector cascade. Advances in Neural Information Processing Systems, 2002, 14(1-2): 1311-1318.
- [105] S.Z. Li, Z.Q. Zhang. FloatBoost learning and statistical face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(9): 1112-1123.
- [106] J. Bentley. Programming pearls: algorithm design techniques. Communications of the ACM, 1984, 27(9): 865-873.
- [107] R.E. Schapire, Y. Freund, P. Bartlett, W.S. Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. The Annals of Statistics, 1998, 26(5): 1651-1686.
- [108] J. Hummel. Where View-based Theories Break Down: The Role of Structure in Shape Perception and Object Recognition. Cognitive Dynamics: Conceptual Change in Humans and Machines, 2000: 157-185.
- [109] F. Attneave. Some Informational Aspects of Visual Perception. Psychological Review, 1954, 61(3): 183-193.
- [110] R.O. Duda, P.E. Hart. Use of Hough Transformation to Detect Lines and Curves in Pictures. Communications of the ACM, 1972, 15(1): 11-15.
- [111] D. Douglas, T. Peucker. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature. The Canadian Cartographer, 1973, 10(2): 112-122.
- [112] Y.S. Gao, M.K.H. Leung. Line segment Hausdorff distance on face matching. Pattern Recognition, 2002, 35(2): 361-371.
- [113] Y. Ke, R. Sukthankar, M. Hebert. Event detection in crowded videos. IEEE 11th

International Conference on Computer Vision, 2007: 1424-1431.



## 攻读博士学位期间主要的研究成果

- [1] Weize Zhang, Jean-Marc Vézien, Ruofeng Tong, Patrick Bourdot, Jinxiang Dong. Hough Transformed Contour Line Segments for Object Detection. Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP 2010). (to appear)
- [2] Weize Zhang, Ruofeng Tong, Jinxiang Dong. Boosting 2-Thresholded Weak Classifiers over Scattered Rectangle Features for Object Detection. Journal of Multimedia, 2009, 4(6): 397-404.
- [3] ZHANG WeiZe, TONG RuoFeng, DONG JinXiang. Boosted cascade of scattered rectangle features for object detection. Science in China Series F: Information Sciences, 2009, 52(2): 244-251.
- [4] Weize Zhang, Ruofeng Tong, Jinxiang Dong. Z-AdaBoost: Boosting 2-Thresholded Weak Classifiers for Object Detection. Second International Symposium on Intelligent Information Technology Application, 2008, V2: 839-844.
- [5] 张维泽, 林剑波, 吴洪森, 童若锋, 董金祥. 基于改进蚁群算法的物流配送路径优化. 浙江大学学报(工学版), 2008, 42(4): 574-578.
- [6] Weize Zhang, Ruofeng Tong, Jinxiang Dong. Assembly Sequence Planning in VM System. 11th International Conference on Computer Supported Cooperative Work in Design, 2007, 908-913.
- [7] Zhihui Jiao, Weize Zhang, Ruofeng Tong. A Method for Accurate Localization of Facial Features. Education Technology and Computer Science, 2009, V3: 261-264.
- [8] 计算机软件著作权: 童若锋, 廖胜辉, 张维泽等. 基于医疗逆向工程的计算机辅助植牙系统. 编号: 软著登字第 044116 号, 登记号: 2005SR12615, 2005 年 10 月 25 日.



## 致 谢

衷心感谢我的导师董金祥教授和童若锋教授，感谢他们对我博士阶段学习和科研上的悉心指导和帮助。董老师渊博的学识、儒雅的学者风范和平易近人的人格魅力令我终身难忘，童老师严谨的治学态度，踏实的科研作风，诲人不倦的师者风范让我终生受益。论文的完成离不开童老师的全力指导与严格要求。童老师敏锐的洞察力及对关键问题的把握能力使我在研究中少走了许多弯路，而童老师对于科研工作的执着精神和忘我的工作态度也同样鼓舞着我。感谢大课题组的林兰芬教授、唐敏副教授、蔡铭副教授的指导和帮助，三位老师平易近人、谦逊和蔼，给实验室带来了自由、宽松和愉快的工作学习氛围。感谢学院的陈丽、贺蓓、章方铭和姚青老师，在行政和管理方面给我的热心帮助。

感谢中国留学基金委提供的奖学金，资助我在法国巴黎十一大交流学习一年，完成了本文的部分研究工作。感谢 LIMSI-CNRS 实验室的 Jean-Marc Vézien 教授和 Patrick Bourdot 教授在此期间给我的指导和帮助。两位教授风趣幽默、坦率直爽的性格让我在异国感受到文化差异带来的乐趣而不是困惑。

