

分类号_____

密级_____

UDC _____

编号_____

中国科学院研究生院

博士学位论文

基于 1 范数最小化学习的人体检测算法研究

徐冉

指导教师 焦建彬 教授

中国科学院研究生院

申请学位级别 博士学位 学科专业名称 计算机应用技术

论文提交日期 2011 年 4 月 论文答辩日期 2011 年 5 月

培养单位 中国科学院研究生院

学位授予单位 中国科学院研究生院

答辩委员会主席 _____

独创性声明

本人郑重声明：所提交的学位论文，是本人在指导教师的指导下，独立进行研究工作所取得的研究成果。尽我所知，文中除特别标注和致谢的地方外，学位论文中不包含其他人或集体已经发表或撰写过的研究成果，也不包含为获得中国科学院研究生院或其它教育机构的学位或证书所使用过的材料。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

签名：_____日期：_____

关于学位论文使用授权的说明

本人完全了解中国科学院研究生院有关保管、使用学位论文的规定，其中包括：
①学校有权保管、并向有关部门送交学位论文的原件与复印件；②学校可以采用影印、缩印或其它复制手段复制并保存学位论文；③学校可允许学位论文被查阅或借阅；④学校可以学术交流为目的，复制赠送和交换学位论文；⑤学校可以公布学位论文的全部或部分内容。

（涉密的学位论文在解密后应遵守此规定）

签名：_____导师签名：_____日期：_____

摘 要

人体检测是计算机视觉与模式识别领域中的研究热点，具有较强的理论意义及较大的应用价值，受到国内外研究人员的广泛关注。图像传感器捕捉到的人体目标具有一定的姿态与视角变化，在嘈杂背景中呈现复杂的模式，这使得人体检测成为目标检测方向中的一个难题。

要使计算机能够自动地识别或检测人体目标，需要赋予它透过图像内容的变化区分人体目标的本质特征与非本质特征的能力。本文首先提出一种基于 1 范数最小化学习 (L1-norm minimization learning, LML) 的优化模型，然后在此基础上提出一种线性分类算法 (Linear LML, L-LML)。该算法利用 1 范数的稀疏性，从 VC 维最小化的角度构建线性分类器同时实现特征选择。在分类器构建过程中，L-LML 算法通过对分类器的法向量建立 1 范数最小化约束模型，然后根据训练样本的特征分布计算分类器的阈值，从而完成线性分类器的学习。由于法向量的稀疏性，对特征向量实施法向量加权，突出了特征向量各元素之间的差异性，实现了特征向量的稀疏表示，达到了特征选择的目的。考虑到样本分布的复杂性及非线性可分等特点，本文在 LML 优化模型的基础上，提出了一种基于 1 范数最小化学习的非线性学习算法 (Non-linear LML, NL-LML)。该算法首先对样本实施一种由核函数诱导的非线性变换，将样本映射到高维空间，并对样本重新进行表示，然后利用 LML 模型对变换样本进行训练。这种由核函数诱导的非线性变化简单有效，与 LML 优化模型结合可以实现非线性分类。NL-LML 也可以实现特征的稀疏表示。不过与上述 L-LML 不同，此过程是在变换特征空间而非原空间进行的。

为实现图像块状特征的选择及提高目标检测效率，本文基于 LML 优化模型设计了一种级联分类器 (Cascaded LML, CLML)。该分类器中的每一级为多个弱分类器加权投票组合构成的强分类器。每一个弱分类器对应一个图像块状特征，对弱分类器的选择即是对特征块的选择。弱分类器的法向量和阈值分别通过 LML 优化模型和 min-max 罚函数模型训练得到。在强分类器的构建过程中，本文提出使用整数规划模型，从全局最优化的角度确定有投票权的弱分类器，从而实现弱分类器的最优组合方式。

本文使用梯度方向直方图特征 (HOG) 及变尺度梯度方向直方图特征 (v-HOG) 表示人体目标，所提到的所有算法均在公开的人体测试集 INRIA 与 SDL 上进行了验证。

关键词：1 范数最小化, 人体检测, 特征选择, 分类器

Abstract

Research on Human Detection via L1-norm minimization learning

Ran Xu (Computer Application Technology)

Directed by: **Jian-bin Jiao** (Professor)

Human detection is a hot topic in the computer vision and pattern recognition, which has a strong theoretical significance and a wide application. It has attracted many researchers' attention. A human body has the variations of multi-posture and multi-view, which is collected by image sensors. It displays a complex pattern under a noisy background and it is a difficulty in object detection.

To automatically detect or recognize humans by computers, we should give the computers the ability to discriminate the essential and nonessential features of the images. This dissertation first proposes a L1-norm minimization learning (LML) optimization model and then designs a linear classification method based on the LML model (L-LML). The L-LML method utilizes the sparseness of L1-norm to design a linear classifier and simultaneously to achieve feature selection from the perspective of direct VC-dimension minimization. During constructing the classifier, L-LML method uses L1-norm minimization model to optimize the normal vector and compute the threshold according to the feature distribution of training samples. Owing to the sparseness of the obtained normal vector, the L-LML weights the feature vector by using the normal vector, which highlights the variation among the elements in the feature vector and achieves the goal of feature selection.

In consideration of the complicated and nonlinear distribution among samples, we propose a nonlinear classifier (NL-LML) based on the LML optimization model. First, the NL-LML method performs a nonlinear transformation induced by a kernel function, which projects the samples into a high dimensional space. Then, the samples are reformulated and are trained by the LML model. The nonlinear transformation induced by kernel functions is simple and effective, which is combined with LML model to perform a nonlinear classification. The NL-LML method also can achieve the sparse representation of features. The NL-LML method implements the sparse operation in the transformed feature space instead of the original one. Therefore, it is different from L-LML method.

To select the block features from an image and improve the detection efficiency, a cascade L1-norm minimization learning (CLML) classifier based on the LML model is designed. In each level, the CLML is constructed by a strong classifier, which consists of

some weighted combination of weak classifiers. Each weak classifier corresponds to a block feature and thus selecting the weak classifiers is to choose the feature blocks. A normal vector of a weak classifier is trained by the LML model and a threshold is determined by the min-max penalty function model. During building the strong classifiers, we use integer programming to decide which weak classifiers should be involved to form a strong one from the global perspective in the optimal way.

We use two feature descriptors to represent a human body, which are HOG and v-HOG feature respectively. All proposed methods in this dissertation are validated on two public human datasets respectively, namely SDL dataset and the INRIA dataset.

KEY WORDS : L1-norm minimization, Human detection, Feature selection, Classifier

目 录

摘 要.....	I
目 录.....	V
图目录.....	VII
表目录.....	IX
第一章 绪论.....	1
1.1. 课题背景及意义.....	1
1.2. 国内外研究现状.....	1
1.3. 难点问题分析.....	6
1.4. 本文的研究内容与主要贡献.....	7
1.5. 本文的组织结构.....	8
第二章 人体检测算法综述.....	9
2.1. 特征描述子.....	9
2.1.1. Haar-Like特征.....	9
2.1.2. HOG特征.....	11
2.1.3. v-HOG特征.....	12
2.1.4. MSO特征.....	13
2.2. 特征选择.....	14
2.2.1. 特征选择的定义.....	14
2.2.2. 特征选择的分类.....	18
2.3. 分类器.....	19
2.3.1. SVM分类器.....	19
2.3.2. Adaboost分类器.....	21
2.4. 本章小结.....	23
第三章 基于线性 1 范数最小化学习的人体检测.....	25
3.1. 稀疏表示综述.....	25
3.2. 1 范数研究综述.....	27
3.3. 经验风险与期望风险.....	29
3.3.1. 经验风险与期望风险的定义.....	29
3.3.2. 经验风险与期望风险之间的关系.....	30
3.4. 基于 1 范数最小化学习的优化模型.....	31
3.5. 线性分类器的构建.....	33
3.6. 特征稀疏表示.....	34

3.7.	实验结果	36
3.8.	本章小结	39
第四章	基于非线性 1 范数最小化学习的人体检测.....	41
4.1.	非线性 1 范数最小化学习的思路	41
4.2.	核函数	42
4.3.	非线性分类器的构建	43
4.3.1.	基于核函数的非线性变换.....	43
4.3.2.	非线性变换与LML优化模型结合.....	45
4.4.	特征稀疏表示	45
4.5.	实验结果	46
4.5.1.	IRIS数据集.....	46
4.5.2.	人体数据集.....	47
4.6.	本章小结	51
第五章	基于级联 1 范数最小化学习的人体检测.....	53
5.1.	弱分类器的构建	53
5.1.1.	min-max罚函数的来源.....	54
5.1.2.	min-max罚函数模型.....	55
5.2.	强分类器的构建	57
5.2.1.	整数规划模型.....	58
5.2.2.	整数规划模型的求解.....	59
5.3.	级联分类器的构建	61
5.4.	算法讨论	63
5.5.	实验结果	68
5.5.1.	参数分析.....	68
5.5.2.	性能比较.....	70
5.6.	本章小结	74
第六章	总结与展望.....	77
6.1.	本文工作总结	77
6.2.	未来工作展望	78
参考文献	79
附录 1	89
博士期间发表论文目录	89
致 谢	91

图目录

图 1-1.人体检测算法流程图.....	7
图 2-1.Haar小波示意图.....	10
图 2-2. Haar-Like特征.....	10
图 2-3. 拓展的Haar-Like特征.....	11
图 2-4.HOG特征.....	12
图 2-5.v-HOG特征.....	13
图 2-6.MSO特征.....	13
图 2-7.特征选择流程图.....	15
图 2-8.最大间隔原理.....	20
图 3-1.L-LML算法流程图.....	25
图 3-2.人体HOG特征表示.....	34
图 3-3.稀疏表示和稠密表示.....	35
图 3-4.人体可视化显示.....	36
图 3-5.正反例训练样本.....	37
图 3-6.LML在SDL上测试曲线.....	38
图 3-7.LML在INRIA上测试曲线.....	38
图 3-8.LML检测实例.....	40
图 4-1.NL-LML算法框图.....	41
图 4-2.NL-LML稀疏表示.....	46
图 4-3.NL-LML在SDL上性能比较.....	49
图 4-4.NL-LML在INRIA上性能比较.....	49
图 4-5.NL-LML检测实例.....	51
图 5-1.特征选择示意图.....	61
图 5-2.级联机制示意图.....	61
图 5-3.1 范数与 2 范数的几何解释.....	65
图 5-4.异或样本分布.....	67
图 5-5.分类误差.....	69
图 5-6.级联数目.....	69
图 5-7.CLML特征块展示.....	70
图 5-8.CLML在INRIA上单独性能比较.....	71
图 5-9.CLML在INRIA上性能比较.....	71

图 5-10.CLML在SDL上性能比较.....	72
图 5-11.CLML检测实例.....	74

表目录

表 2-1.距离度量函数.....	17
表 2-2.相关性度量函数.....	17
表 2-3.信息度量函数.....	18
表 2-4.Adaboost算法流程图:	23
表 3-1.L-LML算法流程图	33
表 4-1.IRIS数据集上NL-LML性能	47
表 5-1.CLML算法训练过程	63

第一章 绪论

1.1. 课题背景及意义

图像或视频中的目标检测，意在基于目标表观和轮廓区域等信息，准确地对图像中感兴趣的进行定位，即将目标的定位和识别合二为一。人体检测是目标检测的一个特例，是目标检测中比较典型的研究问题之一。人体检测涉及到图像处理、模式识别和计算机视觉等领域的知识，是分析人体运动视觉信息的基础，如人体目标跟踪、轨迹分析、姿态识别等。在视频监控、智能交通和汽车辅助驾驶系统等应用的驱动下，人体检测越来越受到研究者的重视，在现实中有着很广泛的应用前景。

在国防领域，鲁棒的人体目标检测可用于辅助电子探测设备自动识别危险人体目标，帮助特种部队及空降兵来探测周边区域的情况并保卫军营的安全，防止人为的偷袭和破坏等活动；在民用领域，该技术可用于重点建筑（博物馆、体育场馆、重要桥梁等）及重要地点（核电站、水电站、地铁站、银行等）的安全保卫工作；随着城市规模的扩大，住宅小区逐渐成为安全防范的重点。结合红外/近红外等视频设备，该技术可强化小区安全监控的能力、减少犯罪，从而增强居民的安全感。在智能交通领域，可以及时监控交通事故并可以在十字路口、新建高速路端等进行自动行人检测和安全预警。在汽车辅助驾驶系统中，可以通过车载视频，实时分析路况和检测行人变化，从而警示驾驶员路况信息。

虽然在近年来的图像处理和模式识别领域中，人脸、车牌等其他目标检测方法已经日趋成熟，但是在复杂环境下可靠的人体目标检测算法还有待进一步研究，其原因在于：首先，人体是一个非刚性的、多姿态的、多角度的物体；其次，含有人体目标的图像，其背景一般都是复杂多变的；再次，人体目标很容易被其他人体或者其他物体遮挡。为实现鲁棒、快速地检测人体，人体目标的自动提取和识别显得极为重要。通过运用机器学习与模式识别中的相关知识，使计算机能够自动、准确地检测人体，主要是赋予它透过图像的各种变化区分人体的本质特征与非本质特征的能力。与此相关的问题逐渐成为模式识别、图像处理领域内的研究热点。此外，开展此方面的研究对目标的模式表达及检测方法等领域内核心问题的解决具有重要的理论意义。

1.2. 国内外研究现状

目前关于人体检测算法的研究，在理论上，国内外已经形成了几种比较经典的框架。

从应用的角度，人体检测算法可以简单地分为两类：基于动态视频帧的检测和基于单帧静止图片的检测。由于动态视频可以被划分成多个单帧静止图片，因此动态视频帧检测问题可以转化为单帧静止图片的检测问题。

基于动态视频帧的检测算法主要是结合运动信息来识别“动目标”。目前代表性的基于动态视频帧的检测算法有背景差分法[1-3]、帧差法[4-5]等。这些基于动态视频帧的检测算法在复杂、动态背景下，效果不是很好，原因在于这些算法一般只适用于静止的背景。而且，在静止的背景中，前后帧之间光线的轻微变化及树木的摇摆等多种干扰信息，都极易被判别为人体目标。由于视频中背景的复杂化与动态化，目前基于动态视频帧的检测算法的发展受到了一定的约束，因此在应用方面会受到一些影响。在最近的一些研究中，研究者们更关注基于单帧静止图片的检测。虽然在静止图片的检测中，人体的定位相对繁琐一些，然而一旦在单帧图片上人体检测的研究得到突破，那么它将会在实际应用中取得很好的效果，有非常远大的前景。

基于单帧静止图片的检测方法中，根据人体构建模式的划分不同，可分为整体模型检测与部分模型检测方法。整体模型[8,11,12,36,38]是基于整个训练样本窗口，提取人体的整体或者部分信息作为人体特征，然后采用某种分类准则获得判别函数。部分模型是指将人体进行分块，比如说按着图片的大致比例将人体分为头部区域，躯干区域[13, 17, 18,40]等或者基于人体的自然形态进行分块（Pictorial structure）[41,42]，比如说人的头部，胳膊，腿部等。无论是整体模型检测还是部分模型检测，都主要涉及两部分内容，一是特征表述（特征描述子），它是指从图像中提取出表示人体目标的特征向量，该特征应该尽量对光照、背景、表观等因素的变化不敏感。二是分类器的构建，它主要是使用前面所提取人体的某种特征，按照某种学习准则，获取分类函数的过程。特征表述和分类器，主要是让计算机意识到什么样的模式是属于人体的，什么样的模式是属于背景的。大多数研究者致力于这两方面的研究。下面本节将分别介绍整体模型与部分模型中用到的比较典型的特征描述子与分类器。

在整体模型中，许多特征描述子用来表示人体模式。最明显、最容易想到的能够表示人体模式的特征是形状信息[6, 29]。复杂的人体形状信息通过离散或连续的轮廓表示方式，可以捕捉到人体的某些共性特征。Papageorgiou 和 Poggio[30]等人提出使用 Haar 小波函数作用于训练样本，获得基于灰度差的 Haar-like 特征。随后许多研究者参与到该特征的改进中。如在[7,38]中,完备的 Haar-like 特征分别用来表示人脸与人体特征。虽然 Haar-like 特征对于人脸表示的效果比较好，但它不太适合表示人体目标，原因在于 Haar-like 特征比较适合描述目标的显著区域，如眼睛、嘴巴、眉毛等，不太适合表示边

缘轮廓信息,容易受到目标形态,光照条件及视角的影响。随后,一些考虑到空间位置关系的自适应特征涌现[19,34]。这些特征中一种比较典型的是局部感知域特征[19],它模拟了人类视觉系统中的神经结构。在文献[8]中,作者提出稠密的、重叠的、固定尺度的局部特征描述子——梯度方向直方图 HOG (Histogram of Oriented Gradient)表述人体。该描述子借鉴了 SIFT (Scale Invariant Feature Transform) 特征点中,运用梯度方向直方图表示目标的这一思想。后来,在 HOG 特征的基础上,涌现出一些改进版本的特征[9, 14]。这些改进的 HOG 特征,认为原始的 HOG 由固定尺度、固定位置的特征块组成,其不能很好的把握人体的局部轮廓特性,所以存在着弊端,研究者们引入改进的 HOG 特征,并且取得了比较好的结果。在文献[10],作者使用区域的协方差算子(COV)来表示人体特征。区域中的每个像素点是由灰度值,梯度值,位置等信息组成的特征向量,每个区域的 COV 算子是由位于该区域的所有像素点特征向量构成的协方差矩阵,协方差矩阵可以很好地把握不同位置、不同尺度下人体区域的特征。Mu 等人[11]认为原始的用来表示纹理特征的局部二值模式 LBP (Local Binary Patterns)算子,虽然在人脸识别、纹理检测等方面取得很好的效果,但其不适合于描述人体的轮廓,因此提出使用改进的 LBP 算子来描述人体。在文献[12]中,作者为解决部分遮挡条件下的人体检测问题,采用 HOG 特征和 LBP 特征相结合的方法。LBP 特征可以表述纹理,对单调灰度变化有不变性,当背景比较复杂,有干扰边缘时,HOG 特征将受到很大影响,而此时 LBP 特征可以滤除背景噪声。因此,HOG 特征与 LBP 特征相结合表示人体目标,取得较好的检测效果。在文献[14]中,作者提出了一种新的多尺度方向(MSO)特征描述子,与现有方法不同,该描述子采用了尺度/方向竞争和块装配机制。这种方法可以很好把握人体的整体信息,但是对于人体的姿态和视角变化的容忍度有限。在文献[36]中,基于边缘、纹理、颜色三种信息组合的高维描述子用来表示人体模式。Wu[13]等人提出 Edgelet 特征表示人体模式,每个 Edgelet 特征是一条边,反映着人体局部位置的轮廓细节信息。Liu[44]等人提出使用粒度可变方向划分描述子(Granularity-tunable Gradients Partition, GGP)表示行人,该算子在不同的尺度情况下,可以分别等价于 Edgelet 特征和 HOG 特征等,因此该算子从多层次角度,实现了细节的边缘信息与粗略的轮廓信息的统一,能够很好地把握人体的轮廓。此外,他还在论文[104]中融合了 GGP 算子与时空信息的相关性,对 GGP 算子进行拓展,取得较好的效果。崔新毅在[105]中分别提出三种人体特征描述子,分别是 3 维 Haar 特征、基于梯度信息的 3 维 Haar 特征和边缘方向直方图特征。以上所述这些研究都是从特征的角度,考虑使用何种特征能更好地表示人体目标。

在获得特征表示后,人体检测算法需要对区域中的特征进行分类,即需要构建分类

器,从而判别区域中是否含有人。相比于特征方面的研究,分类器(分类方法)的研究稍微少一些。目前,现有的分类方法可大体分为两种,一种是确定性方法,如模板匹配算法、神经网络算法、SVM(支持向量机)算法、Boosting 算法(Adaboost 和 Logitboost 等),另一种方法是概率推理方法,如贝叶斯后验概率算法等。模板匹配算法主要是使用某种相似性规则(如距离度量等)来测量特征向量与原始模板中特征向量之间的匹配程度。神经网络算法根据不同的网络结构,可以分为很多种。大部分的神经网络方法都是通过最小化网络参数的误差准则来估计最优的判别平面。与神经网络方法不同,SVM[20]算法并没有采用直接最小化误差准则,而是最大化优化判别函数在正反例样本之间的边界(边界最大化原则)。考虑到样本的非线性分布,SVM 方法使用核理论将样本投影到高维空间中,并将核理论中的内积形式运用到优化模型的对偶规划中,进而在高维空间中求解线性分类器。Adaboost[21]算法是 Boosting 算法中典型的代表算法之一,它采用加权投票机制,每个弱分类器看做一个投票委员,通过贪婪的重采样策略,每次选择一个具有最小错误率的弱分类器参与强分类器的投票。理论上,Schapire 等人[75]已经证明 Adaboost 算法的迭代过程是收敛的,并且还证明了 Adaboost 算法也遵循边界最大化原则。概率推理方法主要是使用贝叶斯后验概率,分别计算先验知识与条件概率之积,比较后验概率的值与阈值的关系,实现对人体模式的分类。下面详细介绍下几种分类方法在人体检测中的应用。

在文献[7]中,Viola 等人首次提出级联(Cascade)机制,并将其与 Adaboost 算法结合用于人体检测。该方法能够快速检测人体,在当时获得较高的检测性能。在文献[10]中,因为 COV 算子是正定对称矩阵,作者证明这些对称正定矩阵位于连通的黎曼流形上,并且认为常规的机器学习方法不能处理矩阵问题,因此将所有数据点投影到同一类中心点的切向量空间中,使用 Logitboost 算法进行分类。在文献[8, 11, 18, 33]中,线性或者核函数 SVM 方法用来分类人体目标。在文献[9]中,作者首先使用线性 SVM 训练得到弱分类器,然后使用级联 Adaboost 算法组合这些弱分类器形成强分类器,最后使用级联的强分类器检测人体。在文献[31]中,作者提出了一种针对自适应局部感知模型特征的多层的神经网络算法,用于人体检测。除此之外,文献[39]中提出一种加性分类器的组合用来检测行人。在文献[36]中,作者首先提出使用部分最小二乘(PLS)方法处理高维空间中的样本,降维后的样本用来训练 SVM 分类器,从而实现行人窗口分类。孙庆杰在文献[46]中提出模板匹配的方法检测行人。他使用基于侧影的人体模型,主要是分析人体的四肢和躯干间的几何位置关系,使用改进的彩色边缘信息作为各个部分的底层特征,最后抽象出高层特征,并将各个部分的拓扑结构与基于侧影的人体模型进行拟合

匹配。朱文佳在文献[47]中为了提高人体检测的速度，首先改进了 HOG 特征，将 36 维 HOG 特征块精简为 8 维特征，然后使用 fisher 判别法，训练每个 Adaboost 弱分类器的阈值，结合级联机制实现快速的人体检测。徐堃在论文[106]中，也是使用改进的 Adaboost 算法训练弱分类器，并使用 fisher 判别准则对选出的弱分类器进行二次训练，最后形成强分类器，实现视频序列中的人体检测。

上述研究是从整体模型角度，分别提取特征和设计分类算法。下面介绍从部分模型角度，一些现有的人体检测算法现状。部分模型受到广泛关注的原因之一，在于它能够部分地解决人体检测中存在的一些遮挡、多姿态等疑难问题，这些算法的典型研究可参见文献[13, 15-18, 37, 40-42]等。基于部分模型的人体检测算法的研究在不久的将来，应该会成为人体检测问题的主要研究方向。

在文献[13]中，作者将人体样本分成三个部分：头部，胸部和腿部，对每一部分使用实值的 Adaboost 算法训练，最后使用联合的似然函数组合各个部分的响应，来判断图像或者视频中是否有人。这种将人体模型分块的思想能比较好的解决遮挡问题，但同时也带来了一些错判。文献[15]中，也是将人体区域分块，然后根据每个块之间的约束关系，来寻找人体目标，并且作者提出两种方法寻找人体目标，一个使用投影算法另一个使用概率推理的方法。文献[16]中，作者使用从上至下的概率分割算法，通过局部和全局的人体的特征信息，能够在拥挤的人群中检测出每个人体目标。在文献[17]中，作者将特征信息以及几何约束关系信息使用逻辑程序语言进行编码，使用逻辑推理的方法检测行人。在文献[18]中，人身体被分为四个部分，每个部分使用 SVM 进行训练，并且作者提出四个部分得到的 SVM 分类器的自适应组合来检测行人。Lin.等人[37]提出使用基于部分的局部策略与基于模板的全局策略并在贝叶斯概率框架下搜索人体。[40]作者认为人体是由几部分可形变模板的组合而成的，每一部分使用隐 SVM 进行判别，最后通过几何关系学到各部分的一个整体模型。在文献[41,42]中，作者们都是基于 Pictorial structure 对人体进行部分模型的分解，这种分解的好处在于它能按照人体的自然形态，捕捉到人体部分的信息。文章[42]还进一步地基于这种部分模型检测的方法，对人体的姿态进行了识别。

Munder 等人[20]对比了几种特征及分类器，结论是：局部轮廓特征是最有效特征表达，而 SVM 分类器与级联 Adaboost 是最有效的人体分类器。其中级联分类器的主要优点是较快的检测速度，SVM 的主要优点是对目标模式变化的鲁棒性。在其后续的综述论文中[45]，Munder 等人的结论是：HOG 特征加 SVM 分类器获得了最好的性能，而 Haar 小波特征加级联 Adaboost 分类器性能稍差，但是速度高出其它方法很多。本文的作者认

为 SVM 分类器虽有较好的分类性能，但不能恰当地选择特征。和 SVM 分类器相比，级联 Adaboost 算法虽然可以达到更高的速度，但是 Adaboost 算法中弱分类器的训练有些繁琐，因为它需要通过观察正反例样本的特征分布来确定弱分类器的阈值。

1.3. 难点问题分析

综上所述，人体检测算法在近几年比较热门，层出不穷。目前现有的人体检测算法，虽然能达到一定的检测性能，但是因为人体姿态多样性、遮挡等问题，加上检测算法对实时性、鲁棒性、准确性等要求，使得现有的检测算法仍有待改进。例如，一些算法误检测或者漏检测率比较高；一些算法虽然在检测精度上有了很好的提高，但是检测时间有些长。如何在复杂背景下，兼备速度与精度从而提高人体检测算法的实时性、准确性、鲁棒性是目前人体检测的难点。

人体检测算法的难点分析如下：

1. 在训练过程中，人体目标的表达方式不统一。目前大多数人体检测算法都是基于监督学习方法，即通过现有的训练样本，让计算机学习到人体统一的表达模式。然而因实际生活中人体姿态的多样性、视角变化、衣服样式等使得训练样本的模式并不是很统一，因此即使再好的分类学习方法也只能得到比较典型的一些人体特征，如头部特征、肩膀特征等。

2. 在检测过程中，人体姿态、视角变化、服饰等表现因素与训练过程中学习得到的人体模型有差别，另外人体目标与摄像机的距离的远近也会在一定程度上改变人体目标的特征，很大程度上的影响着检测的精度。这些问题的产生，对人体特征描述子以及分类器的设计提出了更高的要求。总而言之，目标的特征表示与分类算法的设计，在很大程度上决定了人体检测算法的检测精度。

3. 人体目标的部分及完全遮挡。人体目标有可能受到背景或者其他人体目标的遮挡。部分遮挡造成人体目标的部分特征缺失，而这种特征缺失是随机性的，计算机很难捕捉到这种缺失信息，对此只能依赖未被遮挡的特征信息。这种信息随机性的缺失，对特征的表述及分类方法提出了更高的要求。对于人体目标的部分遮挡，一些算法提出使用图像分割或深度信息等手段解决，但是暂时还没有很好的算法可以完全的解决这一问题。

4. 背景模式与光照亮度的影响。实际应用中的背景比较复杂，存在各式各样、形态各异的物体，例如：路边树木、栏杆等，这些信息和学习到的人体竖直模式比较相似，此外，背景光线亮度变化等这些干扰信息，如同一目标在不同的光线亮度条件下会产生

明显不同的图像模式。因此，背景模式的变化与光照亮度的影响是实际应用系统中阻碍性能提升的瓶颈。

1.4. 本文的研究内容与主要贡献

特征表示和分类器构建，是人体检测算法的两个关键要素。目前，大多数人体检测算法也都从特征描述子或者分类器的角度作为算法改进的出发点。本文主要从特征描述子的特征选择和分类器设计的角度，构建了鲁棒的、快速的、精确的人体检测系统，一定程度上解决了由于人体姿态、衣着、光线、遮挡与背景所带来的干扰问题。本文的人体检测算法是基于监督学习的框架，分别由训练过程和检测过程两部分组成，其中虚框标注的部分是本文重点的研究部分，其研究内容如下：

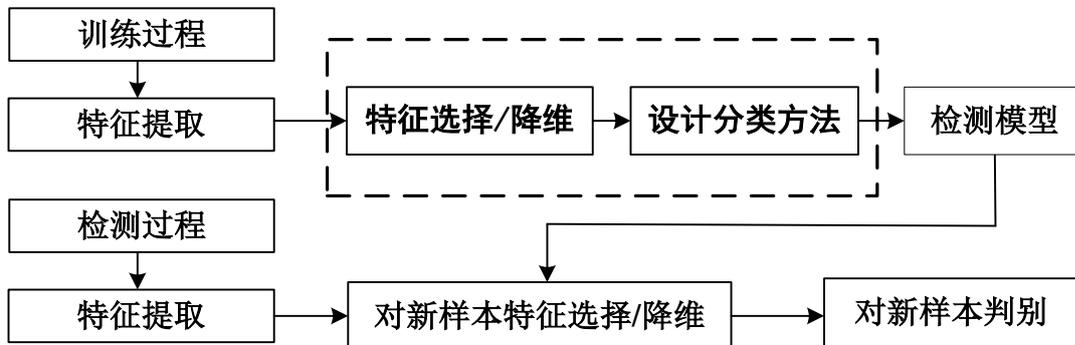


图 1-1.人体检测算法流程图

本文主要研究内容及贡献总结如下：

★ 提出了一种基于 1 范数最小化学习(LML)的优化模型，并在此基础上建立了线性人体检测(L-LML)算法。LML 优化模型是依据 VC 维最小化的原则建立的。L-LML 算法意在利用 LML 优化模型中 1 范数的稀疏特性，同时实现对特征维度的选择与分类器构建。它采用所提的 LML 模型训练出线性分类器的法向量，然后，通过使用正反例样本特征分布的加权组合，计算得到线性分类器的阈值，从而完成线性分类器的训练。该算法因为对特征进行了稀疏化表示（特征选择），使得算法在检测过程中对多视角、多姿态的人体仍然具有较好的性能，并对部分遮挡问题不敏感。

★ 提出了一种基于 1 范数最小化学习的非线性人体检测(NL-LML)算法。针对样本在特征空间中呈现出复杂的非线性分布，本文提出一种基于 1 范数最小化的非线性人体检测算法。对于低维空间中不是线性可分的样本，该算法假设在高维空间中近似认为样本线性可分，首先对样本实施基于核函数诱导的非线性变换，然后结合 LML 优化模型，在变换空间中训练线性分类器的法向量，同时构建线性分类器，实验结果验证了 NL-LML 算法可以实现样本的非线性判别，并且其效果优于 L-LML 算法。

★ 设计了一种基于 1 范数最小化学习的级联人体检测(CLML)算法。与前两种方法

不同，CLML 算法采用不同位置、不同尺度下的块状特征表示人体，意在同时实现块状特征的选择与分类器的设计。该算法采用弱分类器加权投票机制形成强分类器，每个块状特征对应一个弱分类器，弱分类器是通过所提的 LML 优化模型和 min-max 罚函数模型训练得到的。该算法提出使用整数规划从全局的角度选择一些弱分类器形成强分类器，目的是实现块状特征的最优选择。最后 CLML 算法结合级联机制，实现对人体快速、准确、鲁棒的检测，实验结果表明，该算法的性能优于现在一些流行的人体检测算法。

1.5. 本文的组织结构

第一章，绪论。主要论述了人体检测算法研究的背景及意义，简略地描述了现在国内外人体检测算法的研究现状以及基于静止图片中人体检测算法现存在的难题，进而列出本文的研究目的和主要研究内容，最后总结了本文的贡献。

第二章，人体检测算法综述。详细地介绍了基于静止图片的一些人体检测算法，重点是从典型特征描述子、特征选择方法与目前比较受欢迎的分类器角度，来讨论现有人体检测算法的优点与不足。

第三章，基于线性 1 范数最小化学习的人体检测。本章主要论述了稀疏表示的发展历程，介绍了 1 范数近年在模式识别领域的研究进展，提出了基于 1 范数最小化学习的优化模型，分析了 1 范数最小化学习的优化模型与 VC 维、期望风险、经验风险之间的关系，最后阐述了 1 范数最小化学习如何实现对人体模式进行特征选择，设计了适用于人体检测的线性分类器，并在人体数据集上验证了所提算法的性能。

