密级:_____

中国科学院大学
University of Chinese Academy of Sciences

硕士学位论文

基于特征融合与混合分类器的行人检测

作者姓名：梁吉祥

指导教师:叶齐祥 副教授 中国科学院大学

学位类别:工学硕士

学科专业:计算机应用技术

研究所:中国科学院大学电子电气与通信工程学院

2013 年 4 月

Pedestrian Detection Based on Feature Fusion and Mixture of Piecewise Models

By

Jixiang Liang

A Dissertation/Thesis Submitted to
The University of Chinese Academy of Sciences
In Partial Fulfillment of the Requirement
For the Degree of
Master of Computer Application Technology

College of Electronic, Electrical and Communication Engineering

04, 2013

中国科学院大学直属院系
研究生学位论文原创性声明

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

作者签名：

日　　期：

中国科学院大学直属院系

学位论文授权使用声明

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

涉密的学位论文在解密后适用本声明。

作者签名：　　导师签名：

日　　期：　　日　　期：

摘　要

图像与视频中的目标检测是计算机视觉领域的重要研究内容之一。 行人检测，作为目标检测的一个典型案例，在智能视频监控、智能交通和辅助安全驾驶系统等应用领域具有重要的应用价值。但是，在室外环境，目前的行人检测方法仍然面临着特征表达瓶颈，受到多视角、多姿态等问题的困扰。本文从目标检测特征表达与分类器模型两个方面进行研究，解决行人检测特征描述以及视角和姿态等关键问题，以提高行人检测算法的性能及鲁棒性。

本文的主要工作如下：

1） 评测了几种最常用的行人检测特征描述子，在对特征进行模式分析的基础上，提出了一种新的行人检测特征描述子——HOG_SURF 特征描述子。在 SDL 以及 INRIA 行人检测数据集上的测试结果表明，该特征描述子在达到最好特征 HOG_LBP 性能的前提下，维数显著降低、效率显著提高；

2） 针对行人检测中的多视角、多姿态问题，分析和实现了几种不同的正例样本划分策略，并在 INRIA 数据集上评测了样本划分策略的性能。在此基础上，提出了一个基于 ECOC 编码的混合分段模型，并在多个数据集上与现有方法进行了对比实验，验证了该模型处理多视角和多姿态问题的有效性；

3） 进行了行人检测的应用研究探索，将所研究的行人检测算法应用到两个原型系统中：基于深度信息与图像信息融合的行人检测系统，基于背景建模的监控视频中的行人检测。验证了行人检测算法的有效性。

关键词：目标检测，行人检测，特征融合，ECOC 编码，分段混合模型

Abstract

- **Jixiang Liang** (Computer Application Technology)

Directed by: **Qixiang Ye** (Associate Professor)

Object detection in images and video is one of the fundamental problems of the computer vision community. Pedestrian detection, as one of the most representative cases of object detection, has a broad of application background from intelligent video surveillance, intelligent traffic to automatic driving assistant systems. Many pedestrian detection methods proposed in recent years, the robustness is still an open problem, in particularly when there are multi-view and multi-posture pedestrians in a natural scene. The goal of the thesis is to solve the multi-view and multi-posture problems from the perspectives of feature descriptors and classification methods. The contributions are summarized as follows:

1) We evaluated four kinds of representative local descriptors including HOG, Haar-like, SURF and LBP for pedestrian representation, on which we try to find the best combination of the descriptors by analyzing the complementarity of them and evaluating the performance with cross-validation experiments. Experiments on two public pedestrian datasets show that the combination of HOG and SURF descriptors (named HOG_SURF descriptor) is the best one, reporting comparable performance to the state-of-art HOG_LBP descriptor as well as having a much lower feature dimensionality and then much higher computational efficiency.

2) We compared several strategies for positive samples splitting and evaluate the performance of these strategies on the INRIA pedestrian dataset. Based on the result, a new pedestrian detection method on Error Correcting Output Code (ECOC) classification of manifold sub-classes is proposed. Experiments on three datasets show that our proposed method improves the state-of-the-art.

3) We applied the proposed pedestrian detection methods into two applications. The first application is a fast pedestrian detection system using laser and image data fusion, and the other is a software prototype for detecting objects in surveillance video. Both of them show the effectiveness of our proposed methods.

KEY WORDS：Object Detection, Pedestrian Detection, Feature Fusion, ECOC Coding, Pedestrian Detection Applications

目　录

图目录

表目录

第一章 绪论

1.1 课题背景和研究意义

图像与视频中的目标检测是计算机视觉领域的重要研究内容之一，具有重要的理论意义和应用价值。行人检测是是目标检测中比较典型的研究问题之一。理论方面，行人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。应用方面，行人检测算法在智能视频监控、智能交通和汽车辅助驾驶系统中有着很广泛的应用前景。

近年来，在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取得非常大的进展。 其中人脸检测与车牌检测算法更是走向了实际应用。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，其困难与原因在于：

1). 人体是一个非刚性的、多姿态的、多角度的物体；

2). 含有人体目标的图像，其背景一般都是在室外，受到自然环境的干扰；

3). 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、快速地行人检测，目标的特征描述、自适应的视觉模型显得极为重要。Dalal 和 Triggs[1][25][48]于 2005 年提出基于 HOG[1]特征与支撑向量机（SVM）结合的行人检测算法，使得行人检测性能有了较大提升。他们当时建立的一个很有挑战的行人检测数据集—INRIA[38][48]数据集，一直沿用到今天。X.Wang[2]等人于 2009 年结合 HOG 与 LBP[2]提出了一种新的组合特征 HOG_LBP 特征描述子，并且提出了一种解决部分遮挡问题的办法，使得行人检测性能又有了较大的提升。P. F. Felzenszwalb[3][34]等人于 2010 提出了基于部件的形变模型(DPM[3])，解决了特征对齐和模式分散的问题，该模型在 INRIA 行人检测数据集上的性能较以前的方法有了质的提高（约 20%）。

尽管如此，人体目标检测的问题并没有被完全解决。Piotr Dollar [4]等人在他们的行人检测文献综述中指出，在存在部分遮挡以及低分辨率的情况下的行人

[1] Histogram of Oriented Gradient，中文译为：梯度方向直方图

[2] Local Binary Pattern，局部二值模式，一种常用的纹理特征

[3] Discriminatively Trained Part based Models

检测仍然没有完全解决。尤其是在驾驶环境中、低分辨率情况下的性能仍然非常低（有的低于 50%）。此外虽然行人检测的理论研究已日渐成熟，但是还没有走向实际应用，实际环境下的行人检测需要与其他的信息融合以实现更为准确、鲁棒的检测。因此，开展基于特征融合与混合分类器的行人目标检测具有理论意义和应用价值。

本论文受到了以下研究课题的资助：

1、 “基于多源数据的飞行器进近威胁目标检测跟踪及行为预测”，国家自然科学基金重点项目（课题编号：61039003），2011.01-2014.12。

2、 “飞行器威胁目标识别与图像鲁棒匹配理论与方法”，国家 973 计划子课题（课题编号：2010CB731804-2），2010.01-2014.12。

3、“多视角多姿态人体目标检测研究”国家自然科学基金面上项目（项目编号：61271433），2013.01-2016.01。

1.2 国内外研究现状

行人检测所采用的方法与技术一方面继承自早期的比较经典目标检测如：车牌检测、人脸检测；另一方面在行人检测方面新的特征、模型以及分类器方法也可以直接推广到其它的目标检测。所以行人检测技术的发展与目标检测是大体一致的，下面主要从目标检测的角度介绍国内外的研究现状。

目标检测主要涉及两部分内容，一是特征表述即特征描述子，它是指从图像中提取出表示目标的特征向量，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目标的某种特征，按照某种学习准则，获取分类函数的过程。特征表述和分类器的主要功能是让计算机意识到什么样的模式是属于目标的，什么样的模式是属于背景。目前，很多研究者致力于这两方面的研究。

在特征描述方面，Papageorgiou 和 Poggio[5]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。如在文献[6][26]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。在文献[1][25][48]中，作者提出稠密的、重叠的、固定尺度的 HOG 局部特征描述子表述人体。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思想。后来，在 HOG 特征的基础上，涌现出一些改进版本的特征[7][8][27]。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，存在不能很好的把握人体的局部轮廓特性的弊端，研究者们引入改进的 HOG 特征，并且取得了比较好的结果。文献[9]中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[10]认为原始的用来表示纹理特征的 LBP 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[2]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[8]中，作者提出了一种新的多尺度方向（Multi-scale Orientation 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[11]]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[12]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。文献[13]提出了一种 SIFT 特征的改进特征用以实现快速鲁棒的图像匹配。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。

在分类器的构建方面，大致可以分为两种：第一种是基于概率的方法（Probabilistic Method），另一种是基于判别的方法（Discriminative Method）[9]。早期的目标检测像：车牌检测识别、人脸检测识别大多采用基于概率的方法，比较经典的有模板匹配、神经网络、Fisher 判别等；后期的目标检测如：行人检测、车辆检测、人脸检测等基本转向了基于判别的方法，比较常用的基于判别的方法有支撑向量机、Boosting 等，大量的研究表明基于区分的方法要优于基于概率的方法[9]。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[14][28][29]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[15][30][31][32]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[16]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。

随着研究的深入，对于问题的描述变得越来越复杂，表现为目标函数非凸性、约束条件极大、隐变量的引入等几个方面。模型的复杂使得求解变得异常困难，为此许多学者在这方面做了大量的工作也取得了大量的理论成果。文献[17][33]中，作者通过求解 L1-Norm 最小化学习方法实现对目标的稀疏表达，并取得了优异的性能。文献[18]中作者通过将特征空间 X 与输出空间 Y 联合起来构成一个新的空间（X,Y），在这个空间中进行分类器的学习称之为结构 SVM 巧妙的将多类分类问题转换成二分类问题。文献[19]中作者引入含隐变量的结构 SVM 在理论上给出了一种求解的算法。文献[3][34]中作者通过构建一个包含隐变量的视觉模型，通过训练基于部件可形变混合模型，取得了最佳结果。

尽管人体目标检测已取得了相当大的进展，但是距离普遍意义上的实际应用仍有一定距离。在部分遮挡以及低分辨率条件下的检测就仍然有很大的提升空间，在多个行人检测公开测试数据集上如 TUD[35], Caltech[36][37]，当前最好的基于静止图像的检测算法的表现也不尽如人意。Pascal[38][39]每年的目标检测方面的比赛也仍然具有相当大的挑战。

1.3 本文研究内容

本文开展了以下三个方面的研究：1).人体目标的特征表达方面，研究现有的

目标描述子之间在人体检测方面的互补性关系，为后续的研究找出一个比较好的特征描述子；2).在现有的理论框架下，结合目前最好的行人检测模型，提出一个有力于解决遮挡问题的视觉模型，并求解测试其性能；3).针对实际的应用环境，通过多源信息融合的方法，测试人体目标检测的实际性能，并探讨可能的信息融合方法。其中第一部分和第二部分的内容是本文的主要内容，第三部分的内容本文将只做简要介绍。

在特征描述子方面，根据 Marr 的计算视觉理论，视觉目标分为三级表象描述：描述图像的密度变化及其局部几何关系；描述可见表面的方位、轮廓、深度及其他性质和三维形状表象。视觉目标可以通过其轮廓来表示和计算，当前的研究也证明了这一点。目前常用的视觉局部特征：Haar-like 特征、MSO（Multi-scale Orientation）特征、SURF（Speed Up Robust Feature）特征、SIFT 特征、HOG 特征、LBP 特征、COV（协方差）特征、LRF（Local Receptive Field）特征等等。上述的特征各自都有优势和不足，但是特征之间的互补性讨论的并不够充分，这是因为特征和所采用的模型及方法有着密不可分的关系，不同的特征可能适用于不同的方法，这样就给特征的评估带来了很大的困难。因为特征描述是目标检测的基础，所以仍然有许多研究人员在做特征改进方面的工作。例如：ICCV 2009 年的文章[2]就指出了在 INRIA 行人检测数据集上 HOG 特征与 LBP 特征的组合就取得了优异的性能。本文选择了 HOG、SURF、LBP、Haar-like 等几种具有代表性的特征在 SVM 的模型和理论框架下进行了特征的评测，探讨这几种特征在行人描述上的互补性关系，得出了一些实验性的结论，并发表了国际会议论文[55]。

在视觉模型和分类器方面，视觉模型是描述目标特征向量在特征空间中分布的数学模型，由于行人目标的特殊性质——非刚体、多视角、多姿态等特性。行人的特征描述在特征中间中的分布可以描述成一个数学流形，反例样本是这个特征空间中正例样本的补集，分类器的目标是区分这个特征空间中的正例样本和反例样本。线性 SVM 分类器模型认为正例样本和反例样本在特征空间中线性可分，即正例样本和反例样本分别构成凸的流形。P. F. Felzenszwalb 等人认为由于行人存在多姿态、多视角以及部分遮挡等问题，正例样本不构成凸的流形，但是反例样本仍假设构成凸的流形，为此他们提出了基于部件可形变混合模型，形变能够解决特征与位置的对齐，混合模型是将正例样本进行分段划分。本文在研究各种正例样本分段方法的基础上，提出了基于 ECOC 编码的分

段混合模型，该模型与线性 SVM 分类模型相比性能有较大的提升，该工作已发表在 IEEE 汇刊[56]。

在行人检测应用研究方面，本文作者参与了实验室另外两个课题的部分工作，实际环境中行人检测的应用环境往往不是基于单幅图像的检测，视频中的目标检测和监控环境下的行人检测更有实际意义。在这种情况下，行人检测不仅仅是通过静止图像中目标的边缘、纹理特征来进行行人检测，还需要与目标的其他特征进行融合以实现更鲁棒的检测。本文将简单介绍两个应用实例，一个是基于激光信息与图像信息融合的行人检测系统，另一个是基于背景建模的监控视频中的行人检测。这两项工作作者都是部分参与其中第一项工作已经完成并以第而作者发表了一篇国际会议论文[57]，第二项工作目前仍在进行中。

1.4 本文的组织结构

第一章，绪论。介绍了基于特征融合与混合分类器的行人目标检测的研究背景和意义，分析了国内外行人检测的研究现状，总结了行人检测尚未解决的一些问题，列出了本文的主要研究内容。

第二章，行人检测相关研究综述。分析了行人检测常用的特征描述子以及常用的分类方法。

第三章，行人检测特征融合研究。论述了研究行人检测特征描述子的意义，在评测了几种主要的行人检测特征描述子的基础上，研究这几种特征描述子之间的互补性关系，并提出了一种 HOG_SURF 特征。

第四章，行人检测分段混合模型。针对行人检测中目标的多视角、多姿态问题，通过简单线性 SVM 分类器模型不能得到很好的检测效果，在分析和研究各种样本分段方法的基础上，提出了一种基于 ECOC 编码的分段混合模型，与线性 SVM 分类模型相比有效的解决了行人检测的多视角和多姿态问题。

最后总结了本文的主要工作，分析了论文的不足、列举了仍然存在难点问题，展望了未来的研究方向。

第二章 行人检测相关研究综述

2.1 行人检测特征

图像处理的过程中人们经常提到一个概念—图像特征。那么究竟什么是图像特征？图像特征分为哪几类？常用的行人检测特征又有哪几种呢？本章将简要介绍一下图像特征的分类和几种行人检测常用的特征描述子。

什么是图像特征，并没有一个准确的定义。理想的特征描述子应该具有：可重复性、可区分性、集中以及高效等特性；还需要能够应对图像亮度变化、尺度变化、旋转和仿射变换等变化的影响。特征描述子又是计算机视觉应用中的一个重要组成部分，例如对图 2.1 中两幅图像进行图像匹配。我们能想到的第一类特征就是图像中的那些特殊位置，比如建筑物的角点处，这些局部特征通常被叫“关键点特征”或者“兴趣点”，或者还可以叫“角点”，它们通常用其位置点周围像素块的表现来描述。另一类特征是“边缘”，基于它们的方向和局部表现（边缘结构），可以对这些不同类型的特征进行匹配，而且这些特征还可以作为图像序列中出现物体边界和遮挡情况的理想指示器。边缘可以聚成长的曲线和直线片段，这些线段可以直接用于匹配或者用于分析寻找消失点，从而获取摄像机的内部参数和外部参数[20]。

图 2. 两组待匹配的图像

由此可以看出图像特征大致可以分为三类：点状和区域状特征、边缘特征、线段（几何）特征。在计算机视觉领域，SIFT（旋转尺度不变特征）、HOG（梯度统计直方图特征）、SURF（快速鲁棒特征）、MSO（块主方向特征）、HAAR-like（HAAR 小波特征）、LBP（局部二值模式特征）是比较有代表性的，而其中 SIFT 特征不直接用于行人检测，但是它与 HOG、SURF、MSO 这三种特征有着密切的关系。后三种特征都是在 SIFT 的基础上发展而来的；而 HAAR-like 特征与 LBP 特征是人脸检测和识别方面经典的特征描述子。

2.1.1 SIFT 特征

SIFT 特征点是一种广泛使用的图像特征，可用于物体跟踪、图像匹配、图像拼接等领域。提出 SIFT 角点的 David Lowe 教授[21][22]已经用 C 语言和 matlab 实现了 SIFT 的检测，并开放了源代码[4]。SIFT 特征的提取分为以下步骤：

1. 构建尺度空间。通过尺度空间内的多层表示来保证尺度不变性；

2. 采用 LoG（Laplacian of Guassian）近似寻找感兴趣特征点。使用 LoG 能够很好的找到感兴趣特征点（Interesting Point），但是需要大量的计算量，所以通过计算图像 DoG（Difference of Gaussian）的极值点来近似；

3.去除不好的特征点。边界和低亮度区域是不好的特征点，提高效率和鲁棒性，可以通过 Harris Corner 检测算子实现；

4. 为特征点赋主方向值。并且依照这个方向做进一步的计算，步骤有效的消除方向的影响，使 SIFT 特征具有旋转不变性；

5. 生成 SIFT 特征。通过区域梯度方向统计得到 SIFT 特征的描述子。

尺度空间的构造是用第一幅图像逐渐生成一组模糊后的图像，然后将原始图像的尺寸缩小一半，再逐渐生成下一组模糊后的图像以此类推。然后对相邻尺度的图像做差分计算得到 DoG 图像，其过程如图 2.2 所示。