感谢人工智能所实验室提供的舒心的科研环境和优越的软硬件条件。最让人难忘的是有幸与来自五湖四海的同学共同学习工作。特别是合作过的同学，包括彭文、钱归平、廖胜辉、赵岚、王敏科、韩伟、丁濛、谢迪、丛林、颜洁清、姚原岗、赵杰伊、张小红、张贊、杜鹏、焦智慧、林江等，对我的研究工作给出了很多有益的建议的热忱的帮助。感谢其他师兄弟姐妹们给实验室带来那么多的快乐和精彩。

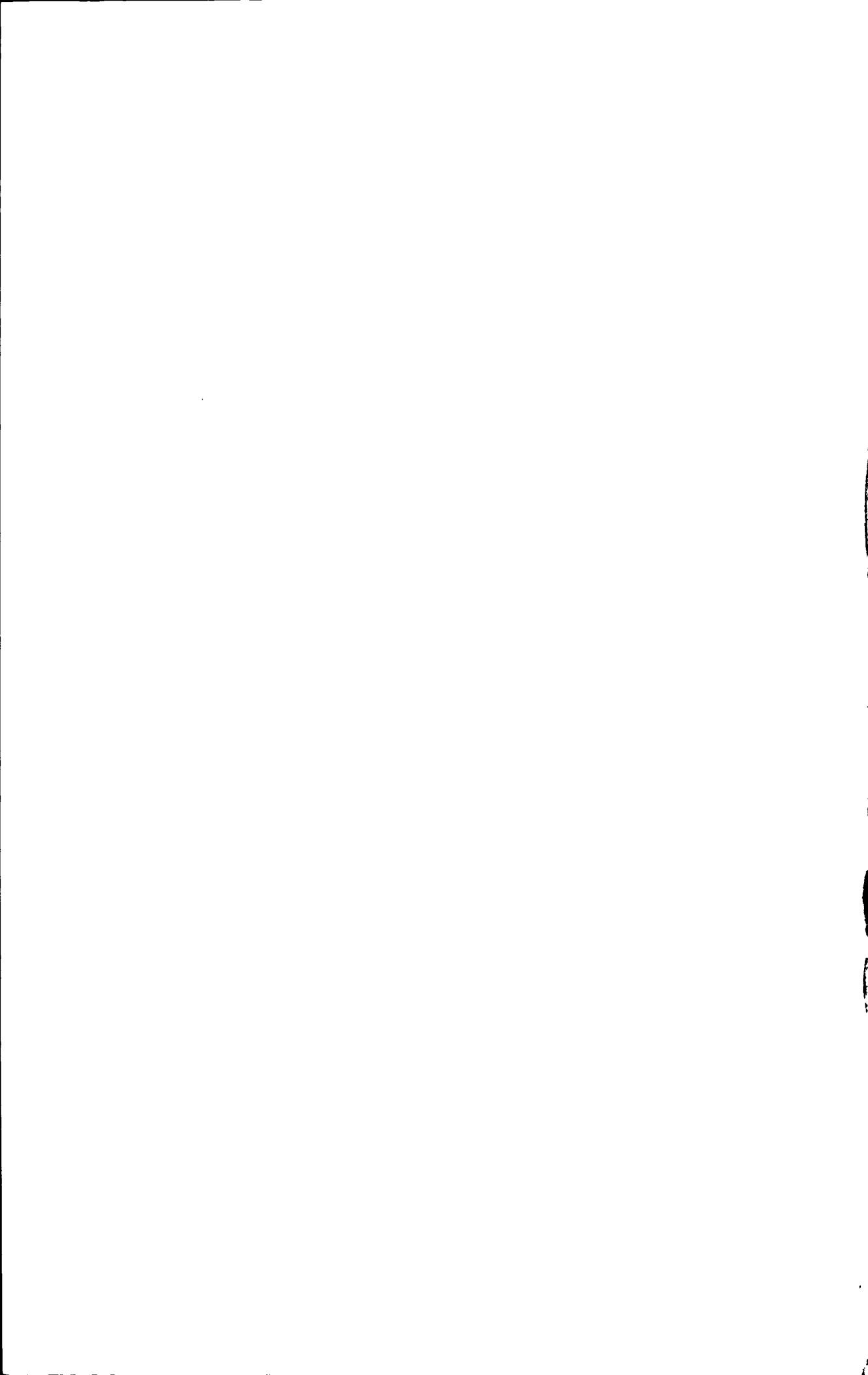
特别感谢我的父母和兄嫂们，感谢这些年来一如既往的给予我全面有力的支持和无微不至的关心。感谢侄子侄女们给我生活中带来的欢声笑语和美妙时光。同样非常感谢其他亲人朋友对我的关心和帮助。

衷心感谢所有给予过我帮助的老师、同学和朋友！

张维萍

2010年4月

于浙江大学



# 基于简单局部特征学习的物体检测方法

作者： 张维泽  
学位授予单位： 浙江大学



文献链接

本文链接: [http://d.g.wanfangdata.com.cn/Thesis\\_Y1853379.aspx](http://d.g.wanfangdata.com.cn/Thesis_Y1853379.aspx)