第四章，基于非线性 1 范数最小化学习的人体检测。针对样本的非线性分布，本章主要探讨了基于 1 范数最小化学习优化模型的本质特性，得出其不能直接在对偶规划中引入内积形式实现非线性判别这一结论，进而提出结合基于核函数的非线性变换，建立了一种适合于 1 范数最小化的非线性分类模型，并在 IRIS 数据集与人体数据集上进行了验证。

第五章，基于级联 1 范数最小化学习的人体检测。针对样本的块状特征，本章提出了使用 LML 优化模型和 min-max 罚函数模型分别构建弱分类器的法向量和阈值，然后系统地介绍了 min-max 罚函数模型的来源、与对策论之间的关系及相应的解法。对所训练得到的弱分类器，本章提出使用整数规划模型学习强分类器，从全局角度完成对块状特征的选择与分类。随后本章介绍了关于整数规划模型的求解方法，通过引入级联机制，构建了一个级联的 LML 分类器。最后，本章探讨了基于级联 1 范数最小化学习的人体检测算法与其他算法的区别，并在人体数据集上验证了其效果。

最后是对现有工作的总结和对未来工作的展望，以及讨论使用分块的人体检测算法实现对人体的部分遮挡的处理、提高算法的鲁棒性和精准性等。

第二章 人体检测算法综述

人体检测是计算机视觉、模式识别等领域的热点研究问题之一。目前，如何提高人体检测算法的实时性、鲁棒性和准确性等方面的研究，成为学者们追求的目标。本章将重点介绍和综述一些经典的特征提取方法，特征选择方法及分类方法。

2.1. 特征描述子

这一部分主要介绍几种具有代表性的人体特征描述子：**Haar-Like** 特征、**HOG** 特征、**v-HOG** 特征和 **MSO** 特征。

2.1.1. Haar-Like特征

M. Oren [49]等人提出了使用 Haar 小波作为人体的 Haar-Like 特征。该算法提出使用监督学习方法的思想，从而消除了由于手动建模与运动信息引起信息不准确的影响。它最大的成功在于首次运用机器学习的思想，通过对训练样本特征提取获得行人的各种小波表示模板，再利用一些机器学习与模式识别中的方法让计算机“记住”这些小波模板，实现对图像窗口进行判断。下面将详细介绍算法中使用的 Haar-like 特征及其演化版本。

Haar-Like 特征是通过图像使用 Haar 小波变换，将图像的像素灰度空间变化为小波系数空间，并得到不同尺度上的小波系数，获取了图像在不同层次上的细节特征，从而形成一个完备的特征字典，字典中的每个特征反映图像局部区域在不同方向上的灰度差。M. Oren 的合作者 Papageorgiou 等人[30]，在行人检测中使用两个尺度上的小波函数对图像进行 Haar 小波变换，选择能够表现类内紧致性的小波系数作为人体特征。后来，Viola 等人又对原始的 Haar-Like 特征进行了拓展，实现了对图像更精细的描述。该特征分别在人脸、人体等目标上进行了实验，在人脸检测上的效果好于人体检测的效果[7,35]。

在不同的分辨率空间下，Haar 小波的形式有所变化。分辨率空间在数学上是指：令子空间序列 $V^0 \subset V^1 \subset \dots \subset V^j \subset V^{j+1} \dots$ 表示多分辨率空间，每一个更大的子空间比邻近的子空间包含着更多的细节知识，子空间 V^j 的基底是一些尺度函数，具有如下形式：

$$\phi_i^j = \sqrt{2^j} \phi(2^j x - i) \quad i = 0, \dots, 2^j - 1. \quad (2.1)$$

$$\text{其中 } \phi(x) = \begin{cases} 1 & \text{如果 } 0 \leq x \leq 1 \\ 0 & \text{其他} \end{cases} \quad (2.2)$$

定义小波子空间 W^j 是两个连续多分辨率空间的正交补子空间， $V^{j+1} = V^j \oplus W^j$ ，子空间

W^j 的基底如下: $\psi_i^j = \sqrt{2^j} \psi(2^j x - i) \quad i = 0, \dots, 2^j$, 其中函数 $\psi(\cdot)$ 有如下定义:

$$\psi(x) = \begin{cases} 1 & \text{如果 } 0 \leq x < \frac{1}{2} \\ -1 & \text{如果 } \frac{1}{2} \leq x < 1 \\ 0 & \text{其他} \end{cases} \quad (2.3)$$

形象化地, Haar 小波函数的图形表示如图 2-1 所示:

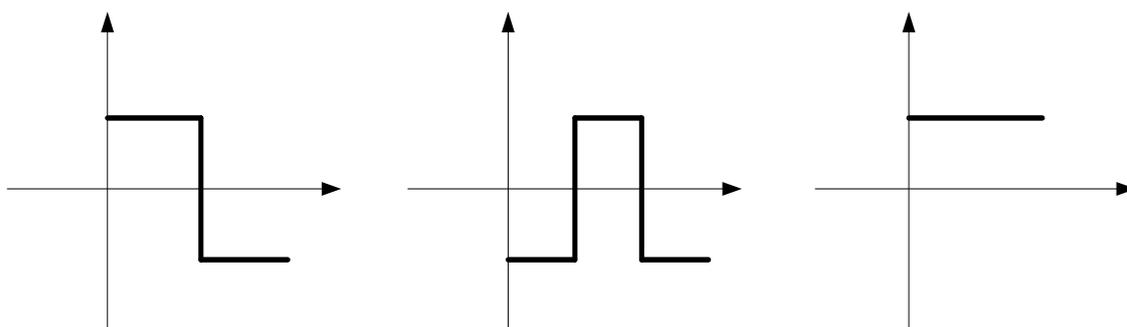


图 2-1. Haar 小波示意图

小波函数作用于图像之后, 得到局部区域的块状 Haar-Like 特征。在灰度图像上, 块状特征是同一子区域中所有像素灰度和, 减去其他子区域的灰度和, 如图 2-2 所示。对于 2 个矩形所组成的特征块 (见图 2-2A,B), 分别对两个矩形区域的所有像素灰度值求和, 然后再相减来提取块状特征的特征值。这些相邻的矩形区域都有相同的大小、形状、以及相同的排列方式。对于由 3 个矩形组成的特征块 (见图 2-2C), 其特征值是中间灰色矩形灰度值之和的 2 倍减去两个白色矩形的灰度值之和。图 2-2D 是由 4 个矩形组成的区域, 按照某一对角方向的两个矩形灰度值相加, 再减去另一对角方向的两个矩形灰度值之和, 可以得到该区域的特征值。

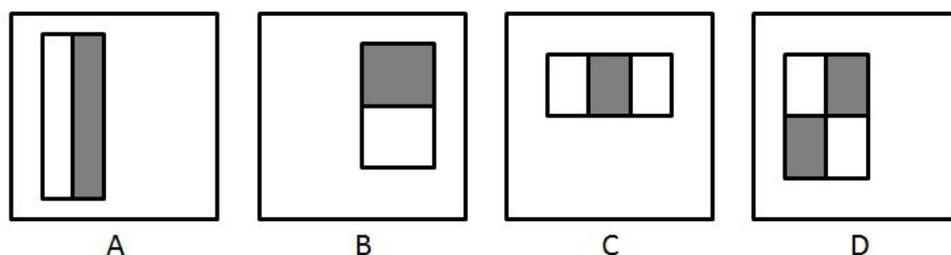


图 2-2. Haar-Like 特征, (A) 和 (B) 为两个矩形的特征, (C) 为三个矩形的特征, (D) 为四个矩形的特征。

后来, Viola 等人又对上面基本的 Haar-Like 特征进行了拓展, 便有了现在更丰富、完备的 Haar-Like 特征, 见图 2-3。

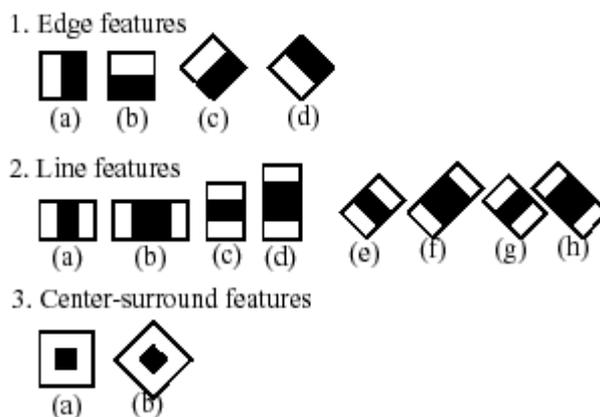


图 2-3. 拓展的 Haar-Like 特征

即使在一个 24×24 大小的图片中，这些完备的矩形特征集合中元素的个数非常多，超过了 180 000 个。对于这么多的特征块，如果相对独立的计算每一个矩形区域的特征值，那么计算量也是十分巨大的。为了解决这个问题，Viola 等人[7]提出了“积分图像”的办法计算特征块，这一方法节省了特征的计算时间，提高了检测速度。

2.1.2. HOG特征

HOG 特征是梯度方向直方图 (Histograms of Oriented Gradients) 的简写，在 2005 年由 Dalal[8]等人提出，他们借鉴了 SIFT[55]算法的思想，采用梯度方向直方图这一统计信息去表示人体目标。HOG 与 SIFT 的区别在于：HOG 对图像局部块的大小、单元的大小以及梯度方向空间的划分等细节与 SIFT 算法不同。另外，HOG 算法采用 4 个单元组成一个块，每个块采用滑动步长重叠的方式统计直方图，最后串联形成所有的特征向量。

这些简单的改进使得 HOG 特征取得了非常好的效果：在 MIT [30]人体测试库上表现近乎完美。后来 Dalal 等人收集了一些更具有挑战性的人体测试数据——INRIA 人体测试集。目前，INRIA 人体测试集是人体检测算法中常用的、比较有挑战性的数据集之一。HOG 特征不仅可以表示人体这一目标，还可以表示车辆、摩托车等物体。即使在今天，HOG 特征的优越性能仍受到很多研究者的青睐，而且很多研究者在此基础上提出改进版的 HOG 特征算法。

对于 64×128 大小的训练样本，HOG 特征的提取过程如下：一张训练图像划分成很多个单元 (cell)，每个单元是 8×8 像素大小的区域，相邻的 4 个单元 (田字形结构) 构成一个块 (block)。每两个块之间的间隔步长是一个单元的长度，采用滑动的方式形成重叠的块状描述子，因此 64×128 大小的训练样本有共有 $8 \times 16 = 128$ 个 cell， $7 \times 15 = 105$ 个 block。将人体样本区域划分好之后，Dalal 等人首先计算每个 cell 中像素点的梯度方向，并将其向 9 个方向进行直方图投影。每个直方图是 20 度的间隔，9 个方向形成对 0-180

度半圆形空间的一个划分，至于属于 180-360 方向的角度，可以采用取绝对值的转换方法，将其变换到 0-180 度空间中。这样，每个 cell 形成的特征向量是 9 维，依次连接每个 block 中的 4 个 cell 的特征，每个 block 便是一个 36 维的向量。对于一个训练样本中所有的块的特征向量依次串联起来，便形成了一个 $36 \times 105 = 3780$ 维的特征向量。

HOG 特征使用局部区域的梯度方向直方图这一统计信息，可以很好地表征局部区域内目标的边缘或梯度结构，体现了区域的局部轮廓，进而可以表征目标的形状。由于是在局部区域进行统计，所以 HOG 特征对辐射差异和配准误差有较强的鲁棒性。图 2-4 形象化地显示了 HOG 特征的形成方式。

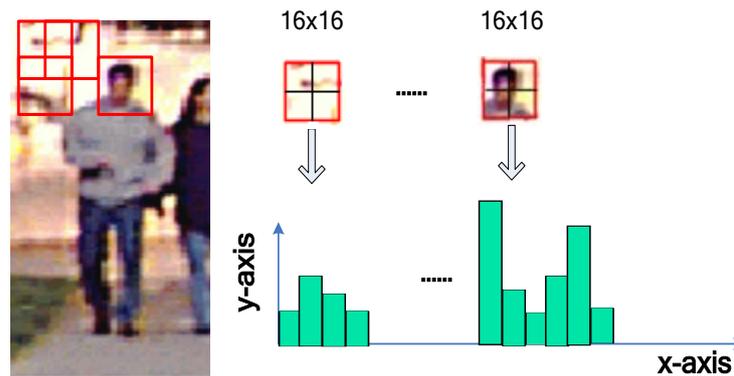


图 2-4. HOG 特征

2. 1. 3. v-HOG特征

v-HOG (variable-size HOG) 特征[9]是基于 HOG 特征演变出来的一种变尺度块状特征。HOG 特征是固定位置、固定尺度块上的梯度方向直方图特征，v-HOG 特征是由变尺度、非固定位置块上的梯度方向直方图组成的。在文献[9]中，Zhu 等人根据特征选择与分类器的结果，挑选出哪些 v-HOG 块可以用来表示人体目标。v-HOG 块的长宽比一般为 1:1、1:2、2:1，块的大小从 12x12 变化到 64x128 的像素区域，所有块仍然像 HOG 特征那样是由 2x2 个 cell 组成，每个 cell 里面统计的是一个 9 维的梯度方向直方图，每个块中 cell 的大小随着块的大小而改变，但是无论 cell 大小如何变化，仍然是往 9 个方向进行投影，因此每个 cell 得到 9 维梯度直方图的向量，相应地每个 v-HOG 特征块的特征向量是 $4 \times 9 = 36$ 维。图 2-5 显示了 v-HOG 特征的提取过程。

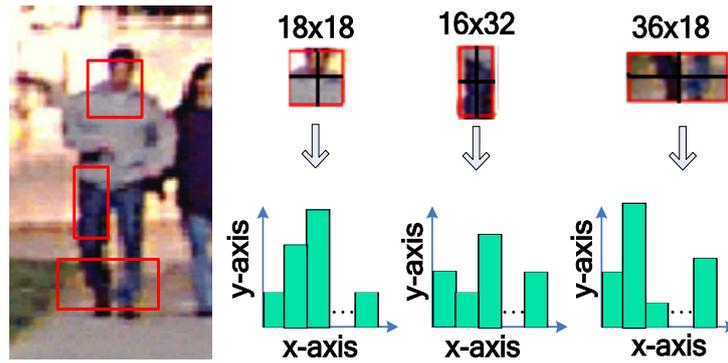


图 2-5. v-HOG 特征

2.1.4. MSO特征

MSO (Multi-scale orientation) [14]特征是受到 Haar-like 特征与 HOG 特征的启发，融合了两种特征的优点，能够更好地捕捉人体模式的变化和提高了计算效率。MSO 特征分为单方向、多方向两种形式。单方向 MSO 特征由尺度不同、位置各异的块组成，它以图像的颜色特征为基础，通过对特征块像素的颜色值进行累加计算，得到块的方向特征用以描述块的整体方向。多方向 MSO 特征结合了 HOG 的思想，对块内所有像素的梯度方向进行统计得到一个 8 维特征向量。多方向 MSO 特征可以很好地把握图像的整体信息，而且由于其块的维数小于 HOG 特征块的维数，所以具有很快的运算速度。

MSO 特征使用 8 个尺度扫描训练图像，分别得到在 8 个尺度下的正方形特征块，如图 2-6(a)所示。在某个尺度上单方向 MSO 特征的特征值计算，可分为两步：首先，计算正方形区域的梯度方向；其次，将梯度方向映射成固定的特征值编码。将正方形区域的划分成水平和竖直两个方向的四个子区域，分别计算两种划分下的水平灰度差 dx 和竖直灰度差 dy 。水平灰度差 dx 的计算，首先需要将正方形分为左右两个等份，分别计算两个小区域的所有像素的灰度值之和，得出水平灰度差；再将正方形分为上下两等份，用同样的方法计算出竖直灰度差，如图 2-6(b)所示。

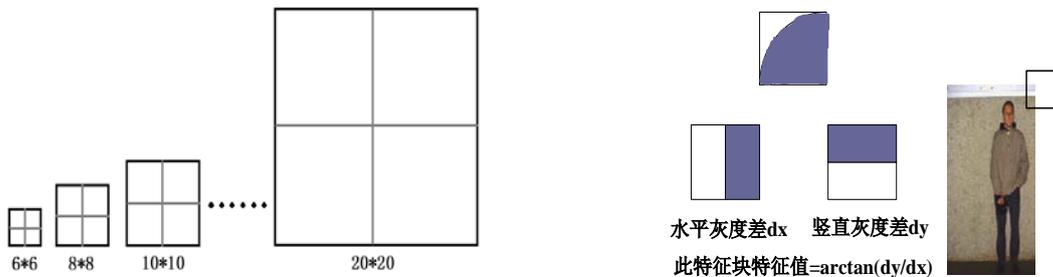


图 2-6. MSO 特征 (a)多尺度的特征窗口及其划分； (b)单方向 MSO 特征的提取

然后，根据所求出的 dx 和 dy 计算出这个特征区域内的梯度方向 Ori_Rect ：

$$Ori_Rect = F(\arctan(dy/dx)) \quad (2.4)$$

其中, $F(\cdot)$ 为一个离散的映射, 取值在非负整数集合 $\{0, 1, \dots, 8\}$ 中。它等价于将范围在 $0^\circ \sim 360^\circ$ 度的每个梯度方向 $\arctan(dy/dx)$ 分别向 9 个方向进行投影。当 $\arctan(dy/dx)$ 的值在 $0^\circ \sim 20^\circ$ 或 $180^\circ \sim 200^\circ$ 之间时, $F(\arctan(dy/dx))$ 的值为 1; 当 $\arctan(dy/dx)$ 的值在 $21^\circ \sim 40^\circ$ 或 $201^\circ \sim 220^\circ$ 之间时, $F(\arctan(dy/dx))$ 的值为 2, 依此类推, 可以得到所有块的梯度方向。

相对于单方向 MSO 特征, 多方向 MSO 特征主要是对所选的特征区域, 使用一个 8 维的直方图作为此区域的特征值。首先利用水平和垂直方向的 Prewitt 算子 $[-1, 0, +1]$ 分别计算特征区域内像素点 (x, y) 的水平和竖直灰度差 dx 和 dy , 然后计算区域中每个像素位置的梯度方向, 最后将所有像素点的梯度方向投影到 8 个方向上形成直方图向量。

2.2. 特征选择

随着文本分类、组合化学和生物信息学等领域数据大规模的增长, 数据中包含成百上千维的特征, 大部分数据间存在冗余性或者不相关性。由于数据的维度过高, 导致所需要的训练样本数目呈指数型增长, 即产生模式识别领域中谈及的“维数灾难”问题。因此对数据进行降维是一个非常迫切和具有挑战性的研究课题。

数据(简称为特征)降维常用的两类方法是特征变换(特征提取)和特征选择。特征变换是指将原有特征通过某种线性变换(主成分分析, PCA) [60] 或者非线性的变换(低秩矩阵) [61], 得到一种低维的新特征。特征的可理解性比较重要, 但特征变换使得特征的理解性变差, 因为即使很简单的线性组合也能构造出新特征。相反地, 特征选择是从原始特征集中, 按着某种标准选择出特征子集, 并未改变原始特征的信息, 增加了特征的可理解性。

2.2.1. 特征选择的定义

特征选择最早的研究始于上个世纪 60 年代, 当时的范围主要涉及统计学和信号处理等领域, 所面临的特征维数并不高, 维数灾难现象不是很明显。随着 90 年代大规模数据的出现和机器学习领域的发展, 特征选择问题受到学者们的广泛关注。

目前, 特征选择还没有比较统一的数学定义, 因为特征选择需要和具体的应用背景和研究问题相结合。针对不同的问题如分类[82, 84-85, 86-87]、回归[70-71, 92]等, 特征选择可笼统地定义为: 在满足一定的准则下, 寻找原始特征集中“最好”的某一特征或者某一特征子集。从其定义上, 可以看出特征选择的本质是对原始特征进行排序或者加权,

最重要的特征排在前几位（权重很大）。特征选择的目的是主要有三个方面：一是改进分类器的预测性能；二是对特征的生成过程有一个更好的认识；三是减少存储空间的使用和计算费用，提高预测速度。

特征选择算法的实现过程一般需要四个要素[101-103]，它们分别是：原始样本集、搜索策略、评价函数和终止条件。特征选择算法首先需要输入原始样本，特征维数等信息，然后根据某种搜索策略，寻找候选特征子集，并判断该特征子集是否能使评价函数的性能有所改进，最后与终止条件进行比较，判断是否获得最优解，从而执行下一步的方案。特征选择算法的流程图如图 2-7 所示。

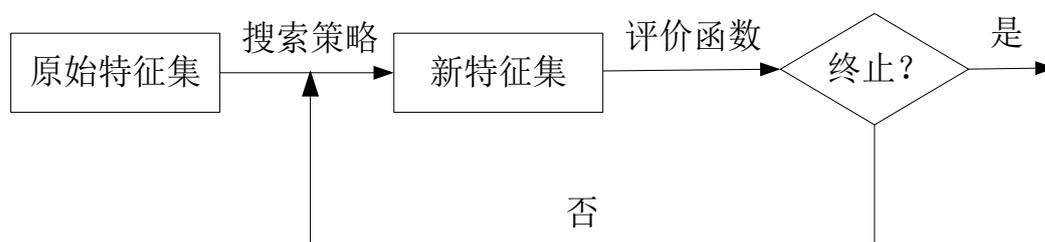


图 2-7.特征选择流程图

下面重点介绍下特征选择算法的搜索策略和评价函数这两个基本要素。

一、搜索策略

根据不同的搜索方向，可以得到不同的搜索策略。搜索方向主要有[64]：前向搜索、后向搜索、双向搜索和随机搜索等。前向搜索是指在当前特征子集的基础上，增加一个或多个新的特征；后向搜索是指删除当前特征集中的一个或多个特征；双向搜索是指先删除一些特征，再增加若干个新的特征；随机搜索是指依据概率分布，随机的选择一个或者多个特征。根据上述的几种搜索方向，特征选择的搜索策略大致分为以下三大类[62-65]：

- 全局最优搜索策略：它是指从全局的角度，选择最优的特征子集。典型的代表方法有枚举法和分支界定法，其中枚举法是从当前候选子集出发，采用逐一举例法访问搜索空间中的每个状态，以找到最优子集。理论上，给定一个原始特征集和评价函数，枚举法可以找到最优的特征子集，然而这种方式计算费用比较高。另一种典型的全局搜索法是分支界定法[66]，它利用评价函数的单调性，以树状结构的形式对搜索空间进行划分。树中的每个节点对应一个候选子集，若某节点的评价函数值小于当前候选集的值，则将该节点及其子节点集删除，再考虑其他节点，这样迭代的求解直至获得最优解为止。这种方法比枚举方法效率要高一些。

- 序列搜索策略：顺序地向当前特征子集中添加或者删除某些特征，这样迭代地搜索

满足终止条件的特征子集。该策略的典型代表方法有：顺序前向搜索，顺序后向搜索等。某些特征一旦被添加到当前特征子集，或者从子集中被删除后，不再考虑对该特征进行任何操作，因此序列搜索策略容易陷入局部极值。为解决此问题，提出了双向的增 m 减 n 搜索法，即搜索方向不再是单向的，可以根据评价函数灵活地增减，然而增 m 减 n 搜索法的缺点在于 m 和 n 的大小难以确定。为克服此缺点，Pudil 等提出了顺序浮动搜索算法[93]。该算法可以浮动地改变 m 和 n 的值，减少了不必要的回溯并在需要时增加回溯的深度。后来，Somol 等人提出了自适应浮动搜索算法[94]，根据当前特征子集大小来控制搜索空间的大小，自适应的浮动搜索减小了算法陷入局部最优的可能性。

- 启发式搜索策略：指在搜索特征子集时，使用启发式算法选择特征，并以一定的概率选择特征进入候选特征子集。典型代表算法有遗传搜索[79]、模拟退火搜索等。遗传搜索在特征选择算法中应用很广，因为它速度比较快，不易陷入局部最优解等 [77]，但是理论上，它不能保证找到最优解。模拟退火搜索算法因为它的计算量较大，初始温度以及迭代次数等参数的值难以确定等因素，使得它的应用不是很多。从文献[80-81]的实验结果可见，在上述所提搜索算法中，自适应浮动搜索和遗传搜索是性能较好的两种搜索算法。

二、评价函数

评价函数，是指能够区分特征“好”、“坏”的度量函数，依照评价函数的某种优化准则在原始的特征集中挑选特征。因此，评价函数决定保留哪些特征，抛弃哪些特征。常见的优化准则如下：

- ◆ 从原始特征集中找到某个特征子集，使得评价函数的值达到最优；
- ◆ 在评价函数值大于某一个给定的阈值条件下，寻找一个最小特征子集；
- ◆ 从原始特征集中寻找一个最小特征子集，并且同时使得评价函数的值尽可能最优。

上述三种准则体现了不同的优化目的，总体说来评价函数的定义是特征选择的重要因素。下面简单介绍下几种常见的评价函数。

● 距离度量函数

距离度量函数利用统计模式识别中一些常见的距离等度量方式作为评判标准。从是否引入随机因素的角度，距离度量函数大体可分为确定距离和概率距离。确定距离通常是指欧式空间中的距离，如欧氏距离[82]和马氏距离等。概率距离是指引入概率形式来衡量距离等度量方式。

表 2-1. 距离度量函数

距离类别	距离名称	公式
确定距离	欧氏距离	$d(x_i, x_j) = \sqrt{\sum_k (x_{ik} - x_{jk})^2}$
	马氏距离	$d(x_i, x_j) = (x_i - x_j)^T C^{-1} (x_i - x_j)$
	Chebychev距离	$d(x_i, x_j) = \max_k x_{ik} - x_{jk} $
概率距离	Bhattacharyya距离	$d = -\ln \int_x p(x Y_i)^{1-s} p(x Y_j)^s dx$
	Kolmogorov距离	$d = \int_x (p(x Y_i) - p(x Y_j)) dx$

- 一致性度量函数

一致性是指：给定两个样本，若它们的特征值均相同，且其类别标号相同，则称它们是一致的，否则是不一致的[99]。样本集合的不一致性是指该数据集中不一致的样本数与样本总数之间的比例。一致性度量利用样本集合的不一致性反映特征的重要性程度[100]。如果去除某一特征后，数据集的不一致性明显增大，那么就认为该特征很重要；否则认为它不重要。一致性度量函数的优点是能获得一个较小的特征子集，但它对噪声数据敏感，且只适合离散特征。

- 相关性度量函数

相关性度量[83, 85]利用类别与特征之间的统计相关性质度量特征的重要程度，即如果已知两个变量是统计相关的，那么就可以利用其中一个变量的值来估计预测另一个变量的值。目前，常用的统计相关系数，如 t 检验、Pearson 相关系数、Fisher 分数[83]等用来表达特征相对于类别可分离性的重要性程度。

表 2-2. 相关性度量函数

相关度量函数	公式
Pearson相关系数	$r = \text{cov}(X_i, Y) / \sqrt{\text{var}(X_i) \text{var}(Y)}$
Fisher分数	$\text{Fisher}(x_k) = \sum_i n_i (m_k^i - m_k)^2 / \sum_i n_i (\sigma_k^i)^2$

● 信息度量函数

信息度量[68,88-90]主要利用信息熵、信息增益、互信息等手段量化特征的不确定性程度，以判定其包含的信息含量。信息度量的优点在于它是一种无参数的、非线性度量，且它不需要预先知道样本的分布。由于上述优点，信息度量函数在特征选择算法中得到广泛关注。表 2-3 给出一些常见的信息度量函数的表示形式。

表 2-3. 信息度量函数

信息度量名称	公式
信息熵	$H(x_i) = \sum_k p(x_{ik}) \log_2 p(x_{ik})$
互信息	$MI(x_i, y) = \sum_{x_i} \sum_y p(x_i, y) \log_2 [p(x_i, y) / p(x_i)p(y)]$
信息增益	$IG(X) = -\sum_y p(y) \log_2 p(y) - \sum_x [p(x) \sum_y p(y x) \log_2 p(y x)]$

● 误差度量函数

特征选择的目的之一是期望选择后的特征能够提高学习算法的性能，学习算法的性能一般由误差函数来体现。针对分类问题而言，分类误差是衡量特征选择算法优劣的标准之一。目前，有许多特征选择算法直接采用分类误差衡量特征的重要性，如 Stoppiglia [91] 等人提出使用均方误差准则作为分类误差函数，通过前向搜索逐步添加能够使误差函数改变最多的特征。Huang 等[67]使用遗传搜索与互信息熵结合的封装式算法获取特征子集，并能明显提高分类模型的性能。Neumann 等[69]提出新颖的特征选择方法，其融合核函数约束与 l_p 范数正则项构建分类目标函数，取得较好的分类性能。Schapire 等人[75]提出 Boosting 算法进行特征选择方法，其采用重采样前向搜索策略和误差最小为评价函数。

2.2.2. 特征选择的分类

根据特征选择与学习算法的结合方式，特征选择算法大体上可以分为三类：过滤式 (Filter)、封装式 (Wrapper)、嵌入式 (Embedded) 算法。

● 过滤式特征选择

过滤式特征选择[68,76,78]的评估标准独立于学习算法，直接对数据集本身进行处理，未考虑和学习算法有关的信息。通常是选择与评价函数相关度大的特征或者特征子集，一般认为相关度较大的特征或者特征子集会以高概率提高学习算法的精度。过滤式特征

选择的评估方法很多，如距离度量、信息增益、相关性以及不一致性等等（这即是前面评价函数中所介绍的度量准则）。过滤式特征选择因为运行效率较高而适用于大规模数据集，但 Kohavi[93]等人指出因为其脱离了学习算法的任何信息，因此找到的特征不一定能改进学习算法的性能。

● 封装式特征选择

封装式特征选择由 John 等人提出[94]。该算法的核心思想是：如果特征选择不与学习算法相结合（如 Filter 特征选择），那么在特征选择之后，仍然会产生和学习算法无关的冗余信息。最重要的是，不同学习算法偏好不同的特征子集，因此他们认为特征选择过程不能脱离学习算法，应该以学习算法性能作为特征评价的标准。他们把学习算法看成是一个黑盒子，特征选择的标准是算法的分类性能（针对分类问题而言），一般通过交叉验证的方法看分类器的性能。然而，封装式特征算法并未涉及到学习算法的具体细节，比如说目标函数的形式等，它的处理方式只看当前候选特征子集对算法性能的改进，没有对学习算法有任何限制。因此，封装式特征方法比较简单、通用。

● 嵌入式特征选择

嵌入式特征选择是指：特征选择过程嵌入到学习算法里，作为学习算法的某个组成部分。嵌入式特征选择算法很早就有，比如在决策树的训练中，如 Breiman 的 CART 算法[95]和 Quinlan 提出的决策树与 bagging 和 boosting 相结合的算法[96]等。这些算法在每一个节点选择分类能力最强的特征，然后基于选中的特征进行子空间分割，继续此过程，直到满足终止条件，可见决策树生成的过程也就是特征选择的过程。后来，一些嵌入式特征选择方法通过估计目标函数值的改变，指导变量在特征空间的移动。Rakotomamonjy [97]等人对线性 SVM 的目标函数，采用有限差分方式观察目标函数值的改变，从而指导向量的特征选择。Guyon[98]等人提出使用保持对偶变量值不变的情况下，估计目标函数有限差分的变化，实现对核函数 SVM 的特征进行选择。

2.3. 分类器

这一部分主要介绍两种非常经典的分类器：SVM 分类器与 Adaboost 分类器。

2.3.1. SVM分类器

Vapnik 提出的支持向量机（Support Vector Machine, SVM）是一种基于最大边界原则的学习方法[20,50,52,53]，后来 Vapnik[54]证明最大边界原则的目的是最小化结构风险（测试误差的上界）。目前 SVM 已广泛被用于分类、识别等模式识别领域中。最大边界

原则的核心思想是找到最优线性超平面,该超平面要尽量正确地将两类样本分隔开,并且使得两类样本集中最近样本的间隔最大,如图 2-8 所示。

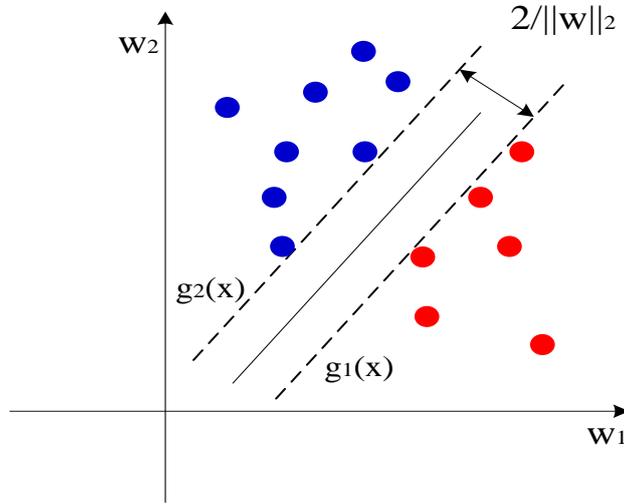


图 2-8. 最大间隔原理

简单起见,先考虑线性可分的两类样本 $\{x_i, y_i\}, i=1, \dots, N$, 其中 x_i 为第 i 个样本的特征向量, y_i 为样本 x_i 的类别标号, 设线性超平面的方程形式为: $g(x) = w^T \cdot x_i + b$ 。 w 是线性超平面的法向量, b 是超平面的阈值。对线性可分的正反例样本, 拟构造两个平行的线性超平面, 通过调整线性超平面方程的阈值, 分别要求对正例样本有 $g_1(x) = w^T \cdot x_i + b \geq 1$, 对反例样本有 $g_2(x) = w^T \cdot x_i + b \leq -1$ 。这样两条平行超平面之间的“间隔”为 $2/\|w\|_2$ 。要使分类间隔最大, 即要使 $\|w\|_2$ 最小, 并要求样本满足一定的约束条件:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y_i[(w^T \cdot x_i) + b] - 1 \geq 0 \quad i=1, \dots, N \end{aligned} \tag{2.5}$$