通过 DoG 图像可以找到显著的特征点，分为两个子步骤：

1. 在 DoG 图像中找到极大或极小像素点；

2. 找到子像素级的极大极小值点。

第一步是找到粗糙的极大极小值点。其过程通过扫描每个像素并且检测所有的邻接像素点。邻接像素点不仅包括当前图像中的邻接像素，而且包括上一

[4] http://www.cs.ubc.ca/~lowe/keypoints/

层和下一层图像中的邻接像素，示意图如图 2.3 所示。

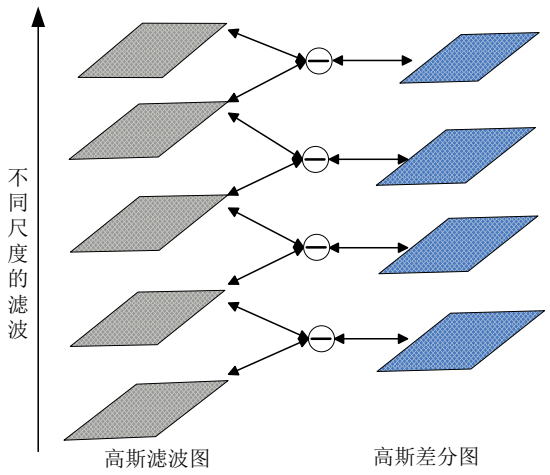


图 2. DoG 计算示意图

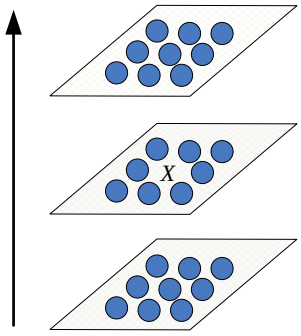


图 2. DoG 图像中的极大极小点示意图

经过上一个操作，我们从每组图像中得到 4 幅 DoG 图像，只需要对中间的两幅 DoG 图像进行极大极小值像素点进行检测。X 标记当前像素点，颜色圈标记邻接像素点。用这个方式，最多检测 26 个像素点。如果 X 是所有邻接像素点的最大值或最小值点，则它被标记为特征点。通常对于非极大或极小值点不需要遍历所有 26 个邻接像素点，少数的几个检测就能够加以排除。请注意在此过程中最高层和最底层的尺度是不需要检测的。这步结束后，所标记的点就是近似的极大极小值点。之所以说是“近似的”是因为极大极小值点都不会恰好在像素点的位置上，它一般位于相邻像素之间的非整数坐标位置。需要通过插值得到子像素的位置。

若子像素点与近似特征点间的偏移量大于 0.5，则按照偏移近似特征点的方向相应改变（移动）特征点，然后再把该点当作近似特征点，重复该操作，直到子像素点与近似特征点间的偏移量小于等于 0.5。

分配特征点的方向：在每个特征点周围计算图像的梯度方向和大小，我们可以得到最显著的方向，并且将该方向赋给该特征点。后面的操作都相对该方向进行计算，确保了旋转不变性。

在特征点附近，创建一个方向收集区域来控制该特征点的影响范围，方向收集区域的大小依赖于它所在图像的尺度，尺度越大，收集区域越大。在方向收集区域中每个像素点的梯度大小和方向用公式（2-1）和（2-2）计算，从而得到另外两幅图，分别是梯度的幅值图和方向图。

$$m(x,y)=\sqrt{\left(I(x+1,y)-I(x,y)\right)^2+\left(I(x,y+1)-I(x,y)\right)^2} \quad (2\text{-}1)$$

$$\phi(x,y)=\arctan\left(\left(I(x,y+1)-I(x,y)\right)/\left(I(x+1,y)-I(x,y)\right)\right) \quad (2\text{-}2)$$

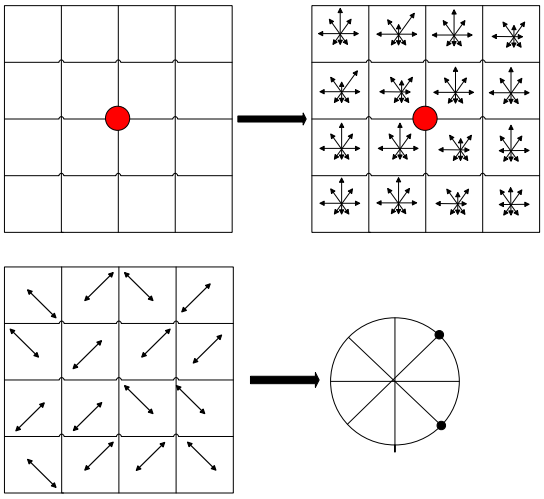


图 2. 特征点周围的窗口分解，并且为每个子窗口创建的 8 位直方图

接下来要为每个特征点创建一个唯一标识它的“指纹”，称为 SIFT 描述子(descriptor)。所生成的 SIFT 描述子既要能让相同场景中图像的特征点能够正确匹配，而且还要让不同场景中图像的特征点能够正确区分。

为了得到这样的SIFT描述子，将特征点周围16×16的窗口分解为16个4×4的子窗口，分解的过程见图 2.4。在每个 4×4 的子窗口中，计算出梯度的大小和方向，并用一个 8 个 bin 的直方图来统计子窗口的平均方向。

梯度方向在 0-45 度范围的像素点被放到第一个 bin 中，45-90 度范围的像素点被放到下一个 bin 中，依此类推。用一个直方图来统计方向收集区域中像素方向。在直方图中，将 360 度的方向分成 8 个 bins，每个 bin 包含 45 度。假设方向收集区域中某个像素点的梯度方向是 18.75 度，把它放入 0-45 度 bin 中，并且加入到 bin 中的量与该像素点的梯度大小成正比。同样，加入到 bin 中的

量依赖于该像素点梯度的大小及其到特征点的距离，这样远离特征点的像素点会加入较少的量到直方图中。

这样每个 4×4 的子窗口都对应一个 8 位的直方图，且直方图中加入的值是像素的用高斯加权后的梯度大小，而特征点周围 16×16 的窗口中包含 16 个 4×4 的子窗口，共有 16×8=128 位。单位化以后得到 SIFT 的描述子。

2.1.2 HOG 特征

HOG 由 Dalal 和 Triggs 于 2005 年提出，并提出了基于 HOG 特征的人体目标检测算法。HOG 特征通过提取局部区域的边缘或梯度的分布，来表征局部区域内目标的边缘或梯度结构，进而表征目标的整体形状。由于是在局部区域统计求得，HOG 特征对小的形变和配准误差有较强的鲁棒性。

Dala 将 64×128 的训练样本按照 8×8 个像素的小块（cell）进行划分，这样就形成了 8×16=128 个 cell。然后再将每相邻的 4 个 cell（田字形结构的 4 个 cell）划分为 1 组（block），通过滑动 block 可以得到多组田字形局部区域特征，每个 block 一次滑动 8 个像素，由此对于一个 64×128 的训练样本便具有 7×15=105 个组。首先，本文对于每个 cell 都按照 SIFT 的方法，将其中的所有像素的梯度方向进行投影，形成每个 cell 各自的梯度方向直方图。

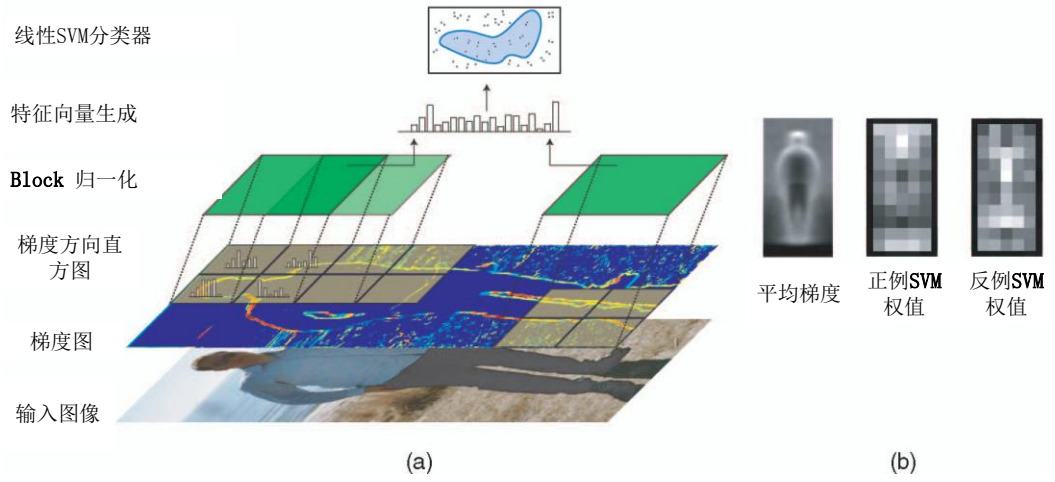


图 2. HOG 特征提取示意图

这里的方向 bins 的数量被设定为 9，而不是 SIFT 中的 8，即每 20 度一个 bin，0-180 与 180-360 的方向采用对等角相等的方法进行归类划分。然后，再将每个 block 中的 4 个 cell 的直方图的数据串联起来。由于每个 cell 的梯度直方图为一个 9 维向量，每个组是一个 36 维的向量。所有的 block 依次串联起来，便形成了对 36×105＝3780 维的特征（对于 64×128 像素的样本共 105 个 block）。

2.1.3 HAAR-like 特征

Haar-Like 特征是由 Haar 小波演变而来的[5][6][26][40]。Haar 小波的形式如图 2.6，而在灰度图像上，可以表现成图 2.7 的“Haar-Like 特征”形式。

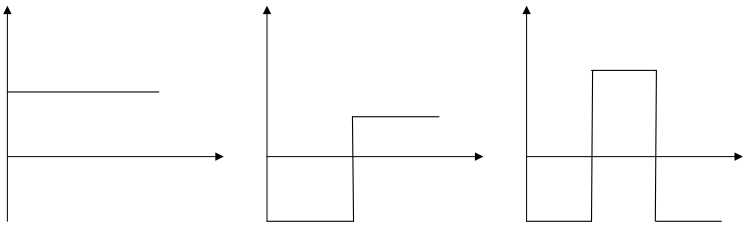
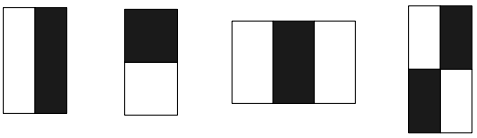


图 2. Haar 小波



(a) (b) (c) (d)

图 2. Haar-Like 特征

特征的值等于灰色矩形框中所有像素的颜色值和减去白色矩形框中的所有像素的颜色值。(a)和(b)为两个矩形的特征，(c)为三个矩形的特征，(d)为四个矩形的特征。对于由 2 个矩形所组成的特征的值，可以通过对两个矩形区域的所有像素分别求和然后再相减来实现。对于由 3 个矩形组成的特征的值是通过使用中间矩形的 2 倍减去两边矩形的和。最后一种是由 4 个矩形组成的本书可以按照对角线的两个矩形相加然后再求差来得到这种特征的值。Viola 等人[6][26]对基本的 Haar 特征进行了拓展，形成了拓展的 Haar-Like 特征，如图 2.8 所示。

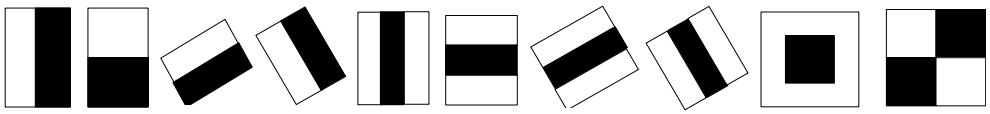


图 2. 拓展的 Haar-Like 特征

2.1.4 MSO 特征

Haar-like 特征反映了一个块内的颜色差异，对目标的颜色是比较敏感的。多尺度方向（MSO）[8]特征则是从大小不同的方块上计算一个方向值用以表示此块的方向。在固定尺度上的单个特征的计算中，每个特征区域都是一个 m× m

的正方形，对于这个正方形特征的特征值的计算，分为两步：

1.计算正方形区域的粗梯度方向（因为这里是在大尺度上求解梯度，相对于原始定义的梯度而言更加粗糙，因此这里将其命名为“粗梯度”）。

2.将梯度方向映射成固定的特征值编码。

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

$$SUM\_A = \sum_{x1<x<x2,1<y<y2} PixelValue(x,y) \tag{2-3}$$

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

$$\begin{aligned} dx &= SUM\_R - SUM\_L \\ dy &= SUM\_D - SUM\_U \end{aligned} \tag{2-4}$$

其中，SUM_R, SUM_L, SUM_D 和 SUM_U 分别表示为正方形特征区域的右半部、左半部、下半部和上半部的灰度值的累加。

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

$$Ori\_Rect = F(\arctan(dy/dx)) \tag{2-5}$$

其中，函数 F 为将角度转换为数值 0，1，2，…，8 的函数，如：当 Ori_Rect 的值在 1~20 或 181~200 之间的时候，则 Ori_Rect 的值为 0；当 $\arctan(dy/dx)$ 的值在 21~40 或 201~220 之间的时候，则 Ori_Rect 的值为 1。

2.1.5 SURF 特征

Herbert Bay 于 2006 年提出了 SURF（Speed UP Robust Feature）特征，SURF 特征是 SIFT 特征的加速版，因此，SURF 是被用来描述特征点的一种特征。SURF 特征的提取与 SIFT 特征类似，但也有些区别。SURF 特征提取的步骤也是：1、建立图像金字塔、2、通过 Hessian 矩阵检测特征点、3、通过 HAAR 小波响应确定主方向、4、提取 HAAR 响应为基础的 SURF 特征描述子。有关 SURF 特征的提取和计算的具体内容参见原作者的论文[13][23]，这里主要对比 SURF 与 SIFT 特征的不同点。

表 2- SIFT 特征与 SURF 特征的不同点

特征	SIFT	SURF
尺度空间	DOG 与不同尺度的图片卷积	不同尺度的 box filters 与原图片卷积
特征点检测	先进行非极大抑制，再去除低对比度的点。再通过 Hessian 矩阵去除边缘的点	先利用 Hessian 矩阵确定候选点，然后进行非极大抑制
方向	在正方形区域内统计梯度的幅值的直方图，找 max 对应的方向。可以有多个方向	在圆形区域内，计算各个扇形范围内 x、y 方向的 haar 小波响应，找模最大的扇形方向
特征描述子	16*16 的采样点划分为 4*4 的区域，计算每个区域的采样点的梯度方向和幅值，统计成 8bin 直方图，一共 4*4*8=128 维	20*20s 的区域划分为 4*4 的子区域，每个子区域找 5*5 个采样点，计算采样点的 haar 小波响应，记录 ∑dx,∑dy,∑|dx|,∑|dy|，一共 4*4*4=64 维

SURF：金字塔仅仅是用来做特征点的检测。在计算描述子的时候，HAAR 小波响应是在原图像（利用积分图）计算得到的。而 SIFT 是在高斯金字塔上计算得到特征描述子的（注意不是高斯差分金字塔）。

论文[24]对三种方法给出了性能上的比较，源图片来源于 Graffiti dataset，对原图像进行尺度、旋转、模糊、亮度变化、仿射变换等变化后，再与原图像进行匹配，统计匹配的效果。效果以可重复出现性为评价指标。

表 -2 SIFT、PCA-SIFT、SURF 性能比较

方法	时间	尺度	旋转	模糊	光照	仿射变换
SIFT	一般	最好	最好	一般	一般	好
PCA-SIFT	好	好	好	最好	好	最好
SURF	最好	一般	一般	好	最好	好

由此可见，SIFT 在尺度和旋转变换的情况下效果最好，SURF 在亮度变化下匹配效果最好，在模糊方面优于 SIFT，而尺度和旋转的变化不及 SIFT，旋转不变上比 SIFT 差很多。速度上看，SURF 是 SIFT 速度的 3 倍。

2.1.6 LBP 特征

局部二元模式（Local Binary Pattern，LBP)是 Ojala 等[44][45][46]提出的一种有效的纹理描述方法，该方法通过 LBP 算子来提取灰度图像中局部相邻区域的纹理特征。Ojala 等给出的 LBP 算子形式是一个固定大小为3×3 的矩形块，包含了中心点和周围八个邻域的共九个灰度值，计算 LBP 特征值时，将四周邻域的 8 个灰度值与中心像素点的灰度值相比较，大于等于中心灰度值的子块表示为 1，否则表示为 0，然后

为各个邻域像素点赋一子不同的权值，计算得到的十进制数值作为该3×3 矩形块的特征值，最后计算整幅图像中每个独立特征值的出现概率，并以直方图的形式作为该图像的纹理特征的描述。LBP 特征的计算过程如图 2.9。

6	5	2
7	6	1
9	8	7

(a)

1	0	0
1		0
1	1	1

(b)

1	2	4
128		8
64	32	16

(c)

图 2. LBP 特征计算示意图

图中，(a)是当前算子覆盖区域的灰度值分布图，中心点的灰度值为 6。根据 8 个邻域像素点灰度值和中心点灰度值的大小比较，得到该区域的二值图(b)，然后按图(c)所示为不同邻域像素点赋予不同的权值，即特征为:11110001。将其换算成对应的加权十进制数为$2^0+2^4+2^5+2^6+2^7=241$，即为当前矩形块的 LBP 特征。

2.2 行人检测分类器

2.2.1 SVM 分类器

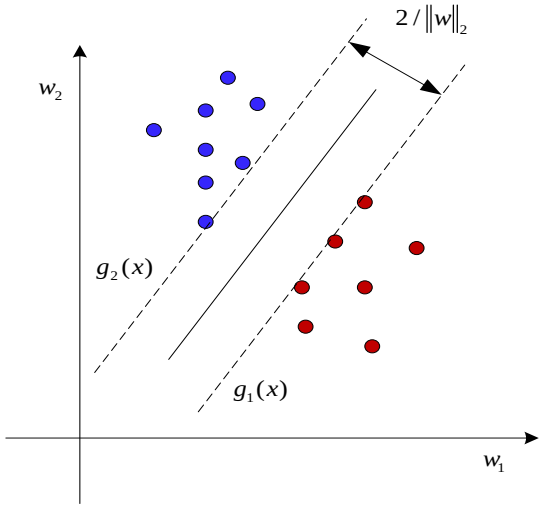


图 2. 最大间隔原理

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[14][28][29]，具有最小化结构风险（测试误差的上界）[43]。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 2.10 所示。

简单起见，先考虑线性可分的两类样本$\{x_i, y_i\}, i=1,\cdots,N$，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：$g(x)=w^T\cdot x+b$。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$g_1(x)=w^T\cdot x+b\geq 1$，对反例样本有$g_2(x)=w^T\cdot x+b\leq -1$。这样两条平行超平面之间的“间隔”为$2/\|w\|$。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:

$$\begin{aligned} &\min \quad \frac{1}{2}\|w\|^2 \\ &s.t. \quad y_i[(w^T\cdot x_i)+b]-1\geq 0 \quad i=1,\ldots,N \end{aligned} \tag{2-6}$$

上述优化模型是一个凸二次规划，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对，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(2-7)所示。

$$\begin{aligned} &\min \quad \frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\xi_i \\ &s.t. \quad \begin{cases} y_i[(w^T\cdot x_i)+b]\geq 1-\xi_i, \\ \xi_i\geq 0, \quad i=1,\ldots,N \end{cases} \end{aligned} \tag{2-7}$$

在式(2-7)中：ξ_i表示第i个样本被误分的程度。C是平衡误分程度与最小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。

$$L(w,b,\xi,a_i,t_i)=\frac{1}{2}\|w\|^2+\sum_{i=1}^{N}\xi_i(C-a_i-t_i)-\sum_{i=1}^{N}a_iy_i(w^T\cdot x_i+b)+\sum_{i=1}^{N}a_i \tag{2-8}$$

其中a_i, t_i为Lagrange系数。根据Wolfe对偶定理，式(2-8)分别对求w, a_i, b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a_i，这个规划比原规划要更方便求解，根据对偶变量a_i与原变量w,b的关系，可以求解到法向量w和阈值b。

$$\begin{aligned} &\min \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_ia_jy_iy_jx_ix_j-\sum_{i=1}^{N}a_i \\ &s.t. \quad \begin{cases} \sum_{i=1}^{N}a_iy_i=0 \\ 0\leq a_iy_i\leq C, \quad i=1,\ldots,N. \end{cases} \end{aligned} \tag{2-9}$$

若样本不是线性可分的的，则使用上述模型分类正确率率会降低。SVM理论巧妙的采用了“核函数”（Kernel function）[29]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（2-9）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。核函数的引入巧妙的将高维计算变成了内积计算，极大地提高了计算效率。

2.2.2 Adaboost 分类器

Adaboost[15][30][31][32]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 采用贪婪的迭代方式，每次迭代选择一个最好的弱分类器，最后弱分类器进行线性加权组合，形成一个强分类器。

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 $h_i(x)$ 有一个特征 f_i，一个阈值 θ_i，及用来指示不等式符号方向的函数 p_i：

$$h_i(x)=\begin{cases}1 & if\ (p_i f_i(x) < p_i\theta_i)\\ 0 & otherwise\end{cases} \qquad (2\text{-}10)$$

Adaboost 算法中关于每个弱分类器，需要学习一个最佳的阈值 θ_i。一个弱分类器的训练（特征 f_i）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。

第三章 行人检测特征融合研究

基于 HOG 特征与支撑向量机（SVM），Dalal 等人在 MIT 行人检测数据集上取得了很好的结果，推动了他们建立一个更加具有挑战性的行人数据集——INRIA 行人检测数据集。与 MIT 行人检测数据集相比较，INRIA 数据集的难点主要在于以下三个方面：1、多视角多姿态。MIT 数据集上的行人基本上是平视、正面的居多，而 INRIA 数据集上的行人视角变化大、姿态变化多；2、部分遮挡。MIT 数据集行人之间基本上不存在相互遮挡，而 INRIA 数据集上的行人人群比较多，行人之间存在着相互遮挡，这可能降低了特征和分类器的性能；3、更复杂的背景。MIT 数据集的背景比较单一，而 INRIA 数据集的背景包括了街道、树林、天空等等接近真实世界复杂背景。

如上所述，行人检测有两方面的只要研究内容：一个方面是寻找更鲁棒性的特征描述；另一方面是构建更好的分类器。本章将集中于特征描述子的研究。

3.1 特征提取

针对 INRIA 数据集中的难点问题，X.Wang 等人于 2009 年提出了一种组合特征 HOG_LBP 特征。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，如树枝以及竖直栅栏等，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。但是 LBP 特征的维度太高、提取 LBP 的计算量太大，这些都给 HOG_LBP 描述子的大范围应用带来了困难。

受 X.Wang 等人研究的启发，作者希望通过对现有特征描述子的分析和实验验证的方法，找到一种比 HOG 与 LBP 特征描述子更合适的特征组合，用以进行更鲁棒地行人检测。为了更好的体现检测性能与特征描述子之间的直接关系，所有的特征提取都采用与 HOG 特征相似的提取方式、统一采用 linear-SVM 分类器进行行人检测。下面具体介绍这项工作的内容：本文评测了四种局部特征：HOG、HAAR-like、SURF、LBP。根据实验框架的需要，有些特征的提取与作者的方式相比有一些细微的改动，但是与原作者的实现方式相比没有本质改变。下面分别具体介绍本文中的这几种特征的提取方法：

HOG 特征：Dalal 和 Trigg 于 2005 年提出 HOG 特征描述子，本文采用原

作者相同的特征提取方法。首先计算图像每个像素的梯度方向以及梯度幅值，梯度方向在 0-180 度范围内量化成 9 个 bins，梯度方向大于 180 度的减去 180 度之后量化到相应的 bin。例如：角度为 15 度或者 195 度则对应为第 1 个 bin，如果是 52 度或者 232 度则对应为第 3 个 bin。然后确定一个统计区域，我们称之为 cell，按 bins 统计该区域的梯度，并且用每个像素的梯度幅值进行投票，这样每个区域(cell)就得到了一个 9 个柱形的统计直方图。最后我们按照 Dalal 的建议将相邻的 2×2 个 cell 组成一个 block 用二范数归一化，并且相邻的 block 重叠 50%。取 cell 的大小为 8×8 像素这样一个 64×128 的样本就会产生 105 个 block 一共 $105\times2\times2\times9=3780$ 维的特征描述向量。具体的 cell 和 block 的构成如图 2.5 所示。需要指出的是：本文中的 HOG 特征是可以在彩色图像上提取的，因此对于 R、G、B 三个通道都要分别计算梯度图，每个像素点的梯度的确定是取三个通道中梯度幅值最大的梯度作为该像素点的梯度。

HAAR-like 特征：HAAR-like 特征是由 Haar 小波演变而来的，本文采用最简单的矩形 HAAR-like 特征如图 3.2(d)所示，每个 cell 提取两维特征 Dx, Dy 计算公式如下：

$$\begin{aligned} Dx &= \sum_{X\in left\ subcell} I(X) - \sum_{X\in right\ subcell} I(X) \\ Dy &= \sum_{X\in up\ subcell} I(X) - \sum_{X\in down\ subcell} I(X) \end{aligned}. \tag{3-1}$$

其中 Dx 为左边白色区域图像像素值之和减去右边黑色区域图像像素值之和，Dy 为上边白色区域图像像素值之和减去下边黑色区域图像像素值之，如图 3.2(d)。对于彩色图像分别在 R、G、B 通道上计算 Dx 和 Dy 取 $(Dx)^2+(Dy)^2$ 最大的通道对应的 Dx 和 Dy 为这个 cell 的特征。最后与 HOG 特征类似，每 2×2 个 cell 组成一个 block 做二范数归一化，这样一个 64×128 的样本的一共有 $105\times2\times2\times2=840$ 维。

SURF 特征：Herbert Bay 于 2006 年提出 SURF 特征，SURF 特征是 SIFT 特征的加速版，因此常常是被用来描述特征点的一种特征。本文中利用了 SURF 特征中的一个统计特性每一个 cell 提取一个四维 $(\sum Dx,\sum Dy,\sum|Dx|,\sum|Dy|)$ 的统计特征，其中 Dx 和 Dy 为 HAAR-like 的两维特征。因为这里用到了 SURF 特征描述的 4 维统计特性，因此仍将这种特征称之为 SURF 特征。为了获得统计特征将一个 cell 区域又划分成更小的 2×2 个子区域每个区域分别计算两维 Dx 和 Dy 的 HAAR-like 特征，对 4 个子区域进行统计得到每个 cell 的 4 维的特征描述。最后与 HOG 特征类似按 block 进行二范数归一化，一个 64×128 的样本一共 $105\times2\times2\times4=1680$ 维，如图 3.2（e）所示。

LBP 特征：局部二元模式（Local Binary Pattern，LBP)是 Ojala 等[44][45][46]提出的一种有效的纹理描述方法，该方法通过 LBP 算子来提取灰度图像中局部相邻区域的纹理特征。最基本的 LBP 特征描述子的计算方法在第二章中已经有了简单的介绍，但是第二章中介绍的 LBP 特征描述子并不直接用做行人检测的特征描述子。以下说明为什么原始的 LBP 描述子不直接用来进行行人检测：

一个基本的 LBP 算子如第二章中的图 2.9 所示。为了适应不同尺度的纹理特征，Ojala 等[45][46][52]对 LBP 算子进行了改进，将 3×3 邻域扩展到任意邻域，并用圆形邻域代替了正方形邻域，采用双线性插值算法计算没有完全落在像素位置的点的灰度值。改进后的 LBP 算子允许在半径为 R 的圆形邻域内有任意多个像素点，如图 3.1 所示。符号 LBP_R^P 表示在半径为 R 的圆形邻域内有 P 个像素。

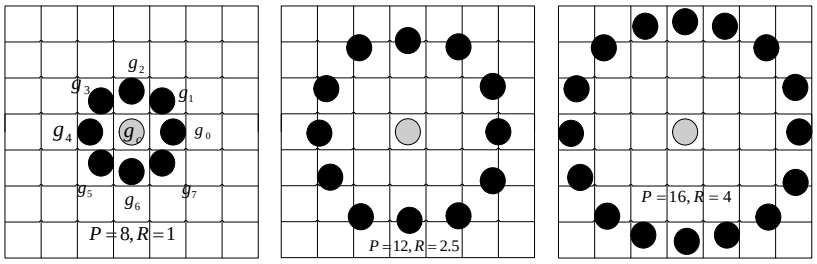


图 3. 不同 P，R 值的 LBP 算子形式

注意到随着采样点数的增加，二进制模式的种类会急剧增加。如 3×3 邻域内 8 个采样点，对应有 2^8 种二进制模式；5×5 邻域内 20 个采样点，对应有 2^{20} 种二进制模式；7×7 的邻域内 36 个采样点，对应的二进制模式种类多达 2^{36} 种。显然，如此多的二值模式对于提取纹理是不利的。

为了解决二进制模式过多的问题，提高统计性，Ojala 等[45][46][52]还利用“等价模式类”(uniformpatterns)对 LBP 进行了改进。当某个局部二进制模式所对应的循环二进制数从 0 到 1 或从 1 到 0 最多有两次跳变时，该局部二进制模式所对应的二进制就称为一个等价模式类，如 00000000，11111111，10001111 都是等价模式类，除等价模式类以外的模式都归为另一类，称为混合模式类。通过改进，局部二值模式的种类大大减少。等价模式类占总模式中的绝大多数，利用这些等价模式类和混合模式类的直方图，提取更好的特征。

本文所使用的 LBP 特征描述子为改进后的 LBP 描述子，与 HOG_LBP 描述子中的 LBP 相同，一个描述子被划分成了 59 个等价类模式，具体等价模式类的划分可参考文献[45][46][52]，因此每个 cell 的维度为 59 维的统计特征。这里的一个 cell 的

大小与前面三种特征的 cell 是不一样的，也是因为 LBP 特征的维度太高的原因，所以这里一个 cell 的大小为16×16 像素，相当于前面几种特征的一个 block 大小，归一化的时候也是一个 cell 的 59 维特征做二范数归一化，所以一个64×128 的样本的特征维度为仍然具有105×59=6195 维，如图 3.2（f）所示。

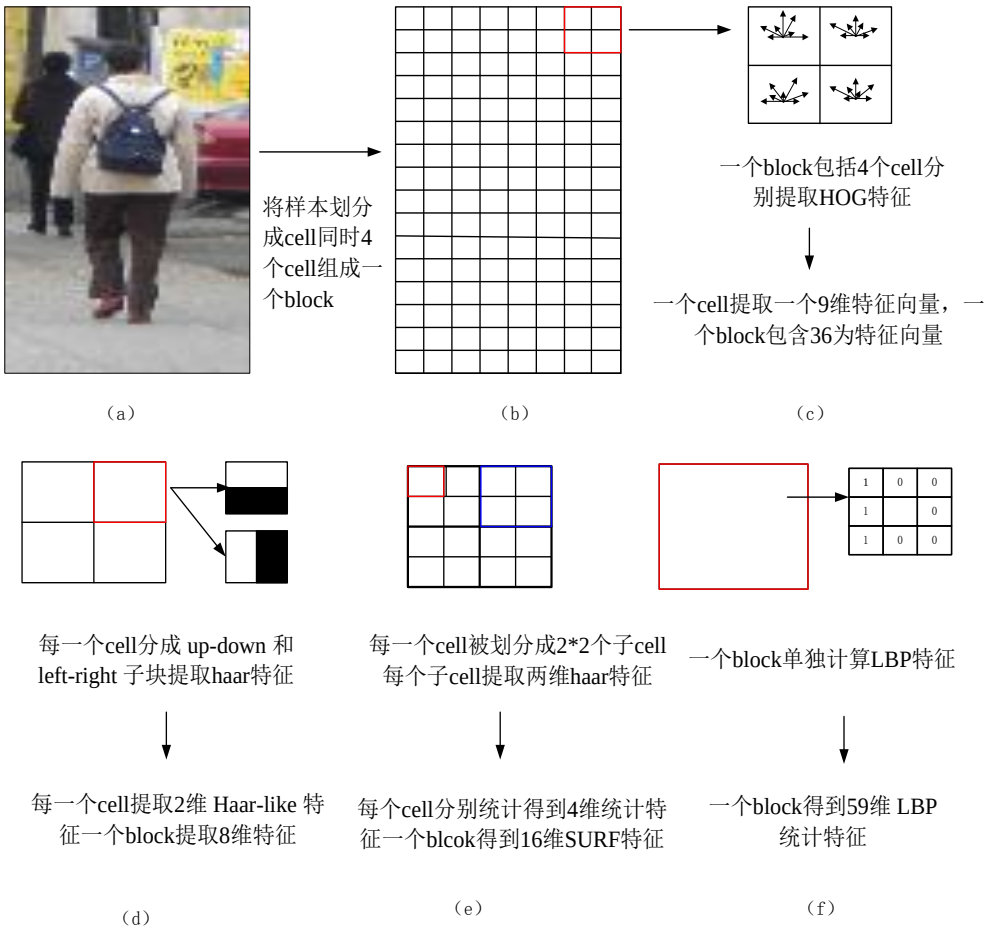


图 3. 特征提取示意图

3.2 模式分析

上一节介绍了本文所评测的几种局部特征描述子的特征提取方法，本小节分析这几种特征描述子对不同模式的区分能力，得出特征描述子之间的互补性关系。分析不同特征对模式的区分能力时，作者从几种简单的模式分析着手，通过在这几种典型的模式上提取相应的特征；然后将特征向量映射到特征空间，通过简单的对比观察这几种特征对不同模式的区分能力。图 3.3 描述了这几种特征描述子在四种典型模式下的分类性能。这四种模式中第一种模式代表“随机噪声的平坦区域”，第二种模

式代表“竖直条形纹理”，第三种模式代表“突变的边缘”，第四种模式代表“渐变的平坦区域”。这几种模式中最能反映“行人”特性的是第三种模式，因为行人目标与背景之间在图像中的表现往往是颜色（灰度）的突变。相应的第四种模式最能反映渐变模式（如天空），因为渐变模式表现是一大片存在颜色（灰度）渐变的区域，而第二种模式能反映出高梯度（树枝或者栅栏）的模式，第一种模式则是比较普遍的包含随机噪声的平坦区域模式。

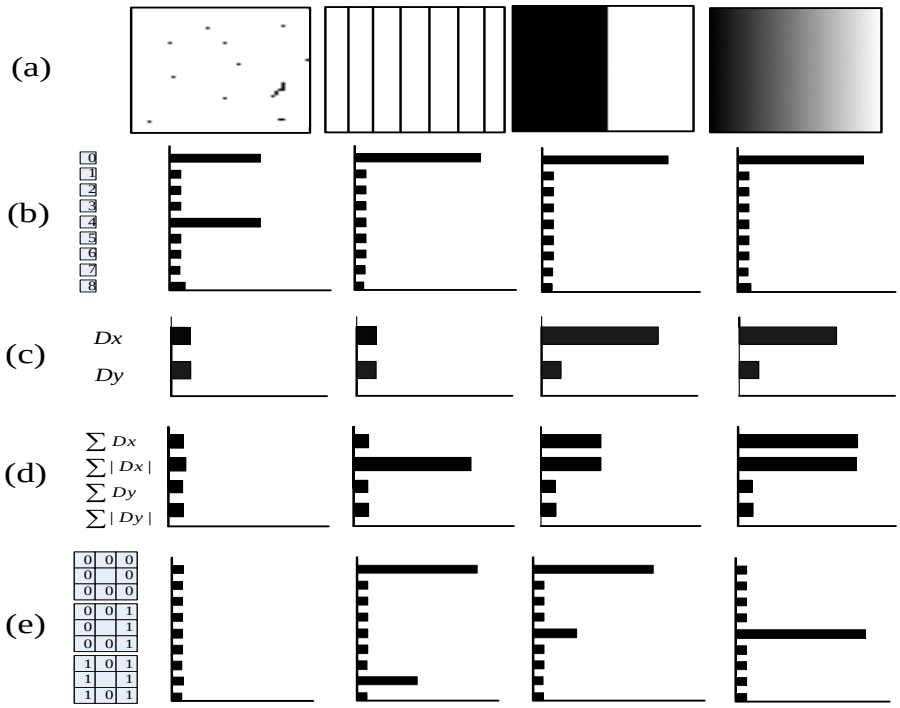


图 3. 特征对不同模式的分类性能分析

四种特征描述子中除了 HAAR-like 特征之外其它三种特征都是一种统计直方图特征，以上图 3.3.b-e 分别为 HOG、HAAR-like、SURF、LBP 这四种特征在四种模式下提取的特征描述子。对于 HOG、SURF、LBP 三种特征来说，图 3.3.a 中的四种不同的模式可以看成是图像区域，这里的特征描述子是单个的 cell 在这四块图像区域中通过“滑动窗口”提取的统计特征；对于 HAAR-like 特征来说，图 3.3.a 中的四种不同模式可以看成是一个 cell 区域，特征描述子就是在这个 cell 上提取的 HAAR-like 的两维特征。从图 3.3b 中可以看出 HOG 特征的 8 维（代表 8 个梯度方向）统计特征描述，图 3.3c 是 HAAR-like 特征的 2 个维度的描述，图 3.3d 是 SURF 特征的 4 个统计特征维度的描述，图 3.3e 本来因该是 LBP 特征的 59 个统计特征维度描述，但是由于画图的关系，只给出了这几种模式中出现的 3 个维度的描述。

由图3.3 中的各个特征的描述子可以看出 HOG 只区分第一种模式和第三种模式、其它两种模式与第三种模式不能相互区分开；HAAR-like 特征区分不了第四种模式和

第三种模式，能够区分第二种和第一种模式；SURF 特征能区分第三种和其他三种模式；LBP 同样能够完全区分第三种模式和其他三种模式。图中只是列举了几种比价简单和典型的模式，对于行人检测而言，需要区分的模式比以上列举的这几种模式要多很多。但是通过这几种简单的模式分析，可以发现在这几种特征描述子之间还是存在着一些互补性关系，而且仅仅通过一种特征描述很难区分第三种模式和其他所有模式，由此也能看出 HOG 特征在进行行人检测分类时将会遇到瓶颈。