上述优化模型是一个凸二次规划, 是连续可微的, 可以通过求解其对偶规划来得到法向量 w 和阈值 b 的解析式。在实际中, 样本的分布比较复杂, 若两类样本不是线性可分的, 或者由于噪声的影响使得样本分布不是线性的, 那么上述优化模型是无解的, 即不存在一条线性超平面将正反例样本全部分开, 因此需要对每个样本加入误差扰动项, 相应的优化模型如式 (2.6) 所示。

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i[(w^T \cdot x_i) + b] \geq 1 - \xi_i, \\ \xi_i \geq 0, \end{cases} \quad i=1, \dots, N \end{aligned} \tag{2.6}$$

在式(2.6)中： ξ_i 表示第*i*个样本被误分的程度。 C 是平衡误分程度与边界之间的惩罚因子。为方便求解，将上述凸规划模型转换为其对偶规划。可以通过引入Lagrange函数对原变量求偏导的方式，将原模型转换成对偶规划。

$$L(w, b, \xi, a_i, t_i) = \frac{1}{2} \|w\|^2 + \sum_{i=1}^N \xi_i (C - a_i - t_i) - \sum_{i=1}^N a_i y_i (w^T \cdot x_i + b) + \sum_{i=1}^N a_i \quad (2.7)$$

其中 α_i, t_i 为Lagrange系数。根据Wolfe对偶定理，式(2.7)分别对 w, α_i, b 求偏微分之后，令其等于0。然后将原变量用对偶变量进行表示，回代到优化模型中，就可以把上述问题转换为一个较简单的对偶问题。可见对偶规划中只有*N*个对偶变量 α_i ，这个规划比原规划要更方便求解，根据对偶变量 α_i 与原变量 w, b 的关系，可以求解到法向量 w 和阈值 b 。

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j x_i x_j - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^N \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq C, \quad i=1, \dots, N. \end{cases} \end{aligned} \quad (2.8)$$

不是所有的正反例样本都是线性可分的，若样本分布是非线性的，那么虽然上述的模型可以训练求解得到一个线性超平面，但是非线性的样本强加使用线性分类器，得到的识别率和分类率会大大降低。SVM理论认为：对于低维空间中不是线性可分的样本，可以将样本映射到一个高维空间中去，在高维空间中样本可以被认为线性可分的。但是一般样本的特征向量本身已经处于高维空间中，再对样本进行升维会极大增加运算费用。SVM理论巧妙的采用了“核函数”(Kernel function)来解决这类问题。核函数是样本间的某种内积形式，反映的是在高维空间中，对样本的相似性的一种度量。结合SVM的线性优化模型(2.8)式中样本的内积 $x_i \cdot x_j$ ，将其替换为核函数的形式，得到SVM关于高维空间中样本的对偶规划，这个对偶规划和原来的样本单独升维到高维空间，再使用Lagrange函数求解出来的对偶规划一致。同理，通过求解对偶规划，便得到SVM的对偶变量，从而也即得到在高维空间中的法向量和阈值。因此，核函数的引入，非常巧妙的将高维计算变成了样本内积的计算，这样极大地减少了计算费用，而且实验结果验证一般的核函数SVM要比线性SVM结果要好。

2.3.2. Adaboost分类器

Adaboost[7, 35, 21, 51]算法是Boosting系列算法的一种，意在将弱学习算法提升为强学习算法。Adaboost算法中的弱学习算法，也称为弱分类器，强学习算法称为强分类器。

Adaboost 采用贪婪的迭代方式，每次迭代选择一个最好的弱分类器，最后将所有迭代中所选择出的弱分类器进行线性加权组合，形成最后的强分类器。

在每次迭代中，Adaboost 对每个训练样本赋予一个权重，这样在每次迭代中所有的样本权重形成一套概率分布。分类误差是权重的组合，每次迭代中选择分类误差最小的弱分类器，并调整每个训练样本的权重。权重调整的原则是更加重视被误分的样本，因此被误分的样本权重较大，被正确分类的样本权重减少。这样每次迭代随着权重的增加，算法训练会集中到更难训练的样本上。最后，每次迭代选择出的弱分类器的加权投票形成强分类器，并且每个弱分类器按其在训练集上的权重作为强分类器中的权重。

每一个弱分类器对应着某一个特征，在选择哪些弱分类器形成强分类器的同时，也即完成了特征选择的功能。对于每一个特征，相应的弱分类器学习得到一个最佳的分类函数，使得训练样本的错误分类数达到最小。对于一个弱分类器 $h_j(x)$ 有一个特征 f_j ，一个阈值 θ_j ，还有一个用来指示不等式符号方向的函数 p_j ：

$$h_j(x) = \begin{cases} 1 & \text{if } (p_j f_j(x) < p_j \theta_j) \\ 0 & \text{otherwise} \end{cases} \quad (2.9)$$

Adaboost 算法中关于每个弱分类器，需要学习一个最佳的阈值 θ_j 。一个弱分类器的训练（特征 f_j ）是在当前权重分布的情况下，确定其最优阈值，使得这个弱分类器对所有训练样本的分类误差最低。阈值的确定方式如下：

对于每个特征 f_j ，计算所有训练样本的特征值，并将其排序，通过观察特征值的分布，可以为这个分类器（特征）确定一个最优的阈值。具体来说，对排好序的表中每个特征值，分别计算下面四个值：

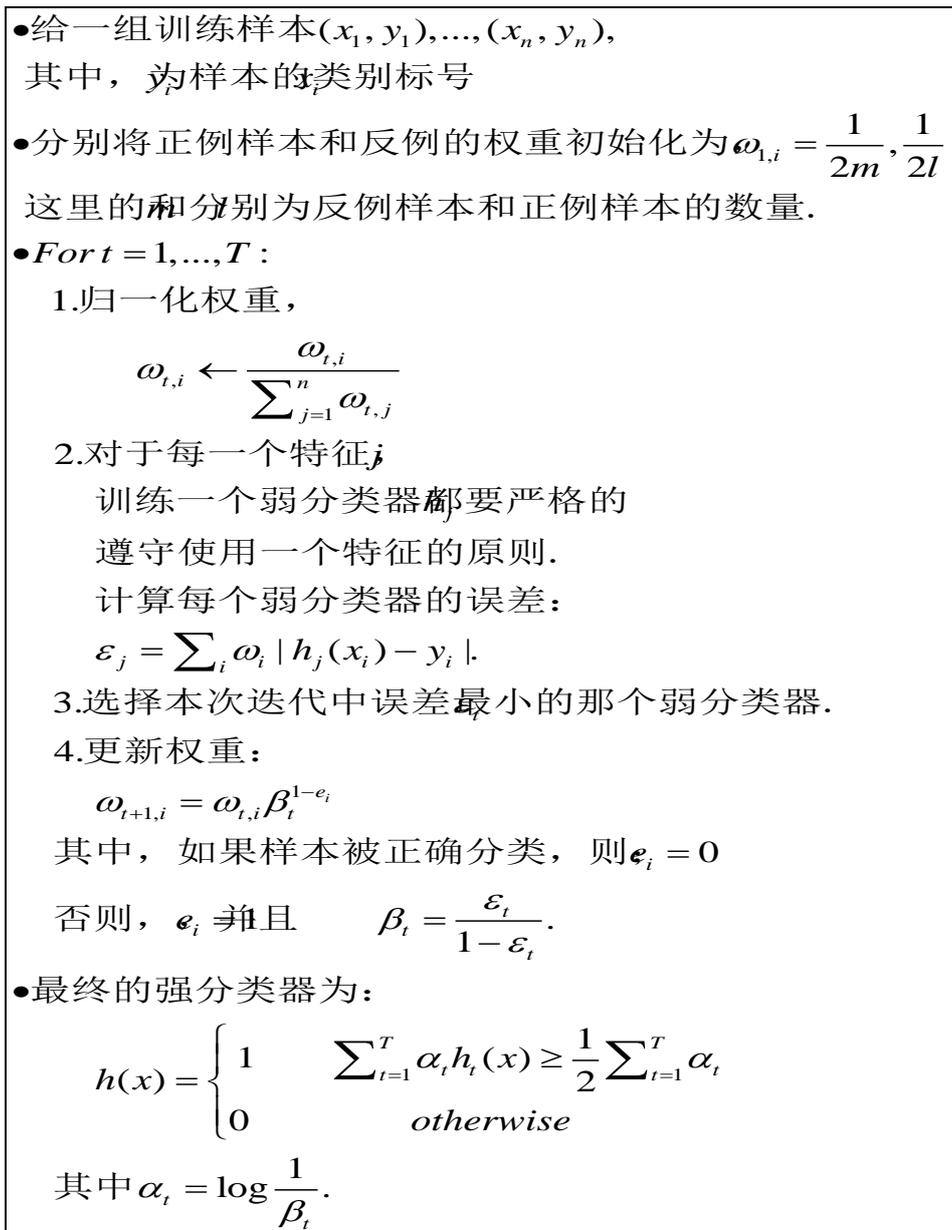
- 1) 全部正例样本的权重之和 W^+ ；
- 2) 全部反例样本的权重之和 W^- ；
- 3) 在此特征值之前的目标样本的权重之和 S^+ ；
- 4) 在此特征值之前的非目标样本的权重之和 S^- ；

这样，当选取当前特征值和它前面的一个特征值之间的实数作为阈值时，所得到的弱分类器就在当前元素处把样本分开。那么这个阈值所带来的分类误差为：

$$e = \min (S^+ + (W^- - S^-), S^- + (W^+ - S^+))$$

于是，通过把对所有排好序的特征值扫描一遍，选择分类误差最小的那个阈值作为弱分类器的最佳阈值。

表 2-4. Adaboost 算法流程图:



2.4. 本章小结

目前, 目标的特征描述主要包括对颜色、轮廓、纹理等信息的提取。本章主要介绍了这几方面最具有代表性的特征描述子, 如 Haar-Like、HOG 以及 v-HOG 和 MSO, 它们均在人体检测方面取得了很好的效果。Haar-Like 特征描述子是基于图像灰度信息的一种特征描述。它统计了图像中不同尺度、位置上特征块的灰度差。该特征简单、可以把握目标的局部轮廓, 然而它对于颜色和光照变化等信息比较敏感。HOG 描述子是基于局部区域的梯度方向直方图, 它对被检测物体的形变信息有很强的容忍能力, 这一点可以很好应对不同物体的检测, 但是, 基于固定大小块的 HOG 描述子往往会忽略被检测物体

的整体信息，且运算速度略慢。 v -HOG 特征打破了 HOG 特征固定尺度、固定位置的提取方式，形成了一系列变尺度、不同位置的块状信息。MSO 特征结合了 Haar-Like 和 HOG 两种特征的优点，分别针对粗检测和精检测过程，形成单方向与多方向的 MSO 块状描述子。

特征选择是为了提高分类效率，减少运算费用的常用数据降维手段之一。本章详细地介绍了特征选择算法包含的四个关键要素，重点介绍了搜索策略和评价函数这两个要素。搜索策略和评价函数一般可以确定特征选择算法的形式与效果。除此之外，本章还介绍了三类典型的特征选择方法：过滤式特征选择、封装式特征选择和嵌入式特征选择。三类方法不是割裂独立的，可以结合具体的应用背景选择任意两种方式的组合。特征选择算法经过多年的研究和发展，针对小规模的数据集已取得较好的效果，但是针对大规模的数据集，效果还是有待改进。

目前比较受欢迎人体检测的分类方法中，典型的方法有 SVM 算法和 Adaboost 算法。上述两种方法都是基于统计学习理论，并且都分别证明所建立的优化模型或者学习方法在概率意义下近似正确。SVM 算法是通过求解凸二次规划确定一个最优线性超平面的法向量和阈值，巧妙地运用核函数的方法实现非线性的判别，但是未对特征进行选择。该方法简单实用，并且已有很多开源代码供学者使用。Adaboost 算法主要采用加权投票机制，由所选择的弱分类器的线性组合构成强分类器，意在通过对特征进行选择，同时实现对人体模式的判别。Adaboost 算法中弱分类器阈值的确定有些繁琐，并且通过贪婪的迭代方式选择出的特征还是存在一定的冗余信息。该方法一般使用的时候，需要结合级联机制，能够达到和 SVM 同样或者更好的效果，级联机制的 Adaboost 算法，速度要比 SVM 算法快很多。

第三章 基于线性 1 范数最小化学习的人体检测

本章主要是针对分类问题，设计一种基于 1 范数最小化的线性分类(L-LML)算法，并将该算法用在人体检测（二值分类）问题上。L-LML 算法意在利用 1 范数的稀疏性，对高维人体特征向量中的特征值进行选择，并构造线性超平面（分类器）实现对人体的快速检测。本章首先介绍了稀疏表示的发展历程，阐述了 1 范数已经成为稀疏表示的主要方式，提出了基于 1 范数最小化学习的优化模型(LML)，引入了经验风险、期望风险和 VC 维等概念，分析了 LML 优化模型建立的原则（根据 1 范数最小化与 VC 维之间的关系），最后设计了 L-LML 算法并展示了其在人体检测上的效果。图 3-1 显示了 L-LML 算法用于人体检测的流程图。

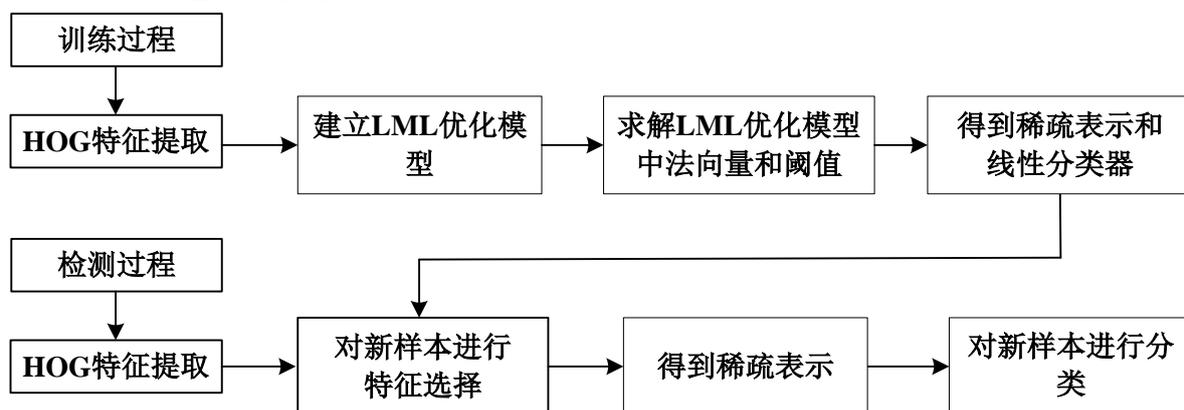


图 3-1.L-LML算法流程图

3.1. 稀疏表示综述

稀疏表示(Sparse Representation)是近几年来信号处理，机器学习和模式识别等领域中的一大热点研究问题。从近几年的模式识别与计算机视觉领域的各种会议期刊文献中，便可看出稀疏表示这一问题是属于非常前沿的研究问题，并且 ICCV'09 与 CVPR'10 中分别开设了一系列的专题报告。下面本节围绕稀疏表示的发展历史，详细介绍下为什么它这么吸引学者的眼球。

早在上个世纪 93-97 年[115-117]的时候，就有一些稀疏表示的研究成果发表。首先是 Chen 提出使用 1 范数最小化方法建立优化模型解决回归问题，并使用基追逐法去求解。后来非常著名的 Lasso[115]算法也发表在统计学期刊上。Lasso 算法也是针对回归问

题，建立最小二乘误差和 1 范数正则化的约束模型，并提出使用贪婪算法去求解。但是，在当时这个模型的真正意义并没有引起很多人的注意，大多数人也没有意识到 1 范数和稀疏性之间的联系。

随着信号处理领域压缩感知理论[22-23, 56, 110]的提出，稀疏表示越来越受到学者们的重视。下面内容介绍了压缩感知理论的诞生及其是如何进行稀疏表示的。

信号采样是从模拟信源获取数字信息的必经之路。过去，指导信号采样的理论基础一直是著名的 Nyquist 采样定理[110]。它指出，只有当采样速率达到信号带宽的两倍以上时，才能由采样信号精确重建原始信号。但是，随着信号的带宽变得越来越大，根据 Nyquist 采样定理，那么相应的信号采样速率必须越来越高。现在，针对信号传输速度的高要求和存储空间有限等问题，常见的解决方案是对信号压缩。以那么高的速率对信号采样，然后再对信号实施压缩变换，这样极大的浪费了资源，因为高速率采样后的数据在压缩过程中只保留很少的信息，抛弃了很多冗余信息。从这个角度考虑，研究者们思考是否存在另一种基于信息采样的理论框架，使得采样的同时能实现信号的压缩，并且只需远少于 Nyquist 采样定理所要求的采样数目就可精确或近似精确重建原始信号，于是压缩感知采样理论便在这种迫切的实际需求中诞生了。

压缩感知(compressed sensing)或压缩采样(compressive sampling)的某些抽象结论源于 Kashin 创立的范函分析和逼近论[112]，后来由 Candès、Tao、Romberg、Donoho 等人系统化地创立。该理论否认了 Nyquist 采样定理，也即信号采样速率不再取决于信号带宽，并且指出采样速率在很大程度上取决于两个要素，信号本身的稀疏性和非相干性。换句话说，该理论主要论述：当信号在某个变换域是稀疏的或可压缩的，可以利用与变换矩阵非相干的测量矩阵将变换系数线性投影为低维观测向量，同时这种投影保持了信号重建时所需的信息，通过进一步求解稀疏优化问题就能够从低维观测向量精确地或高概率精确地重建原始信号。

压缩感知理论主要涉及变换域中的过完备字典、满足非相干性的测量矩阵和快速鲁棒的信号重建三个问题。这三个问题的核心都是信号是否可以进行稀疏表示，如果能进行稀疏表示，应该使用何种模型去表示，包括模型的解法等一系列问题。压缩感知最初的稀疏性重构模型，使用的是 0 范数。某一个向量 x 的 0 范数 $\|x\|_0$ 指向量 x 中不为 0 的元素的个数。

$$\begin{aligned} \min_x \quad & \|x\|_0 \\ \text{s.t.} \quad & y = Ax \end{aligned} \tag{3.1}$$

上述的 0 范数优化模型中， A 为观测矩阵， x 为原始信号经过稀疏变换后的表示系

数， y 是接受端对信号进行重构之后的结果。关于模型 (3.1)，其解的存在以及是否唯一，这些都是值得考究的，所以在用这个模型做稀疏表示之前，没有什么理论保证。幸运的是，在 2003 年，Donoho & Elad [107] 等学者证明，如果矩阵 A 满足： $\sigma(A) \geq 2 \|x\|_0$ ，那么 0 范数最优化问题具有唯一解。 $\sigma(A)$ 指的是矩阵 A 中最小的线性相关的列向量集所含的向量个数。然而，虽然证明出模型的解是存在唯一的，但是上述优化模型的求解过程是 NP-hard 的。

随着对稀疏表示问题的深入研究，终于在 2006 年，Tao 和 Candes [113-114] 合作证明了在限制等距性质(Restricted Isometry Property, RIP)条件下 [112]，0 范数优化问题与模型 (3.2) 中 1 范数优化问题具有相同的解，其中 RIP 条件指稀疏信号在观测矩阵作用下必须保持几何性质相一致。

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (3.2)$$

在 RIP 性质提出之后，稀疏表示终于在理论上和实践中得到统一。后来很多研究学者也注意到稀疏表示的优点，竞相地将其用于模式识别和计算机视觉等领域中。目前稀疏表示的实现主要有三种方式：直接优化法，迭代收缩方法和 l_p 范数正则化方法 [109]。

直接优化法：这种方法是指直接使用 0 范数最小化作为稀疏性追求的目标函数，优化模型与 (3.1) 优化模型类似。这类优化问题的求解一般都使用贪婪算法去求解，典型的代表算法有匹配追逐法 [117] 和正交匹配追逐法 [119] 等。

迭代收缩优化方法：这种方法是前面所提到的著名 LASSO 方法，该方法可以通过不断松弛模型 (3.3) 中的 ε 因子，来实现迭代收缩求解。

$$\begin{aligned} \min_x \quad & \|Ax - y\|_2 \\ \text{s.t.} \quad & \|x\|_1 \leq \varepsilon \end{aligned} \quad (3.3)$$

l_p 范数正则化优化方法：这个方法目前研究比较热门， l_p 范数，当 $p=1$ 时，即为本文所谈到的 1 范数。因为 Candes 等人理论上证明 1 范数的稀疏性，所以现在关于 1 范数的研究比较多。当 $0 < p < 1$ 时，有些研究者表明此时的 l_p 范数比 1 范数更加有利于实现稀疏性 [118]。目前 l_p 范数正则化的常见求解方法是使用基追逐 [116] 方法，软阈值方法 (常见的有坐标梯度法和 Bregman 迭代法) 等。

3.2. 1 范数研究综述

范数是某种距离度量。1 范数是 l_p 范数中的一种，常见的范数还有 2 范数，0 范数等。

某一向量的 1 范数是指向量中每一分量的绝对值求和。对于 1 范数的研究早在范数理论创建的时候，许多数学爱好者就对其性质进行了分析[120]，但是其一直未受到重视。在实际应用中 2 范数一直都受到学者的青睐，原因在于 1 范数的连续但非光滑性质使得其求解算法比较复杂。直到上节所提到的稀疏表示问题的产生以及压缩感知理论在信号处理领域中的出现，1 范数终于因其优良的性质而得到研究者的青睐。压缩感知理论中最早考虑使用 0 范数去重构信号，但由于 0 范数的求解是 NP-hard 的问题，所以该理论使用 1 范数来代替 0 范数，而且理论上已经证明 1 范数可以作为 0 范数的一个近似估计。

此外，在模式识别领域，很多学者应用 1 范数的稀疏特性去判别分类。这里面比较典型的应用有图像重建、人脸识别、图像对齐等。早期的关于 1 范数在模式识别领域应用的文献，要数 John Wright 和 Yi Ma[24] 等人提出的 SRC (Sparse Representation Classification) 算法。该算法使用 1 范数最小化的稀疏表示识别人脸。该问题是一个多类回归问题，约束方程假设一张测试人脸样本 y 可以由训练样本集 A 进行线性的稀疏表示，目标函数是表示系数 x 的 1 范数最小化。SRC 算法使用 1 范数的目的是希望这种线性表示尽可能的稀疏，最后根据重构误差最小来实现对人脸的识别。

$$\begin{aligned} \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (3.4)$$

假设有 k 类不同人的脸，同一个人的不同表情、不同光照下的图像构成一个集合 A_i ，所有人的脸构成训练样本集合 $A = [A_1, A_2 \cdots A_k] \in R^{mn}$ ，上述优化模型可等价写成：

$$\begin{aligned} \tilde{x} = \arg \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & y = Ax \end{aligned} \quad (3.5)$$

考虑到噪声与误差的影响，那么 (3.5) 式重写成：

$$\begin{aligned} \tilde{x} = \arg \min_x \quad & \|x\|_1 \\ \text{s.t.} \quad & \|Ax - y\|_2 \leq \varepsilon \end{aligned} \quad (3.6)$$

最后计算残差 $r_i(y) = \|y - A\delta_i(\tilde{x}_i)\|_2$ ，残差最小的标号则是测试样本 y 的类别。

$$\text{class}(y) = \arg \min_i r_i(y) \quad (3.7)$$

该算法最巧妙的地方就是通过稀疏表示，可以消除遮挡和噪声对人脸图像的影响，详细的关于算法的实验部分本节不再介绍，请参见[24]。

从压缩感知到人脸识别问题，这些问题的本质思想，都是通过稀疏表示来实现对物体的识别，属于回归问题。本文受到上述研究的启发，研究基于 1 范数最小化的分类问题，

本文研究的算法与 Yi Ma 等人提出的回归模型以及压缩感知中重构部分的想法不同，主要是借鉴了 SVM、Adaboost 和凹函数支持向量机[120]等分类算法，结合 1 范数的稀疏性设计人体检测模型。下面首先要介绍下所提的 1 范数最小化学习模型在理论上的依据，即讨论它和期望风险，VC 维等评测分类算法性能指标之间的关系。

3.3. 经验风险与期望风险

考虑一个分类问题，如何建立一个决策函数 f （这里称决策函数为“假设”），使得其能够更好地处理分类问题和具有更好的推广能力。下面本节引入衡量一个决策函数推广能力的标准。这个标准，学者们很自然地会联想和分类误差相结合。但是是否在训练样本集上的分类误差率越低越好呢？伴随着这些问题，下面需要引进几个概念，它们分别是损失函数、期望风险、经验风险、VC 维、学习算法在概率意义下近似正确（PAC）等。

3.3.1. 经验风险与期望风险的定义

设样本集 $X \subset R^n$ ， $Y = \{-1, 1\}$ ， (x, y) 是随机向量，其中 $x \in X, y \in Y$ ，记函数 $P(\bar{x}, \bar{y}) = P(x \leq \bar{x}, y \leq \bar{y})$ 为 $X \times Y$ 上的概率分布。损失函数是评价预测准确程度的一种度量，若根据某个假设的结果，定义损失函数：

定义 3.1 (损失函数): 记 $(x, y, f(x)) \in X \times Y \times Y$ 是一个三元组，这里 $f(x)$ 是一个预测值，若映射 $c: X \times Y \times Y \rightarrow [0, \infty)$ 对任意的 $x \in X, y \in Y$ ，都有 $c(x, y, y) = 0$ ，则称 c 是一个损失函数。比较常用的一种损失函数是 0-1 损失函数，其定义如下：

$$c(x, y, f(x)) = \begin{cases} 0 & \text{如果 } y = f(x) \\ 1 & \text{其他情况} \end{cases} \quad (3.8)$$

定义 3.2 (期望风险): 设 $P(x, y)$ 为 $X \times Y$ 上的概率分布， $c: X \times Y \times Y \rightarrow [0, \infty)$ 为给定的损失函数，那么假设 $f(x)$ 关于概率分布 $P(x, y)$ 的期望风险是指：

$$R[f] = E[c(x, y, f(x))] = \int_{X \times Y} c(x, y, f(x)) dP(x, y) \quad (3.9)$$

从上式可以看出期望风险其实是指假设 $f(x)$ 的“平均”损失程度。

定义 3.3 (学习算法在概率意义下近似正确，PAC 原则): 如果对任意给定实数 $\delta (0 < \delta < 1)$ 和 $\varepsilon (0 < \varepsilon < 1)$ ，对 $X \times Y$ 上的任意概率分布 $P(x, y)$ ，存在 l_0 ，使得对于长度大于等于 l_0 的按分布 $P(x, y)$ 选取的独立同分布的训练集 $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$ ， $l \geq l_0$ ，由学习算法得到的 $f(x)$ 满足 $P\{R[f] \leq \varepsilon\} \geq 1 - \delta$ 成立，则称算法在概率意义下近似正确。

因此可以看出，若使学习算法能够在概率意义下近似正确，那么则寻找一个决策函

数让期望风险比较小。如果已知损失函数和样本的概率密度分布 $P(x, y)$ ，当然可以通过最小化期望风险求得决策函数。但是不幸的是，现实应用中，样本的分布非常复杂，它们的概率分布是未知的，已知的只是训练集 T ，因此想通过最小化期望风险这个标准来构造决策函数是不可能的，所以必须通过别的手段去评价决策函数的好坏。接下来，本节引入经验风险的概念。

定义 3.4 (经验风险): 设任意给定训练集 $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$ ，并且给定损失函数 c ，所谓决策函数 $f(x)$ 的经验风险是指：

$$R_{emp}[f] = \frac{1}{l} \sum_{i=1}^l c(x_i, y_i, f(x_i)) \quad (3.10)$$

所以衡量一个学习算法好坏的直观想法是，对决策函数集进行一定范围的限制，然后在这个范围内寻求使得经验风险最小的假设 $f(x)$ 。那么是不是一味的追求经验风险最小，便会得到最优的 $f(x)$ 使得它对于样本的期望风险也最小呢？也即会使得假设函数 $f(x)$ 对于测试样本效果非常好呢？这个答案是否定的。因为期望风险不是单调的。下面来介绍下期望风险与经验风险之间的关系。

3.3.2. 经验风险与期望风险之间的关系

本节首先介绍 VC 维[50]的概念：VC 维是统计学习理论中比较中重要的一个概念，VC 维最初由 Vapnik 和 Chervonenkis 提出的，意在评判一种学习算法（决策函数）的表达能力。假设集 F 是指决策函数集，VC 维是建立在点集被 F 打散的基础上。下面引入一个点集被 F “打散”的概念。

定义 3.5 ($N(F, Z_m)$): 设 F 是一个假设集，即其定义域是 $X \subset \mathbb{R}^n$ ，取值为 -1 或 1 的若干函数组成的集合。令 $Z_m = \{x_1, \dots, x_m\}$ 是由 X 中 m 个点组成的集合。考虑当决策函数 f 取遍 F 中所有可能的假设，产生一个 m 维向量 $(f(x_1), \dots, f(x_m))$ 。记 $N(F, Z_m)$ 为上述 m 维向量中不同的向量的个数。

定义 3.6 (Z_m 被 F 打散): 设 F 是一个假设集，如果 $N(F, Z_m) = 2^m$ ，则称 Z_m 被 F 打散。

定义 3.7 (VC 维): 定义 F 的 VC 维为： $VCdim(F) = \max\{m : N(F, Z_m) = 2^m\}$ ，如果 $\{m : N(F, Z_m) = 2^m\}$ 是一个无限集时，定义 $VCdim(F) = \infty$ 。

举个简单的例子来说，一个 2 维平面上，对于任意不共线的 3 个点，一条直线的 VC 维是 3。推广到一般情况，对于一个线性分类面 $y = w^T x + \theta$ ，其中 $w \in \mathbb{R}^n$ 是分类面的法向量， θ 是分类面的常数项（阈值），则线性分类面的 VC 维 = $\dim(w) + 1 = n + 1$ 。

下面的定理证明了 VC 维与期望风险（测试误差）之间存在如下关系。

定理：令 F 的 VC 维 h ，若 $N > h$ ，且 $h(\ln(2N/h) + 1) + \ln(4/\delta) \geq 0.25$ ，则对于任意概率分布 $P(x, y)$ 和任意的 $\delta \in (0, 1]$ ， F 中的任意假设 f 都可以使得下列不等式至少以 $1 - \delta$ 的概率成立。

$$R[f] \leq R_{emp}[f] + \sqrt{\frac{h(\ln(2N/h) + 1) - \log(\eta/4)}{N}} \quad (3.11)$$

从上述定理可以得知，期望风险 $R[f]$ 的上界是经验风险 $R_{emp}[f]$ 与 VC 维之和的某个函数，其中不等式(3.11)右边的两项之和也被称为结构风险。而此处的经验风险即对应着训练误差，期望风险对应着测试误差。当经验风险（训练误差）一定时，最小化 VC 维，会导致结构风险变小，因此会使得测试误差的上界变小，这样有利于决策函数的推广能力。

3.4. 基于 1 范数最小化学习的优化模型

众所周知，SVM、Adaboost 等分类器设计的准则是基于边界最大化原则，边界最大化的目的是使得期望风险的上界变小。本文提出基于 1 范数最小化的优化模型 LML，直接追求 VC 维最小化，是从另一个角度最小化期望风险的上界。

下面证明所提 LML 优化模型是依据 VC 维最小化原则建立的。首先给出一个一般的 1 范数最小化学习的优化模型。

$$\begin{aligned} \min \quad & \|w\|_1 \\ \text{s.t.} \quad & \text{linear constraints of } w \end{aligned} \quad (3.12)$$