从实际的检测结果中更能看出 HOG 特征描述子在某种程度上遇到的瓶颈。如图 3.4 所示，是 HOG 特征和 linear-SVM 分类器检测出来的一些错误的窗口。

图 3. HOG-SVM 方法错误分类的窗口

图 3.4 是在实验的两个数据集上，采用 HOG 特征以及 linear-SVM 分类器进行实际的行人检测时出现的误检的窗口实例。可以看出第一行的图片与我们上面分析的四种模式中的第一种模式“随机噪声的平坦区域”是对应的，第二行的图片与第二种模式“竖直条形纹理”是对应的，第三行的图片与第四种模式“渐变的平坦区域”是对应的。这从实验上验证了单独使用 HOG 特征在这四种模式时存在的瓶颈问题。

从特征描述子的本质来分析，我们知道 HOG 特征是梯度统计直方图，它关注的是梯度信息也就是图像上的“边缘”信息，而并不关心边缘是否与“实体”相关联。我们可以想象一下如果我们用一根铁丝将其弯折成一个人的形状，那么对于 HOG 特征来说它与一个真正的行人的边缘特征是一致的，会不加区分的对待。而实际上行人在图像上的表现是不仅仅是有边缘信息还是有“实体”也就是有“厚度”的，这中有厚度的纹理信息通过 HOG 特征不能很好的描述。因此通过以上的分析我们知道

HOG 特征与其它特征的组合是提高特征鲁棒性的一个可行的也是必要的研究内容。

3.3 特征融合方法

3.2 节从几种简单模式着手，分析了 HOG 特征存在的瓶颈问题，说明了研究 HOG 特征与其它特征组合，也就是特征融合的必要性。本节将给出 HOG 特征与其它的特征融合的两种方法。

X.Wang 等人在论文[2]中提出了一种直观的特征融合的方法，这种方法将提取的 HOG 特征与 LBP 特征直接“串接”成一个新的特征向量，即 HOG_LBP 特征描述子。假设 HOG 特征描述子的特征向量是 x_1，LBP 特征向量是 x_2 则“串接”之后新的特征向量为 $x=(x_1,x_2)$。

论文[2]中作者还提出了一种“遮挡概率图”的概念，本文将在分析这个概率的基础上提出一种新的融合的方法。这里先介绍一下论文[2]所提的“遮挡概率图”的概念，并分析这种概念图的物理意义。如图 3.2 所示我们将 64*128 的样本网格化，划分出 105 个 block，每个相邻的 block 之间相互有 overlap。在每个 block 上提取相应的特征采用二范数归一化，将样本上所有 block 上提取的特征级联成一个特征向量用来描述该样本的特征称之为特征描述子，这种特征描述子是直接在图像像素上提取得到的，因此也叫初始特征描述子。如果对所有 block 上提取的特征通过某种加权得到新的特征描述子，这样的描述子称之为中层特征描述子。根据文献[2]中的方法，将这种中层特征通过 Mean-shift 的方法进行分割可以得到“遮挡概率图”。获得这个“遮挡概率图”的方法是：

1、将直接从样本上提取得到的特征描述子经过 linear-SVM 训练得出分类模型，对于每一种特征对应的分类器可以写成如下形式：

$$\mathrm{f(x)=w}^T x+b \tag{3-1}$$

其中 $x=(x_1,x_2,\cdots;x_{105})$ 对应的 w^T 也可以表示成 $w^T=(w_1^T,w_2^T,\cdots,w_{105}^T)$ 的形式。为了计算的方便，根据文献[2]的方法我们将 b 也表示为 $(b_1,b_2,\cdots,b_{105})$。这种形式的描述是为了将 x,w,b 这三个量分别与图 3 种的 block 划分对应。

那么根据文献[3]中的描述，中层特征的特征向量 v 可以表示为：

$$v=\begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & w_{105} \end{pmatrix} * \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{105} \end{pmatrix} \tag{3-2}$$

2、中层特征向量 v 可以描述为 $v=(v_1,v_2,\ldots v_{105})$，$v$ 的每一个分量为对应的一个 block 的得分。文献[2]的作者认为如果 $v_i<0$ 就表示这个 block 对应的图像区域有遮挡，通过对样本窗口中所有 block 的得分也就是 v_i 的值采用 Mean-shift 分析的方法可以将样本窗口分割成有遮挡和没有遮挡的部分。有遮挡部分的图像区域对应的 block 的得平均得分值小于零，而无遮挡图像区域对应的 block 的平均得分值大于零。

本文作者通过对无遮挡部分单独训练一些部件检测分类器，重新对检测窗口进行分类以此提高检测精度。本文作者受“遮挡概率图”的启发（更直观的说是 block 的得分 v_i 能够指示样本是否被遮挡），提出了另外一种特征融合的方法：通过中层特征与初始特征融合的方式得到一个新的特征描述子称其为结构特征描述子。

为了从直观上理解这种中层特征 v 我们在样本上提取 HOG 特征，用 linear-SVM 训练得到一个分类器 $f(x)=w^Tx+b$，然后提取中层特征 v。我们将 w^T 和 v 可视化之后得到如图 3.5 所示的效果：

图 3. 中层特征 v 以及分类器权值 w^T 可视化效果图

图 3.5(a) 和(c)为正例样本，图 3.5(b)和(f)是 w^T 的可视化效果，图 3.5(c),3.5(g)是中层特征 v 可视化的效果，从效果图我们可以看出在行人边缘的 block 的得分值普遍大于零。本文第二种特征融合方法首先利用 HOG 特征训练分类器，提取中层特征 v，将 v 与另一种初级特征通过“串接”形成一个新的特征描述向量 $\tilde{x}=(x,v)$，我们称特征向量 $\tilde{x}$ 为结构特征描述子。对于结构特征向量我们同样可以训练一个 SVM 的分类器：$f(x)=\beta^T\tilde{x}+b$。

3.4 特征融合实验及结论

上一节我们介绍了特征融合的两种方式，本节将介绍特征融合的实验数据集以及实验的框架和结论。本文对于特征描述子的研究分为三个步骤，第一步，将四种特征描述子 HOG、SURF、LBP 以及 HAAR-like 的特征，统一采用 HOG 局部特征描

述子的特征提取方式，3.1 节介绍了特征提取的细节。将样本上提取的特征通过 linear-svm 做交叉验证得到四种特征描述子在行人检测方面的分类能力。第二步，分析四种特征描述子在不同模式下的分类性能，发现特征之间存在补充性关系。通过两种特征融合的方式验证特征之间的互补性关系，这两种融合方式一种是直接在原始特征上融合（直接融合），一种是我们称之为结构性特征融合。同样通过交叉验证来得到这些融合特征在行人检测方面的分类性能。第三步，在 INRIA 测试数据集上验证实际检测性能。

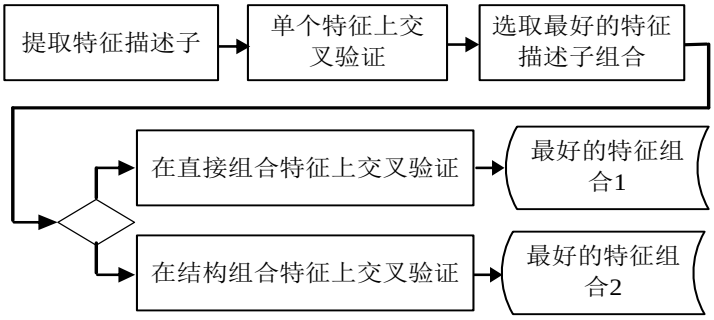


图 3. 特征融合实验框图

3.4.1 实验数据集

本文选取了两个公开的行人检测样本集，它们分别是SDL[8][47]和INRIA[38][48]行人检测数据集。读者可以在文献[38][47]中提供的网址中下载相关数据。以下是这两个数据集中部分的行人训练样本：

图 3. INRIA 和 SDL 部分正反例训练样本图示

图 3.7 中第一排的样本来自 INRIA 行人检测数据，第二排的样本来自 SDL 行人检测数据集。其中 INRIA 数据集标定的正例样本 1238 个通过镜面对称之后一共有样本 2476 个，SDL 数据集有 4300 多个正例样本，主要包括正面和侧面两种视角的样本。反例包括 INRIA 数据集提供的 1239 幅反例图片以及 SDL 数据集提供的 6000 余个反例样本集，所有的样本窗口大小为64×128个像素，每个样本左右两边的 “空

白边”（margin）为 12 个像素。

3.4.2 实验结论

实验分为三个步骤：第一步，单独评测四种特征描述子在行人检测方面的分类能力，在样本上提取的特征通过 linear-SVM 做交叉验证得出四种特征描述子的分类能力。第二步，将样本上提取的特征通过两种融合方式得到的新的特征描述子，分别通过 linear-SVM 做交叉验证得出各种特征组合的分类能力。这两步评测的标准都是交叉验证的准确率，交叉验证的参数为惩罚因子$C=0.01$折叠次数$V=5$。这两步具体的实验结果如下图：

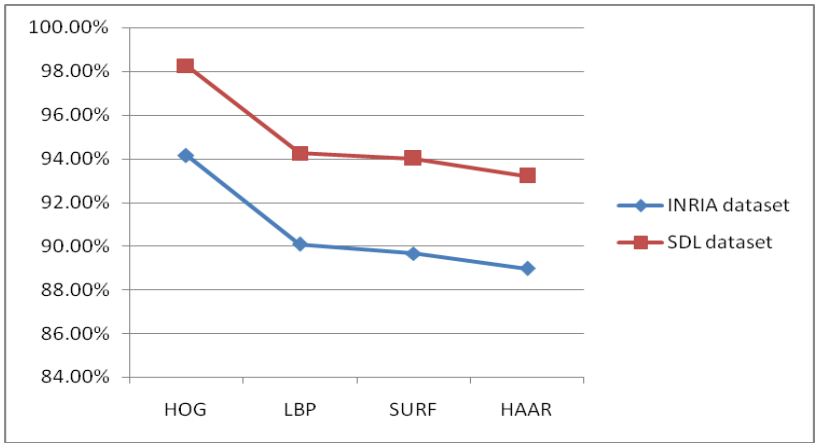


图 3. 单个特征交叉验证实验结果

本文选取了两个公开的行人训练样本集，它们分别是 SDL 行人检测数据集和 INRIA 行人检测数据集。图 3.8 为四种特征单独描述时的分类性能，从交叉验证的结果可以看出四种特征描述子中，HOG 特征在两个数据集上的分类效果最好，其次是 LBP 特征，而且这些特征在 SDL 数据集上的性能普遍比 INRIA 上的性能高。

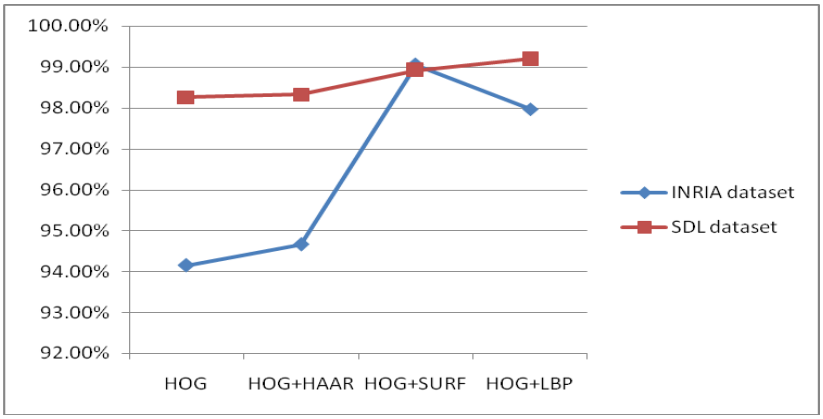


图 3. HOG 与其它特征直接融合实验结果

图 3.9 为 HOG 特征与其他几种特征通过直接“串接”的融合方式获得新的特征描述子的分类性能，从交叉验证的结果可以看出在直接融合的方式下，HOG 特征与 SURF 特征的组合在 INRIA 数据集上效果最好，而且性能提升了近 5 个百分点，其次是 HOG 特征与 LBP 特征的组合，在 SDL 数据集上 HOG 特征与 LBP 特征的组合的性能比 HOG 特征与 SURF 特征的组合的性能高 0.5 个百分点。

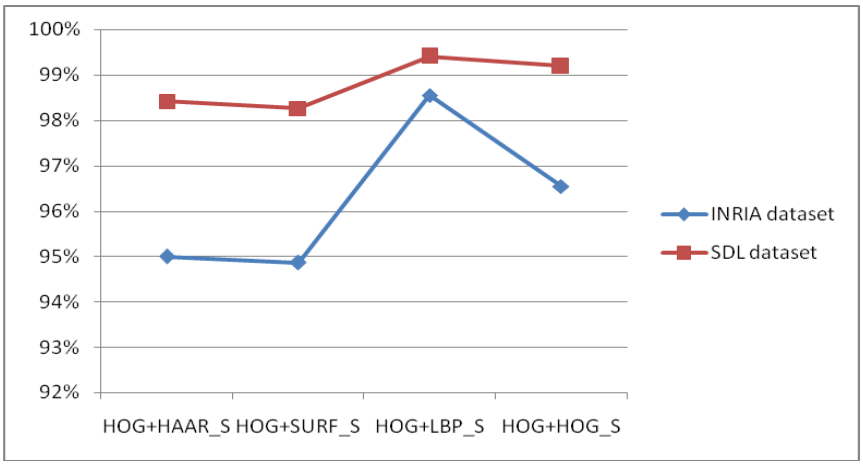


图 3. HOG 与其它特征结构融合实验结果

图 3.10 为 HOG 特征与其他几种特征通过中层特征融合的方式获得的结构特征描述子的分类性能，从交叉验证的结果可以看出在中层特征融合方面，HOG 特征与 LBP 特征的性能最好，但是比直接融合的性能低 1 个百分点左右。

总的来看在直接特征融合的方式下 HOG+SURF 特征的交叉验证结果与 HOG+LBP 的交叉验证性能非常接近。在中层特征融合方面，HOG 特征与 LBP 特征的组合性能最好，但是比直接融合的效果差，并且需要进行二次训练。因此在第三步，在 INRIA 测试数据集上进行行人检测测试的时候，我们只采用了直接特征融合的方式。从特征的维度来看，在直接融合的情况下 HOG+SURF 特征的维度为 5460 维而 HOG+LBP 特征的维度为 9975 维，从交叉验证的结果来看，HOG_SURF 特征是一种比 HOG_LBP 特征更合适的行人检测特征描述子。

实验的第三步，在 INRIA 测试数据集上验证特征描述子的实际检测性能。从前两步已经得出 HOG 特征与 SURF 特征以及 LBP 特征的互补性最好，这两种特征组合方式最终的测试性能如何，作者在 INRIA 测试数据集上做了相关的比较，实验的性能曲线如图 3.11 所示：

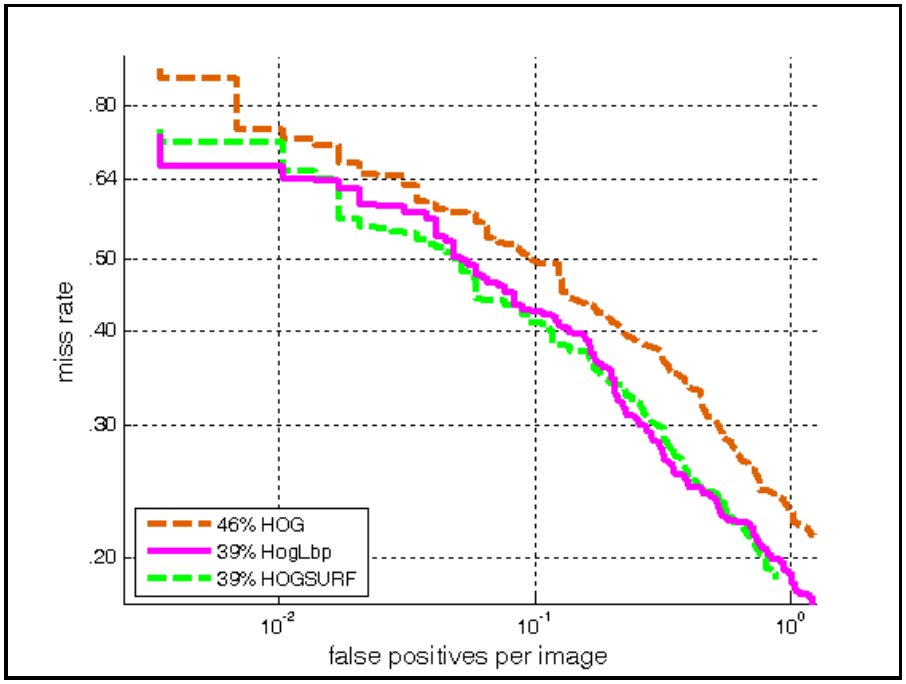
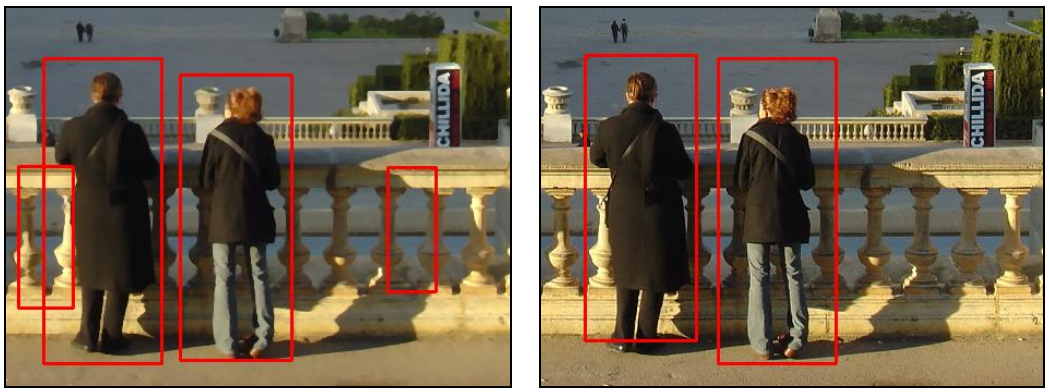


图 3. INRIA 测试数据集上的测试结果

图 3.11 所示的评测曲线是目前行人检测评测的标准，有关行人检测评测的方法可以参考文献[4]，在训练分类器的过程中，采用了难样本挖掘的技术，有关资料可以参考论文[3][34]。从性能曲线可以看出 HOG_SURF 和 HOG_LBP 特征描述子都比 HOG 特征描述子有不少性能提升，且 HOG_SURF 特征的性能与 HOG_LBP 特征的性能接近，但是 HOG_SURF 特征的维度为 5460 维而 HOG_LBP 特征的维度为 9975 维。因此可以得出结论：在行人检测特征描述子方面，HOG_SURF 特征比 HOG_LBP 更合适。以下图 3.12 是 INRIA 行人检测数据集上部分检测结果示例，对比图的左边为 HOG 特征的检测结果，右边为 HOG_SURF 特征的检测结果：