其中 $\|w\|_1 = \sum_{j=1}^n |w_j|$ ， w 表示一个线性超平面的法向量， w_j 表示向量 w 的第 j 维分量。

下面来看下 1 范数最小化与 VC 维之间的关系。一个线性超平面的 VC 维 h 是等于其法向量的维数加 1，即 $h = \dim(w) + 1 = n + 1$ 而且根据上述 VC 维、训练误差与测试误差之间的关系：

$$R_{train} + \sqrt{\frac{h(\ln(2N/h) + 1) - \log(\eta/4)}{N}} \quad (3.13)$$

其中 R_{train} 表示训练误差，也称为经验风险，与式中的经验风险 $R_{emp}[f]$ 一致。1 范数最小化已经证明是 0 范数最小化的近似最优估计。0 范数最小化是指寻找一个有着最少非零元素的向量解。因此线性超平面法向量的 1 范数最小化意在寻求一个法向量有着最少的非零元素。这意味着法向量的维数尽可能地低，也即 n 的值尽量小，如果 n 很小，那么这将导致 h 的值尽量小，也即要求 VC 维尽量小，对应于式(3.13)中的结构风险变小，

即测试误差的上界变小。因此 1 范数最小化可以导致 VC 维最小，进一步保证测试误差上界减少，保证决策函数的泛化性能提高。

下面将详细介绍所建立的 LML 优化模型，其优化目标函数和约束方程如下：

Model I:

$$\begin{aligned} \min_{w, \xi_i} \quad & \|w\|_1 + C_1 \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i \cdot h_w(x_i) \geq \alpha - \xi_i, \\ \xi_i \geq 0, & i = 1, \dots, N \\ h_w(x_i) = w^T \cdot x_i \end{cases} \end{aligned} \quad (3.14)$$

上述优化模型中， w 表示线性超平面的法向量。考虑到样本不一定是线性可分的，对第 i 个训练样本引入松弛变量 ξ_i ，松弛变量 ξ_i 可以用来衡量第 i 个训练样本被错分的程度。 N 是训练样本的总数， C_1 是一个预定义的参数，用来平衡 1 范数最小化与错分程度之和最小化哪个更重要。 x_i 表示第 i 个训练样本的特征向量， y_i 表示第 i 个训练样本的类别标号。 α 是一个预定义的参数， α 与所有的 ξ_i 用来保证正例样本与反例样本具有可分性。

SVM 采用的是 w 的 2 范数的平方最小，本文采用的是 w 的 1 范数最小。1 范数和 2 范数虽然在向量空间是等价的，都表示某种距离度量，但是在优化理论上，二者有着本质的区别：1 范数，是向量 w 所有分量的绝对值之和，而 2 范数是向量 w 所有分量的平方和开根号。所以对向量 w 求最小化的时候，2 范数最小化要求向量 w 的每一维均匀地减少，而 1 范数的“稀疏性”可允许 w 的某个分量很大，其余的分量很小。1 范数的这种特性激发作者利用 1 范数最小化来对特征进行选择，并建立线性分类器。

优化模型 Model I 是一个凸规划问题。虽然 1 范数和 2 范数都是凸函数，但是二者的求解方式完全不同，2 范数光滑可微可以使用 Lagrange 对偶算法求解，1 范数不是可微的，只能通过其他的一些算法(如内点算法)求解。因此它没有像 2 范数那样，直接通过 Lagrange 函数求偏导从而获得原始变量解析式的求解算法。1 范数的求解问题可以通过一个简单的变换，首先先引入两个新向量 u 和 v 将其转换成线性规划来求解。严格地说，这并不是数学上所说的线性规划，因为变量 u 和变量 v 之间存在一定的关系，并不独立。

在引入两个新向量后，进行如下的替换，令 $w = u - v$ ，其中 $u \geq 0, v \geq 0$ 。向量 u 和 v 与原向量 w 之间的关系如下： $u_j = (w_j)_+$ ， $v_j = (-w_j)_+$ ，其中 $(\cdot)_+$ 表示“正部分”操作算子，定义如下： $(w_j)_+ = \max\{0, w_j\}$ 。这样，有 $\|w\|_1 = I_n^T u + I_n^T v$ ，其中 $I_n = [1, 1, 1, \dots, 1]^T$ 是一个 n 维的单位向量。因此，将新向量代入到式 (3.14) 中，得到如下的优化模型：

$$\begin{aligned} \min \quad & I_n^T u + I_n^T v + C_1 \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i \cdot (u - v)^T x_i \geq \alpha - \xi_i \\ u \geq 0 \\ v \geq 0 \\ \xi_i \geq 0 \end{cases} \end{aligned} \quad (3.15)$$

上述(3.15)优化模型和原来的 Model I 优化模型等价, 对于该模型的求解可以使用内点算法, 具体关于优化模型转换的细节以及求解算法, 可以参见文献[26]。

3.5. 线性分类器的构建

最后通过求解优化模型, 得到线性超平面的法向量 w 。从而得到一个线性分类器, 该线性分类器也即上面所述的线性超平面的符号函数。其表述形式如下:

$$g(x) = \text{sign}(h_w(x) - \theta) = \text{sign}\left(\sum_{j=1}^n w_j x^j - \theta\right) \quad (3.16)$$

其中 x^j 表示测试样本 x 的第 j 维特征。 θ 是分类器 $g(x)$ 的阈值。阈值 θ 是通过正例样本中最小的 $h_w(x_i)$ 与反例样本中最大的 $h_w(x_i)$ 的线性组合得到的。其计算公式如下:

$$\theta = \eta \left(\min_{x_i \in \text{positives}} h_w(x_i) \right) + (1 - \eta) \left(\max_{x_i \in \text{negatives}} h_w(x_i) \right) \quad (3.17)$$

其中 $i=1, 2, \dots, N$, 并且权重因子 $\eta \in [0, 1]$ 。L-LML 算法的流程图请参见表格 3-1。

表 3-1. L-LML 算法流程图

输入: 训练样本集 $\{(x_i, y_i)\}_{i=1..N}$, $y_i \in \{-1, +1\}$

- 令 $\alpha = 1.0$, 求解优化模型 (3.15)
- 对于 $k=1..K$, 进行如下迭代, $K \leq N/2$
- 随机选择一个正例样本 x_k^+ , 一个反例样本 x_k^- 将其放进优化模型中。
 - 对于 $l=1, 2, \dots, L$, 进行如下循环:
 - $u_{l+1} = u_l + \lambda_l d_l$, $v_{l+1} = v_l + \gamma_l t_l$, 其中 λ_l, γ_l 是迭代步长; d_l, t_l 是迭代方向;
- 获得 u, v 两个向量。

Output: 弱分类器的权重向量

$$g(x) = \text{sign}(h_w(x) - \theta) = \text{sign}\left(\sum_{j=1}^n w^j x^j - \theta\right)$$

3.6. 特征稀疏表示

对于一个人体的 HOG 特征，图 3-2 显示了原始的 HOG 特征值随着特征维数变化的取值，可以看出原始的 HOG 特征是一种稠密的特征描述子，存在一定的冗余信息。因此，有必要对 HOG 特征进行选择。通过对原始向量空间中的 HOG 特征进行 LML 优化后，本章得到权重向量以及一个人体训练样本的加权特征向量（即权重与人体样本向量的相应维度进行乘积，得到一个同样维度的向量，称为加权特征向量）。当设定一个阈值是 0.001 时，加权特征向量中约 75% 的元素值都是小于这个给定的阈值，其余的 25% 是主要的成份，这一稀疏表示过程可以视为特征选择的过程。

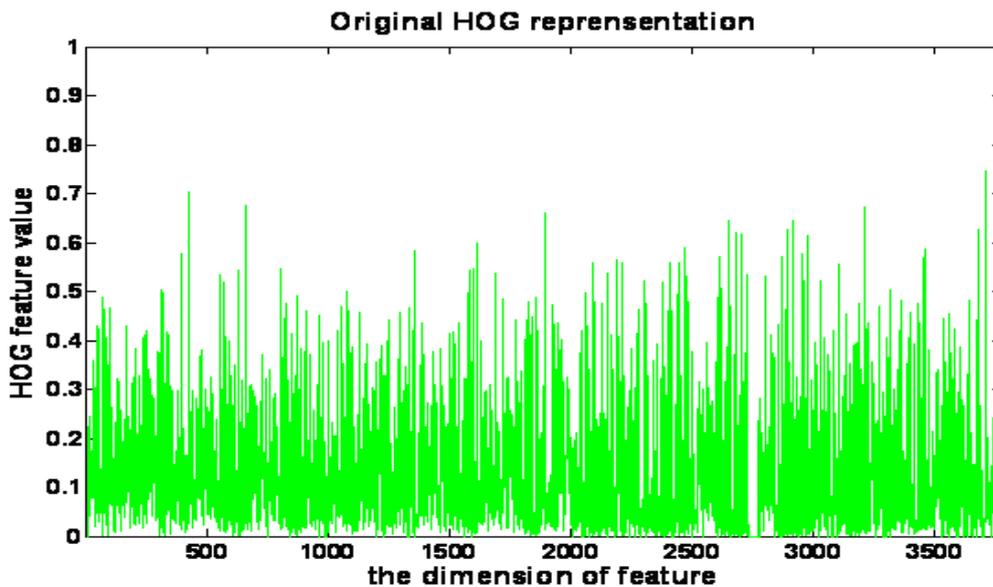


图 3-2. 人体 HOG 特征表示

图 3-3 中显示了所提 L-LML 方法得到的加权特征向量稀疏表示与线性 SVM 中采用 2 范数获得的加权特征向量稠密表示的对比图。根据实验观察，使用所提的 L-LML 方法的加权特征向量中至少有 75% 是稀疏的，那么当遮挡或者多姿态出现在这些稀疏权重的地方，遮挡和多姿态将会对检测结果造成很小的影响或者几乎无影响。换句话说稀疏表示能够减轻遮挡或多姿态等问题造成的困扰，并且使得人体检测算法更加鲁棒。

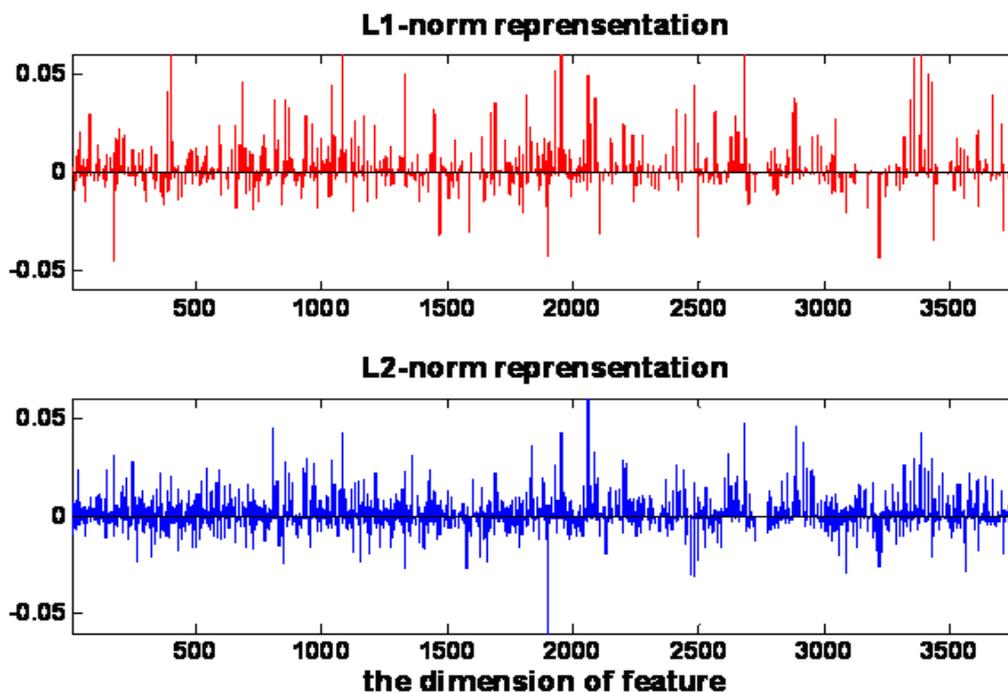


图 3-3.稀疏表示和稠密表示

下面结合 64×128 大小的人体训练样本，给出不同位置、同一尺度下 HOG 块状特征及 L-LML 和线性 SVM 算法所学到加权 HOG 特征的可视化显示。可视化的显示方式如下：对于一个人体的原始 HOG 特征，水平方向有 7 个特征块，竖直方向有 15 个特征块，每个块的大小是 16×16 像素，每个块由 4 个 cell 组成，因为 HOG 是采用重叠计算的滑动块形式，所以本节参照 Dalal 等人的显示结果，只选择每个 cell 中的主方向展示，那么 4 个 cell 只有 4 个主方向，并且它们分别向 0-9 方向进行投影，每个块的中心点位置绘制所有 cell 的主方向，并且线段的长度不超过 4 个像素。

对于加权图的展示，如图 3-4，将已经训练得到 L-LML 算法和线性 SVM 算法的权重，分别对同一个人体的原始 HOG 特征进行加权显示，从左到右依次显示了训练样本、原始的 HOG 特征、SVM 加权之后的人形图和 L-LML 加权之后的人形图展示。从图 3-4(c) 和(d)，(g)和(h)比较可知，L-LML 算法得到的权重向量和稀疏响应向量要比 SVM 的权重向量和响应向量稀疏很多，而且可以观察到，L-LML 算法得到的稀疏响应向量，大部分集中在中间位置，这对应着图片中人体的位置，其余部分都是背景信息。并且从图上还可以得知 L-LML 算法的稀疏响应向量是从 SVM 权重向量中挑选得到的。SVM 在背景信息部分的权重响应还是非常大，所以如果背景部分噪声很多，那么势必会影响检测结果。从这一点也可看出，稀疏表示的特征选择的作用是去除冗余信息，尽可能地减少冗余信息将有用的信息淹没的可能性。

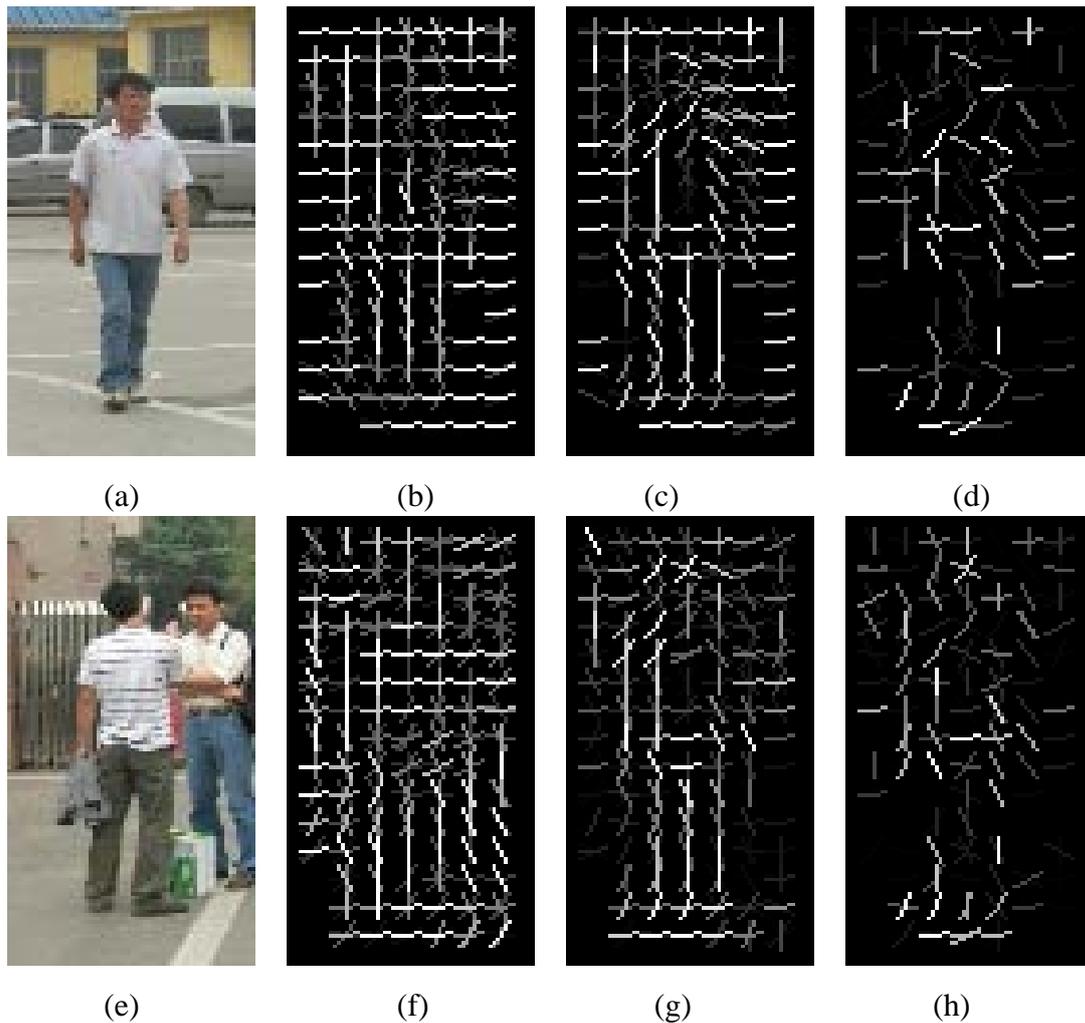


图 3-4. 人体可视化显示: (b)和(f)分别是原始的 HOG 特征; (c)和(g)是线性 SVM 加权之后的显示结果; (d)和(h)是 L-LML 加权之后的显示结果;

3.7. 实验结果

本章使用大概 1300 个正例训练样本, 3700 个反例训练样本, 正例训练样本都是来自于 MIT[30]和 SDL[28]的训练样本集。反例训练样本来自于 INRIA 上的反例图片集。图 3-5 展示了本章中所用的一些正例样本与反例样本, 可以看出, 所选择的正例训练样本主要是来自于正面视角的人体或者近似于正面视角的人体, 然而令作者惊讶的是, 所得到的人体检测模型可以处理多视角问题, 并对遮挡问题比较鲁棒, 实验结果支持本章的这一结论。



图 3-5.正反例训练样本

验证 L-LML 算法的测试数据集有两个，一个是作者所处实验室自己采集的 SDL 数据集，该集合里面有 59 副测试图片，这 59 副图片背景和人体的姿态相对比较简单。另外一个数据集是由 Dalal 等人采集的并公开发布的 INRIA 测试集，该集合有 288 副测试图片，虽然测试图片中的人体大部分是出于直立状态，但是因为背景的复杂性、人体姿态的多变性、视角的多样化以及遮挡等问题使得这些测试图片非常具有挑战性。

Recall Rate 和 False Positives Per Window (FPPW) 这两个指标是衡量检测算法的性能的常用指标，本章使用这两个指标定性地估计 L-LML 算法的性能。两个指标的定义分别如下：

$$\text{RecallRate} = \frac{\text{正确的检测结果个数}}{\text{全部人体的个数}} \quad \text{FPPW} = \frac{\text{虚假检测结果个数}}{\text{图像中总窗口数目}} \quad (3.18)$$

本章认为检测区域与真实的物体区域重叠达到 90% 以上，则认为是一个正确的检测结果，否则认为是虚假的检测结果。在训练 L-LML 分类器的时候，当经验地确定参数 $\eta = 0.87$ 和 $\theta = 1.1$ 时，可以得到一组 Recall Rate 和 FPPW 的值，该组值是一个比较好的折中值。不同组的 Recall Rate 和 FPPW 的值，可以绘制出图 3-6 和图 3-7 中的曲线。在图 3-6 和图 3-7 中，本章展示了 L-LML 算法的性能与 SVM 算法的性能比较曲线。图 3-6 中展示的是在 SDL59 副测试图片上的结果，图 3-7 是在 INRIA288 副测试图片上的结果。从两个测试集上的结果展示了所提 L-LML 方法优于线性 SVM 分类器。

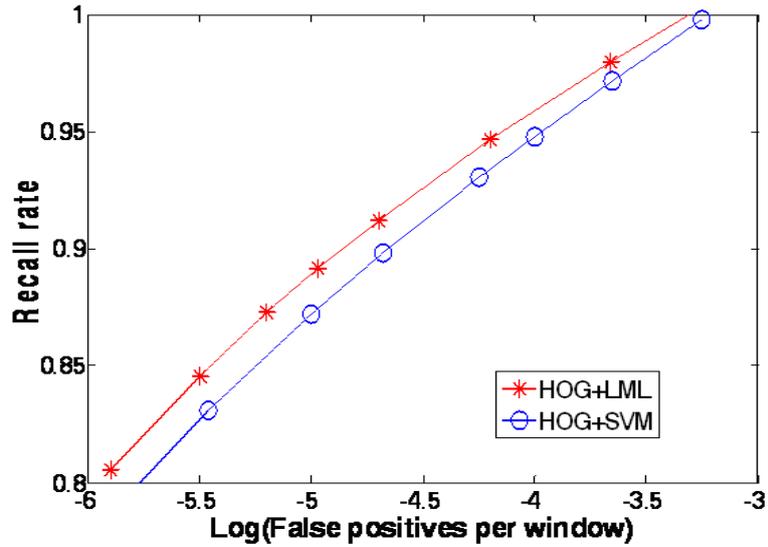


图 3-6.LML 在 SDL 上测试曲线

除此之外，本章还比较了 L-LML 算法与 SVM 算法的效率。L-LML 在奔腾 IV 3.0 GHz CPU 下，处理一副 320x240 的图片大概需要 8 秒钟，并且本章未使用积分图计算 HOG 特征等优化处理手段，这个速度要比 SVM 分类器快 6 倍左右。

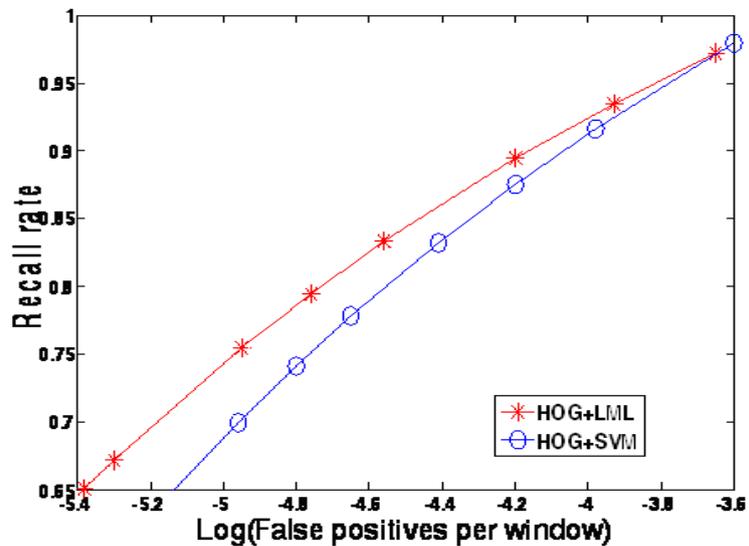


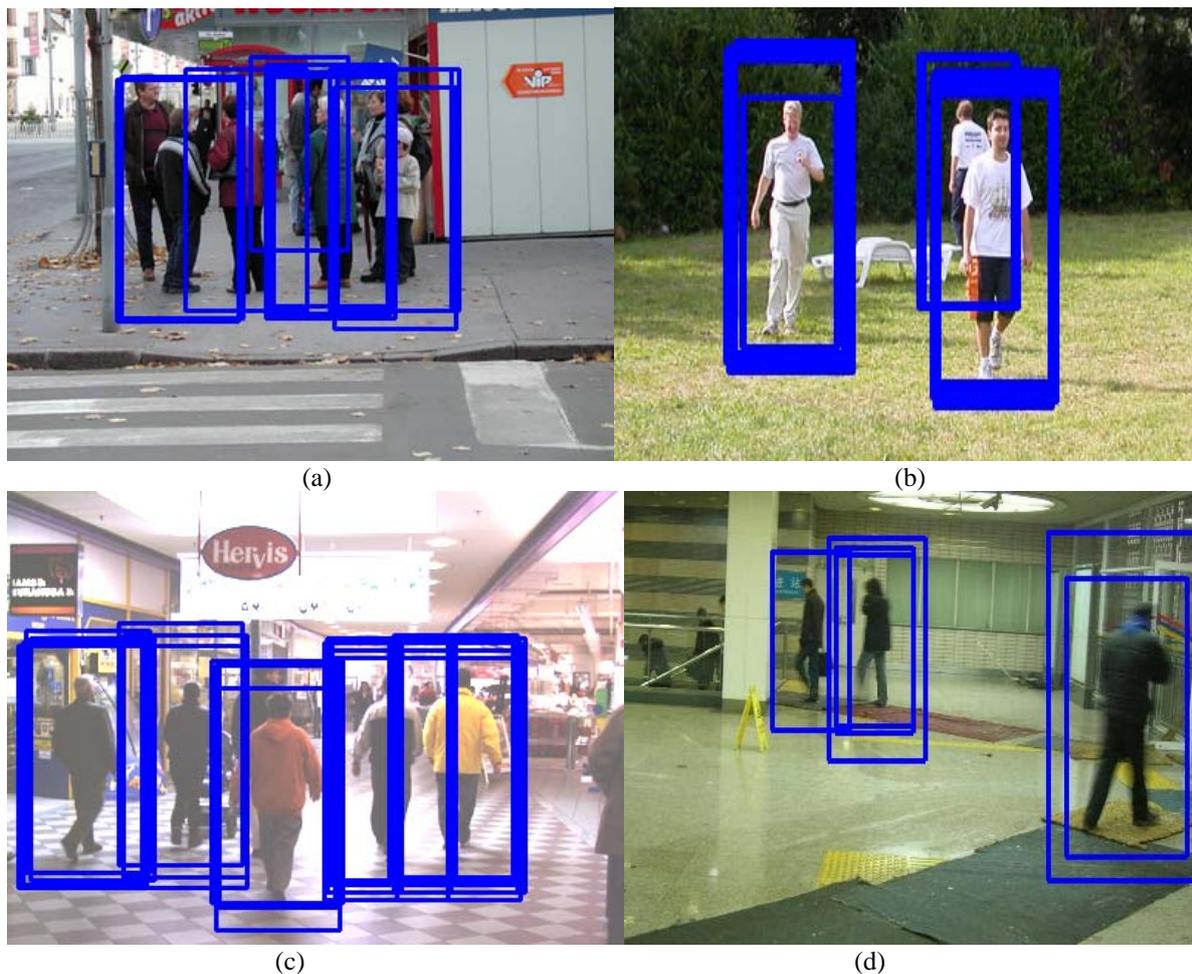
图 3-7.LML 在 INRIA 上测试曲线

在图 3-8 中，展示了使用 L-LML 方法的检测结果。从图 3-8(a)到图 3-8(j)，所有的人被正确地定位，无论他们是否处于部分遮挡状态或者多姿态。在图 3-8(a)中，那个穿黑夹克的人被部分遮挡，但是 L-LML 算法仍然正确将其定位。在图 3-8(b)中，从左数的第二个人，他的背面呈现在图像中，虽然他被遮挡，但是 L-LML 算法可以成功地检测到该目标。在图 3-8(c)中，从图像的左边数第三个人，他被前面穿红色衣服的人遮挡了，

但是仍然能够被 L-LML 算法找到。在图 3-8(d), 在地铁中无论正视角还是侧视角的人体, 全部被检测到。在图 3-8(f)中, 这个骑自行车的人处于非竖直的弯腰姿态, L-LML 算法正确将其检测到。在图 3-8(h)中, 白色的石像被错误的检测为人体。图 3-8(b)、3-8(d)和 3-8(j)来自于 SDL59 测试集, 其余的展示图片来自于 INRIA 测试集。

3.8. 本章小结

本章提出一种基于 1 范数最小化的人体检测算法 (L-LML)。本章首先论述了稀疏表示, 1 范数目前的应用进展以及所提模型 LML 理论上的依据, 然后介绍了 LML 优化模型以及该优化模型的转换形式, 并构建了线性分类器。随后, 本章展示了 L-LML 与 SVM 算法在人体表示上的差别, 可见 L-LML 算法能够达到稀疏表示的结果, 即实现特征选择的功能。除此之外, 本章还展示了 L-LML 算法在人体检测集上的效果。从实验结果来看, 特征选择功能对于减轻人体检测中遮挡与多姿态等问题的影响是非常重要的。L-LML 算法在分类器构建的时候, 实现了特征选择的功能, 因此该算法对于遮挡与多姿态等问题比较鲁棒。



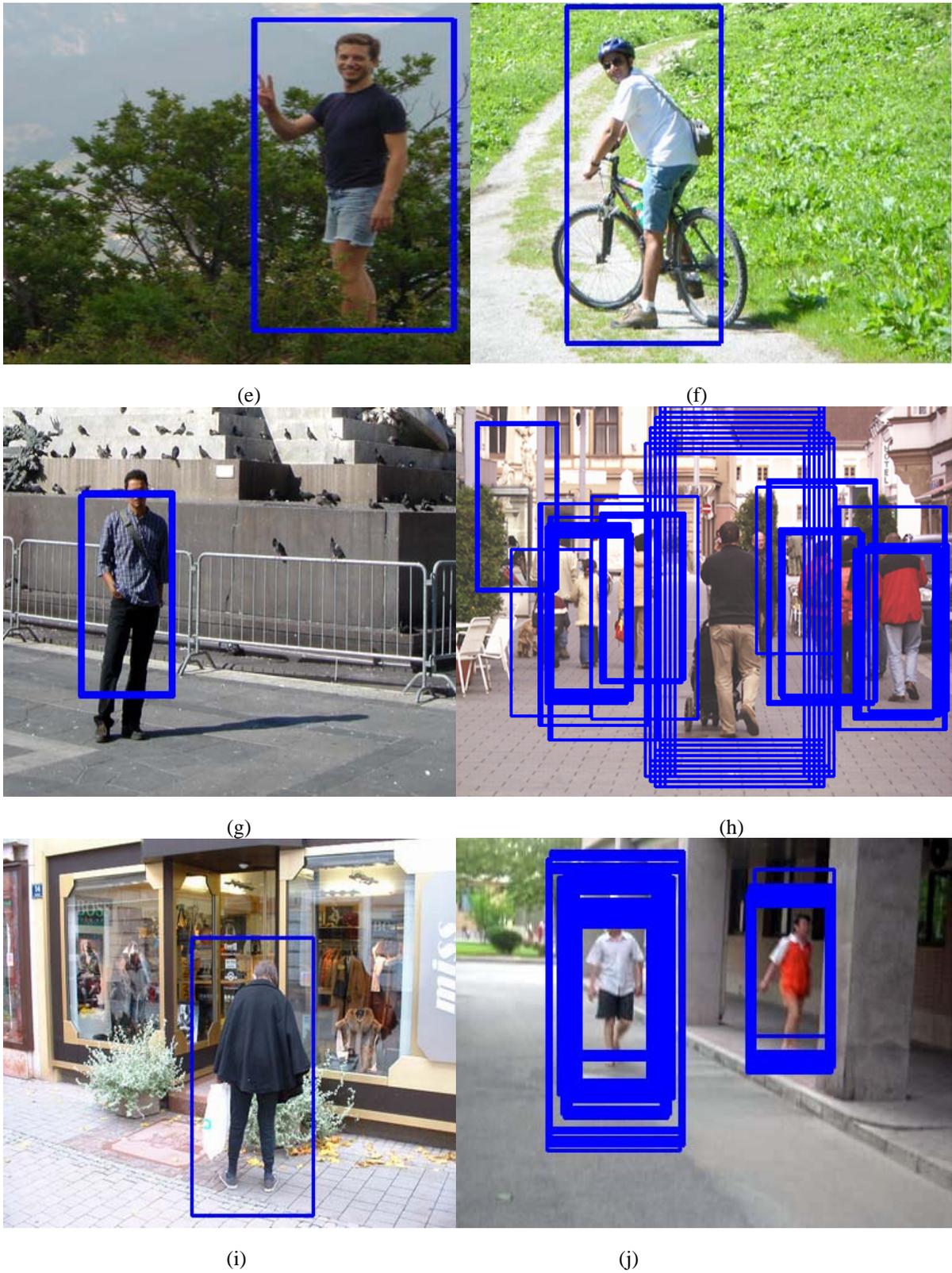


图 3-8.LML 检测实例

第四章 基于非线性 1 范数最小化学习的人体检测

本章在上一章提出的 LML 优化模型基础上，设计了一种非线性的 1 范数最小化学习 (NL-LML) 算法。该算法将一种基于核函数的非线性变换与 LML 优化模型结合，实现对样本的非线性判别。本章首先思考了基于 1 范数最小化的非线性判别如何实现，是否能像 SVM 那样引入核函数的内积形式到其对偶规划模型中，然后介绍了一些典型的核函数，并在此基础上引入了一种基于核函数的非线性变换，将该变换与 LML 模型结合训练判别函数的法向量与阈值，随后展示了在变换空间中人体样本的稀疏表示，并在 IRIS 数据集和人体数据集上展示了 NL-LML 算法的效果。