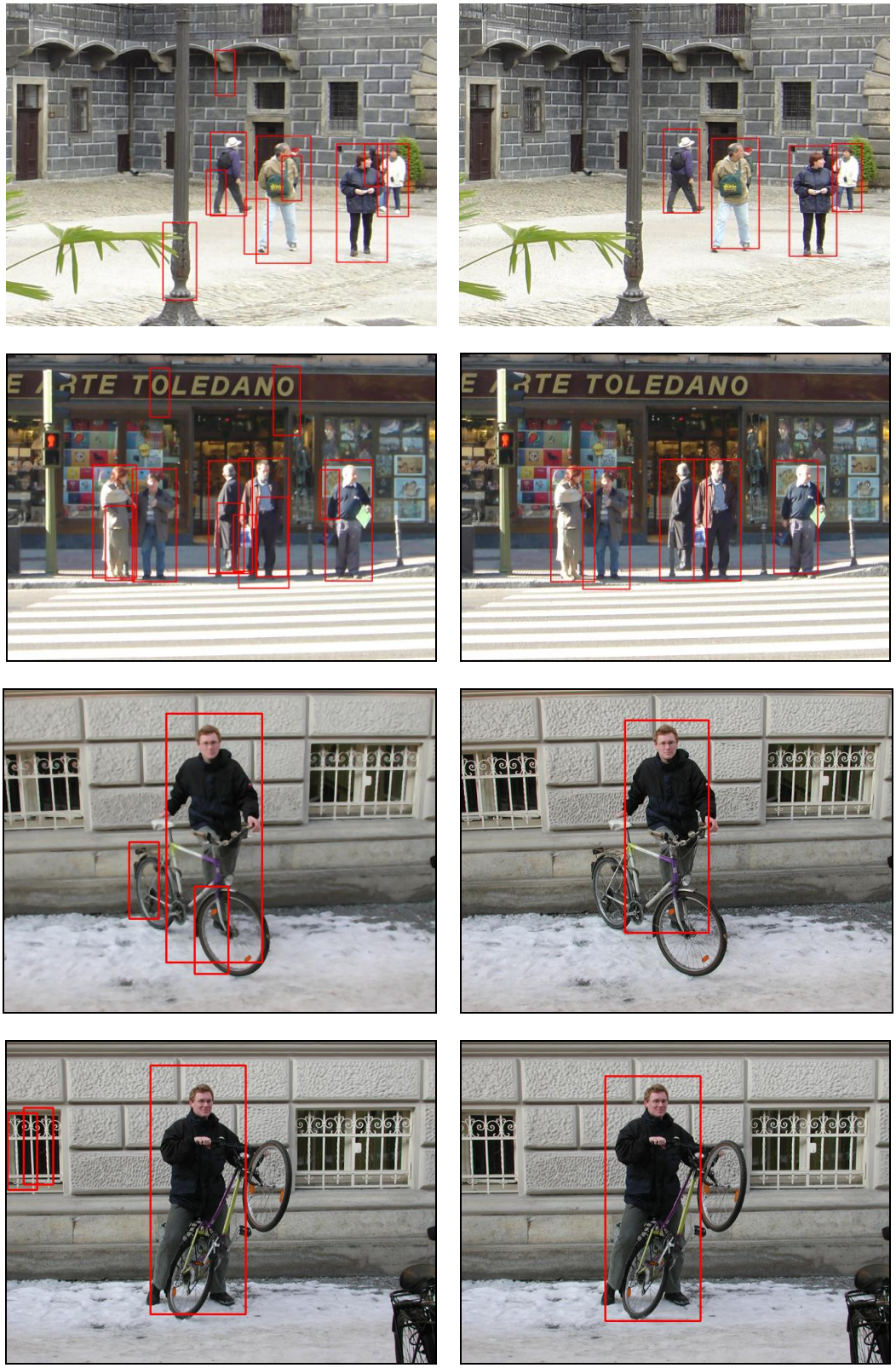


图 3. INRIA 数据集上部分检测结果

第四章 行人检测分段混合模型

第三章介绍了如何在样本上提取特征，研究了几种特征之间的互补性关系。样本提取特征之后，如何在特征空间中构建分类面，区分正例样本和反例样本是一个监督学习的问题。本章主要介绍本文在行人检测视觉模型和分类器构建方面所做的工作，在 4.3 小节还将简单介绍一下行人检测的应用实例。

这种分类器的学习方法大致可以分为两种：第一种是基于概率的方法（probabilistic method），另一种是基于判别的方法（discriminative method）。近年来，目标检测领域的研究表明基于判别的方法要优于基于概率的方法。本文研究分类器的构建主要也是基于 SVM 的判别方法。

文献[1][25][48]于 2005 年通过在 HOG 特征空间中构建一个 linear-SVM 的分类器，在早期的行人检测数据集 MIT 行人检测数据集上取得完美的分类性能达到 95%以上，于是作者建立一个更具有挑战的行人检测数据集—INRIA 行人检测数据集。后来又陆续出现了更有挑战性的行人检测数据集。如 Daimler[37][41]，ETH[37][42]，Caltech 与 TUD-Brussels 行人检测数据集等等。其中，在 INRIA 这个数据集上的研究有两个比较重要节点：一个是 2009 年 HOG_LBP 特征描述子的提出从特征的角度提升了性能；第二个是 2011 年 P. F. Felzenszwalb 等人提出的形变模型（discriminatively trained part based models）从方法的角度使得检测性能有了质的提高。应该注意到，而这两项工作使用的分类器都是基于 SVM 的分类器。

有关支撑向量机（SVM）的相关理论在 2.2.1 小节中我们做了简单介绍，针对数据非线性可分的问题，在 2.2.1 节中介绍的方法是通过“核函数”的办法来解决的，本章将介绍另一种在数据非线性分布时解决分类问题的方法。文献[49]中，作者提出了分片支撑向量机的理论，首先将特征空间剖分成若干子空间，在每个子空间中基于支撑向量机构造一个最优分类面，然后，将各个分类面链接起来构成一个分片最优分类面以逼近理论上的最优分类超曲面。同时，文中还从理论上分析探讨了其推广能力的界，为分片支撑矢量机模型提供了坚实的基础。这个方法很好的解决了特征空间中，正反例样本线性不可分的问题，但是其难点是如何划分子空间。

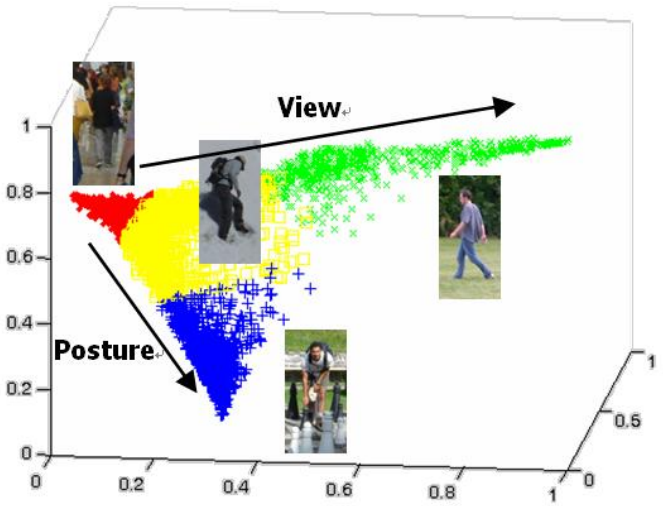


图 4. 流形空间中正例样本分布示意图

与文献[49]类似，本文中介绍的分段线性 SVM 也是要解决特征空间中，正反例样本线性不可分的问题。与[49]中方法不同的是本文中的分段线性 SVM 不是进行子空间剖分而是通过聚类的方法，先对样本空间中的正例样本进行聚类分成几个类别。然后利用每个类中的正例样本和所有的反例样本构建这个类的分类器，最后对所有类别通过构建 ECOC 编码分类器的方式对未知样本进行分类。这么做的理由是：如图 4.1 所示，行人检测是一个特定的分类问题，作为行人的正例样本存在不同的视角和姿态问题，它在特征空间中的分布可以描述为一个高维空间的流形。如图 4.1 不同的姿态和视角可以通过聚类的方法聚成几个类别，并且各个类之间还存在一些邻接关系，因此我们首先将正例样本进行聚类，划分出不同视角和姿态的类别，每个类别的正例样本与反例样本分别训练一个线性的 SVM 分类器，然后利用 ECOC 来描述各个类之间的邻接关系从而构造了一个分段线性分类器。高维空间中的聚类问题是模式识别中的难点问题，本章 4.1 节将介绍作者实现的几种样本划分的方法，4.2 节将介绍 ECOC 的基本理论以及基于 ECOC 编码的混合分类器的构建，4.3 介绍基于流形聚类和 ECOC 编码混合分类器的实验结果，4.4 节将介绍应用实例。

4.1 正例样本分段方法

视觉模型是实现鲁棒性目标检测的重要一环，好的视觉模型能有效的提高检测性能。最简单的视觉模型是简单窗口模型（Sample Window），通常也叫整体模型，即将整个人体目标标准化到样本大小，通过尺度变换和滑动窗口的方

式在图像上进行目标检测。MIT 实验室 Mohan and T. Poggio[12][50][51]等人在早期的行人检测时提出了一个基于部件的视觉模型，作者将人体样本划分为几个部件（Components or Parts），每个部件分别提取相应的特征并训练分类器，最终将各个分类器经过投票得出最终的分类结果。实际上这种部件模型的理论基础就是分片支撑向量机，每一个部件的特征空间可以看成是整个样本特征空间的一个子空间，这样的好处就是能够处理行人目标的遮挡和部分的多姿态，多视角问题。基于部件的可形变的视觉模型与基本的部件模型的不同之处在于，各个部件可以有相对位置的移动，而这种相对移动通过一个参数模型来描述，并通过训练学习得到参数作为移动位置大小的约束。这样不仅能解决目标的遮挡问题，还能够解决相同类别目标个体之间的差异。比如同样是行人目标的胖瘦、高矮不同。形变模型能够将形变后的目标通过“形变”与标准样本进行位置对齐，从而保证同样位置的特征是近似的，进而解决模式分散的问题。

在本文中，我们提出了一种错误纠正码（Error Correcting Output Codes，ECOC）的方法来解决行人检测中的视角与姿态的问题。由于各种视角和姿态的人体目标可以看成不同视角和姿态的行人聚类而成的高维流形，因此 ECOC 的分类器比一般的基于视角的分类器在处理人体的多视角和多姿态方面更具有鲁棒性和准确性[54]。这个方法的难点是如何对正例样本进行划分，早期行人检测公开数据集在标定的时候是不区分行人的视角和姿态问题的，把所有的行人目标都标定为“行人”也就是正例窗口。但是在实际检测过程中我们发现这种多视角和多姿态问题给线性 SVM 分类器带来了很大问题，为了解决这种问题又需要将样本根据视角和姿态进行划分。正例样本的划分第一种方法是标定的时候我们就根据不同视角和姿态进行标定，这个方法比较直观，但是样本的分布是一个连续的“高维流形”，人在主观标定的时候不能很好的标定各中视角和姿态。第二种方法是通过聚类的方法将原有的样本聚类成不同的视角和姿态，这个方法比较简单，但是高维空间的聚类效果却不一定能达到预想的结果。以上的两种方法各有优缺点，我们在接下来的三个小节将介绍作者实现的三种不同的样本划分的方法，其中第一种是基于流形空间的聚类方法，第二种是基于标定样本的长宽比例的聚类的方法，第三种是通过标定典型的样本学习得到各个视角和姿态的分类模型，然后利用分类模型进行样本划分的方法。

4.1.1 基于流形聚类的分段方法

LLE[53]一种最近提出的非线性降维算法，也是一种非监督学习算法。它能

够使降维后的数据保持原来的拓扑结构，已经广泛的应用于图像数据的分类与聚类、文字识别、多维数据的可视化、以及生物信息学等领域中。LLE 算法能保持流形的邻域不变性，是其它线性方法都不能比拟的，因此 LLE 算法可以应用于样本高维特征空间中的聚类。LLE 算法操作简单，且算法中的优化不涉及到局部最小化。该算法能解决非线性映射，但是，当处理数据的维数过大、数量过多、涉及到的稀疏矩阵过大时不易于处理。如果数据分布在整个封闭的球面上，LLE 则不能将它映射到二维空间，且不能保持原有的数据流形。所以我们在处理数据中，首先假设数据不是分布在闭合的球面或者椭球面上。

LLE 算法利用线性重构的局部对称性找出高维数据空间中的非线性结构，并在保持各数据点临近位置关系情况下，把高维空间数据点映射为低维空间对应的数据点。其计算步骤包括：计算、寻找数据点或邻居数据点、构造数据点及计算权值矩阵，并通过权值矩阵计算低维向量。以一个多视角的人体流形为例，伴随着特征维数的降低，沿着流形逐步变化的样本点具有重要的意义。综上可知，通过降维运算后而创建的人体流形就涵盖了多个视角、姿态的人体，为解决多视角多姿态问题提供了条件。

本文采集了众多包含着不同视角和姿势的人体样本来构建人体流形。采用 HOG 特征来表述这些样本。每一个样本被 3780 维特征所表述，HOG 特征抽取的详细内容，已经在本文的第二章中表述过。假定 n 样本点 $\{x_i\}, i=1,\ldots,n$，LLE 算法有以下三步组成。首先，需要创建一个临界图。一对样本点 (x_i, x_j) 被连接了起来（其中，x_i 是 x_j 的 k 近邻）。第二，LLE 将会计算在正例训练样本中任意两个数据点之间的最短路径。对于每一对临界图中非临界的样本点，LLE 将会在临界图中通过距离最大的操作找到最短路径。当任意两对之间的测量距离都被正确计算出来时，一个初级的流形就这样建立了起来。LLE 同时将高维样本映射到一个低维空间，并保持流形领域不变性。对于样本 x_i 的 k 近邻计算权值 $w_{ij}(j=0,1,...,K-1)$，这些权值就会更好的线性重构 x_i；接着，LLE 会计算 w_{ij} 在低维嵌入空间上最好重构向量 z_i。最后经典的多尺度放缩(MDS)被用来重新构建一个低维嵌入数据这些数据很好保持了流形的本质几何特征。在这个低维嵌入空间上进行样本聚类。

聚类的目的就是在流形上寻找线性的簇。对于一个线性的簇，大量研究证明现有分类器都可以很好的将其分类，比如，本文所采用的 Linear SVM 分类

器。采用一个由低到高的分层聚类去构建流形中线性的聚类（“linear clusters”），分层聚类算法是一个聚类或者是样本合并的过程。在此过程中聚类被不断的合并，留下来的聚类越来越少。在实施样本的合并算法过程中，需要考虑以下两个方面的因素：1）保持尽可能少的聚类以使得 ECOC 编码简单和基础分类器更少；2）保证尽可能多线性的聚类以使得基础的线性分类器得较好的检测效果。基于以上的矛盾聚类的数目需要从实验中获得。

假设包含有L聚类的流形 $M=\{C_1,C_2,\cdots,C_L\}$ 其中 $C_l|_{l=1}^{L}=\{z_1^{(l)},z_2^{(l)},\cdots,z_{nl}^{(l)}\},(\sum_{l=1}^{L}nl=N)$ 表示流形上的聚类，nl 表示第l个类的样本数目。z 表示一个嵌入空间上的样本。首先，需要计算两个样本(z_p,z_q)在嵌入空间的欧式距离 $D_g(z_p,z_q)$ 及侧底线距离 $D_e(z_p,z_q)$。接着，定义一个各对之间的非线性比率 $R(z_p,z_q)=D_g(z_p,z_q)/D_e(z_p,z_q)$。因为测量距离总是不小于欧式距离。那么 $R(x_p,x_q)\geq 1.0$ 总是存在。为了测量流形上所有聚类非线性度，定义一个非线性的度量函数来表示所有聚类非线性比率：

$$R=\frac{1}{L}\sum_{l=1}^{L}R(C_l)=\frac{1}{L}\sum_{l=1}^{L}(\frac{1}{Nl^2}\sum_{n=1}^{Nl^2}R(z_p,z_q)) \tag{4-1}$$

在执行从下到上的聚类算法的过程中，聚类的数目在减少，然而非线性度量 R 将会增加。聚类的数目 L 和非线性 R 可以构成如下图所示的曲线。

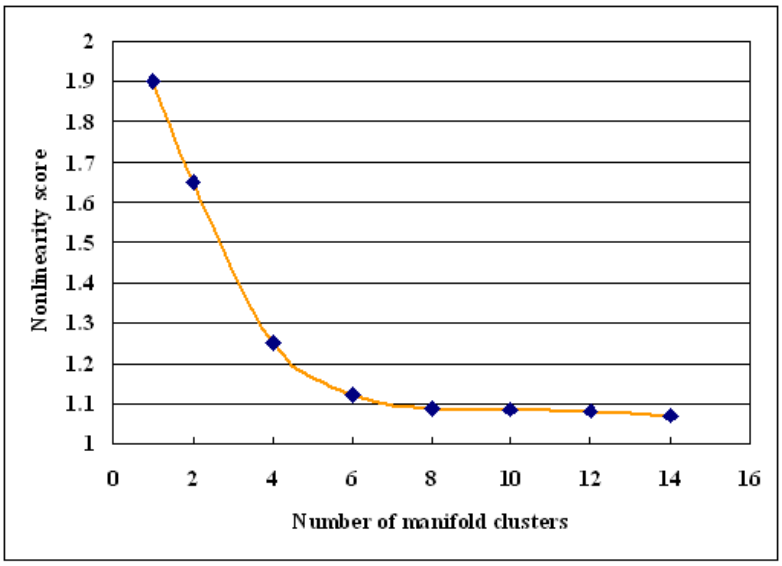


图 4. 非线性度量随聚类数的变化曲线

为了确定给定训练样本的聚类数目，选择曲线上满足 $\frac{dR}{dL}>T$ 对应的类别数，T 表示最大的非线性度，根据实验选为 1.2 即类别数选为 4。

4.1.2 基于标定样本长宽比的分段方法

基于部件的可形变的视觉模型可以部分的解决多姿态多视角的问题。但是部件模型更多的是针对遮挡问题，而形变模型是解决特征对齐的问题。因此 P.F. Felzenszwalb 等人在他们的实验框架中加入了“混合模型”的概念，混合模型主要是解决多姿态多视角问题。混合模型一共有 M 个子模型，每一个子模型又是有 P 个部件模型组成。有关混合模型的更多内容读者可以参考文献[3][34]，这里就不再过多展开了。作者混合模型中的 M 个子模型其实就是一种分段模型，每个子模型代表了一个视角或者姿态，这些子模型对应的正例样本的获取，作者也给出了一种按标定样本长宽比的统计直方图划分样本的方法。