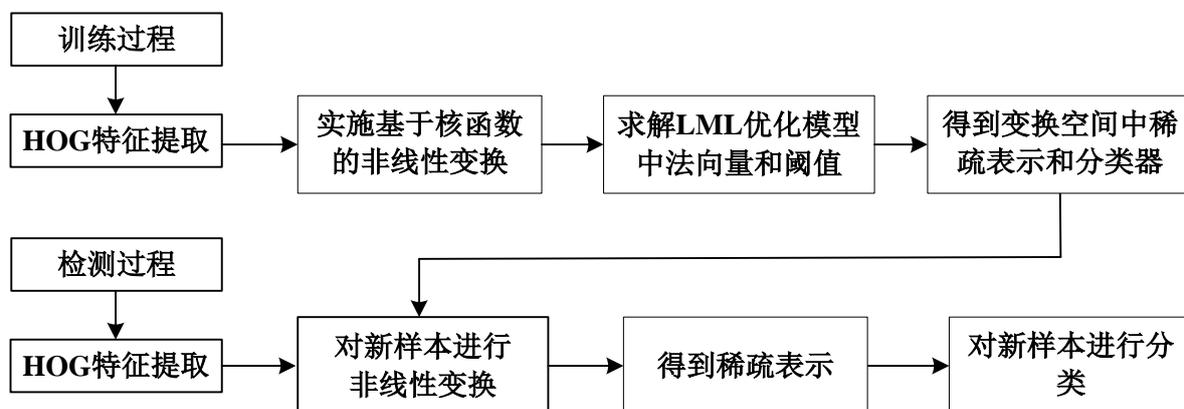


图 4-1.NL-LML 算法框图

4.1. 非线性 1 范数最小化学习的思路

在实际生活中，样本（数据点）分布通常是比较复杂的，有些样本并不是线性可分的，虽然可以通过在 LML 优化模型引入松弛变量的形式容许一些非线性可分样本的存在，但是如果仍然采用线性分类的模式，不能保证训练数据集上的样本分类效果（训练误差），若强行地将训练得到的线性分类算法用于测试数据集，可想而知，分类结果会如何。因此，针对非线性的样本分布，需要设计非线性的分类模型。

对于目前非常成功的 SVM 方法，它采用了核函数的形式实现非线性分类。SVM 首先认为在低维空间不是线性可分的样本，当其升维到高维空间中，可以用线性超平面近似分开。因此 SVM 先构造一个映射函数，将样本映射到高维空间，然后在高维空间中使用线性超平面去分类。幸运的是，在线性 SVM 的对偶规划中，存在样本的内积形式，

核函数是一种内积形式的度量，所以内积可以合理地由某种核函数来代替，这样即使对样本升维，在高维空间中的内积计算量没有比原空间中优化模型的计算量多很多。

但是针对所提的 LML 模型，从目标和约方程 (3.15) 中，可以得出 LML 优化模型转换成一个线性规划的形式，线性规划的对偶规划仍然是一个线性的。因此像内积形式的二次项并不存在于 LML 的对偶规划中，所以直接在 LML 优化模型的对偶规划中引入核函数的形式，这种方式是不可行的。并且如果盲目地将样本升维到一个高维空间，然后再用 LML 优化模型去计算，这样不仅增加计算费用，而且理论上由于维数灾难的原因，这种想法违背了模式识别中的经典理论。因此本章重新开始思考基于 LML 的非线性分类。

本章工作主要受到[43]研究工作的启发，提出了使用非线性变换与 LML 优化模型结合的非线性分类方法。在文献中[43]，作者提出核函数可以被看成是一个映射，该映射可以视为特征的一种变换模式，并且他们已经证明在高维空间中，如果正例样本与反例样本之间存在边界是 r ，那么使用核函数映射得到的新正例样本与新反例样本在映射空间中，仍以概率 $1 - \delta$ 存在边界为 $0.25r$ 的误差最多是 ε 。从上述论断，可以认为核函数可以被看做是一种特征变换，而且这种特征变换并不影响高维空间中样本的可分性，仍使不同类别的样本间保持着一定的间距。

4.2. 核函数

从上述介绍可知，核函数之所以受到极大的欢迎，是因为其内积性质既实现了样本的相似性度量，又减少了计算费用。下面从数学上正式引入核函数的表达形式。首先，定义核函数之前，需要引入矩阵算子，半正定性等概念。

$X = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_m}\}$ 是一个有限点集，令 $K = (K(\overline{x_i}, \overline{x_j}))$ 是一个 $m \times m$ 的矩阵。

定义 4.2.1 (矩阵算子): 对一个 m 维的向量 $u = (u_1, u_2, \dots, u_m)^T \in R^m$ ， $Ku \in R^m$ ，并且其分量由下式表示：

$$[Ku]_i = \sum_{j=1}^m K(\overline{x_i}, \overline{x_j})u_j \quad i = 1, 2, \dots, m \quad (4.1)$$

则称 $K = (K(\overline{x_i}, \overline{x_j}))$ 为一个矩阵算子。

定义 4.2.2 (半正定性): 如果对于任意的一个 m 维向量 $u = (u_1, u_2, \dots, u_m)^T \in R^m$ ，和一个矩阵算子 K 若有：

$$u^T Ku = \sum_{i,j=1}^m K(\overline{x_i}, \overline{x_j})u_i u_j \geq 0 \quad (4.2)$$

则称矩阵算子 K 是半正定的。

定义 4.2.3 (核函数): 若一个矩阵算子 K 是半正定的, 则存在从 $X = \{\overline{x_1}, \overline{x_2}, \dots, \overline{x_m}\}$ 到 R^m 的映射 ϕ , 使得 $K(\overline{x_i}, \overline{x_j}) = (\phi(\overline{x_i}) \cdot \phi(\overline{x_j}))$, $i, j = 1, 2, \dots, m$, 其中 (\cdot) 是 Hilbert 空间 R^m 上的内积, $K(\cdot, \cdot)$ 是被称为一个核函数。

从上述定义可知, 核函数其实是在 Hilbert 空间上的一种内积运算, 因此选取不同的映射 ϕ , 然后在其上定义内积运算即可形成核函数。目前常见的核函数有以下几种:

$$\text{多项式核函数: } K(x, \overline{x}) = ((x \cdot \overline{x}) + c)^d \quad \text{其中 } c \geq 0 \quad (4.3)$$

$$\text{径向基核函数: } K(x, \overline{x}) = \exp(-\sigma \|x - \overline{x}\|^2) \quad \text{其中 } c \geq 0 \quad (4.4)$$

$$\text{傅里叶核函数: } K(x, \overline{x}) = \frac{1 - q^2}{2(1 - 2q \cos(x - \overline{x}) + q^2)} \quad 0 < q < 1 \quad (4.5)$$

$$\text{Sigmoid 核函数: } K(x, \overline{x}) = \tan(c(x \cdot \overline{x}) + d) \quad \text{其中 } c \geq 0, d < 0 \quad (4.6)$$

4.3. 非线性分类器的构建

下面将介绍基于 1 范数最小化的非线性分类方法 (NL-LML)。首先假设在低维空间不是线性可分的样本, 那么它们在高维空间中可近似线性可分 (这与 SVM 的假设一致)。NL-LML 方法首先对样本实施由核函数诱导的非线性变换。所有的样本被变换到另一个空间中去, 这样形成一个核矩阵。因为变换特征存在冗余和相关性等无用信息, 所以在变换空间中, 有必要提取核矩阵的主成分 (正交基), 使用正交基重新表示变换空间中的样本。这样对新样本, 再次使用 LML 优化模型来训练变换空间中线性超平面的法向量。

4.3.1. 基于核函数的非线性变换

数学上, 令训练样本的特征向量构成的集合记为: $X = \{x_1, x_2, \dots, x_N\}$ 。对于一个原始样本点 $x \in X$, 可构造如下一个非线性变换:

$$\phi: x \rightarrow R^N, \text{ 并且令 } \phi(x) = (K(x_1, x), K(x_2, x), \dots, K(x_N, x))^T$$

其中 $K(\cdot, \cdot)$ 是一个核函数。一般来讲不同形式的核函数作用在训练集上, 将形成不同的核矩阵。核矩阵里面的第 i 行第 j 列的元素为 $K = (K_{ij}) = K(x_i, x_j)$ 。核矩阵定义如下:

$$K = (K_{ij}) = \begin{pmatrix} K(x_1, x_1), K(x_1, x_2), \dots, K(x_1, x_N) \\ K(x_2, x_1), K(x_2, x_2), \dots, K(x_2, x_N) \\ \vdots \\ K(x_N, x_1), K(x_N, x_2), \dots, K(x_N, x_N) \end{pmatrix} \quad (4.7)$$

核矩阵 K 是由所有样本的内积形成的。样本的核矩阵对应一个线性空间，在该空间中，需要寻找该空间的主成分，即正交基，以便每个样本都可以用正交基重新表示。由于核矩阵的半正定性质，所以求解其正交基是比较容易的，可以使用 Cholesky 分解或者 SVD 分解[57]获得主成分。下面分别介绍 Cholesky 和 SVD 分解，SVD 分解比 Cholesky 分解的用途更广，因为它不仅对于正定、半正定矩阵可以操作，对一般的矩阵也能进行如下操作。

定理 4.3.1 (Cholesky 分解): 设 $A = (a_{ij})$ 是正定矩阵，存在唯一的对角元全为正数的下三角矩阵 $G = (g_{ij})$ ，使得 $A = GG^T$ 。对于矩阵 $G = (g_{ij})$ 的每个元素的计算公式如下：

$$\begin{cases} g_{ii} = \sqrt{a_{ii} - \sum_{j=1}^{i-1} g_{ij}^2} \\ g_{ki} = \frac{a_{ki} - \sum_{j=1}^{i-1} g_{kj}g_{ij}}{g_{ii}} \quad k = i+1, i+2, \dots \end{cases} \quad (4.8)$$

定理 4.3.2 SVD (奇异值) 分解: 设 A 是 $m \times n$ 实矩阵，存在阶数分别为 m 和 n 的正交矩阵 U 和 V ，满足 $A = USV^H$ ，这里， S 是对角元为 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ 的 $m \times n$ 阶长方形对角矩阵。其中对角元 $\sigma_1, \sigma_2, \dots, \sigma_n$ 称为矩阵 A 的奇异值。

下面介绍下为什么一个实矩阵可以进行奇异值分解，顺便引出如何计算矩阵的奇异值。

由于 $A^T A$ 矩阵的对称性、非负性，设 $A^T A$ 有 n 个特征值，其特征值都是非负的，记为： $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > 0$ ， $\lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$ ，且存在 n 阶正交矩阵 V ，使得

$$V^H (A^T A) V = D \quad (4.9)$$

其中 D 是以 $\lambda_1, \lambda_2, \dots, \lambda_n$ 为对角元的 n 阶对角矩阵。令 $V = [v_1, v_2, \dots, v_n]$ ，每个 v_i 是矩阵 V 的列向量。从式(4.9)可知，矩阵 V 是由矩阵 $A^T A$ 的特征向量所组成，也即 v_1, v_2, \dots, v_r 是 $A^T A$ 关于非零特征值 $\lambda_1, \lambda_2, \dots, \lambda_r$ 相应的单位特征向量， $v_{r+1}, v_{r+2}, \dots, v_n$ 是 $A^T A$ 关于零特征值 $\lambda_{r+1}, \lambda_{r+2}, \dots, \lambda_n$ 对应的特征向量。记 $W = AV$ ， $S^H S = D$ ，其中 $W = (w_1, \dots, w_r, \dots, w_n)$ ， $w_i \in \mathbb{R}^m$ ， S 是对角元为 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$ 的 $m \times n$ 阶长方形对角矩阵，易得出 $\sigma_i = \sqrt{\lambda_i}$ 。记 $U = (u_1, u_2, \dots, u_m)$ ，因为矩阵 U 也是酉交矩阵，所以 $US = AV$ ，相应地，当 $w_i \neq 0$ 时，对应有

$$u_i = w_i / \sigma_i = (Av_i) / \sigma_i, \quad i = 1, 2, \dots, r, \quad (4.10)$$

其余对应 $w_i = 0$ 时，可以选择 $u_{r+1}, u_{r+2}, \dots, u_m$ 使得矩阵 $U = (u_1, u_2, \dots, u_m)$ 成为一个正交阵。所以从上述分解过程，很容易得到矩阵 A 的奇异值分解。首先，先求出 $A^T A$ 的特征值

$\lambda_1, \lambda_2, \dots, \lambda_r$ 和其对应的特征向量 v_1, v_2, \dots, v_r ，然后每个特征值开方即得到矩阵 S ，至于矩阵 U ，它的列向量可依据 (4.10) 求出，因此很容易求出矩阵 A 的奇异值分解。

通过分解，可以得到核矩阵的特征向量，取前 d 个特征向量 $\{\eta_1, \eta_2, \dots, \eta_d\}$ 作为核矩阵的正交基，也即是核矩阵的主成分。 d 个特征向量组成的矩阵记为： $V = \{\eta_1, \eta_2, \dots, \eta_d\}$ ，然后可以获得样本 x_i 在新的空间（变换空间）中的表示，也即指将样本 x_i 的映射投影到矩阵 V 上。表示方程如下：

$$\Psi(\phi(x_i)) = V^T \cdot \phi(x_i) \quad (4.11)$$

4.3.2. 非线性变换与LML优化模型结合

对所有的样本进行重新表示之后，NL-LML 使用 LML 优化模型在变换空间训练线性超平面的法向量。如下是新样本在变换空间中的 LML 模型，和原来的 LML 优化模型相比，只是修改了其中部分变量形式。

$$\begin{aligned} \min_{w, \xi_i} \quad & \|w\|_1 + C_1 \sum_{i=1}^d \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i \cdot h_w(x_i) \geq \alpha - \xi_i, \\ \xi_i \geq 0, & i = 1, \dots, d \\ h_w(x_i) = w^T \cdot (V^T \cdot \phi(x_i)) \end{cases} \end{aligned} \quad (4.12)$$

在求解得到变换空间中的法向量之后，建立一个线性分类器，该分类器的阈值由下式确定：

$$\theta = \eta \left(\min_{x_i \in \text{positives}} h_w(x_i) \right) + (1 - \eta) \left(\max_{x_i \in \text{negatives}} h_w(x_i) \right) \quad (4.13)$$

最后的分类器的形式如(4.14)式：

$$g(x) = \text{sign}(h_w(x) - \theta) = \text{sign}\left(\sum_{j=1}^d w_j \cdot \Psi(\phi(x)) - \theta\right) \quad (4.14)$$

4.4. 特征稀疏表示

对于人体训练样本的特征表示，本章使用 HOG 特征作为其特征描述子，提取过程请参见第二章 2.12 节。对 HOG 特征进行上述非线性变换之后，使用公式 (4.12) 得到变换空间中的法向量，使用公式 (4.13) 得到阈值。由于 LML 优化模型的特征选择功能（已在 L-LML 方法中论述），所以在变换空间中，类似地也得到了新样本在变换空间中的稀

疏表示。在下图 4-2 展示了新样本在变换空间中的稀疏表示，此处非线性变换的核函数采用的是径向基核函数。从图 4-2 中，可以看见样本的特征向量中只有一小部分是主要特征，其余的都近似接近于 0，这一过程突出了样本特征向量元素之间的主要成分，尽量使非主要成分接近于 0，不起作用，因此这一过程可以看成是稀疏核的特征选择过程。也正是因为 NL-LML 算法的这种特性，将大大减少核变换所引起的计算费用。

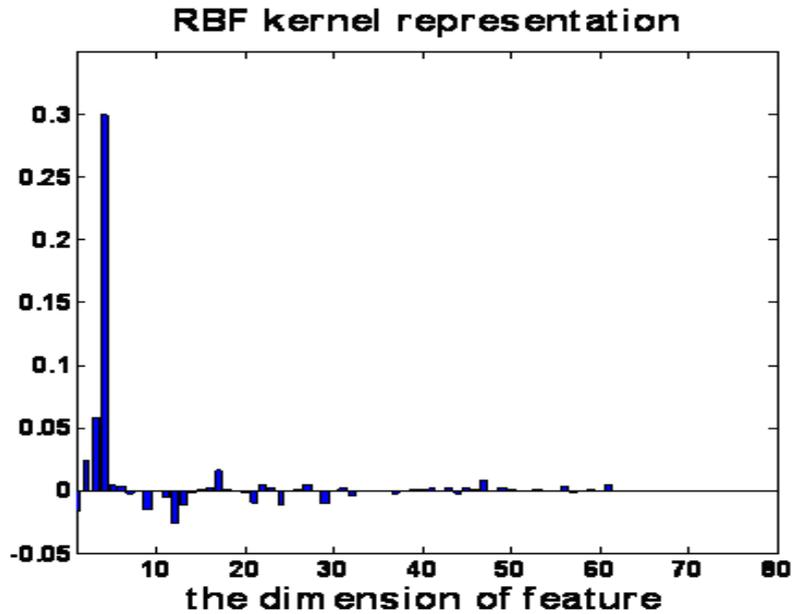


图 4-2.NL-LML 稀疏表示

4.5. 实验结果

NL-LML 算法主要是侧重于非线性分类判别，并未使用其他的人体检测技巧，比如说级联机制等。因此，本章主要将 NL-LML 算法和一些与之类似的算法进行实验对比，比如说线性 SVM 和核函数 SVM，以及前面一章提出的 L-LML 算法。本章一共使用 3 个测试集合，一个是 IRIS 数据集，另外两个是人体数据的测试集。其中 IRIS 数据集是一个比较简单的著名的花朵数据集，该数据集中的第二类和第三类样本是非线性可分的。本章使用 IRIS 数据集的主要目的在于测试所提出 NL-LML 方法是否可以处理非线性分布的样本。至于其他两个人体检测的测试集合，主要是验证 NL-LML 处理高维空间中人体样本的分类性能。

4.5.1. IRIS数据集

IRIS 数据集包含三种花，分别称做 *Iris Setosa*, *Iris Versicolor* 和 *Iris Virginica*。每种花有 50 个样本，每个样本的特征向量维度是 4 维，分别描述了花的四种属性即萼片长度、

萼片宽度、花瓣长度、花瓣宽度。将 IRIS 数据集中随机选择的 60 个样本作为训练集，其余的 90 个样本作为测试集。另外首先选择训练样本集中的一种花作为正例样本，其余两种花作为反例样本，这样一共要进行 3 次分类器的训练。本章并未对 IRIS 数据集的样本做任何预处理，直接使用其 4 维的特征向量。本章比较了线性 SVM、多项式核函数 SVM、径向基核函数 SVM、L-LML 算法与 NL-LML 算法的性能，在表 4-1 中显示了各种算法的分类误差率。简单起见，记 Iris Setosa 这类花的标号是 1，Iris Versicolor 花的标号是 2，Iris Virginica 花的标号是 3。当分类 Iris Setosa 和另外两种花时，Iris Setosa 为正例样本，其余两种为反例样本，简记为 IRIS 1V23，同理当分类 Iris Versicolor 和其余两种花时，Iris Versicolor 为正例样本，其余两种为反例样本，简记为 IRIS 2V13，依此类推。

可以看见 NL-LML 方法优于线性 SVM 和 L-LML。当分类 IRIS 2V13 时，无论采用哪种核函数，NL-LML 方法要优于核 SVM 方法。当分类 IRIS 3V12，核函数 SVM 方法要优于 NL-LML 方法。因此本章只能得出结论说在 IRIS 数据集上，NL-LML 取得和 SVM 方法相当的性能，可以实现非线性的数据分类。

表 4-1. IRIS数据集上NL-LML性能

组合模式	核函数	d	L-LML、NL-LML	SVMs
IRIS 1V23	Linear	4	0.0000	0.0000
IRIS 2V13	Poly	10	0.0444	0.0556
IRIS 2V13	RBF $\sigma=1.0$	10	0.0112	0.0223
IRIS 3V12	Poly	10	0.0223	0.0112
IRIS 3V12	RBF $\sigma=1.0$	10	0.0444	0.0223

4.5.2. 人体数据集

在 IRIS 数据集上验证完 NL-LML 算法是否能处理非线性数据之后，下面进一步验证 NL-LML 方法在人体数据集上的性能。本章使用大概 1100 个正例训练样本来自于 MIT[30] 和 SDL[28] 的训练样本集，大概 3800 个反例样本是从 INRIA 反例图片上选择出的。在第三章中显示了一些正例、反例训练样本，可以看到大部分的正例训练样本来自于前视角或者近似于前视角，但是 NL-LML 算法所得到的检测模型能够处理多姿态和多视角问题，请参见本章实验结果。两个测试数据集分别是 SDL 人体多视角测试集，该测试集有 140 副测试图片；另一个测试集是 INRIA 测试集，该测试集与第三章实验中用的测试集是同

一个集合。SDL140 副图片是在原有的 59 图片的 SDL 测试集中挑选了包含一些多姿态和侧视角的图片，另外又重新拍摄了一些多姿态和多视角的图片。SDL140 测试集目前和 INRIA 测试集相比，也非常有挑战性，其一是图片的背景很复杂多样、其二是新增加的图片涵盖了多视角、多姿态的人体。

因为多项式核函数与径向基核函数在 IRIS 数据集上的分类结果，性能差不多。所以在人体数据集上，本章只选择了径向基核函数为 NL-LML 算法中的核函数，并将其用于人体模式的高维特征处理。对于径向基核函数中的参数 σ ，当其从 0.01 到 10 进行变化时，考虑到训练误差，参数 σ 取 0.01 时获得的人体检测结果比较好。另外，非线性空间的维数，也即特征向量组成矩阵 V 的列维数 d 是根据最大特征值与最小特征值的比例来确定的。确定准则是：在满足矩阵正定性的前提下，选择尽可能少的正交特征向量。

当进行人体检测时，需要对图像窗口进行多尺度的扫描，然后对每个图像窗口使用已经训练好的分类器进行判别。类似于前一章，Recall rate 和 False Positives Per Window (FPPW) 仍然作为评测算法性能的标准。本章使用开源的 LibSVM 实现的线性 SVM 和核 SVM。图 4-3 显示的是在 SDL140 测试集上的结果，图 4-4 显示在 INRIA288 测试集上的结果。从实验结果可以得知，所提的 NL-LML 方法在两个测试集合上优于第三章的 L-LML 方法，线性 SVM 方法及核函数的 SVM 方法。值得一提的是，NL-LML 方法在 IRIS 数据集上获得和核函数 SVM 方法相当的性能，但是在人体测试集合上，NL-LML 方法要稍好于核函数的 SVM 方法。其原因在于两点：首先，当一个人体特征向量处于高维的特征空间中，特征选择对于减少冗余和干扰信息起到非常重要的作用，而 NL-LML 方法在变换空间中的特征选择使其结果好于核函数 SVM 的原因之一。其次，在 IRIS 数据集上当分类 IRIS2v13 时，NL-LML 方法要比 SVM 好，这可能说 NL-LML 和核函数 SVM 算法比，两者可能都有各自适用的样本分布。

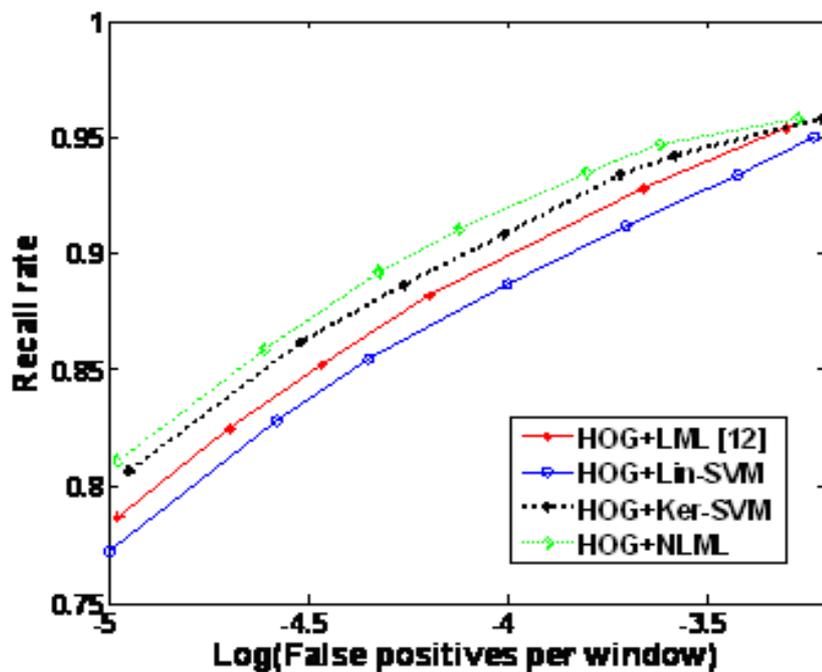


图 4-3.NL-LML 在 SDL 上性能比较

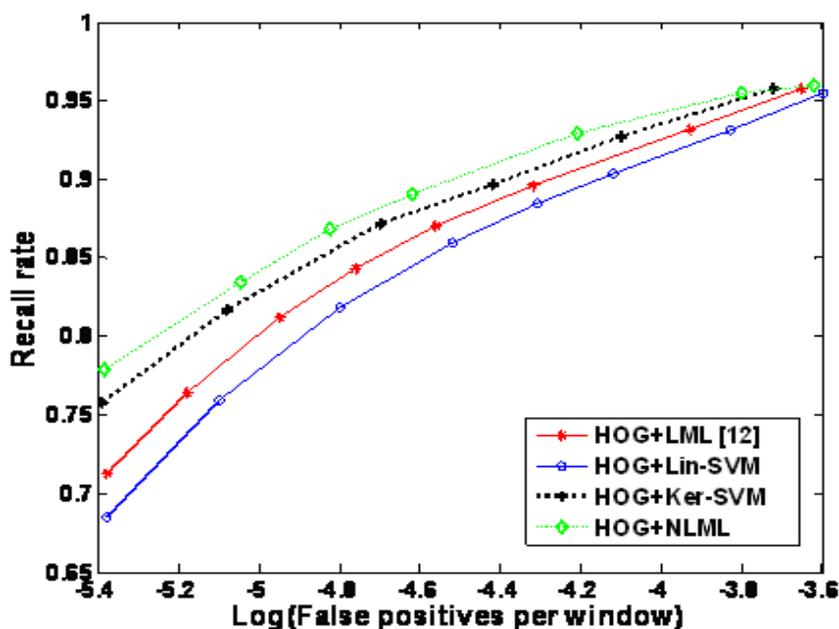
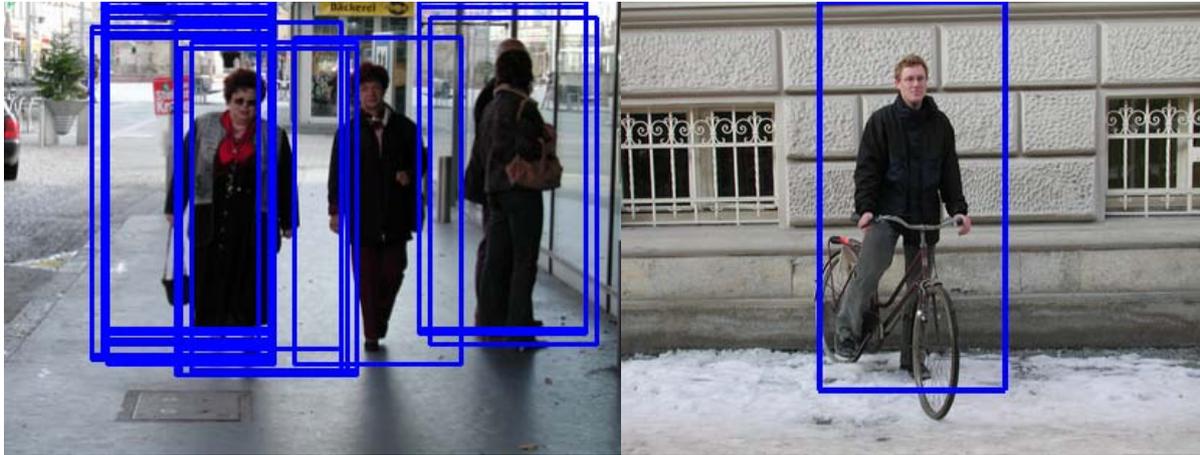


图 4-4.NL-LML 在 INRIA 上性能比较

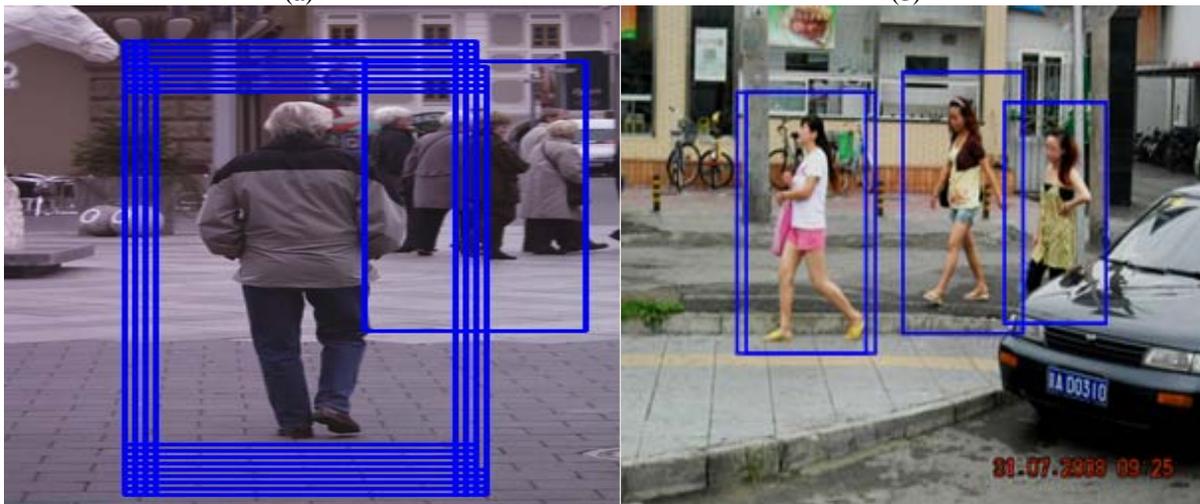
在图 4-5 中,展示了 NL-LML 算法的检测结果。从图 4-5(a)到图 4-5(f),尽管图片中的背景很复杂,并且存在人体的多种姿态、多种视角等多种影响因素,大部分人体能够被正确定位。在图 4-5(a)中,人体模式中有一些正视角的模式,还有一个侧视角的模式,处于所有视角模式的人体都能够被正确定位。在图 4-5(b)中,虽然骑自行车的人腿部姿态的变化,导致腿部特征信息的丢失,但是 NL-LML 算法仍检测到骑着自行车的人。在图

4-5(d)中, 虽然那个穿绿色衬衫的女孩被前面的小汽车遮挡, 但是那个女孩仍然被检测到了。在图 4-5(f)中, 在那位女士和婴儿车之间有一些虚假的错误窗口, 这些窗口覆盖了女士的部分姿态, 这些误检测可能是由于目标复杂的结构, 和一些干扰信息造成的。在图 4-5(h)中, 只检测到了三个小朋友, 那个大人被漏检了。这些误检和漏检问题以后将从特征的角度研究以避免这类问题的出现。



(a)

(b)



(c)

(d)



(e)

(f)