样本标定在各个数据集不尽相同，对于行人检测的评测方法也各有区别，这就给大家的算法评测带来了困难。Piotr Dollar 等人于 2011 年给出了一个行人检测的评测框架，并对行人数据标定给出了一个标准，是总结近些年行人检测工作的一篇比较完善的文献，相关内容读者可以参照文献[4][36]。与 INRIA 数据集类似，标定的时候 Piotr Dollar 等人提出的标准都是不留边的（No Margin）标定，所不同之处在于 INRIA 数据集对于远近视角，是否存在遮挡问题没有相应的标定，而 Piotr Dollar 等人提出的标准对这些都进行了标定。我们要介绍的通过标定样本长宽比的统计直方图划分样本的方法中的样本标定准则与 INRIA 数据集的标定准则是一样。标定、训练样本与评测窗口的示意如图 4.3 所示。

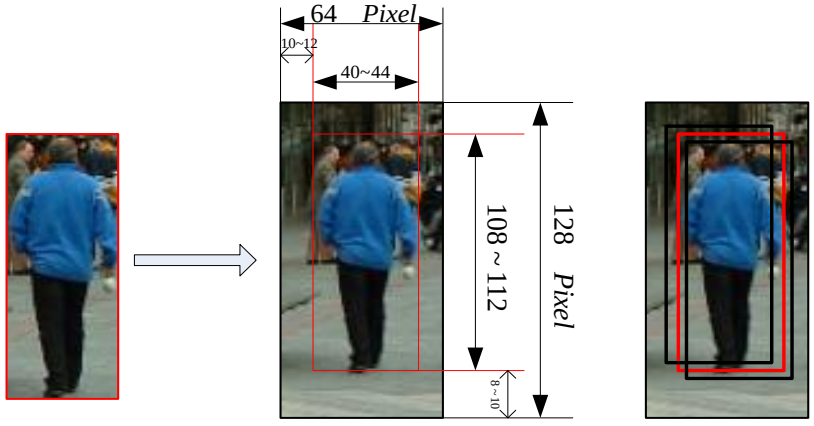


图 4. 标定与样本窗口关系示意图

图 4.3 左边第一幅图就是数据集标定的行人窗口。中间的图片就是根据 Dala

的实验给出的样本窗口，一般左右留有 10~12 个像素的边，上下留有 8~10 个像素的边，样本标准化到 64*128 个像素的大小。右边这幅图片中，红色的矩形框是标定的窗口，两个黑色的矩形框是可能的检测窗口，是否检测成果的评测标准是：红色矩形框与黑色矩形框相交的面积比上两个矩形框相并的面积大于某一个阈值即为检测成功，否则是错误检测窗口。

在做样本集的划分时用到的是第一幅图片中的样本窗口信息，假设标定样本窗口的宽为W 高度为H，则样本高度比上样本宽度$r = H/W$ 会随着样本的视角和姿态而改变，通过r 的统计直方图我们可以将样本划分成几个等分的类别。

4.1.3 基于学习的分段方法

样本集的划分我们还有另外一种方法，就是通过人为标定的办法标出一些典型视角和姿态的样本，通过这些样本学习这几个视角的分类器，然后通过这些初始的分类器将整个样本集划分成几个子集，最后利用划分的正例样本的子集重新训练各个视角的分类器。

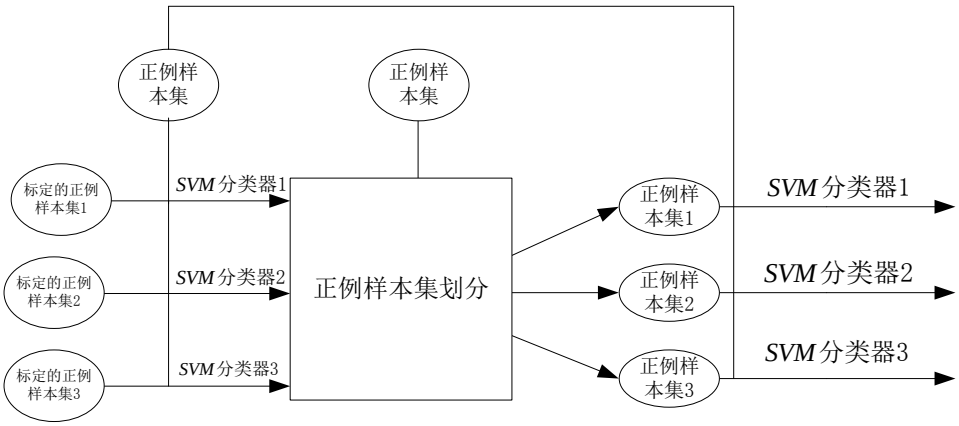


图 4. 基于学习的分类方法示意图

如图 4.4 描述了基于学习的样本划分方法，标定的正例样本集 1，2，3 是人为标定的少量样本，它们分别代表几个典型的视角或者姿态。利用初始分类器进行正例样本集的重新划分时规则如下：设初始标定了K 个正例样本集，训练得到K 个初始分类器 $f_k(x)$，第i 个样本 x_i 的类别l 的计算方法如下：

$$l = \arg\max_{1 \le k \le K}(f_k(x_i)) \tag{4-2}$$

在上述分类过程中反例样本集在分别与标定的正例样本集训练初始模型以及与分类后的正例样本集训练最终分类器的过程中保持不变。

4.1.4 分段方法小结

前三小节分别介绍了三种正例样本的分段方法，这三种方法究竟哪一种分段的方法更合适呢？本小节将给出作者的一个实验结果供读者参考，为了简单起见我们用以上的三种方法将正例样本都聚成四个类别，各个线性分类器的混合方法也用如下最简单的方式：正例样本被聚成四个类，每一个类的正例样本分别与反例样本训练得到一个线性 SVM 分类器，对于待分类的样本，任意一个分类器将其分为正例则认为这个样本是正例样本。

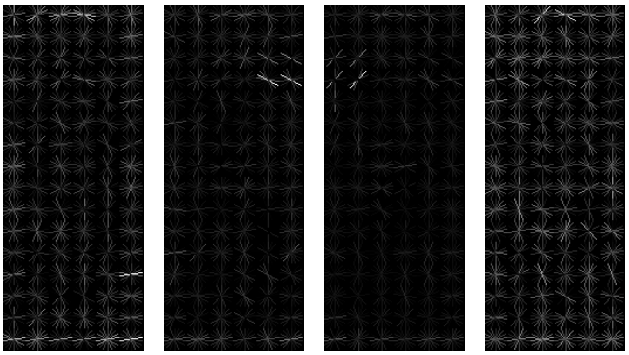


图 4. 基于流形聚类的分段模型 W 可视化

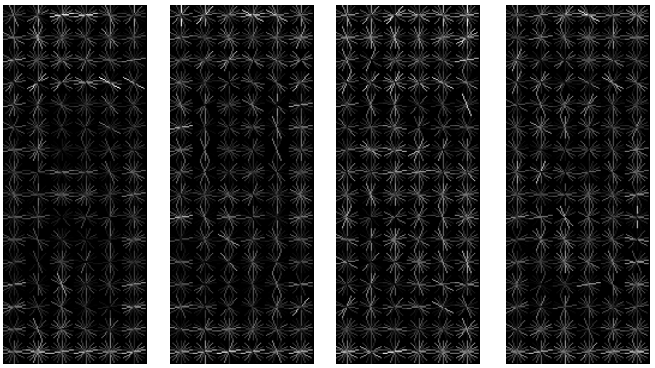


图 4. 基于学习的分段模型 W 可视化

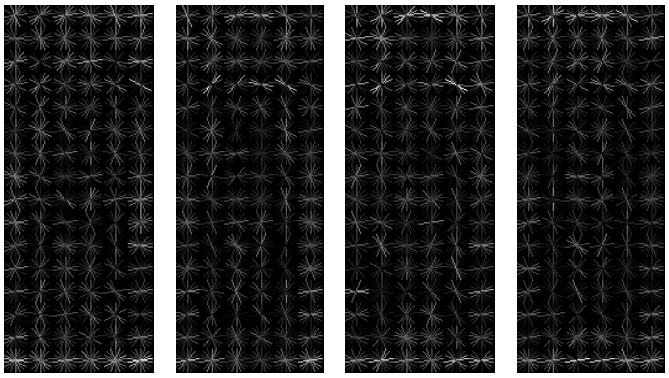


图 4. 基于样本长宽比的分段模型 W 可视化

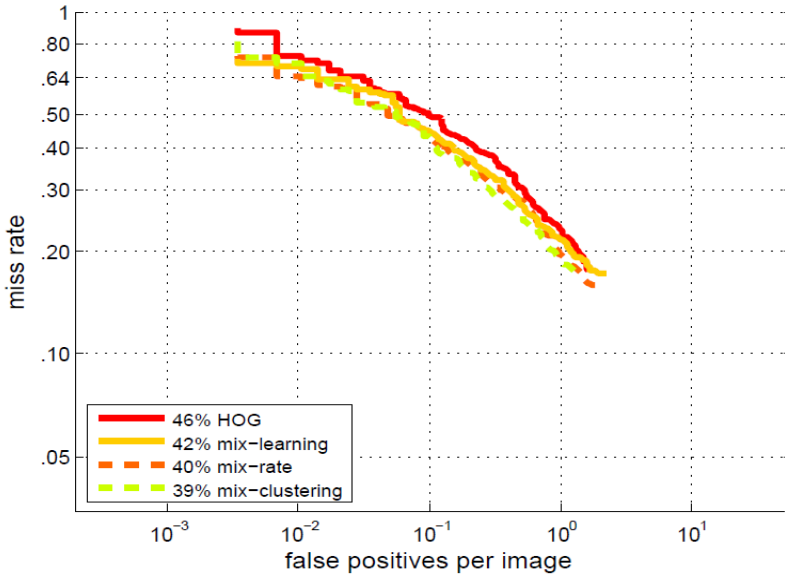


图 4. 各分段模型 INRIA 上的检测性能曲线

如图 4.5、4.6、4.7 是三种不同分段方法得到的分段线性模型的权值 w 的可视化效果图。图 4.8 是三种分段混合模型在 INRIA 测试集上的性能曲线，从中可以看出基于聚类的方法与单一的线性 SVM 分类器相比性能提升最多，是最有效的样本划分方法。

4.2 ECOC 算法概述

前面三小节介绍了三种不同的样本划分方法，样本划分的目的是解决视角和姿态的问题。通过以上三种方法实现了对样本集的划分，可以实现对每个视角和姿态训练单独的 SVM 分类器，新的问题是检测的时候如何确定使用哪一个视角的分类器。常用的办法是通过各个混和的分类器进行投票，文献[12][50][51]都是采用了这种方式，本文采用了另一种混合方法：错误纠错编码（ECOC）。

以 $M=\{(x_1,y_1\},...,(x_m,y_m)\}$ 为训练样本，每一个实例 x_i 都属于一个特征空间 X，每一个 y_i 都表示某一个分类的标签。学习一个多类分类器的目的就是去构造一个函数： $H:X\to Y$ 即将实例 x_i 映射到某一个分类上 $y\in Y$。我们创建一个长度为 B 的码，如表 4-1 所示，没有任何行或者列有相同或者互补的码。对于每一列，一个基分类器需要确定一个超类，即一个群集或者流形上两个相邻聚类的集合，整体而言 B 个基础分类器 $\{f_0,f_1,\cdots,f_{B-1}\}$ 组合成一个集成分类器。

对于表 4-1 任何一列，如果有 1 个非零的元素，那么相应的分类器就将对应的样本分类标号为 1 其余类别的样本标号为 0。如表 4-1 中 $\boldsymbol{b_4}$ 对应的分类器

$f_4(x)$ 就将聚类 1 和聚类 2 中的样本分类为 1 其余类别的样本分类为 0，这就组成本文 ECOC 的基础分类。在参考文献[54]，Kong etc 已经证明 ECOC 分类方法可以减少基础学习算法的方差和偏差，但是其他的投票方法仅仅能够减少学习算法的方差，这可能也就是 ECOC 可以获得较好检测效果的原因。

表 4-人体样本聚类的分类编码

	$\boldsymbol{b_0}$	$\boldsymbol{b_1}$	$\boldsymbol{b_2}$	$\boldsymbol{b_3}$	$\boldsymbol{b_4}$	$\boldsymbol{b_5}$		$\boldsymbol{b_{B-1}}$
非人体 (c_0)	0	0	0	0	0	0	…	…
聚类 1 (c_1)	1	0	0	0	1	0	…	…
聚类 2 (c_2)	0	1	0	0	1	1	…	…
聚类 3 (c_3)	0	0	1	0	0	1	…	…
聚类 4 (c_4)	0	0	0	1	0	0	…	…
…	…	…	…	…	…	…	…	…
聚类 $K(c_K)$	0	0	0	1	0	0		…

根据 ECOC 表，定义 $\{b_0, b_1, \cdots, b_{B-1}\}$ 长度为 B 的编码向量，对应于表 4-1 中的行，对于 K 个类，构成了 K 个长度为 B 的编码串 $\{S_0, S_1, \cdots, S_{K-1}\}$。对于这 B 位编码，需要 B 个基础分类器 $\{f_0, f_1, \cdots, f_{B-1}\}$，如果 s_i 的第 j 个位为 1，则 $f_j(x)=1$，否则 $f_j(x)=0$。在训练过程中，首先通过聚类的方法将正例样本划分为 K-1 个聚类结果 $\{C_1, C_2, \cdots, C_{K-1}\}$ 并将反例样本划分为一个单独的类别 C_0 这样就得到了 K 个聚类 $\{C_0, C_1, \cdots, C_{K-1}\}$ 根据表 4-1 中的码表将这 K 个聚类中的样本进行不同方式的混合训练 B 个基础分类器即 $\{f_1(x), f_2(x), \cdots, f_B(x)\}$，混合训练时正例样本的选择如上一段所分析的，例如分类器 $f_4(x)$ 对应的编码为 $\boldsymbol{b_4}$，则训练 $f_4(x)$ 分类器时正例样本为 $\{C_1, C_2\}$ 其余的聚类中的样本均为反例样本。

对于分类一个新的样本 $\tilde{x}$，通过每一个基础分类器对样本 $\tilde{x}$ 测试，得出这个样本的一个二值化的编码向量 $S(\tilde{x}) = (f_0(\tilde{x}), f_1(\tilde{x}), \cdots, f_{B-1}(\tilde{x}))$。接下来，基于 Hamming 距离计算哪个编码最接近于 $S(\tilde{x})$。计算编码 S_k 与 $S(\tilde{x})$ 的 Hamming 距离公式如下：

$$D_{C_k} = \|S_k - S(\tilde{x})\| = \sum_{j=1}^{B-1} \left| S_k^j - f_j(\tilde{x}) \right| \qquad (4\text{-}3)$$

根据 ECOC 分类器的规则将样本 $\tilde{x}$ 分类为类别 $C(\tilde{x})$，具体公式如下：

$$C(\tilde{x}) = \arg\min_{c_k} \left\{ D_{c_k} \mid k = 0,1,...,K-1 \right\} \qquad (4\text{-}4)$$

通过 Hamming 距离，可以将一个测试样本 $\tilde{x}$ 划分到任意一个正例类或者反例类。

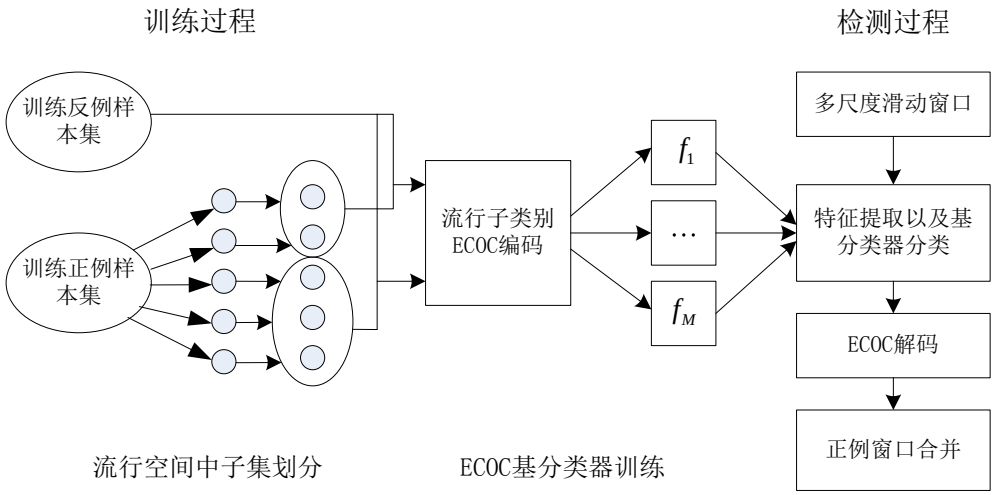


图 4. 分段线性 SVM 训练检测框图

如图 4.9 所示是基于 ECOC 编码混合分类器模型的行人检测框图，框图分为两个过程分别是训练过程和检测过程，其中训练过程又分为两步，第一步样本划分聚类；第二步 ECOC 编码和基础分类器训练。

4.3 ECOC 编码混合模型实验结论

以上分别介绍了正例样本分段方法以及基于 ECOC 编码的分段模型混合算法。本小节将介绍基于流形聚类和 ECOC 编码的混合分段模型的实验结论。

如图4.10所示我们评测了SDL数据集上基于ECOC编码的混合分段模型在将正例样本分成不同子类别情况下的分类性能，图4.10坐标的纵轴为检出率横轴为检测精度。从表4-1我们可以看出当正例样本的子类别增多时，流形空间上描述子集之间的邻接关系变得更复杂，需要的ECOC基分类器也变得更多。从图4.10可以看出，当正例样本划分成8个子类别时检测性能比划分成2个子类别要高的多，但是与6个子类别相比性能提升的并不多。以上结果说明以下两点：1）当正例子类别的数目增多时，各个子类别的样本在特征空间中紧凑性和线性可分性更好因此检测性能提高；2）随着正例子类别数目的增加，各个类别中的样本数目减少这也影响了整个分类器的性能。因此，鉴于以上两个因素，我们通过实验的方法找