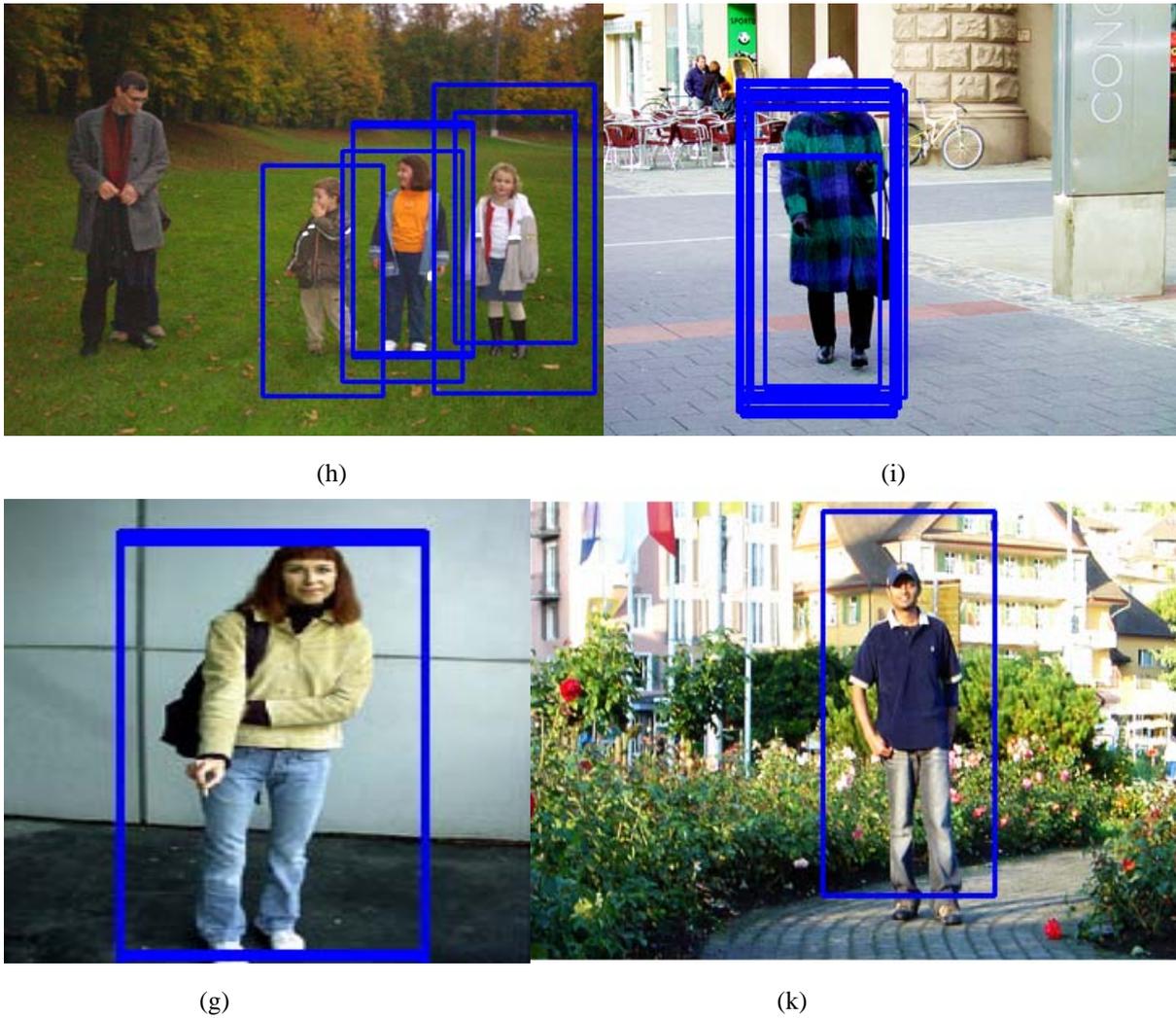


图 4-5.NL-LML 检测实例

4.6. 本章小结

本章提出了一种基于 1 范数最小化的非线性人体检测算法 (NL-LML)。本章首先介绍了 1 范数的固有性质使得其不能像 2 范数那样, 直接在模型中使用核技巧来实现样本的非线性判别。本章的工作, 主要从核函数作为特征变换的角度出发, 探讨了由核函数诱导的非线性变换与 LML 优化模型相结合的方法来实现非线性的判别。值得注意的是, NL-LML 算法在新的变换空间中, 人体模式的表示同样具有稀疏性。NL-LML 算法在 IRIS 数据集和人体数据集上分别得到了验证, 测试曲线说明了该算法的优越性能。最后本章还展示了 NL-LML 算法在人体检测集上的实例结果。

第五章 基于级联 1 范数最小化学习的人体检测

本章提出了一个新的基于 1 范数最小化学习的级联(CLML)人体检测方法。CLML 方法构建了一系列的强分类器,所有的强分类器是串行的级联结构。在级联的每一级中,CLML 方法采用加权投票机制,每个弱分类器充当投票者的角色,一些弱分类器的线性加权组合构成一个强分类器。该方法融合了块状特征的选择与分类器的设计,在其训练中无论是弱分类器的设计,还是强分类器的形成,其依据是统计学习理论中的 VC 维最小化这一原则。

在本章,弱分类器的学习阶段使用 LML 优化模型和 min-max 罚函数模型分别用来构建弱分类器的法向量和阈值,然后详细地介绍了 min-max 罚函数模型的来源、与对策论之间的关系及相应的解法。在强分类器的学习阶段,本章提出整数规划优化模型来学习强分类器,这相当于 LML 优化模型在整数空间中的重新描述,随后介绍了关于整数规划模型的求解方法。最后,结合级联机制,本章构建了一个级联的 LML 分类器。

5.1. 弱分类器的构建

在本章,人体模式的特征表示是块状的 v-HOG 特征。v-HOG 特征,已经在 2.1.3 节中描述过。简单来说,v-HOG 特征的每个特征块的长宽比一般为 1:1、1:2、2:1,每个块仍由 2x2 个 cell 组成,每个 cell 里面统计的是一个 9 维的梯度方向直方图,那么 4 个 cell 最后得到块状特征的维数是 36 维。每个弱分类器对应一个 v-HOG 块状特征,因此对弱分类器的选择,即可以实现对特征块的选择,这一点与 Adaboost 算法中的相同。每个弱分类器是一个线性超平面(线性分类器),第 k 个弱分类器的形式如下:

$$g_k(x) = \text{sign}(h_k(x) - \theta_k) = \text{sign}(w_k^T x - \theta_k) \quad (5.1)$$

从式(5.1)可知,第 k 个弱分类器的未知参数有两个,一个是法向量 $w_k \in R^n$,另外一个为阈值 $\theta_k \in R$ 。对于法向量 w_k ,使用第三章提出的 LML 优化模型去求解,相应的训练样本是块状特征。至于阈值 θ_k ,本章使用 min-max 罚函数优化模型来求解,下面详细介绍下该优化模型以及使用其作为阈值优化准则的目的。

为了满足一个高的检测率,若采用优化模型 LML 训练得到的阈值,会使得在级联训练过程中正例样本丢失。而且,如果采用 Adaboost 算法中通过观察每一维特征的特征值分布,结合最小误差来判断该特征的阈值,再结合所有正反例样本最后确定一个最优

的阈值，这种方式一是比较繁琐，二是也不是很准确。因此本章引入 min-max 罚函数模型。

5.1.1. min-max 罚函数的来源

Min-max 罚函数模型来源于对策论（博弈论）。对策论（Game theory）[58]属于数学的一个分支——运筹学的研究内容，主要是由冯若依曼与奥斯卡·摩根斯特恩等人将现实中很有趣的经济生活问题转换成明确的数学问题。对策论是研究两个或两个以上参加者在对抗性或竞争性局势下权衡利弊并采取行动，做出有利于己方决策的数学理论。下面简单描述下，对策论中涉及到的一些基本术语：

(1)局中人：在一场竞赛或对策中，每一个有决策权的参与者称为一个局中人。只有两个局中人的对策称为“二人对策”，而多于两个局中人的称为“多人对策”。

(2)策略：一局对策中，每个局中人都拥有选择实际可行的、完整的行动方案，一个局中人的一个可行的行动方案，称为这个局中人的一个策略。如果在一个博弈中局中人都总共有有限个策略，则称为“有限博弈”，否则称为“无限博弈”。所有的可供选择的方案组成策略集。

(3)得失（支付）：一局对策结束时，局中人的结果称为得失。每个局中人在一局对策结束时的得失，不仅与该局中人自身所选择的策略有关，而且与所有局中人选定的一组策略有关，即一局博弈结束时每个局中人的“得失”是全体局中人所选定的一组策略的函数。

(4)对策均衡：均衡是平衡的意思，是指某一稳定的对策结果。

生活中，常见的对策论模型主要是矩阵对策，该对策模型简单直观，其实质是一种离散策略，从而所形成的得失可表示为矩阵形式。为了帮助读者理解，本章举个简单的矩阵对策的例子。

假设一个游戏中，两个局中人（参加者，分别称为 1 和 2），设局中人 1 有 m 个策略，局中人 2 有 n 个策略。设局中人 1,2 分别选择策略 i 和 j 时，局中人 1 从局中人 2 处得到的支付是 a_{ij} ，则支付矩阵是 $A = (a_{ij})_{m \times n}$ 。其中 a_{ij} 是任意实数。局中人 1 得到的支付是 a_{ij} ，则局中人 2 得到的支付是 $-a_{ij}$ ，这种对策称为“二人零和矩阵对策”。局中人 1 希望得到的支付 a_{ij} 越大越好，而局中人 2 则希望得到的支付 $-a_{ij}$ 越大越好，若局中人 1 选择一个策略 i ，其最少可以得到的支付为 $\min_{1 \leq j \leq n} a_{ij}$ ，局中人 1 希望其最少的支付越大越好，即其优化目标是：

$$\max_{1 \leq i \leq m} \min_{1 \leq j \leq n} a_{ij} \tag{5.2}$$

同理，如果局中人 2 选定一个策略 j ，那么局中人 1 从他这得到的最大支付是 $\max_{1 \leq i \leq m} a_{ij}$ ，他希望其最大的支付越少越好。即

$$\min_{1 \leq j \leq n} \max_{1 \leq i \leq m} a_{ij} \quad (5.3)$$

所以能否找到一个理论上的“解”或者“平衡”，使得双方的利益或者所得都是最“合理”、最优的具体策略？冯若依曼证明的最大最小值定理说明，对于零和对策中，无论是纯策略还是混合策略，双方的最合理的解，称为“鞍点”，即有如下等式成立：

$$\max_{1 \leq i \leq m} \min_{1 \leq j \leq n} a_{ij} = \min_{1 \leq j \leq n} \max_{1 \leq i \leq m} a_{ij} \quad (5.4)$$

5.1.2. min-max 罚函数模型

本章所建立的 min-max 模型是连续对策的一种，因为每个策略变量是连续值，并且此处的对策已不再是零和对策，即正例的损失值与反例的获得值不一定相等，所以此处追求的目标与公式 (5.4) 不一样。虽然不能保证鞍点的存在，但是此处的 min-max 模型仍然符合对策论的含义。下面正式提出用以确定弱分类器阈值的 min-max 函数模型。

Model II:

$$\begin{aligned} \min_{\theta_k} & \left(r_1 \left(\sum_{pos=1}^{PN} \max\{0, \theta_k - h_k(x_{pos})\} \right) \right. \\ & \left. + r_2 \left(\sum_{neg=1}^{NN} \max\{0, h_k(x_{neg}) - \theta_k\} \right) \right) \end{aligned} \quad (5.5)$$

其中 $\theta_k \in R$ 是第 k 个弱分类器的阈值，也是式 (5.5) 中唯一一个变量。 x_{pos} 表示正例样本的特征向量， x_{neg} 表示反例样本的特征向量。 PN 和 NN 分别表示正例样本和反例样本的数目。样本的总数 $N = PN + NN$ 。另外 $h_k(x) = w_k^T \cdot x$ ，即其等于第 k 个弱分类器的法向量 w_k 与样本 x 的内积。另外 $r_1 \in R$ ， $r_2 \in R$ 是两个惩罚因子。

函数 $\max\{0, t\} = \begin{cases} 0 & \text{如果 } t < 0 \\ t & \text{其他情况} \end{cases}$ ，它是为了保证其所获得的函数值非负。

首先本章证明 Model II 和对策论中的函数优化是等价的。对于 Model II，首先对正反例两项做等价的分解：

$$\min_{\theta_k} r_1 \left(\sum_{pos=1}^{PN} \max\{0, \theta_k - h_k(x_{pos})\} \right) + \min_{\theta_k} r_2 \left(\sum_{neg=1}^{NN} \max\{0, h_k(x_{neg}) - \theta_k\} \right) \quad (5.6)$$

对于公式(5.6)中第二项， $\min_{\theta_k} r_2 \left(\sum_{neg=1}^{NN} \max\{0, h_k(x_{neg}) - \theta_k\} \right)$

$$\Leftrightarrow \min_{\theta_k} -r_2 \left(\sum_{neg=1}^{NN} \min\{0, \theta_k - h_k(x_{neg})\} \right)$$

$$\Leftrightarrow \max_{\theta_k} r_2 \left(\sum_{neg=1}^{NN} \min\{0, \theta_k - h_k(x_{neg})\} \right)$$

所以公式(5.6)等价于下面的式子:

$$\min_{\theta_k} r_1 \left(\sum_{pos=1}^{PN} \max\{0, \theta_k - h_k(x_{pos})\} \right) + \max_{\theta_k} r_2 \left(\sum_{neg=1}^{NN} \min\{0, \theta_k - h_k(x_{neg})\} \right) \quad (5.7)$$

公式(5.7)中的两部分的求和项, 分别是 min-max 与 max-min 模型, 也即是对策论中对策双方所追求的优化目标。

下面本章从对策论的角度来解释 min-max 罚函数模型为什么适合于求解弱分类器的阈值。首先假设对策中有两个局中人, 一个是“正例样本方”, 一个是“反例样本方”。每一方中有很多个“代表”, 也即对应着每一方中的训练样本, $\max\{0, t\}$ 可以看做是双方的“损失”, 每个样本都希望其损失越小越好。那么模型 Model II 的实际含义是: 目标函数是求最小值, 并且每个 max 函数保证里面的取值是非负的, 所以目标函数最小值的理想状态是 0。正反双方都希望使得各自的目标越接近 0 越好, 也即损失 $\max\{0, t\}$ 越小越好, 所以每一个正例样本希望阈值尽量的小, 使得 $h_k(x_{pos}) \geq \theta_k$, 原因在于当 $h_k(x_{pos}) \geq \theta_k$ 时, $\max\{0, \theta_k - h_k(x_{pos})\} = 0$ 。而反例样本希望阈值尽量的大, 这样很容易使得 $h_k(x_{pos}) < \theta_k$, 即反例方也追求 $\max\{0, h_k(x_{neg}) - \theta_k\} = 0$, 所以局中的两方都希望自己的损失越小越好。所以该过程是正反例样本一起动态地对策, 从而确定最后对双方都尽量有利的阈值。

正例损失, $\max\{0, \theta_k - h_k(x_{pos})\}$ 实际表示的是正例样本的错分程度, 当 $h_k(x_{pos}) \geq \theta_k$ 时, 表示所选择的阈值使得正例样本被正确分类, 错分程度是 0。当 $h_k(x_{pos}) < \theta_k$ 时表示所选择的阈值使得正例样本错误分类。同理, $\max\{0, h_k(x_{neg}) - \theta_k\}$ 表示反例样本被错分的程度。在实际应用中, 在静止图片或者视频中, 包含人体的窗口数目相对于未包含人体的窗口数目, 是非常稀少的, 也即图片中背景或者其他非人体目标的窗口要远多于人体窗口。而且, 在训练样本中人体样本与非人体样本的比例也是不对称的。因此, 可见阈值的变动对于人体模式的影响非常敏感。为了能减少检测过程中人体的丢失率, Model II 引入了两个惩罚因子, 用来控制阈值对于正例样本的影响。如果模型更注重由于正例样本分类所产生的误差, 那么可以设置罚因子 $r_1 > r_2$ 。这样, 如果一个正例样本被错分了, 那么他所造成的损失将远大于一个反例样本被错分的损失, 为使得 Model II 中总体目标函数最小, 那么必须调整阈值使得正例样本被较少的错分。相反地, 如果模型希望反例样本尽量被分对, 那么可以设置罚因子 $r_1 < r_2$ 。关于两个罚因子如何设置, 请

参考本章实验部分显示的关于两个罚因子比例的曲线。

优化模型 Model II 是一个无限制约束的凸规划问题，可以通过引入几个变量将其转换为线性规划问题，具体转换形式如下：

令 $v_{k,pos} = \max\{0, \theta_k - h_k(x_{pos})\}$ 和 $u_{k,neg} = \max\{0, h_k(x_{neg}) - \theta_k\}$ 。它们分别等价于下面的两个约束：

$$\begin{cases} v_{k,pos} \geq 0 \\ v_{k,pos} \geq \theta_k - h_k(x_{pos}) \end{cases} \quad \begin{cases} u_{k,neg} \geq 0 \\ u_{k,neg} \geq h_k(x_{neg}) - \theta_k \end{cases} \quad (5.8)$$

所以 Model II 和下面的优化模型等价

$$\begin{aligned} \min \quad & (r_1 (\sum_{pos=1}^{PN} v_{k,pos}) + r_2 (\sum_{neg=1}^{NN} u_{k,neg})) \\ \text{s.t.} \quad & \begin{cases} v_{k,pos} \geq 0 \\ v_{k,pos} \geq \theta_k - h_k(x_{pos}) \\ u_{k,neg} \geq 0 \\ u_{k,neg} \geq h_k(x_{neg}) - \theta_k \end{cases} \end{aligned} \quad (5.9)$$

可以看见上述的优化模型是变量 $v_{k,pos}$ ， $u_{k,neg}$ 和阈值 θ_k 三个变量的线性规划形式，因此可以使用一些基本的线性规划方法（如内点法，单纯性法等）去求解，从而得到阈值 θ_k 。最后得到弱分类器：

$$g_k(x) = \text{sign}(h_k(x) - \theta_k) = \text{sign}(w_k^T x - \theta_k) \quad (5.10)$$

5.2. 强分类器的构建

每一个 36 维的 v-HOG 块状特征，对应着一个弱分类器。因此弱分类器的数目与块状特征的数目一致。考虑到特征表示中存在的冗余信息和大量特征块产生较高计算费用，让所有的弱分类器都参与强分类器的构建，是不明智的选择，其中这也与文献[27]中关于分类器构建原则的论述“许多要比全部好”相一致。本章提出使用整数规划的方法来选择最少数目的弱分类器形成强分类器。

首先，强分类器与弱分类器之间的关系如下式：

$$G(x) = \begin{cases} 1 & \sum_{k=1}^M a_k (\lambda_k g_k(x) - 0.5) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.11)$$

其中 a_k 表示第 k 个弱分类器的权重， $a_k = \log[(1 - \varepsilon_k) / \varepsilon_k]$ ， ε_k 是第 k 个弱分类器的训练误差。 λ_k 是一个 0/1 的二值变量，对应着第 k 个弱分类器是否被选择。若 $\lambda_k = 1$ 则表示第 k 个弱分类器被选择参与强分类器的构建，若 $\lambda_k = 0$ 则表示第 k 个弱分类器没有被选中参与强分

类器的构建。

5.2.1. 整数规划模型

在上述变量定义的基础上，本章构建一个优化模型 Model III 来确定哪些弱分类器被选择，哪些不被选择。

Model III:

$$\begin{aligned} \min_{\lambda_k, \eta_i} \quad & \sum_{k=1}^M \lambda_k + C_2 \sum_{i=1}^N \eta_i \\ \text{s.t.} \quad & \begin{cases} A\lambda \leq r_3 \cdot \eta \\ \sum_{i \in \text{pos}} \eta_i \leq \sigma_1 \cdot PN \\ \sum_{i \in \text{neg}} \eta_i \leq \sigma_2 \cdot NN \\ \lambda_k, \eta_i = 0/1 \end{cases} \end{aligned} \quad (5.12)$$

其中 η_i 也是一个 0/1 二值变量，它对应着第 i 个训练样本。如果第 i 个训练样本能够被所选择的弱分类器组合正确分类，那么 $\eta_i = 0$ 。若 $\eta_i = 1$ 则表示第 i 个训练样本被所选择的弱分类器组合错分。 M 是弱分类器的数目， C_2 是一个预定义的参数。 N 是样本的总数， $N = PN + NN$ 。 PN 代表正例样本的总数， NN 代表反例样本的总数。

为了便于书写和简洁的表述，此处引入向量的表述方式。引入包含所有二值变量 λ_k 的一个齐次向量 $\lambda = [\lambda_1, \dots, \lambda_k, \dots, \lambda_M, 1]^T$ 。同理，针对所有的二值变量 η_i 引入齐次向量 $\eta = [\eta_1, \dots, \eta_{PN}, \dots, \eta_N]^T$ 。矩阵 A 中的第 i 行元素们对应着第 i 个样本是否被弱分类器的组合正确分类。根据样本是属于正例样本还是反例样本，矩阵 A 中的第 i 行元素有所区别。

$$(A_i) = \begin{cases} (-a_1 g_1(x_i), \dots, -a_k g_k(x_i), \dots, -a_M g_M(x_i), 0.5(a_1 + \dots + a_M)) & \text{如果正例样本} \\ (a_1 g_1(x_i), \dots, a_k g_k(x_i), \dots, a_M g_M(x_i), -0.5(a_1 + \dots + a_M)) & \text{如果反例样本} \end{cases}$$

如果第 i 个正例样本，并且同时其被 $G(x)$ 正确分类，则相应的变量 η_i 取 0 值可以使目标函数最小，相应地 A 中的第 i 行与齐次向量 $\lambda = [\lambda_1, \dots, \lambda_k, \dots, \lambda_M, 1]^T$ 相乘的结果即是：

$$\sum_{k=1}^M -a_k (\lambda_k g_k(x_i) - 0.5) \leq 0 \quad (5.13)$$

变换下不等式的符号有

$$\sum_{k=1}^M a_k (\lambda_k g_k(x_i) - 0.5) \geq 0 \quad (5.14)$$

相反地，若第 i 个正例样本没有被 $G(x)$ 正确分类，那么矩阵 A 中的第 i 行与齐次向量 $\lambda = [\lambda_1, \dots, \lambda_k, \dots, \lambda_M, 1]^T$ 相乘的结果即是：

$$\sum_{k=1}^M -a_k (\lambda_k g_k(x_i) - 0.5) \leq r_3 \quad (5.15)$$

因为此时正例样本的值不再严格大于等于 0, 此时 η_i 不可能再等于 0, 因为这样方程组必定无解, 所以 η_i 的值是 1。另外, 由于式 (5.15) 左边这个值有可能是大于 1 的, 所以严格要求式 (5.15) 左边的值小于 1 不合理, 因此此处引入松弛因子 r_3 , 其作用是保证被错分的样本并且式 (5.15) 左边的值大于 1 时, 不等式组仍有可行解。同理, 对于反例样本, 上面的推导仍然成立, 只是符号方向相反而已。

上述的解释也即是不等式 $A\lambda \leq r_3 \cdot \eta$ 的含义。 σ_1 是最大的正例误分率, σ_2 是最大的反例误分率, 换句话说, $\sigma_1 \cdot PN$ 是被错分正例样本数目的上界, $\sigma_2 \cdot NN$ 是被错分反例样本数目的上界。在满足上面约束的条件下, Model III 中的目标函数的含义是: 尽量地选择让 λ, η 中的非零分量尽可能地少, 也即希望尽可能选择尽量少的弱分类器参与强分类器的构建。 c_2 是一个平衡因子, $c_2 > 1.0$ 表示对错分的样本数目和弱分类器数目同时最小化时, 应该更加重视错分的样本数目。因为向量 λ, η 中元素的取值要么为 0, 要么为 1, 所以下式(5.16)中左边和右边等价:

$$\sum_{k=1}^M \lambda_k + C_2 \sum_{j=1}^N \eta_j \Leftrightarrow \sum_{k=1}^M |\lambda_k| + C_2 \sum_{j=1}^N |\eta_j| \quad (5.16)$$

进一步地, 目标函数 $\sum_{k=1}^M |\lambda_k| + C_2 \sum_{j=1}^N |\eta_j|$ 可以变换成整数空间中 1 范数的表达式 $\|\lambda\|_1 + C_2 \|\eta\|_1$ 。 $\|\lambda\|_1$ 和 $\|\eta\|_1$ 分别表示向量 λ 和 η 的 1 范数, 因此, 目标函数的最小化可以看成是整数空间中 1 范数最小化的特例, 这样, 所建立的模型可以统一到 VC 维最小化的原则下。

5.2.2. 整数规划模型的求解

Model III 是一个典型的 0/1 整数规划问题, 其中目标函数和约束都是线性的, 因此这是一个线性的整数规划问题。现实生活中很多问题都属于整数规划问题, 如旅行商问题、选址问题、背包问题等, 所以对整数规划的研究比较早。目前其解法[58]有很多种, 比如说分支定界法、0-1 规划的隐数法、割平面法等。本章使用分支定界方法找到其最优解。

1954 年, G. B. Dantzing, D. R. Fulkerson 和 S. M. Johnson 等人在研究旅行商问题时, 首先提出了将问题分解成几个子问题之和的思想, 这是整数规划中分支定界方法的萌芽。后来整数规划方法由 Land, Doiog, Little, Balas 等人提出。下面介绍下分支定界方法的原理, 首先阐述下算法中三个基本概念。

1) 分解:

设对整数规划问题 P , 记其可行解集合为 $T(P)$, 若满足下述两个条件, 则称问题 P 可分解为子问题 P_1, P_2, \dots, P_t 之和:

$$\bigcup_{i=1}^t T(P_i) = T(P) \quad (5.17)$$

$$T(P_i) \cap T(P_j) = \Phi \quad 1 \leq i \neq j \leq t \quad (5.18)$$

2) 松弛

对整数规划问题 P ，若舍弃 P 的某些约束条件，所得到的问题 P' 称为原问题 P 的松弛问题。松弛问题和原问题一般有如下性质：

- 若 P' 没有可行解，则 P 也没有可行解。
- 对于求目标函数的最小值问题， P 的最小值大于等于 P' 的最小值。
- 若 P' 的一个最优解是 P 的可行解，那么它也是 P 的最优解。

3) 探测：

若按照某种规则，已经把问题 P 分解成子问题 P_1, P_2, \dots, P_t 之和，并且每个子问题 P_i 对应的松弛问题为 P_i' ，则有如下性质成立：

- 若 P_i' 没有可行解，则说明子问题 P_i 没有可行解。因此，可以从 P 的分解表上删除子问题 P_i 。
- 若已经得到 P 的一个可行解，其对应的目标函数值记为 f_P 。假设目标函数是求极小值问题，如果某个松弛问题所得到的最小值不小于 f_P ，则证明该松弛问题对应的子问题 P_i 没有更好的解，因此将其删除。
- 若 P_i' 的最优解是 P_i 的可行解，则已求得 P_i 的一个最优解，此时可以先暂停终止对 P_i 的继续分解。
- 如果各个 P_i' 的目标函数值都不小于 f_P ，那么对应于 f_P 的解便是原问题的最优解。

下面详细论述分支定界算法的基本思想：对有约束条件最优化问题的所有可行解（数目有限）空间进行搜索。该算法在具体执行时，把全部可行的解空间不断地分割为越来越小的子集（称为分支），并为每个子集内解的值计算一个下界或上界（称为定界）。在每次分支后，对凡是界限超出已知可行解值的那些子集不再做进一步分支。这样，解的许多子集（即搜索树上的许多节点）就可以不予考虑了，从而缩小了搜索范围。对搜索树上的节点，都有可能作为分枝的选择对象（对求最小值问题而言）。怎样选择搜索树上的节点作为下次分枝的节点呢？有两个原则：

1) 比较当前点的分枝限界法：每次算完界限后，把搜索树上当前所有叶节点的界限进行比较。找出限界最小的节点，此结点即为下次分枝的节点。

2) 优先队列式分枝限界法：从最新产生的各子集中选择具有最小下界的节点进行分枝。这种不断分割的过程一直进行，直到找出可行解为止，该可行解的值不大于任何子集的

界限。因此这种算法可以求得整数规划的最优解。

下图展示了某一级中人体特征选择的示意图。

$$\left(\begin{array}{c} \vdots \\ \vdots \end{array} \right) = \left(\begin{array}{cc} \dots & \dots \\ \dots & \dots \end{array} \right) \left(\begin{array}{c} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{array} \right)$$

图 5-1.特征选择示意图：左图显示所选择的 v-HOG 块，右图等式表明 0、1 整数规划的选择策略

5.3. 级联分类器的构建

级联机制最初是由 Viola 等人[7]提出的。Viola 等人为了实现快速目标检测，融入级联机制到分类器的构建中。级联机制可以理解为分阶段决策，每一级是一个阶段。从整体的结构来看，整个级联过程实质上是一棵退化了的决策树(只有一个分支)，如图 5-2 所示。通过第 1 级的图像窗口才进入第 2 级的判别中，依次类推，只要没有通过任何一级的图像窗口都被认为是反例样本。

分类器的级联形式，即每一级是一个分类器，所有的分类器串行连接组成最后的分类器。其优点在于它可以降低检测所需要花费的时间，并且可以提高检测的性能。它可以在前几级检测尽量保留正确实例的同时排除掉大量的无关窗口。因此可以在前几级使用少量的弱分类器组合排除大量的子窗口，在后几级再利用较复杂分类器的组合来降低误检率。

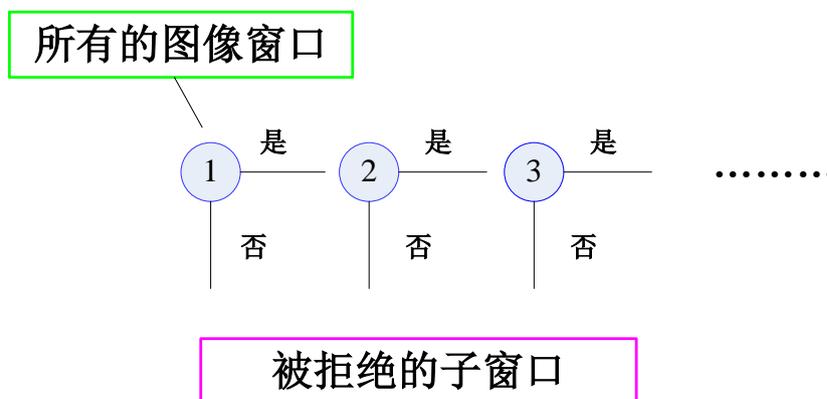


图 5-2.级联机制示意图

在本文作者看来，级联机制的想法其实是 Boosting 算法的一种演变。在 Boosting[59] 法中，首先根据已有的训练样本集设计一个分类器，然后依次地添加多个分量分类器，最后形成一个总体的分类器，它对训练样本集的准确率能够任意的高。下面考虑对于一个两类问题，举例说明 Boosting 算法是如何产生分量分类器的。首先，从大小为 n 的原始样本集 D 中随机抽取 n_1 个样本点，组成样本集 D_1 。根据样本集 D_1 ，训练得到第一个分类器记为 g_1 。现在继续构造第二个样本集 D_2 ，该样本集应该根据分类器 g_1 最富信息的那些样本点组成。更明确地说， D_2 中的样本应该能被 g_1 正确分类，而另一半则是被 g_1 错分。具体的构造方式如下：取 D 中剩余的样本点，并逐个地送入 g_1 进行分类，保留下被 g_1 正确分类的正例样本，保留下被 g_1 错分类的反例样本，这样得到的集合 D_2 便是由 g_1 最丰富的信息构成的。同理构造第三个训练样本集 D_3 时，其构造方式类似于 D_2 的构造，只不过这次挑选样本点的时候，要用 g_1 和 g_2 进行分类，若样本通过 g_1 和 g_2 的判别，即被认为是“正例”（包括虚假正例）样本，那么就把该样本加入集合 D_3 ，再用 D_3 训练第三个分类器 g_3 。比较上述的级联机制和 Boosting 分类器训练机制，本文发现其实两者的本质思想是一致的，但是本文作者仍然很佩服 Viola 等人看到这种本质关系，并且巧妙地将 Boosting 这种思想转化为检测过程中的级联机制。在 Viola 等人提出的级联分类器中，在每一级他们使用的是 AdaBoost 分类器。本章中，将使用所提的 LML 分类器和整数规划方法结合。下面介绍下如何训练级联分类器 CLML。

在每一级，一些弱分类器被选择形成强分类器，因此每一级最后的判别形式只有一个强分类器，并且在下一级，使用另一个强分类器去判别。经过前几级的样本有可能是正例样本，也有可能是反例样本，当然反例样本会越来越来少，那么应该重新补充反例样本（例如从其他图片中筛选一些反例样本）。这一操作其实相当于在原始样本集合 D 的剩余样本中抽取样本。

下面看一下级联机制的检测率与虚假率之间的关系：虚假率 (FPPW) 和检测率 (Recall Rate) 在第三章已经定义过，两者之间是矛盾的整体，随着虚假率的减少，检测率也减少。在训练过程中，假设已经得到一系列的级联分类器，设 f_i 表示第 i 级的虚假率， d_i 表示第 i 级的检测率，因为每一级之间是串行的运行过程，那么经过 K 级训练后，总的检测率和总的虚假率分别如下：

$$D = \prod_{i=1}^K d_i, \quad F = \prod_{i=1}^K f_i \quad (5.19)$$

如在每一级中都要求 $d_i = 0.99, f_i = 0.3$ ，那么经过 10 次级联训练之后， $D = 0.99^{10} \approx 0.9, F = 0.3^{10} \approx 6 \times 10^{-6}$ ，可以看出，随着级数的增加，检测率和虚假率都随着减少，所以有必要在检测率和虚假率之间取个折中值。

在前几级中，本章每一级的最小可接受率是 0.998，每一级的最大虚假率是不能超过 0.3。这与前面的 Model III 一致，相应地，设置参数 $\sigma_1 = 1 - 0.998 = 0.002$ ， $\sigma_2 \leq 0.3$ 。随着级数的增加，反例样本与正例样本越来越相似，那么越来越难以区分，所以虚假率可以相应地增大一些。最后的检测率和最后的虚假率都是随着级数的增加而减少，因此，通过平衡所接受的检测率和误检率之间的关系获得级数是非常有必要的举措。CLML 的训练过程如表 5-1 所示。