到合适的子类别数来平衡这两个因素。

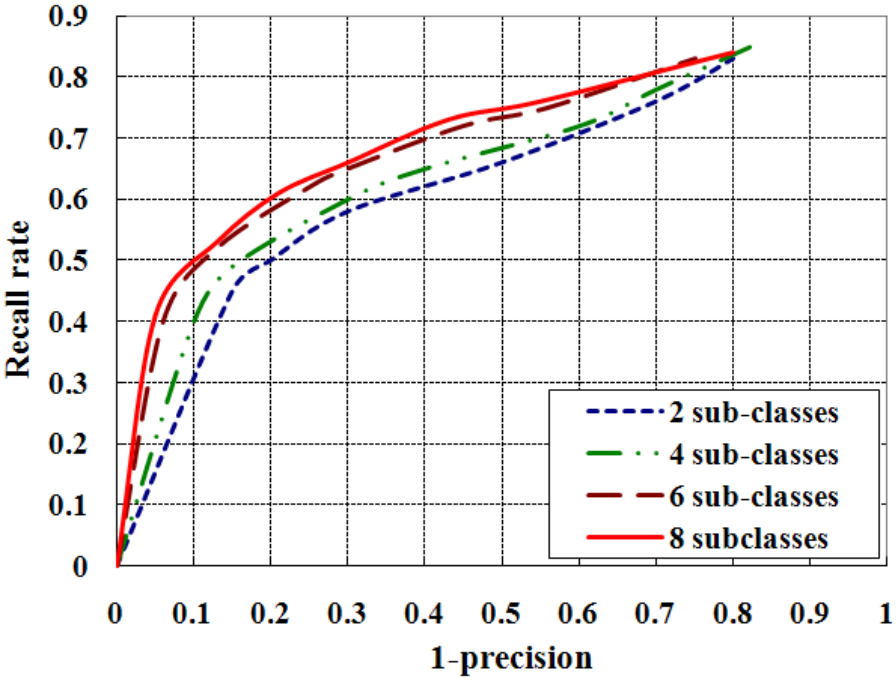


图 4. 检测性能与聚类数的关系

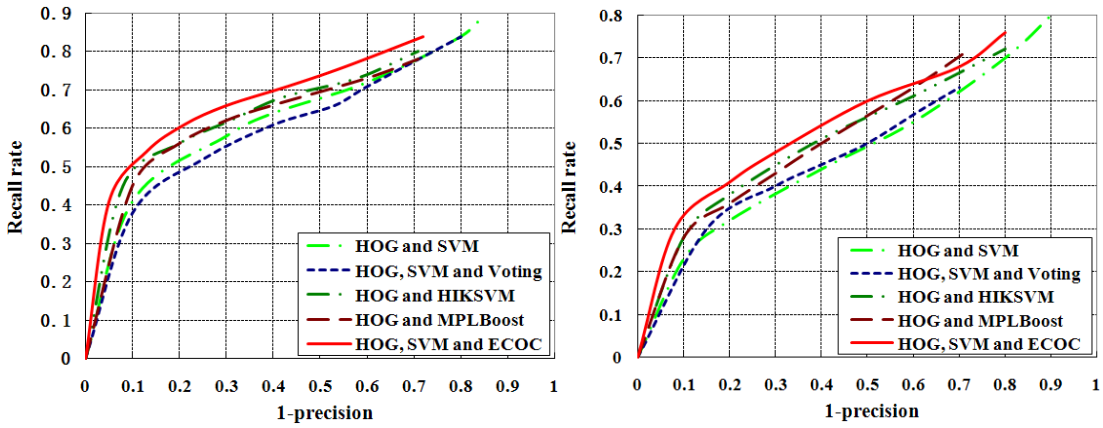


图 4. SDL 以及 TUD-Brussels 数据集各种方法性能对比

为了验证基于 ECOC 编码的混合分段模型的性能，我们的方法与 HOG+SVM[1][25]， HOG+SVM+voting, HOG+HISVM [61] 以及 HOG+MPLBoost (一种基于 SVM 的 boosting 方法) [62]这几种典型的方法进行了实验对比。HOG+SVM+voting 是一种分段线性 SVM 投票方法，每一个类别单独训练一个线性 SVM 分类器通过投票的方式得到最终的分类器，HISVM 是一种高效的核函数非线性 SVM 分类器方法，MPLBoost 是将多个强分类器（SVM 分类器）通过加权投票结合的分类方法。对比实验是分别在 SDL 数据集以及 TUD-Brussels 行人检测数据集上进行的。从图 4.11（a）可以看出我们提出的方法在 SDL 行人

检测数据集上的性能最高，图 4.11(b)显示在 TUD-Brussels 行人检测数据集我们所提的方法检测性能也比其他方法要高，并且在 TUD-Brussels 数据集上当检测精度为 50%时检出率为 60%。以上的评测我们提出的方法将正例样本划分成 6 个子类别，这几种方法都使用了 HOG 特征以及“滑动窗口”检测。

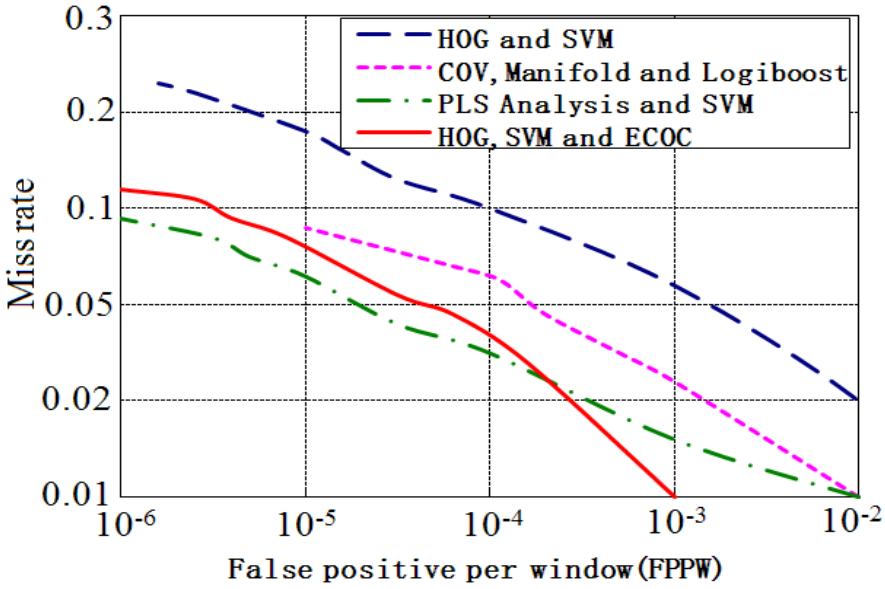


图 4. INRIA 数据集上各种方法性能对比

此外我们还在 INRIA 数据集上与另外几种经典方法进行了对比实验，这次的实验我们将正例样本划分成了 4 个子类别，因为 INRIA 数据集上的训练样本比较少只有 2478 个正例样本。如图 4.12 所示是我们的方法与 HOG+SVM [1][25], COV +Manifold+Logiboost[9], Partially Least Square (PLS) Analysis [11]。这几种方法对比的实验结果。这里的评测方法用的是比较流形的 Miss rate-FPPW(False Positives Per Window)[1]评测标准，其中[9][11]的结果我们直接引用了作者的检测曲线，[1][25]中的实验曲线是我们自己实现得到的，这几种方法所使用的训练样本集都是 INRIA 公开的训练样本集。从图 4.12 的曲线可以看出当 FPPW 为10^{-4}时，我们的方法丢失率为 4%比 HOG+SVM 的 10%要低 6%，比 COV+Manifold + Logiboost 的 7%要低 3%，当 FPPW 为10^{-6}时我们的方法丢失率也要比这两种方法低。但是相对与 PLS analysis 方法我们的方法性能要稍差一些，这是因为 PLS analysis 在分类前进行了特征选择和降维，使得特征表达更鲁棒。

从以上的实验结果，我们还可以得出一个结论，在 SDL 和 TUD-Brussels 这样的数据集中，行人的姿态、视角变化比较明显，我们的方法的优势也更加突出，如果数据集中行人的姿态、视角变化不多如 MIT 行人检测数据集大多数正面视角那么我们这个方法与 HOG+SVM 相比优势就不太明显。

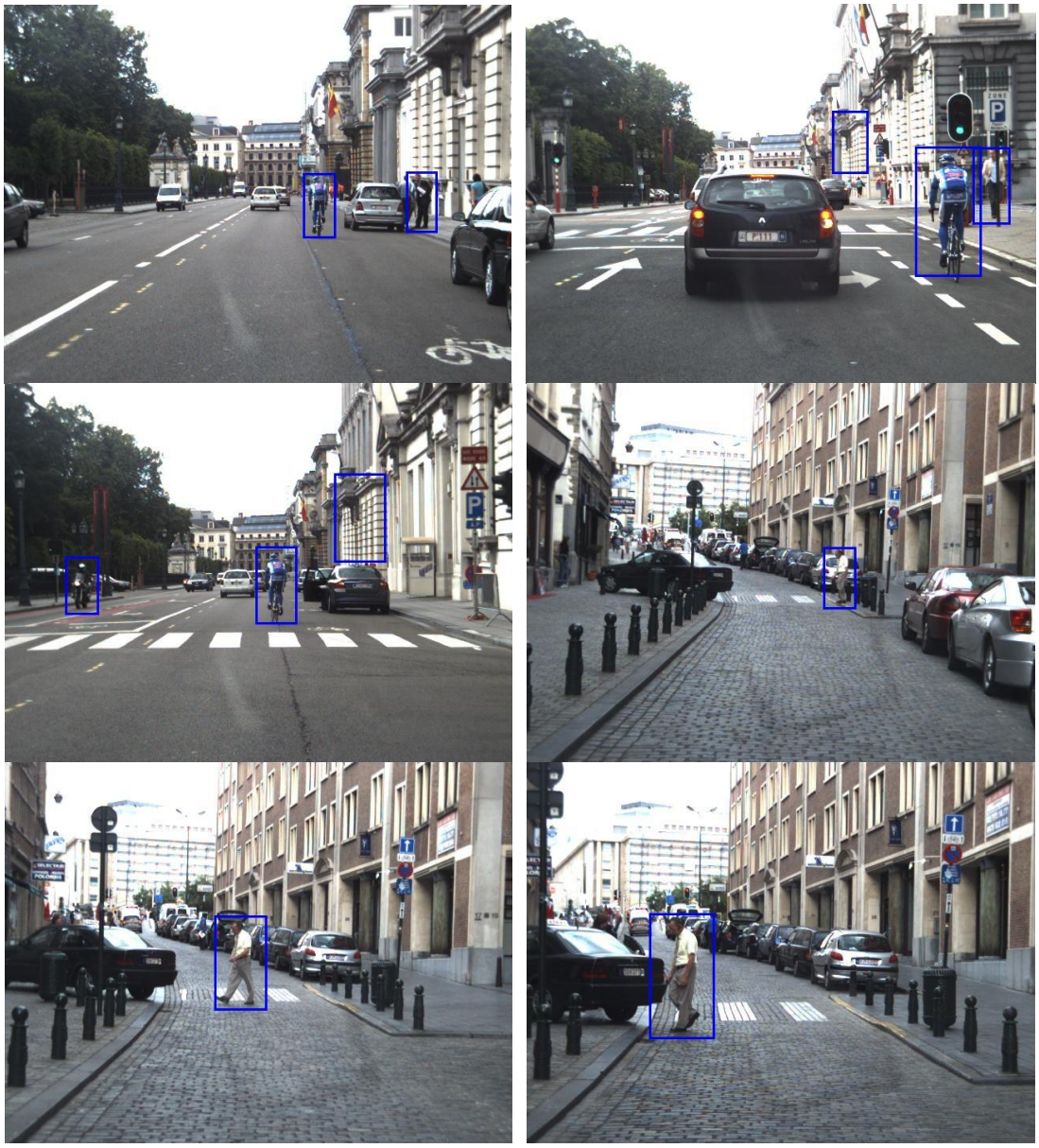


图 4. TUD-Brussels 行人检测数据集部分检测结果

4.4 行人检测应用实例

从第三章开始，我们分别介绍了特征融合以及基于ECOC编码的混合分类器模型，从特征和分类器视觉模型的角度介绍了本文的主要内容。本节内容将简要的介绍一下作者实现的行人检测框架的两个简单应用实例。

行人检测往往是实际的系统中的一个子模块，行人检测更有意义的应用一般是在视频监控或者目标跟踪过程中，为了提高系统的鲁棒性往往还借助了一些其它信息，比如下面要介绍的激光信息与图像信息融合、基于背景建模的行人检测

与跟踪系统。我们这里只做简要的原理介绍，具体的实现过程和研究成果读者可以参考后面给出的文献资料。

根据Marr的计算视觉理论，模拟人的视觉系统，需要三维形状表象信息，而单目视觉系统不能形成三维视觉效果，实际情况下的人体目标检测如果只借助单个摄像头获取的图像信息，有可能不能准确定位目标的位置，这就需要与其他信息进行融合，或者采用双目视觉系统。另外考虑到更普遍的环境如黑夜等，那么单纯的取像摄像头的功能将大大削弱，而激光作为主动光源在黑夜中仍然适用，并且激光器返回目标的距离信息，和图像融合能赋予三维视觉的能力。综合以上考虑，图像与激光信息融合是一个可行的方案。融合的框架如图4.14所示。首先我们通过激光信息得出图像上的感兴趣区域，然后对感兴趣区域进行特征提取，通过事先训练好的视觉分类器，对该区域的目标进行检测。

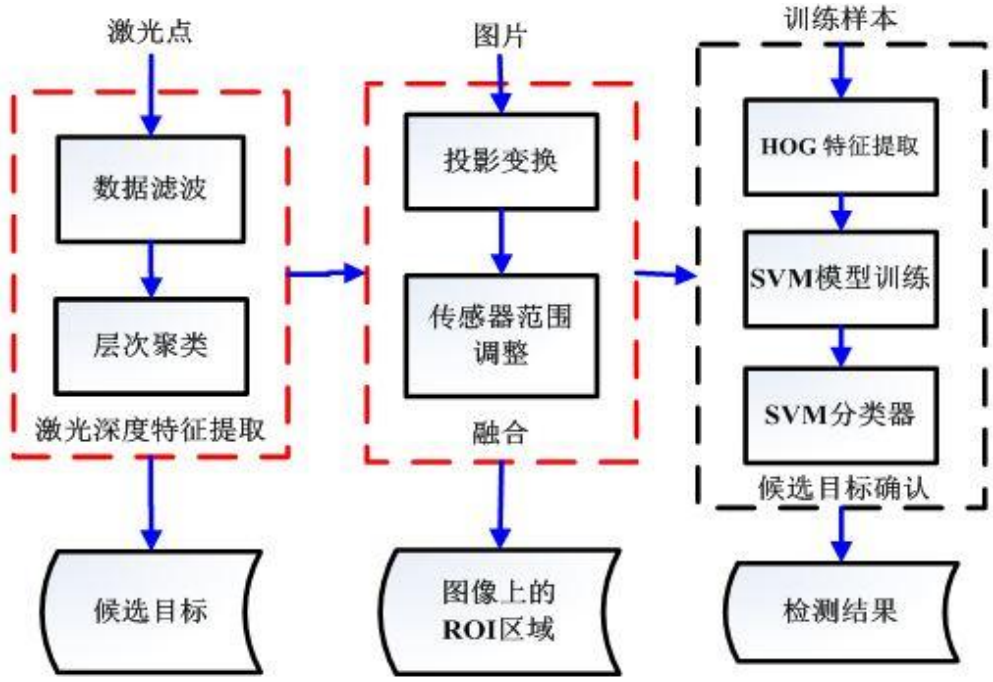


图 4. 激光图像信息融合框图

图4.15为激光图像信息融合行人检测的效果图例，融合的方法和实验的内容具体可以参考文献[57]。图4.15(a)是视频中的一帧图像，图4.15(b)是这帧图像对应的激光测距的回点图，通过对激光回点的处理可以得出激光回点图中的感兴趣区域，图4.15(c)激光回点图通过计算对应到图像帧上的感兴趣区域，图4.15(d)感兴趣区域通过事先训练好的分类器进行行人检测的结果。

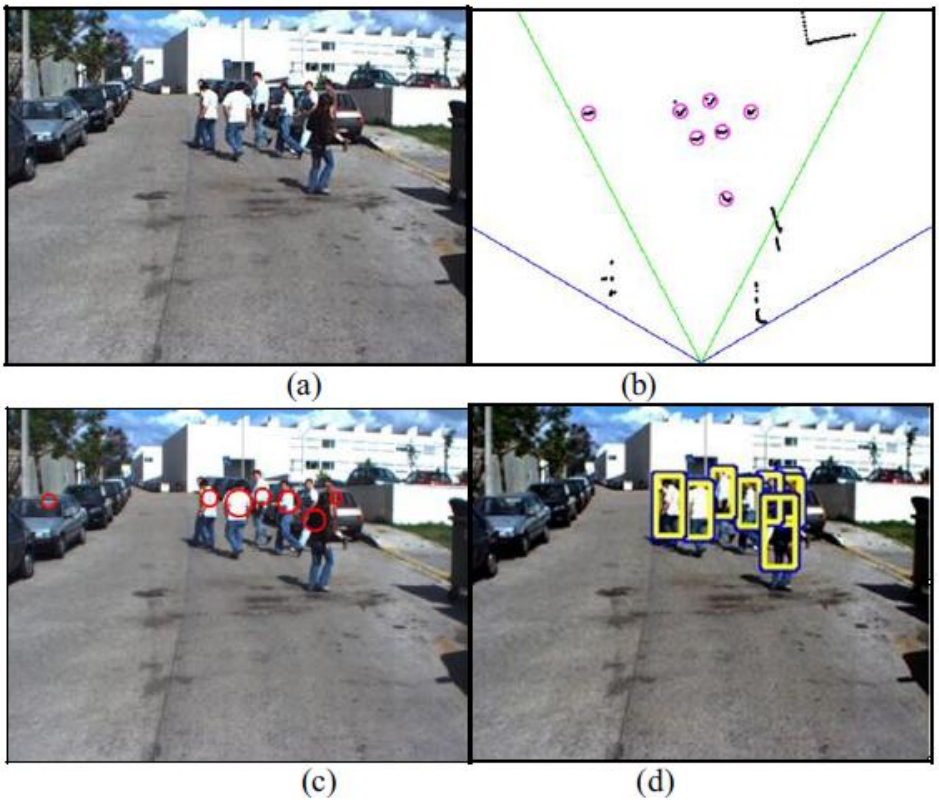
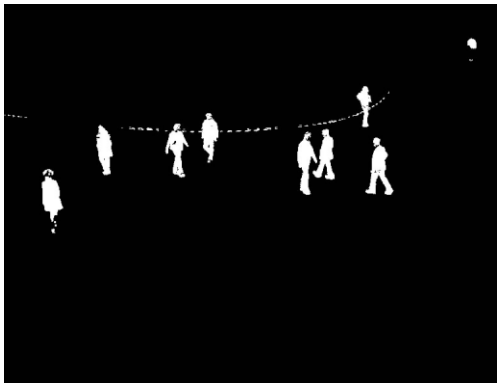