5.4. 算法讨论

这一部分将详细地探究为什么 1 范数最小化优化模型可以得到稀疏解，并且还将比较所采用整数规划算法的 CLML 分类器与级联 Adaboost 分类器的本质区别。

与 SVM 方法中使用 2 范数作为目标函数相比，1 范数对于获得特征的稀疏表示更加有效。某一向量的 2 范数是指该向量中每一维元素的平方和再取根号。目标函数是最小化向量的 2 范数的平方，则要求向量的每一维元素都应该比较小，如果要是某个分量很大，那么其平方之后的结果将是非常大的，这样得到的最小值也会很大。因此，本文认为向量的 2 范数最小化更加注重于向量中每一维元素的“平均变化”，也即在满足约束条件下，每一维元素都应该均匀地变化，这种结果不能产生稀疏解。然而，1 范数的性质恰好与 2 范数相反，因为它是向量中每一维元素的绝对值求和，绝对值可以允许某一维元素值很大，而其余的元素值很小，最小值仍然不变。因此，稀疏表示可以实现特征选择。综上所述 1 范数，而不是 2 范数，可以实现特征选择的功能。如图 5-3 所示，在二维空间中，本章给出一个几何解释来说明 1 范数与 2 范数求解过程的区别。

表 5-1. CLML 算法训练过程

-
- 输入：在第 i 级过程中，最小可接受的检测率 $\sigma_3 = 1 - \sigma_1$ ，最大可接受的虚假检测率 σ_2 。
 - POS：正例样本集合；
 - NEG：反例样本集合；
 - F_{target} ：所设定的总的虚假检测率；
 - f_t ：在第 t 级中虚假检测率的值；
 - D_t 在第 t 级中的检测率的值；
 - 初始化： $t = 0$, $F_0 = 1.0$, $D_0 = 1.0$
 - 当 $F_t > F_{target}$ 进行如下迭代

- 使用正例样本集和反例样本集训练弱分类器，并计算其相应的法向量和阈值。

- $t = t + 1, \Delta_t = 0$

- $\sigma_2 = 0.7 - \Delta_t$

如果 Model III 没有可行解

则增加 $\Delta_t = \Delta_t + 0.1$

否则

1. 求解整数规划 Model III;

2. 使用当前的强分类器估计正例样本集和反例样本集;

3. 在当前迭代的阈值下计算 f_t ;

- 结束循环

- $F_{t+1} = F_t \times f_t$

- $D_{t+1} = D_t \times \sigma_1$

- $NEG \leftarrow \emptyset$

- 估计当前的级联检测子在反例图像上的响应值，比如说：在一些不包含人体目标的图像上，有些样本被误当做人体样本的图像窗口，把这些选中的误分样本再次加入到反例集合 NEG 中

• 结束

• 输出：一个 t 级的级联分类器。

记半径是 α 的 L_p 范数为: $\mathbf{B}_{\alpha p} = \{w \in \mathbf{R}^n: \|w\|_p \leq \alpha\}$ 。在 Model I 中 (LML 模型), 约束函数是线性的, 因此这些约束函数构成一个凸多面体。这些约束构成的凸包也正是可行解的区域。从几何的角度寻找 L_p 范数的最小值, 等价于不断地将 L_p 范数扩张, 直到 L_p 范数的边界接触到可行解的区域, 此时找到一个可行解使得目标函数达到极小值。又因为无论是 LML 优化模型还是 SVM 中所采用的优化模型, 都是凸规划模型, 由凸规划模型的性质可知此时已经求出模型最优解。图 5-3 中可以看见 2 范数达到最小时的最优解, 因为 2 范数的形状是圆形, 那么它与可行解区域边界相交的切点一般都不会位于坐标轴上, 除非可行解的区域边界是与坐标轴平行或者垂直的。然而, 因为 1 范数的形状是菱形, 所以其边界逐渐扩张的时候, 最先接触到可行解区域的点, 要么是落在某个坐标轴上, 要么是其菱形的一个边界与可行区域的边界完全重合。无论哪种情形, 1 范数最小化都会选择落在坐标轴上的点作为最优解。很明显, 坐标轴上的点是具有稀疏性的, 因为其中一维分量是 0, 相当于二维空间中的点退化到坐标轴上了。因此, 从几何角度上

来说，1 范数比 2 范数更容易达到稀疏性。

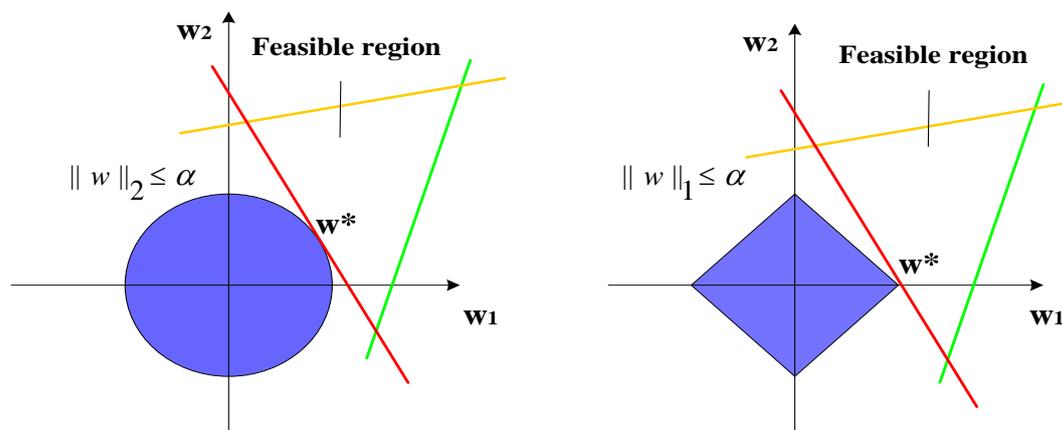


图 5-3.1 范数与 2 范数的几何解释。横轴表示向量 w 的第一维坐标，纵轴表示向量 w 的第二维坐标

值得注意的是，1 范数产生的稀疏特征可以减轻由于遮挡和视角变化导致人体模式部分信息丢失等影响。因此，所提的方法对于视角和遮挡变化不敏感，这一结论已经在实验结果中得到了验证。

CLML 算法与 Adaboost 算法区别在于：

1). 弱分类器的构建方式不同

Adaboost 弱分类器训练的时候，阈值的选取有些复杂，它是通过观察正例样本与反例样本特征值的分布，从而依次计算阈值在哪个区间值范围内，然后再计算每个阈值得到的误差函数值，通过误差函数值的大小进行排序，最后为一个弱分类器选择一个恰当的阈值，这一操作需要对所有的特征值进行一次扫描排序，比较繁琐。CLML 算法使用 LML 优化模型和 min-max 罚函数模型建立弱分类器，采用 min-max 函数计算弱分类器的阈值，能够简便和灵活地获得阈值，比较适合级联机制。

2). 特征选择与强分类器构建方式不同。

Adaboost 算法使用局部贪婪的迭代算法，每次选择一个误差最小的弱分类器加入到强分类器中，这样依次迭代的进行，直至满足要求为止。它的误差是样本权重与损失代价两者的函数（参见表 2-4）。Adaboost 使用重采样策略调整每个样本的权重，然后计算每个弱分类器所引起的误差，从中选择误差最小的那个弱分类器加入到强分类器的组合中，再次调整样本的权重，使得被错分的样本权重变大。CLML 采用整数规划的方法选择弱分类器，尝试从另一个角度选择特征，意在从全局的角度选择最少的分类器形成强分类器，而不是使用局部贪婪的策略。在训练过程中，全局角度和局部贪婪选择角度策略相比，全局的角度时间比较多。然而在训练过程中时间代价的消耗如果能提高在检测时

的精确度和速度，那么这些训练中的耗费应该是值得的。

尽管 Adaboost 算法已经证明其采用的贪婪迭代策略最后会收敛，也即误差的上界在概率意义下是存在的，然而，因为它毕竟是局部搜索算法，所以它不一定总是能获得全局最优的特征组合。换句话说，Adaboost 算法可能会变得不稳定，并且可能会选择更多的冗余特征来达到最后可接受的性能。下面本章举一个简单的异或分类例子来说明 CLML 算法与 Adaboost 算法最明显的区别。而且从这个例子中，可以看出 CLML 算法是致力于选择最少的弱分类器组合形成强分类器，这种方式能满足快速检测的要求和适合级联机制。除此之外，CLML 算法是全局的线性整数规划模型，它的解只有两种情况，一种是无解，一种是有可行解，通过分支定界算法可以得到其最优解，所以算法是收敛的，没有必要证明在训练过程中误差的上界。

异或 (XOR) 分类问题：

首先在一个 2 维空间中，假设有四个样本点，分别属于两类，其分布呈异或的形式，即在对角线方向上的样本点是属于同一类，如图 5-4 所示。分别假设这些样本点的位置以及标号如下：

$$\begin{cases} (x_1 = (+1, +1), y_1 = +1) \\ (x_2 = (-1, -1), y_2 = +1) \\ (x_3 = (-1, +1), y_3 = -1) \\ (x_4 = (+1, -1), y_4 = -1) \end{cases} \quad (5.20)$$

简单起见，设有四个弱分类器（在图 5-4 中显示为四个决策函数），它们的表达形式如下：

$$g_1(x) = \begin{cases} 1 & \text{如果 } x+1 > 0 \\ -1 & \text{其他情况} \end{cases} \quad (5.21)$$

$$g_2(x) = \begin{cases} 1 & \text{如果 } x-1 < 0 \\ -1 & \text{其他情况} \end{cases} \quad (5.22)$$

$$g_3(x) = \begin{cases} 1 & \text{如果 } -x+1 > 0 \\ -1 & \text{其他情况} \end{cases} \quad (5.23)$$

$$g_4(x) = \begin{cases} 1 & \text{如果 } -x-1 < 0 \\ -1 & \text{其他情况} \end{cases} \quad (5.24)$$

从样本的分布、样本的类别标号和决策函数形式来看，最后强分类器的形式是由任意两条平行的决策函数组成的，而且这种组合形式是最优的。当然，如果选择四个的判别函数全部参与形成强分类器的构建，最后也能实现将样本全部分开的结果，但是这种

失去了特征选择的意义，也违背了特征选择的初衷。因此应该选择尽量少的分类器来形成最后的强分类器，实现样本降维。下面来看一下 Adaboost 算法的选择过程。

Adaboost 算法在第一次迭代时，所有样本的权重都相等，那么四个分类器的分类误差是相同的，这样可以随机选择一个弱分类器，比如说选择弱分类器 $g_1(x)$ ，这使得样本 x_4 被错分，和其他样本相比 x_4 的权重变大（实际上被错分的样本权重不变，正确分类的样本权重减少。经过权重归一化之后，相当于错分的样本权重变大），其余的三个样本权重一样。继续下一次的迭代过程，因为三个样本有相同的权重，那么其余三个弱分类器的分类误差也相同，此时可以任意选择一个分类器作为第二次迭代产生的结果。如果选择分类器 $g_2(x)$ ，那么 $g_1(x)$ 和 $g_2(x)$ 的组合可以实现对所有样本的正确划分，并且该种组合也是最优解，结束迭代。但是如果 Adaboost 算法选择 $g_3(x)$ ，那么分类器 $g_1(x)$ 和 $g_3(x)$ 的组合不能将所有的样本划分正确，不幸地是，要想实现对所有训练样本划分正确，那么必须还得继续迭代两次，即得选择弱分类器 $g_2(x)$ 和 $g_4(x)$ 才能实现对训练样本分布的正确把握。从上述例子可以看出，Adaboost 算法因其采用局部贪婪选择的策略，虽然每次迭代选择的一个弱分类器在当前次迭代是最优的，但是并不代表以后是最优的，很可能导致选不出最优分类器的组合形式，所以 Adaboost 算法可能变得不稳定。

相反地，因为所提的 CLML 方法是从全局的角度选择最少的弱分类器数目，从全局的角度来看，CLML 方法很容易选择两条平行判别面的组合形成最后的分类器，因此 CLML 方法可以最优地选择“完全互补型”弱分类器的组合来实现强分类器的构建。

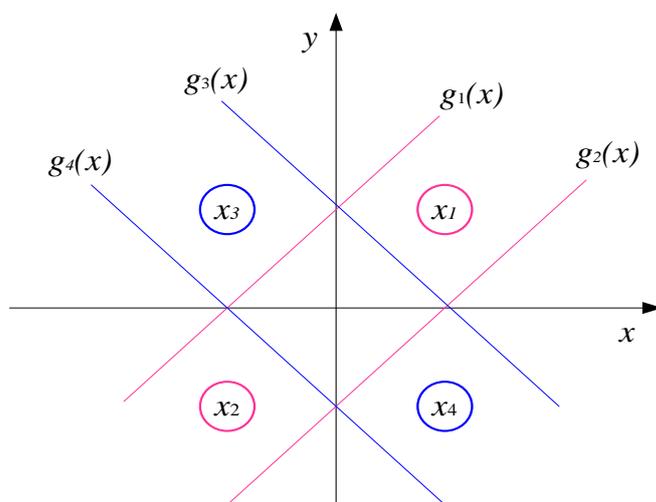


图 5-4. 异或样本分布。有着相同颜色的样本是属于同一标号类的，任意两条相同颜色的判别函数是最优的组合形式。

5.5. 实验结果

本章实验中选择 2400 幅正例图像和 4900 幅反例图像作为训练样本。正例样本主要选自 MIT[30]和 SDL[28]的训练集，反例样本主要是选自 INRIA 的反例数据集。随着级联过程的进行，一些额外的反例样本需要补充到分类器的训练过程中。

本章使用两个人体测试集来验证 CLML 算法的性能，一个是 INRIA 数据集，另一个是 SDL140 数据集，这与第四章的测试数据集合是一致的。下面主要说明实验过程中一些参数的值是如何选择、展示算法的性能对比曲线和检测结果等。尽管所选择的正例样本是来自于前视角和接近于前视角，由于稀疏性的影响，所训练的模型对一些多视角、遮挡的人体模式比较鲁棒，本章的实验结果支持这一论断。

5.5.1. 参数分析

为了处理由于光照不均、外表服饰多样性及背景的复杂性，需要对样本的特征向量和弱分类器的法向量等做归一化处理，归一化的形式如下所示：

$$x_i = x_i / \sqrt{\sum_{j=1}^{36} x_{ij}^2 + \varepsilon} \quad w_k = w_k / \sqrt{\sum_{j=1}^{36} w_k^j + \zeta} \quad (5.25)$$

x_i 表示第 i 个训练样本对应的一个特征块， x_{ij} 表示该特征块的第 j 维特征值。 w_k 表示第 k 个弱分类器的法向量， w_k^j 表示第 k 个法向量的第 j 维的权值。 ε 和 ζ 表示比较小的扰动量，目的是避免分母的值接近 0 时归一化会产生异常值。

当训练弱分类器的时候，有几个比较重要的参数。在 Model II（求解弱分类器的阈值）中，罚因子 r_1 和 r_2 的取值对阈值的影响结果比较大，也即对正例样本与反例样本的分类误差有很大的影响。如果 $r_1 = r_2$ ，此时求出的阈值不倾向正例和反例中的任何一方。但是通常来说，正例样本的因子 r_1 应该要比反例样本的因子 r_2 要大，这样才能保证正例样本在每一级训练的过程中，大量正例样本被分对，进一步地保证在下一级训练时，正例样本的数目不会大幅度地减少。然而，考虑到反例样本的分类精度，也不能过多地增加比例 r_1/r_2 ，因此当比例设定在区间 $(1, 40)$ 比较合适。在图 5-5 中展示了在第一级训练中，当 r_1/r_2 的比例逐渐增加时，正例样本和反例样本分类误差的变化情况。

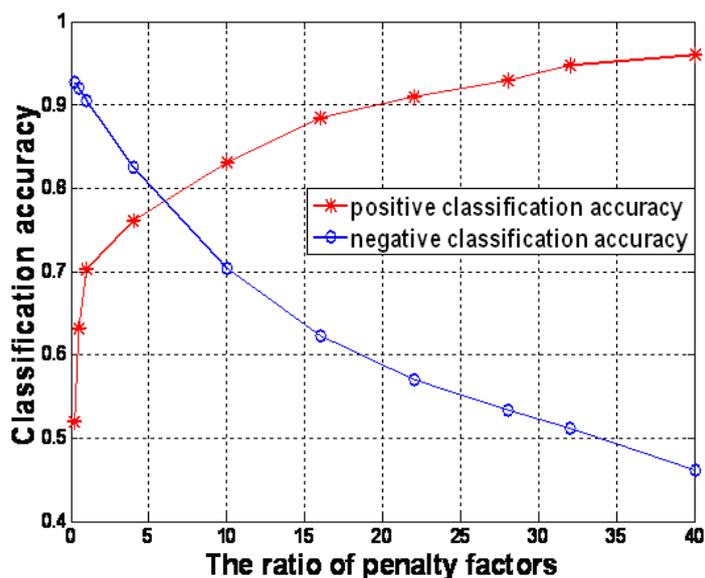


图 5-5.分类误差。

在图 5-6 中，展示了级联分类器数目、拒绝率等结果。在图 5-6(a)中，可以看出大约 6 个级联就可以拒绝接近 95% 的反例窗口。CLML 方法对于一个检测窗口的判别大概需要平均 3.9 个块，这个数目要少于 [9] 中的数目。除此之外，所提的 CLML 方法要快于线性 SVM, 核函数 SVM, 这是因为 SVM 方法使用测试样本与支持向量的内积，相反 CLML 方法由于 LML 的稀疏作用，只是计算相应的非零元素与测试样本的内积，因而可以提高检测速度。此外，CLML 采用了级联机制和整数规划模型，使得它在前几级使用最少的特征块拒绝大量的图像窗口，进一步提高了检测速度。综上所述，CLML 方法中弱分类器的计算时间少于线性 SVM 与核函数 SVM 方法，强分类器的检测速度要快于 [8, 9] 中提到的方法。

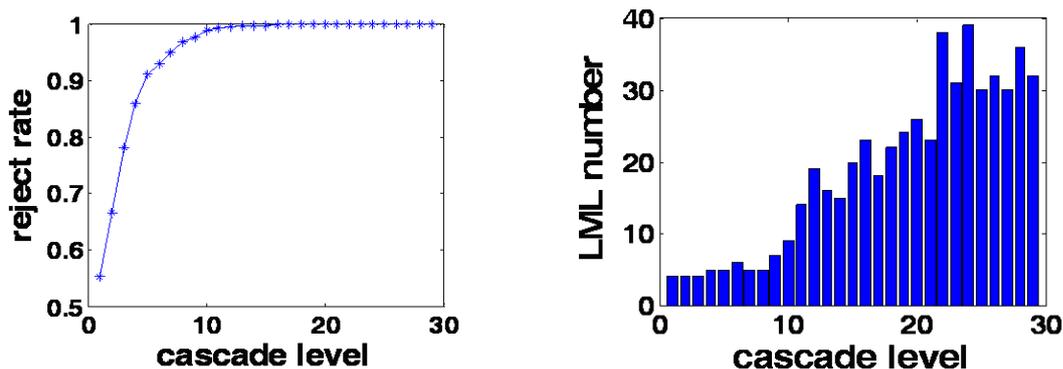


图 5-6. 级联数目 (a). 在所有的级上累加的拒绝率; (b). 每一级 LML 弱分类器的数目。

CLML 方法的每一级都可以得到最紧致的特征块。在图 5-7(a)中，展示了 CLML 方法选择出最好的四个特征块，这些块的大小与位置和文献 [9] 中展示的不同，在他们的论述中，块的大小大概是 36×72 个像素区域。然而，在第一级，CLML 选择出的块大小是 40×40 个像素区域，并且它们的位置处于腿部。在第二级，块的大小是 36×72 个像素区域大概覆

盖了整个人体的轮廓。CLML 方法与[9]中的方法所选择特征块的大小和位置不一样，主要原因在于 CLML 弱分类器的训练、强分类器的学习机制以及样本的选择，这些都与[9]中的方法不同。

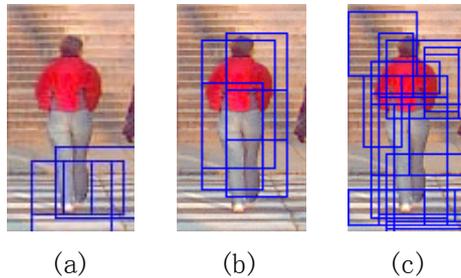


图 5-7.CLML 特征块展示. (a) 第一级中所选择的块, (b)第二级中所选择的块(c) 第 15 级中所选择的块.

5.5.2. 性能比较

下面展示了 CLML 算法和现在一些经典算法比较的性能曲线。这些经典的算法主要包括 Dalal 的 HOG+SVM[8]算法, Zhu 的 v-HOG+SVM+CADaboost (级联 Adaboost) [9]算法和 Tuel 的 COV+Logitboost[10]算法。性能曲线是根据丢失率 (Miss Rate) 和每个窗口的虚假率 (FPPW) 来计算的, 所有的点的绘制是在 log 尺度的坐标下。随着级数的增加, 可以看出虚假率减少的同时, 丢失率逐渐增加, 丢失率和虚假率是相互矛盾的整体, 所以在检测时需要平衡两者之间的关系。虚假率和第三章的定义一致, 丢失率的定义如下:

$$MissRate = \frac{\text{丢失的人体数}}{\text{总的实际人体数}} = 1 - RecallRate \quad (5.26)$$

尽管所提的 CLML 方法与 Zhu 的方法使用相同的 v-HOG 特征来表示人体模式, 但是 CLML 弱分类器的构建, 强分类器的形成方式 (特征选择的原则) 等都与 Zhu 的方法不同。Zhu 使用线性 SVM 方法训练得到弱分类器, 使用 Adaboost 算法将弱分类器组合成强分类器。而所提的 CLML 方法是使用 LML 优化模型训练得到弱分类器, 并且使用整数规划的方式选择弱分类器形成强分类器。从图 5-8 中, 可以看到本章所提的 v-HOG+CLML 要好于 Zhu 的 v-HOG+CADaboost 方法。

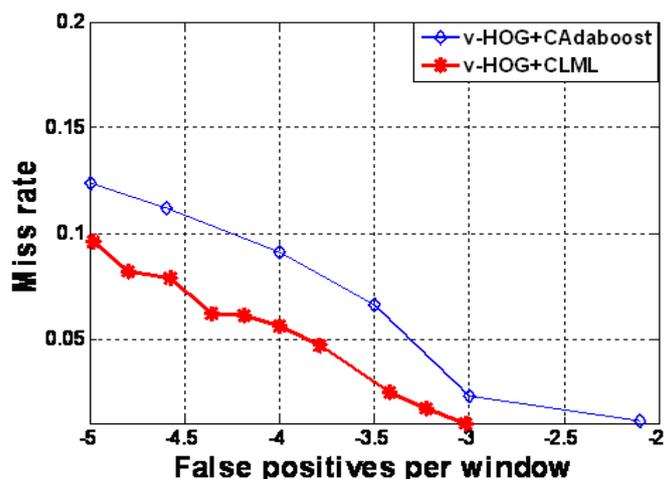


图 5-8.CLML 在 INRIA 上单独性能比较

图 5-9 中展示了 CLML 算法与一些经典检测算法在 INRIA 测试集上的性能对比曲线。图 5-9 是在图 5-8 的基础上,增加了 HOG+SVM[8]和 COV+Logitboost[10]的性能曲线。本章按照 Dalal 的算法编写 HOG 代码,并通过 SVM 的开源代码 LibSVM 实现 HOG+SVM 的方法,可以看到本章重现他们的结果大致和 Dalal 等人提供的曲线图差不多,至于 v-HOG+CADaboost 方法和 COV+Logitboost 方法,因为算法中所涉及参数太多,所以根据这些方法发表在 INRIA 数据集上最好的结果,来绘制两者的测试曲线。从图 5-9 可见,所提的 CLML 方法优于三种现在比较受欢迎的算法。当 FPPW 的值在 10^{-5} 时,本文的方法可以达到 9% 丢失率,这个结果要比同条件下 HOG+SVM 方法低 8% 左右,比 v-HOG+CADaboost 方法大概低 3% 左右,比 COV+Logitboost 方法低 1% 左右。COV+Logitboost 方法中使用协方差矩阵 (COV) 作为特征描述子并且运用流形分析的方法对多姿态、多视角的人体形成初始的聚类,这可能是它与本文所提 CLML 方法性能接近的原因。

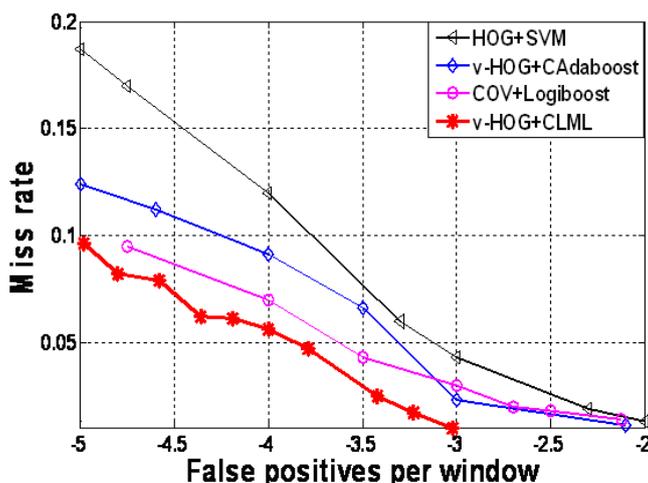


图 5-9.CLML 在 INRIA 上性能比较

此外,为了验证 CLML 方法对视角变化的鲁棒性,本章还在 SDL 数据集上比较了 CLML

和 HOG+SVM 方法的性能，测试曲线的结果如图 5-10 所示。该曲线也是在 log 坐标轴上绘制的丢失率和虚假率。本章未能得到 v-HOG+CADaboost 和 COV+Logitboost 方法的代码，而且由于两种方法的参数比较多，所以本着公平和合理的原则，本章未重现他们的代码，也未将其与 CLML 方法在 SDL 数据集上进行比较。

在图 5-10 中，可以得出 CLML 方法要好于 HOG+SVM 方法，这个曲线展示结果当然也可以说明 CLML 方法的优越性。当 FPPW 取值是 10^{-4} 时，CLML 方法可以达到 4% 的丢失率，这一结果要比 HOG+SVM 方法大概低 7% 左右。

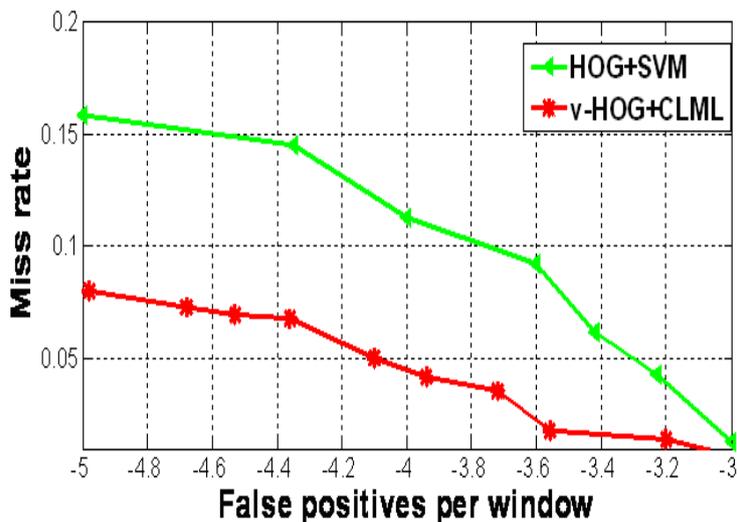
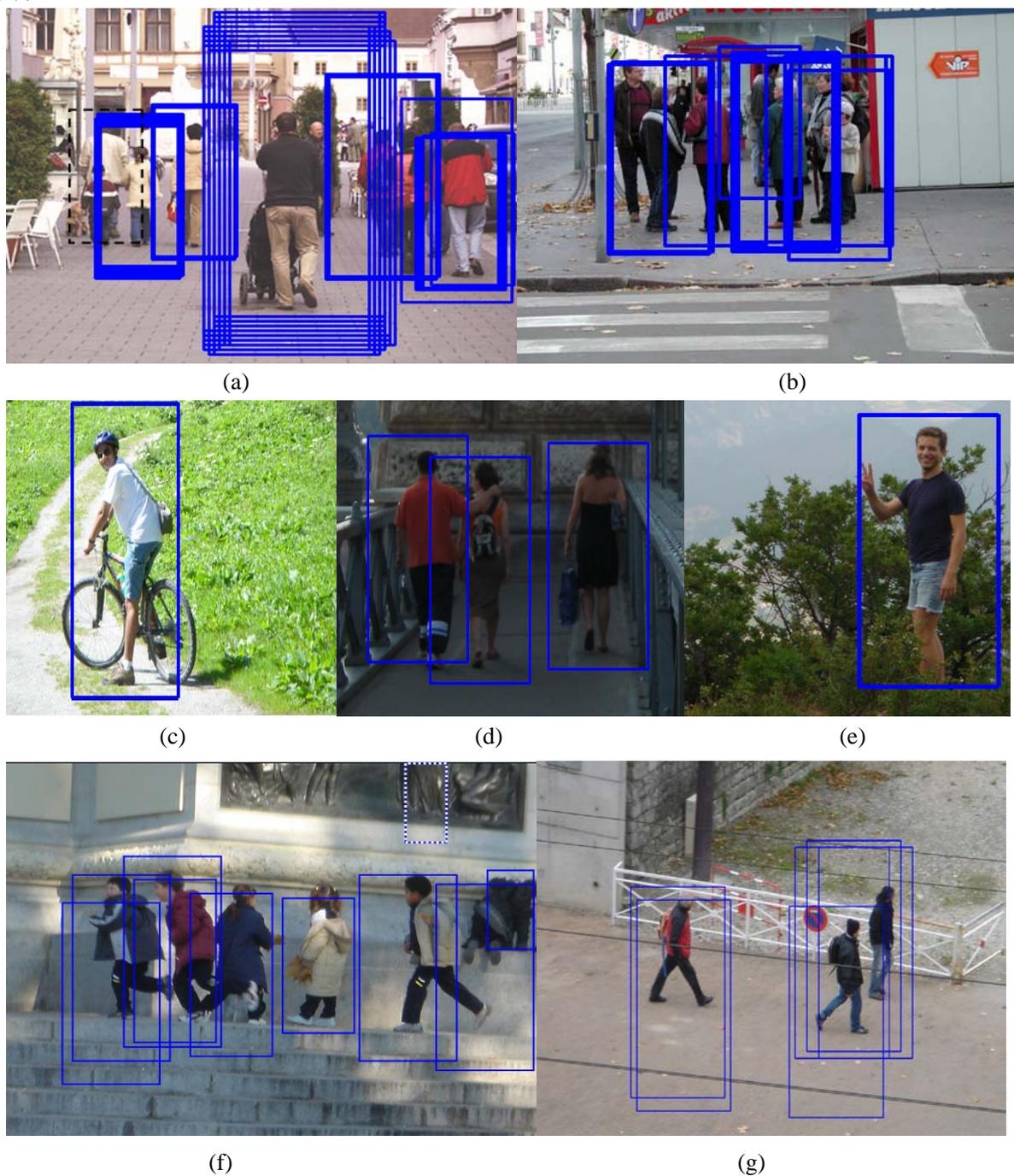
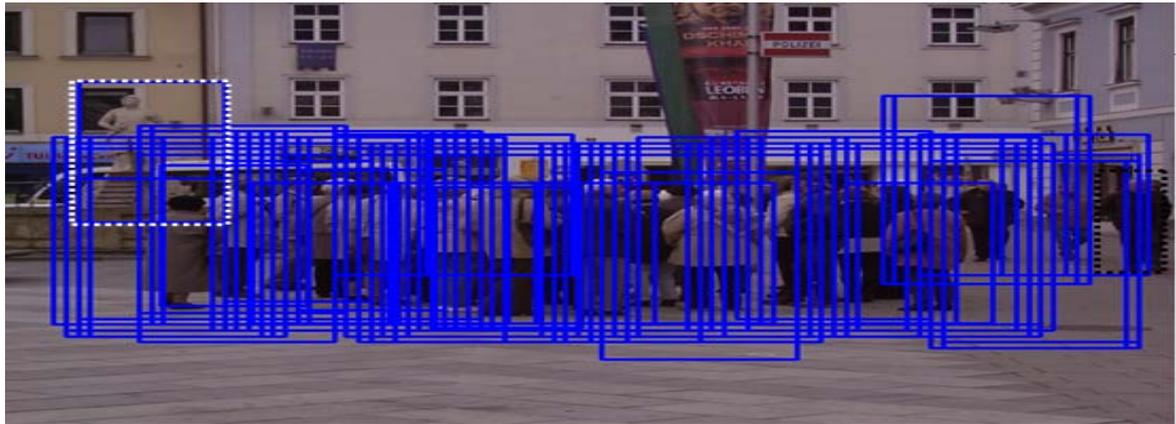


图 5-10. CLML 在 SDL 上性能比较

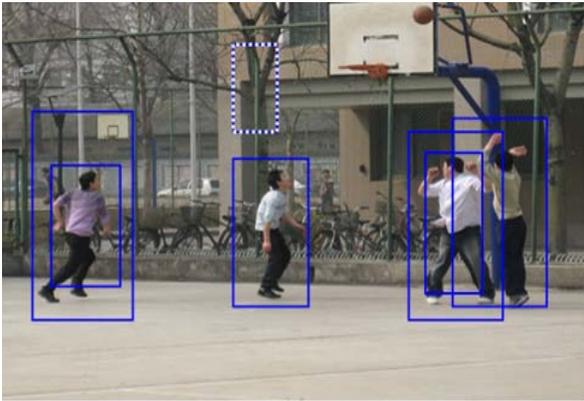
在图 5-11 中，展示了运用 CLML 方法得到的一些检测实例。在检测过程中，该方法采用窗口逐一扫描式并加入多尺度搜索策略。在图 5-11 (a) 中，虽然那个最左边的人，他身体的大部分被右边的蓝色检测框包含，但是实际上他属于漏检的目标。在图 5-11 (b) 中，所有的人体目标都被正确地检测到，特别是从图片左边数第四个人，他穿着黑色夹克衫并被前面的人遮挡，而 CLML 算法仍能够正确发现他。在图 5-11 (c) 中，骑自行车的行人，虽然其姿态处于弯曲状态不是竖直的，但是 CLML 方法可以正确地检测到他。在图 5-11 (f) 中，所有的小朋友无论是处于正视角，还是侧视角，无论是处于跑步姿态，还是处于跪着姿态，都能够被正确定位，但是在这张图片中有一个铜像被错检。在图 5-11 (g) 中，虽然图片是从俯视角度拍摄的行人，但是 CLML 方法仍然检测到了所有俯视视角的行人。在图 5-11 (h) 中，大部分的行人无论其是否处于多姿态或遮挡状态，都能够被正确定位，除了位于图片最右边一个行人。该行人非常靠近图像的边缘地带，轮廓特征不是很明显。另外在这张检测图片中，存在一个误检测，一个白色雕像被认为是行人，这个白色雕像和人体的轮廓非常相似，在静止图片中，若没有一些运动信息，它非常容易被误认为是人体。在图 5-11 (i) 中，由于树干的影响，虽然打篮球的男孩被正确检测到，

但是存在一个错误的检测窗口，误把竖直形状的树枝当成人体，这个与 v -HOG 特征无法区分细树枝与粗人腿有关系。在图 5-11 (j)中,有 5 个可爱淘气的小朋友,他们的姿态变化比较多, 4 个小朋友被正确的定位, 有 1 个小朋友被漏检。在图 5-11 (k)中包含多个人体目标, 他们中部分处于正视角, 部分处于侧视角, CLML 方法能够将所有的人体正确定位。在图 5-11(l)中, 虽然有个女孩被前面的小汽车遮挡, 但是 CLML 算法能够成功将其发现。

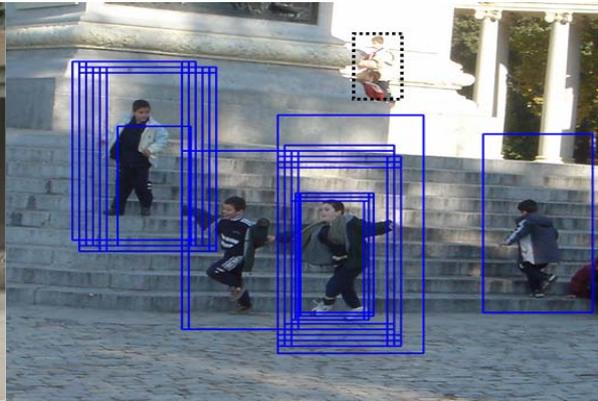




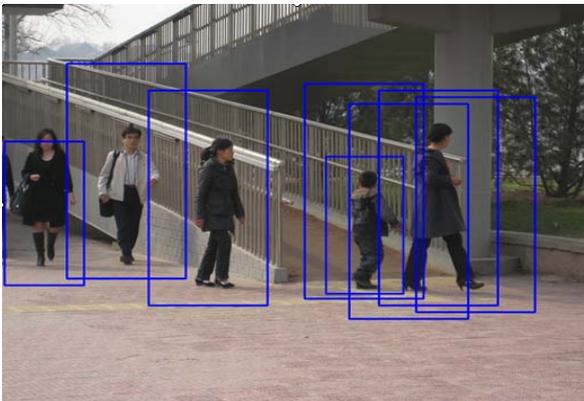
(h)



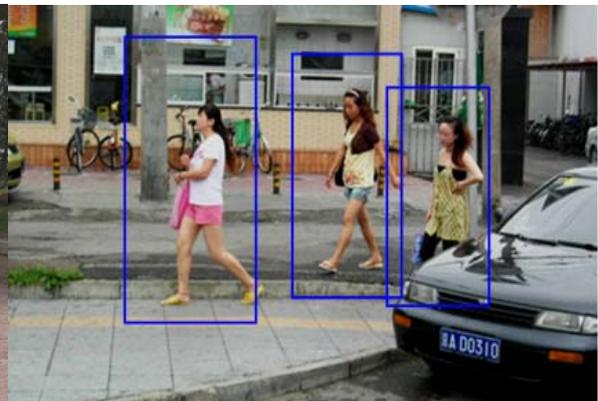
(i)



(j)



(k)



(l)

图 5-11.CLML 检测实例。所检测到的虚假正例使用白色的虚线框标记的，丢失的人体使用的是黑色的虚线框标记的。

5.6. 本章小结

本章提出了一种基于 1 范数最小化学习的级联分类方法，从人体测试集上的实验结果看，该方法的效果要优于现在一些比较受欢迎的人体检测方法。该方法意在实现特征选择，而且是在双重意义下的特征选择。它首先针对块状特征中的每一维特征值，LML 优化模型可以实现对块状特征值进行选择；其次，针对块状特征，整数规划模型从全局角

度选择最紧致的块状特征作为人体模式的表述，同时实现强分类器的构建。无论是弱分类器还是强分类器，本章都是从 VC 维最小化角度建立优化模型，进一步地最小化测试误差的上界。考虑到人体检测算法的特殊性，本章使用了级联机制，并结合上面所述的强分类器，形成最后的级联分类器。

每个特征块中的特征值是通过 1 范数最小化准则选择的，1 范数的稀疏性使得加权特征值在某种稀疏度约束下也是稀疏的。另外，特征块的组合方式是由整数规划完成的，整数规划的目的也是为了选择最少、最紧致的特征块，也即每一级中所选择的特征块也是稀疏的。通过 CLML 方法得到特征块中的特征值以及特征块的数目，都是稀疏的。所以在某种程度上，算法对于视角的变化、遮挡以及多姿态等问题不敏感，这些已经在实验结果中得到很好的说明。

第六章 总结与展望

6.1. 本文工作总结

本文主要是从特征选择与分类器构建的角度，提出了一系列的人体检测算法。所提算法，理论上都是从 VC 维最小化角度设计的分类器模型，意在融合特征选择与分类器构建为一体，从而实现对人体窗口的判别。本文的主要研究成果如下：

1、提出了一种基于 1 范数最小化学习的优化模型。作者首先探讨了 LML 优化模型的设计与 VC 维之间的关系，证明了 LML 优化模型在理论上是通过追求 VC 维最小化来保证测试误差上界的最小化。此基础上设计了 L-LML 算法，意在设计一种线性分类器，能够突出特征向量中元素的差异性，实现对特征向量中元素的差异化选择。实验结果表明，L-LML 算法因为增加了特征选择功能，所以其效果要优于线性 SVM 算法。

2、提出了一种基于 LML 优化模型的非线性分类器 (NL-LML)。该分类器建立在本文提出的 LML 优化模型基础之上，实现对样本的非线性判别。1 范数优化的性质决定了直接在 LML 优化模型基础上实现核函数的内积形式是不可行的。因此 NL-LML 算法引入一种基于核函数的非线性变换，将变换后的样本与 LML 优化模型结合可以实现非线性判别。NL-LML 算法是基于 LML 模型设计的，可以视为在变换空间中对变换样本进行特征选择。

3、设计了一种级联分类器 CLML。在每一级中，通过弱分类器的线性加权组合构成强分类器，每个弱分类器是通过 LML 优化模型和 min-max 罚函数模型训练得到的，关于 min-max 罚函数的含义及其与对策论的关系在论文中亦做了论述和证明。每个强分类器的构建是使用整数规划的方法。强分类器的构建过程相当于 LML 优化模型在整数空间中求解，这一结论在文中进行了解释。因此，所提出的 CLML 分类器中，无论弱分类器的训练还是强分类器的构建都符合 1 范数最小化准则。实验结果表明，与其它一些优秀的人体检测算法相比，所构造的 CLML 算法具有效率高，能够实现对人体快速、准确检测的特点。

实验证明，针对不同的特征选择与复杂样本分布问题，本文所提出的一系列由简单到复杂、由低精度到高精度的检测算法中 (L-LML 算法、NL-LML 算法和 CLML 算法)，CLML 具有最好的检测效果。本文所提出的三种检测算法都是从稀疏性角度考虑的，因而较小地受到人体检测过程中多视角、多姿态和遮挡等因素的影响，具有较好的鲁棒性。

6.2. 未来工作展望

虽然本文算法对于遮挡问题具有一定的鲁棒性，但是，遮挡出现的随机性导致当遮挡部分出现在非稀疏人体特征部分时，算法对遮挡的处理情况仍然不是很乐观。因此，作者关于未来的研究工作提出了以下展望：

1、可以把本文所提出的线性 1 范数最小化学习算法与其他无监督学习技巧相结合，对样本进行初始聚类（如流形聚类分析），从而更加有利于分类器的训练。至于非线性 1 范数最小化学习算法，以后的工作可以着手于参数分析对算法性能的影响和尝试采用该算法对其它数据集进行判别与分析。

2、未来的研究工作拟从人体分块的角度解决遮挡问题。当遮挡出现在人体的某个部位时，基于分块的人体检测算子仍可以检测到其他部位，这样组合所有检测的结果之后，仍然可以检测到处于遮挡状态的人体。由于人体的部分模块和很多物体比较相似，比如说人体的头部，会和球状的路灯、操场上的足球等信息比较相似，人体的腿部、胳膊等部分会和直立的树杆、电线杆等背景信息非常相似，所以分块人体检测现在面临的主要困难是特征描述子的精确性与检测子的鲁棒性。另外，如何通过部分的人体模型组合成整体的人体模型也是分块人体检测算法中的关键问题之一

随着图像处理、计算机视觉与模式识别技术的发展，特征表示与分类方法不断地得到更新与改进，相应地，人体检测算法会不断地完善成熟。相信在不久的将来，人体检测算法能更好地应用到视频监控、智能交通、汽车辅助驾驶系统等方面中。

参考文献

- [1]. Thanarat Horprasert, David Harwood, Larry S. Davis, A statistical approach for real-time robust background subtraction and shadow detection, Vol.99, pp: 1-19, 1999.
- [2]. Beleznai, C., Fruhstuck, B., Bischof, H., Human detection in groups using a fast mean shift procedure, IEEE International Conference on Image Processing, pp:349-352,2004
- [3]. Haga, T., Sumi, K., Yagi, Y., Human detection in outdoor scene using spatio-temporal motion analysis, IEEE International Conference on Pattern Recognition, pp:331-334, 2004.
- [4]. H. Elzein, et al. A motion and shape-based pedestrian detection algorithm. IEEE Intelligent Vehicles Symposium, pp:500-504, 2003.
- [5]. D.-J. Lee, et al. Shape-based human detection for threat assessment. Orlando, International Society for Optical Engineering, SPIE 5438, pp:81-91 2004.
- [6]. Ying Wu and Ting Yu, A Field Model for Human Detection and Tracking, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.28(5), pp.753-765, 2006.
- [7]. P.Viola and M. Jones, Robust Real-time Object Detection, International Journal of Computer Vision, 2001.
- [8]. Dalal, N, Triggs, B., Histograms of Oriented Gradients for Human Detection, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp.886-893, 2005.
- [9]. Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, Fast Human Detection Using a Cascade of Histograms of Oriented Gradients, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, pp.1491-1498, 2006.
- [10]. Tuzel O., Porikli F., Meer P., Human Detection via Classification on Riemannian Manifolds, Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, pp.1-8, 2007.
- [11]. Mu Y. Yan S. Liu Y., Huang T., Zhou B., Discriminative Local Binary Patterns for Human Detection in Personal Album, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol 2328, pp.1-8, 2008
- [12]. Xiaoyu Wang, Tony X Han, Shuicheng Yan., An HOG-LBP Human Detector with Partial Occlusion Handling, Proc. IEEE Int'l. Conf. on Computer Vision, Kyoto, 2009.
- [13]. B.Wu, and R. Nevatia. Detection of Multiple, Partially Occluded Humans in a Single

- Image by Bayesian Combination of Edgelet Part Detectors, Proc. IEEE Int'l. Conf. on Computer Vision, 2005.
- [14]. Qixiang Ye, Jianbin Jiao, Baochang Zhang, Fast pedestrian detection with multi-scale orientation features and two-stage classifiers, IEEE International Conference on Image Processing, pp:881-884, 2010.
- [15]. S. Ioffe and D.A. Forsyth. Probabilistic Methods for Finding People, Int'l. Journal of Computer Vision, vol. 43(1), pp:45–68, 2001.
- [16]. B. Leibe, E. Seemann, and B. Schiele. Pedestrian Detection in Crowded Scenes, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp.878-885, 2005.
- [17]. D. Vinay, J. Neumann, V. Ramesh, and L.S. Davis, Bilattice-Based Logical Reasoning for Human Detection, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2007.
- [18]. A. Mohan, C. Papageorgiou, and T. Poggio, Example-Based Object Detection in Images By Components, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23(4) pp.349-360, 2001
- [19]. S. Munder and D. Gavrilu. An experimental study on classification, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 28(11):1863–1868, 2006.
- [20]. Christopher J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, vol. 2(2), pp.121-167, 1998
- [21]. M.Collins, R.E. Schapire and Y. Singer, Logistic Regression, AdaBoost and Bregman Distances, Machine Learning, vol. 48(1-3), pp.253-285, 2002
- [22]. D. Donoho, For Most Large Underdetermined Systems Of Linear Equations The Minimal L1-Norm Near Solution Approximates The Sparsest Solution, Comm. on Pure and Applied Math, vol. 59(6), pp.797-829, 2006.
- [23]. K Huang, SAviyente, Sparse Representation for Signal Classification, Advances in Neural Information Processing Systems, Proc. of the Twentieth Annual Conf. on Neural Information Processing Systems, 2007
- [24]. A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, Robust Face Recognition via Sparse Representation, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, No. 2, 2009.
- [25]. Michael Jüger, Denis Naddef, Computational Combinatorial Optimization: Optimal or

- Provably Near-Optimal Solutions, Springer Press, 2001
- [26]. A.T. Mario, D. Nowak, J.Wright, Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems, IEEE Selected Topics in Signal Processing, vol.1(4),pp.586-597,2007
- [27]. Zhi-Hua Zhou, Jianxin Wu, W. Tang, Ensembling Neural Networks: Many Could Be Better Than All, Artificial Intelligence. vol. 137(1-2), pp. 239-263, 2002
- [28]. <http://coe.gucas.ac.cn/SDL-HomePage/resource.asp>
- [29]. S. Munder, C. Schnorr, and D.M. Gavrila, Pedestrian Detection and Tracking Using a Mixture of View-Based Shape-Texture Models, IEEE Trans. on Intelligent Transportation Systems, vol. 9(2), pp. 333-343, June 2008.
- [30]. C. Papageorgiou and T. Poggio, A Trainable System for Object Detection, International Journal of Computer Vision, vol. 38, pp. 15-33, 2000.
- [31]. H. Shimizu and T. Poggio, Direction Estimation of Pedestrian from Multiple Still Images, Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 596-600, 2004.
- [32]. J. Alvarez, Th. Gevers, A. Lopez, Learning Photometric Invariance for Object Detection, Int'l. Journal of Computer Vision, vol.90(1), pp. 45-61, 2010.
- [33]. S. Maji, A. Berg, and J. Malik. Classification using intersection kernel SVMs is efficient, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 2008.
- [34]. M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata, Pedestrian Detection with Convolutional Neural Networks, Proc. of the IEEE Intelligent Vehicles Symposium, pp. 223-228, 2005.
- [35]. B.E. Goldstein, Sensation and Perception, sixth ed. Wadsworth, 2002.
- [36]. William Robson Schwartz , Aniruddha Kembhavi , David Harwood , Larry S. Davis, Human Detection Using Partial Least Squares Analysis, Proc. of the IEEE Int'l Conference on Computer Vision, 2009.
- [37]. Z.Lin, L. Davis, D. Doermann, D. DeMenthon, Hierarchical part-template matching for pedestrian detection and segmentation, Proc. of the IEEE Int'l Conference on Computer Vision, 2007.
- [38]. P. Viola, M. Jones, and D. Snow, Detecting Pedestrians Using Patterns of Motion and Appearance, Int'l Journal of Computer Vision, vol. 63(2), pp. 153-161, 2005.
- [39]. Maji, S.; Berg, A.C., Max-margin additive classifiers for detection, Proc. of the IEEE

- Int'l Conference on Computer Vision, 2009
- [40]. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D., Object Detection with Discriminatively Trained Part-Based Models, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.32(9), pp:1627-1645
- [41]. P.F. Felzenszwalb and D.P. Huttenlocher, 2005, Pictorial structures for object recognition, *IJCV*, vol. 61(1), pp.55–79.
- [42]. M. Andriluka, S. Roth, B. Schiele., Pictorial Structures Revisited: People Detection and Articulated Pose Estimation, *Proc. IEEE CVPR*, pp.1014-1021, 2009.
- [43]. M. Balcan, A. Blum, and S. Vempala, Kernels as features: On kernels, margins, and low-dimensional mappings, *Machine Learning Journal*, vol.65(1), pp.79-94, 2004.
- [44]. Liu Y.Z., Shan S.G, Zhang W.C, Chen X.L, Gao W. Granularity-tunable gradients partition (GGP) descriptors for human detection, In: *IEEE International Conference on Computer Vision and Pattern Recognition*, pp.1255-1262, 2009.
- [45]. Markus Enzweiler, and Dariu M, Gavrila Monocular. Pedestrian Detection: Survey and Experiments, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31(12), 2009.
- [46]. 孙庆杰, 静态图像中人体检测技术研究, 中国科学院软件所, 2004 年
- [47]. 朱文佳, 基于及其学习的行人检测关键技术研究, 上海交通大学硕士论文, 2008 年
- [48]. Enzweiler, M., Eigenstetter, A., Schiele, B., and Gavrila, D.M., Multi-Cue Pedestrian Classification with Partial Occlusion Handling, *Proc. IEEE CVPR*, pp.990-997, 2010.
- [49]. M.Oren, C.Papageoriou, P.Sinha, E.Osuna, T.Poggio, Pedestrian Detection Using Wavelet Templates, *IEEE Computer Vision and Pattern Recognition*, pp.193-199,1997
- [50]. 邓乃扬,田英杰.数据挖掘中的新方法---支持向量机, 北京:, 科学出版社, 2004
- [51]. Robert E. Schapire, A Brief Introduction to Boosting, *Proceedings of the 16th international joint conference on Artificial intelligence (IJCAI)*,vol.2, 1999.
- [52]. Christopher J. C. Burges, Advances in Kernel Methods-Support Vector Learning, MIT Press, 1999.
- [53]. Vapik V., *Stastical Learning Theory*, Willey,1998.
- [54]. Vapik V., Chervonenkis A., The necessary and sufficient conditions for consistency in the empirical risk minimization method, *Pattern Recognition and Image Analysis* 1991,vol.1(3),pp:283-305.

- [55]. D.G.Lowe. Distinctive image features from scale-invariant key points. IJCV, 60(2):91-110, 2004.
- [56]. Donoho D.L, Compressed Sensing, IEEE Trans. on Information Theory, vol. 52(4), pp: 1289-1306, 2006.
- [57]. 郑慧娆、陈绍林、莫忠息、黄象鼎, 数值计算方法, 武汉大学出版社, 2002 年
- [58]. 马振华等, 现代应用数学手册—运筹学与最优化理论卷, 清华大学出版社, 2005 年
- [59]. Richasrd O.Duda, Peter E.Hart, David G.Stork, Pattern Classification, 李宏东等人译, 模式分类, 机械工业出版社, 2009.
- [60]. I. T. Jolliffe, Principal Component Analysis, Sprinberg Publishers, 2002.
- [61]. Emmanuel Candes, Xiaodong Li, Yi Ma, and John Wright, Robust Principal Component Analysis, Journal of the ACM, 3, 2011(accepted).
- [62]. Liu H, Motoda H, Feature Selection for Knowledge Discovery and Data Mining, Boston, Kluwer Academic Publishers,1998.
- [63]. Liu H, Yu L.Toward, Integrating Feature Selection Algorithms for Classification and Clustering. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(4):491-502.
- [64]. Molina L C, Belanche L, Nebot A, Feature Selection Algorithms: A Survey and Experimental Evaluation Number, Barcelona, 2002.
- [65]. Somol P, Novovicova J, Pudil P., Notes on the Evolution of Feature Selection Methodology, Kybernetika, vol.43(5):713-730, 2007.
- [66]. Somol P,Pudil P,Kittler J., Fast branch and bound algorithms for optimal feature selection, IEEE Transactions on Pattern Analysis and Machine Intelligence,2004,26(7):900-912.
- [67]. Huang J,Cai Y, Xu X, A hybrid genetic algorithm for feature selection wrapper based on mutual information, Pattern Recognition Letters,2007,28:1825-1844.
- [68]. Zhang D, Chen S, Zhou Z-H., Constraint score: A new filter method for feature selection with pairwise constraints. Pattern Recognition, 2008, 41(5):1440-1451.
- [69]. Neumann J, Schnorr C, Steidl G., Combined SVM-Based Feature Selection and Classification, Machine Learning, vol.61, pp.129-150, 2005.
- [70]. Kononenko I., Estimation attributes: analysis and extensions of RELIEF, in Proc. of the

- European Conference on Machine Learning, pp.171-182, 1994.
- [71]. Robnik-Sikonja M, Kononenko I. Theoretical and Empirical Analysis of ReliefF and RReliefF, *Machine Learning*, 2003, 53:23-69.
- [72]. Liu H, Motoda H, Yu L. Feature selection with selective sampling, in *Proc. of the 19th International Conference on Machine Learning*, 2002, 395-402.
- [73]. Pudil P, Novovicova J, Kittler J, Floating search methods in feature selection, *Pattern Recognition Letters*, 1994, 15(11):1119-1125
- [74]. Somol P, Pudil P, Novovicova J, et al. Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, vol.20(11-13), pp.1157-1163, 1999.
- [75]. Schapire R E, Freund Y, Bartlett Y, et al., Boosting the margin: a new explanation for the effectiveness of voting methods. *Annals of Statistics*, 1998, 26(5):1651-1686
- [76]. K. Torkkola. Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, 3:1415–1438, 2003.
- [77]. Yang J, Vasant H. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems*, 1998, 13:44-49
- [78]. Yu L, Liu H., Feature selection for high-dimensional data: a fast correlation-based filter solution. In *Proc. of the 20th International Conference on Machine Learning*. Menlo Park, CA.: AAAI press, pp.856-863, 2003.
- [79]. Casillas J, Cordon O, et al. Genetic feature selection in a fuzzy rule-based classification system learning process for high dimensional problems. *Information Sciences*, vol.136(1-4), pp.135-157, 2001.
- [80]. Jain A k, Zongker D., Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19(2), pp.153-158, 1997.
- [81]. Kudo M, Jack S., Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 2000, 33:25-41
- [82]. Kira K, Rendell L., A practical approach to feature selection, in *Proc of the 9th International Conference on Machine Learning*, 1992, 249-256.
- [83]. Devijver P A, Kittler J. *Pattern Recognition—A Statistical Approach*, London, Prentice Hall, 1992.
- [84]. Mitra P, Murthy C A, Pal S K. Unsupervised feature selection using feature similarity,

- IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24(3), pp.301-312, 2002.
- [85]. Hall M A. Correlation-based Feature Subset Selection for Machine Learning, Hamilton, New Zealand University of Waikato, 1999.
- [86]. Wu S, Flach P A. Feature selection with labelled and unlabelled data, in Proc of ECML/PKDD'02 workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning, pp.156-167, 2002.
- [87]. Dy J G, Brodley C E. Feature Selection for Unsupervised Learning, Journal of Machine Learning Research, vol.5, pp.845-889, 2004.
- [88]. Forman G. An Extensive Empirical Study of Feature Selection Metrics for Text Classification, Journal of Machine Learning Research, vol.3, pp.1289-1305, 2003.
- [89]. Hall M A, Holmes G. Benchmarking, Attribute Selection Techniques for Discrete Class Data Mining, IEEE Transactions on Knowledge and Data Engineering, vol.15(3), pp.1041-1347, 2003.
- [90]. Yang Y, Pedersen J O. A Comparative Study on Feature Selection in Text Categorization, in Proc of the 14th International Conference on Machine Learning, pp.412-420, 1997.
- [91]. H. Stoppiglia, G. Dreyfus, R. Dubois, and Y. Oussar. Ranking a random feature for variable and feature selection. Journal of Machine Learning Research, vol.3, pp.1399-1414, 2003.
- [92]. R. Tibshirani. Regression selection and shrinkage via the lasso. Technical report, Stanford University
- [93]. Kohavi R, John G H. Wrappers for feature subset selection. In Artificial Intelligence journal, Special Issue on relevance, vol.97(1-2), pp.273-324, 1997.
- [94]. John G, Kohavi R, Pfleger K., Irrelevant features and the subset selection problem. In: Cohen W W, Hirsh H, Eds. The Eleventh International Conference on Machine Learning. 121-129, 1994.
- [95]. Breiman L, Friedman J H, et al., Classification and Regression Trees, Wadsworth International Group, 1984.
- [96]. Quinlan J R., Bagging, boosting, and C4.5, In the 13th National Conference on Artificial Intelligence. Menlo Park, CA: AAAI press, 1996, 725-730
- [97]. A. Rakotomamonjy. Variable selection using SVM-based criteria. Journal of Machine

- Learning Research, vol.3, pp.1357–1370, 2003.
- [98]. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. Gene selection for cancer classification using support vector machines. *Machine Learning*, vol.46(1-3), pp.389–422, 2002.
- [99]. Dash M, Liu H. Consistency-based search in feature selection, *Artificial Intelligence*, vol.151(1-2), pp.155-176, 2003.
- [100]. Arauzo-Azofra A, Benitez J M, Castro J L. Consistency measures for feature selection, *Journal of Intelligent Information System*, vol.30, pp.273-292, 2008.
- [101]. 刘华文, 基于信息熵的特征选择算法研究, 吉林大学, 博士论文, 2010
- [102]. 张丽新, 高维数据的特征选择及基于特征选择的集成学习研究, 清华大学, 博士论文, 2004 年
- [103]. 苏映雪, 特征选择算法研究, 国防大学, 硕士论文, 2006 年
- [104]. 刘亚洲, 基于时空分析和多粒度特征表示的人体检测方法研究, 博士论文, 哈尔滨工业大学, 2009
- [105]. 崔新毅, 视频序列中的人体检测方法研究, 硕士论文, 哈尔滨工业大学, 2007
- [106]. 徐堃, 静态图像中的人体检测技术, 硕士论文, 中国科技大学, 2009
- [107]. D. L. Donoho and M. Elad, Optimally Sparse Representation in General (nonorthogonal) Dictionaries via L1 Minimization, in *Proc. Nat. Aca. Sci.*, vol.100, pp. 2197-2202, 2003.
- [108]. Vincent Guigue, Alain Rakotomamonjy and Stéphane Canu, Kernel Basis Pursuit, *ECML*, 2005
- [109]. Jingyu Yang, Yigang Peng, Wenli Xua, and Qionghai Dai, Ways to Sparse Representation: A Comparative Study, *Tsinghua Science & Technology*, vol.14(4), pp.434-443, 2009.
- [110]. 石光明, 刘丹华, 高大化, 刘哲, 林杰, 王良君, 压缩感知理论及其研究进展, *电子学报*, 卷号 37, 期号 5, 2009.
- [111]. E.J. Candès, The restricted isometry property and its implications for compressed sensing, *Comptes Rendus Mathématique*, vol.346(9-10), pp.589-592, 2008.
- [112]. B.S. Kashin. Diameters of some finite-dimensional sets and classes of smooth functions, *Izv. Akad. Nauk SSSR Ser. Mat.*, vol.41(2), pp.334–351, 1977.
- [113]. E.J. Candès, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math*, vol.59(8), pp. 1207–1223, 2006.

-
- [114]. E.J. Candès and T. Tao, Decoding by linear programming, *IEEE Transactions on Information Theory*, vol.51(12), pp. 4203–4215, 2005.
- [115]. Tibshirani, R. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B* 58, 267-288, 1996.
- [116]. S.S. Chen, Basis Pursuit. Ph.D. Thesis, Stanford University, 1995
- [117]. Mallat, S., and Zhang, Z. Matching pursuit with time-frequency dictionaries. *IEEE Transactions Signal Processing*, vol.41(12), pp.3397–3415, 1993.
- [118]. Zongben Xu, Jian Sun: Image Inpainting by Patch Propagation Using Patch Sparsity. *IEEE Transactions on Image Processing*, vol.19(5), pp.1153-1165, 2010
- [119]. Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal, Matching pursuit: Recursive function approximation with applications to wavelet decomposition. *Proc. of IEEE 27th Annual Asilomar Conference on Signals, Systems and Computers*. Pacific Grove, 40-44, 1993.
- [120]. P.S. Bradley , O. L. Mangasarian, Feature Selection via Concave Minimization and Support Vector Machines, *Machine Learning Proceedings of the Fifteenth International Conference*, 1998.

附录 1

博士期间发表论文目录

已发表文章目录:

- 1、Ran Xu, Baochang Zhang, Qixiang Ye, Jinbin Jao, “Human detection in images via L1-norm minimization learning”, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp: 3566-3569, 2010, Dallas, U.S.A.
- 2、Ran Xu, Baochang Zhang, Qixiang Ye, Jinbin Jao, “Cascaded L1-norm Minimization Learning (CLML) Classifier for Human Detection”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp:89-96, 2010, San Francisco, U.S.A.
- 3、Ran Xu, Qixiang Ye, Jinbin Jao, “Nonlinear L1-norm Minimization Learning for Human Detection”, IEEE International Conference on Image Processing (ICIP), accepted, 2010, Brussels, Europe.
- 4、徐冉, 高随祥, 董萍, 黄菲, “无线传感器网络的可靠性计算”, 计算机应用研究, Vol. 2-3, 2009
- 5、黄菲, 高随祥, 孙波, 董萍, 徐冉, “无线传感器网络传输可靠性计算”, 应用数学学报, Vol. 31(5), 2008

已提交文章目录:

- 1、Ran Xu, Jinbin Jao, Baochang Zhang, Qixiang Ye, “Pedestrian Detection in Images via Cascaded L1-norm Minimization Learning Method”, Pattern Recognition
- 2、Ran Xu, Jinbin Jao, Qixiang Ye, “Linear and Nonlinear L1-norm Minimization Learning for Human Detection”, Pattern Recognition Letters.

致 谢

在攻读硕士学位五年的学习生活中，我经历了诸多坎坷，也付出了艰辛的努力，同时也得到了很大的收获。在毕业论文完成之际，由衷地感谢这五年来曾经给予我无数帮助的老师、同学、朋友和家人。

首先，本课题的研究工作是在导师焦建彬教授的悉心指导下完成的。焦建彬教授在我的学习中，从理论和实验方面给予了大量的、极其有益的建议和具体的指导，并在论文的撰写和审稿中倾注了大量的心血，他诲人不倦的精神和对我的谆谆教导，使我从恩师身上学到了不少可贵的东西，让我受益匪浅。此外，他还在生活和思想上关心我、照顾我，再次表示万分的感谢！尤其是恩师严谨的治学态度、积极的进取精神、平易近人的工作作风将使我终身受益。

其次，特别要感谢叶齐祥老师在三年的学习与生活中对我的极大帮助和支持，他渊博的专业知识、自强不息的学习精神和扎实的动手能力，在理论学习和项目开发过程中给我提供了耐心的引导和帮助。此外，我还要特别感谢数学院的高随祥老师在前两年的学习与生活中对我的指导与支持，他精益求精的治学态度和高屋建瓴的领悟力，在数学基础理论学习过程中给我提供了非常多的帮助。我还要特别感谢北京航空航天大学的张宝昌老师在科研上对我的指导，他孜孜不倦的科研态度是我学习的榜样。

再次，感谢我含辛茹苦的父母，是他们在我面临困难时给我指引方向，在我徘徊迷茫时教会我要懂得去摸索。感谢我的男朋友，在我郁闷沮丧的时候，鼓励我，在我遇到技术困难时，给予我莫大的帮助。感谢同实验室的各位同学，他们在我的论文工作和撰写中，给予了我巨大的支持和帮助。

本项目的完成离不开国内外的技术资料，在此对参考文献的作者、译者以及出版单位表示感谢！

最后，再次向在学习、工作和生活中给予过自己关心、支持与鼓励的所有老师、同学、朋友们表示最诚挚的谢意！

徐冉

2011年4月