图 4. 激光图像信息融合检测效果图

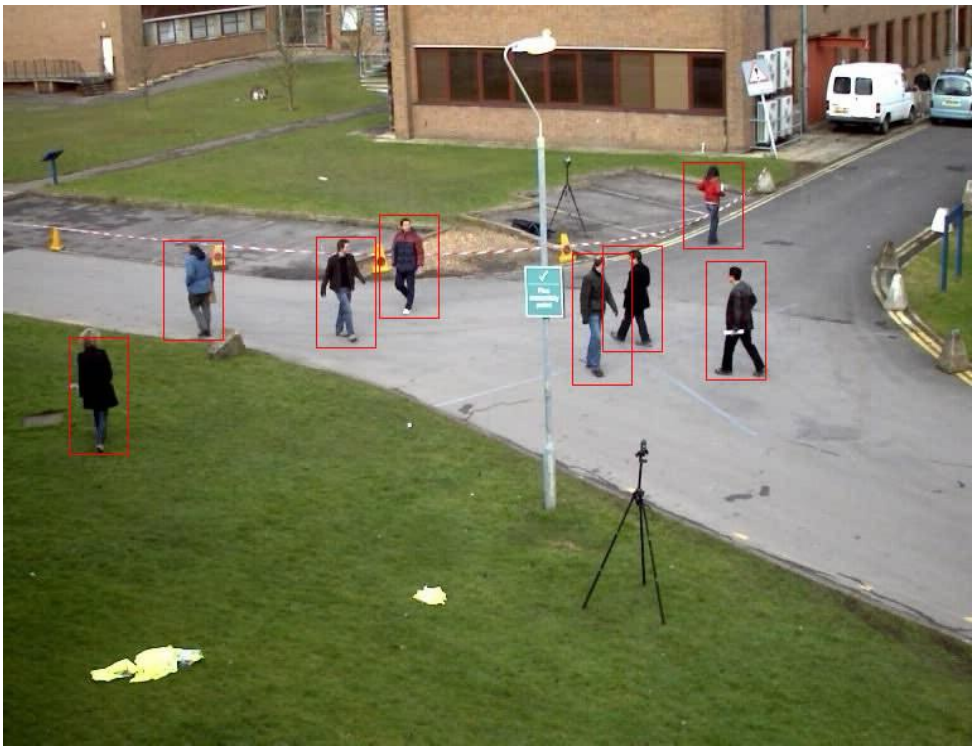
行人检测的另一个应用实例是监控视频中的目标检测和跟踪问题，监控视频有一个特点是背景相对稳定，应用场景也相当广泛，在国内许多的公共场所都装有监控摄像头。这些场景下的应用由于背景相对稳定，我们可以通过背景建模、光流分析等方法提取运动信息与视觉信息融合，从而提高检测的精度。

图4.16是基于背景建模的监控视频中行人检测效果图，(a)是监控视频中的某一帧图像，(b)是原始图像减去背景建模之后的前景图像，(c)是通过连通区域分析，将前景区域对应到原始图像上得到感兴趣区域（ROI），(d)对感兴趣区域通过训练好的分类器进行行人检测之后得到的检测结果图。

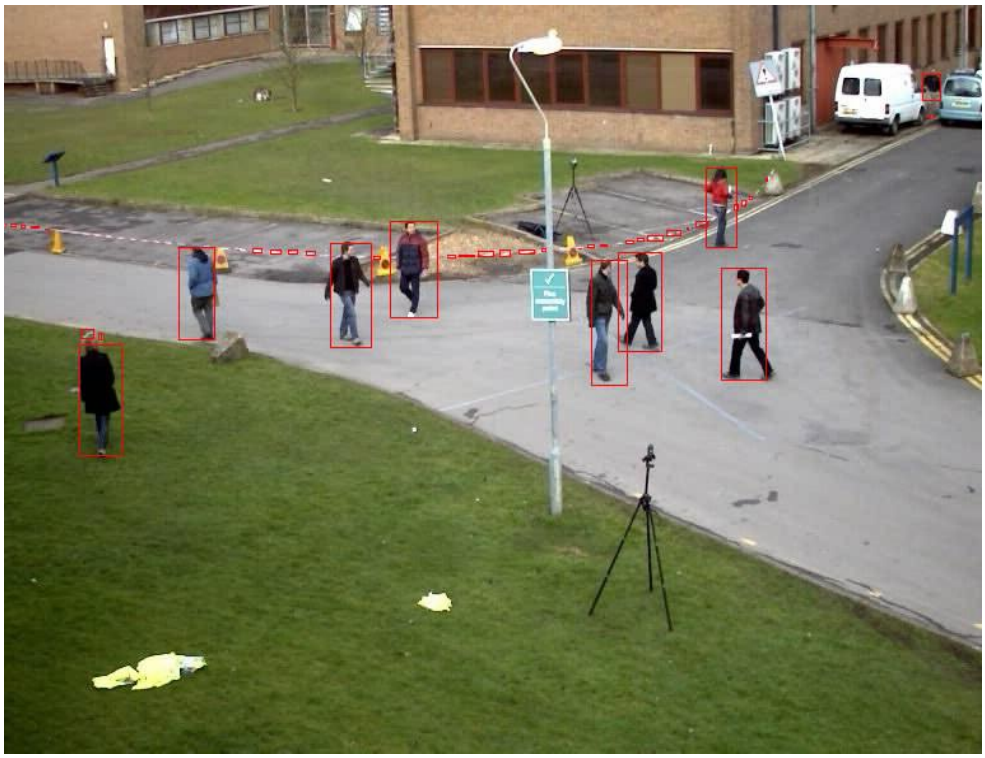
(a)原始图像帧



（b）背景去除之后前景图像



(c) 前景区域对应到原始图像



(d) 行人检测之后的结果

图 4. 基于背景建模的行人检测效果图

总结与展望

作为视觉目标检测的一个重要组成部分，行人检测面临多视角、多姿态以及部分遮挡的难题。本文从目标检测涉及的两部分内容出发，对行人检测特征表达与分类器模型进行了研究，提出了一种新的行人检测特征描述子以及基于 ECOC 编码的分段混合模型，与已有的特征描述子及流形的基于线性 SVM 分类模型相比，在针对多视角、多姿态与部分遮挡问题时性能显著提高。

本文首先简述了行人检测的研究背景和意义、国内外研究现状和本文的研究内容，然后在第二章给出了行人检测常用的特征描述子以及目标检测分类器。第三章介绍了本文在行人检测特征表达方面的工作，在评测了几种常用的行人检测特征描述子的基础上，提出了一种新的特征描述子 HOG_SURF 特征描述子，通过在 INRIA 行人检测数据集上的测试，该特征描述子与现有的最好的特征 HOG_LBP 相比，具有维数低、效率高的特点。第四章介绍了本文在分类器模型方面的工作，混合模型主要针对行人检测中的多视角、多姿态问题。本文在分析和实现几种不同的正例样本分段方法之后，在 INRIA 数据集上评测了几种分段方法的性能，提出了基于 ECOC 编码的混合分段模型。该方法与基于线性 SVM 分类模型相比在解决视角和姿态方面有显著地提高。

虽然本文在行人检测特征以及分类器模型方面取得了一定的研究成果，但是因为时间关系，仍然存在不足。在特征表达方面，评测的特征不全，分析的还不够透彻；在分类器模型方面，没有针对部分遮挡问题。与目前最好基于部件可行变的混合模型相比，在行人检测方面仍有一些差距。如图 5.1 所示是行人检测面临的几个难点问题，第一幅图中的行人存在多姿态的问题比较明显，此外还存在远近的相互遮挡问题；第二幅图像中背景非常复杂且图像的分辨率较小，这对特征表达提出了更高的要求；这两幅图像中的难点问题，本文也已进行了相关的研究工作。第三幅图中的突出问题是部分遮挡的问题，解决部分遮挡的问题目前最好的方法是基于部件检测的可形变模型；第四幅图是在夜晚

的情况下，存在正面高光照时，图像中人的像素上的特征丢失了，此时最好的检测办法应该是激光信息与图像信息的融合。这两幅图像中的难点问题是本文还没有涉及的更广泛的研究难点与热点问题。

图 5. 行人检测难点问题图片，遮挡、复杂背景与高光照

将来基于图像的行人检测可能的改进方向有两个方面，第一个方向是基于部件拓扑约束的混合模型[58]，第二个方向是将特征表达与分类器方法结合的特征学习方法[59][60]。针对低分辨率的行人目标，形变模型无法发挥优势。在此情况下新的分治策略，如部分模型与随机树或者随机森林的结合，可以在一定程度上解决问题。此外，视频中的行人检测方面，基于深度与图像，以及图像与其他多源信息融合的行人检测可能成为新的研究方向。

参考文献

[1] N. Dalal, and B. Triggs, Histograms of Oriented Gradients for Human Detection[C]. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.886-893, 2005.

[2] X.Wang, T.X. Han, and S. Yan, An HOG-LBP Human Detector with Partial Occlusion Handling[C]. *IEEE Int'l Conf. Computer Vision*, 2009.

[3] P. Felzenszwalb, R. Girshick, D. McAllester and D. Ramanan. Object Detection with Discriminatively Trained Part Based Models[J]. *IEEE Trans. PAMI*, Vol. 32, No. 9, pp. 1627-1645, 2010 .

[4] P. Dollar, C. Wojek, B. Schiele, and P. Perona, Pedestrian Detection: An Evaluation of the State of the Art[J]. *IEEE Trans. PAMI*, Jul 28,2011.

[5] C. Papageorgiou and T. Poggio, A Trainable System for Object Detection[J]. *International Journal of Computer Vision*, vol. 38, pp. 15-33, 2000.

[6] P.Viola and M. Jones, Robust Real-time Face Detection[J]. *International Journal of Computer Vision*, 2004.

[7] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, Fast Human Detection Using a Cascade of Histograms of Oriented Gradients[C]. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp.1491-1498, 2006.

[8] Qixiang Ye, Jianbin Jiao, Baochang Zhang, Fast pedestrian detection with multi-scale orientation features and two-stage classifiers[C]. *IEEE International Conference on Image Processing*,pp:881-884, 2010.

[9] O. Tuzel, F. Porikli, and P. Meer, Pedestrian Detection via Classification on Riemannian Manifolds[J]. *IEEE Trans. PAMI*, vol.30, no.10, pp.1713-1727, 2008.

[10] Mu Y. Yan S. Liu Y., Huang T., Zhou B., Discriminative Local Binary Patterns for Human Detection in Personal Album[C]. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, vol 23-28, pp.1-8, 2008.

[11] William Robson Schwartz , Aniruddha Kembhavi , David Harwood , Larry S. Davis, Human Detection Using Partial Least Squares Analysis[C]. *Proc. of the IEEE Int'l Conference on Computer Vision*, 2009.

[12] B.Wu, and R. Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors[C], *Proc. IEEE Int'l. Conf. on Computer Vision*, 2005.

[13] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, SURF: Speeded Up Robust Features[C]. *IEEE Conf. on Europeon Conference on Computer Vision*, 2006.

[14] Christopher J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition[J]. *Data Mining and Knowledge Discovery*, vol. 2(2), pp.121-167, 1998 .

[15] M.Collins, R.E. Schapire and Y. Singer, Logistic Regression, AdaBoost and Bregman Distances[J]. *Machine Learning*, vol. 48(1-3), pp.253-285, 2002.

[16] Schapire R E,Freund Y, Bartlett Y, et al., Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods[J]. *Annals of Statistics*,1998,26(5):1651-1686.

[17]R. Xu, B. Zhang, Q. Ye and J. Jiao. Cascaded L1-norm Minimization Learning (CLML) Classifier for Human Detection[C]. *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Pp.89-96, 2010.

[18] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun, Support Vector Machine Learning for Interdependent and Structured Output Spaces[C]. *International Conference on Machine Learning*, 2004.

[19] Chun-Nam John Yu, and Thorsten Joachims, Learning Structural SVMs with Latent Variables[C]. *International Conference on Machine Learning*, 2009.

[20] Richard Szeliski，计算机视觉—算法与应用 Computer Vision：Algorithms and Applications[M] 清华大学出版社，艾兴海、兴军亮译，2012.

[21] David G. Lowe, Object Recognition from Local Scale-Invariant Features[C]. *International Conference on Computer Vision*, 1999.

[22] http://www.cnblogs.com/xrwang/archive/2010/03/03/ImageFeatureDetection.html

[23] http://blog.csdn.net/jiang1st2010/article/details/6567452

[24] Luo Juan, Oubong Gwun, A Comparison of SIFT, PCA-SIFT and SURF[J]. *International Journal of Image Processing*, vol.3, pp 143-152, 2009.

[25] N. Dalal, Human Detection using Oriented Histograms of Flow and Appearance[C]. *Proc. of the IEEE Conf. on Europeon Conference on Computer Vision*, pp.428-441, 2006.

[26] Rainer Lienhart, Jochen Maydt, An Extended set of Haar-like Features for Rapid Object Detection[C]. *IEEE International Conference on Image Processing*, 2002.

[27] Huixing Jia, Yu-Jin Zhang. Fast Human Detection by Boosting Histograms of Oriented Gradients[C]. *Proc. of the IEEE Conf. on Image and Graphics*, 683 -688, 2007.

[28] 邓乃扬,田英杰.数据挖掘中的新方法---支持向量机[M]. 北京，科学出版社, 2004.

[29] Christopher J. C. Burges， Advances in Kernel Methods-Support Vector Learning[M]. *MIT Press*, 1999.

[30] P.Viola and M. Jones, Robust Real-time Object Detection[J]. *International Journal of Computer Vision*, 2001.

[31] B.E. Goldstein, Sensation and Perception, sixth ed. Wadsworth, 2002.

[32] Robert E. Schapire, A Brief Introduction to Boosting[C]. *Proceedings of the 16th international joint conference on Artificial intelligence (IJCAI)*,vol.2, 1999.

[33] R. Xu, B. Zhang, Q. Ye and J. Jiao. Human Detection in Images via L1-Norm Minimization Learning[C]. *Proc. of the IEEE Conf. on International Conference on Acoustics, Speech, and Signal Processing*, pp. 3566-3569, 2010.

[34] P. Felzenszwalb, D. McAllester, and D. Ramanan. A discriminatively trained, multiscale, deformable part model[C]. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[35] http://www.mis.tu-darmstadt.de/tud-brussels

[36] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian Detection: A Benchmark[C]. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 304–311, 2009.

[37]http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/

[38]http://pascal.inrialpes.fr/data/human/

[39] http://pascallin.ecs.soton.ac.uk/challenges/VOC/

[40] M.Oren, C.Papageoriou, P.Sinha, E.Osuna, T.Poggio, Pedestrian Detection Using Wavelet Templates[C]. *IEEE Computer Vision and Pattern Recognition*, pp.193-199,1997

[41] http://www.gavrila.net/Research/Pedestrian_Detection/Daimler_Pedestrian_Benchmark_D/daimler_pedestrian_benchmark_d.html

[42] http://www.vision.ee.ethz.ch/~aess/dataset/

[43] Vapik V., Chervonenkis A. The Necessary and Sufficient Conditions for Consistency in the Empirical Risk Minimization Method[J]. *Pattern Recognition and Image Analysis*, vol.1(3), pp: 283-305,1991.

[44] T. Ojala, M. Pietikinen, and D. Harwood. A Comparative Study of Texture Measures with Classification Based on Feature Distributions[J]. *Pattern Recognition*, 29(1):51–59, 1998.

[45] T. Ahonen, A. Hadid, and M. Pietikinen. Face Description with Local Binary Patterns: Application to Face Recognition[J]. *IEEE Trans. PAMI*, 28(12):2037–2041, 2006.

[46] T. Ahonen, A. Hadid, and M. Pietikinen. Face Recognition with Local Binary Patterns[C]. *IEEE Conf. on Europeon Conference on Computer Vision*, pages 469–481, 2004.

[47] http://www.ucassdl.cn/resource.asp

[48] N. Dalal. Finding People in Images and Videos[D]. *PhD thesis*, INRIA Rhne-Alpes, Grenoble, France, 2006.

[49] 任双桥，杨德贵，黎湘，庄钊文，分片支撑矢量机[M]，计算机学报，第 32 卷第 1 期，2009 年 1 月.

[50] C. Mohan, C. Papageorgiu and T. Poggio. Example-based Object Detection in Images by Components[J]. *IEEE Trans. PAMI*, Vol. 23, No. 4, pp. 349-361, 2001.

[51] Z.Lin, L. Davis, D. Doermann and D. DeMenthon. Hierarchical Part-template Matching for Pedestrian Detection and Segmentation[C]. *Proc. of the IEEE Int'l Conference on Computer Vision*, 2007.

[52] T. Ojala, M. Pietikainen, and T. Maenpaa, Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns[J], IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971–987, Jul 2002.

[53] S. T. Roweis and L. K. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding[M]. *Science* , 2000:2323–2326P

[54] E.B. Kong, T.G. Dietterich, Error-correcting output coding corrects bias and variance[C]. *In Proceedings of the Twelfth International Conference on Machine Learning*, pp: 313-321，1995

[55] Jixiang Liang, Qixiang Ye, Jie Chen, Jianbian Jiao, Evaluation of Local Feature Descriptors and Their Combination for Pedestrian Representation[C]. *IEEE International Conference of Pattern Recognition, pp 2496-2499*, 2012.

[56] Qixiang Ye, Jixiang Liang, and Jianbin Jiao, Pedestrian Detection in Video Images via Error Correcting Output Code Classification of Manifold Subclasses[J]. I*EEE Trans. Intelligent Transportation Systems*, 2011.

[57] Bo Wu, Jixiang Liang, Qixiang Ye, Zhenjun Han, and Jianbin Jiao, Fast Pedestrian Detection with Laser and Image Data Fusion[C]. *The 6th International Conference on Image and Graphics (ICIG)*，2011.

[58] Wen Gao, Xiaogang Chen, Qixiang Ye, Jianbin Jiao, Pedestrian Detection Via Part-based Topology Model[C]. *In Proc. International Conference on Image Processing (ICIP)*, 2012.

[59] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, Andrew Zisserman, Supervised Dictionary Learning[C]. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[60] Tosic, I., Frossard, P., Dictionary Learning[M]. *IEEE Signal Processing Magazine*, Vol. 28, pp27 - 38,2011

[61] S. Maji, A. C. Berg, and J. Malik, Classification Using Intersection SVM is Efficient[C]. *In Proc. IEEE Int. Conf. CVPR*, 2008, pp. 1–8.

[62] C. Wojek, S. Walk, B. Schiele, Multi-cue onboard pedestrian detection[C]. *In Proc. IEEE Int'l Conf. CVPR*, 2009.

个人简介及发表文章

个人简介

梁吉祥 男 汉族 中共党员

- 2006 年 9 月至 2010 年 7 月 中国科学技术大学 电子工程与信息科学专业 学士
- 2010 年 9 月至 2013 年 7 月 中国科学院大学 计算机应用技术 硕士

曾获荣誉：

2006 年 湖南省 “三好学生”

2007 年 中国科学技术大学 “优秀学生干部”

2012 年 中国科学院大学 “三好学生”

2012 年中国科学院大学 “优秀共产党员”

曾获奖学金：

2007 年 中国科学技术大学 “凡谷励志”奖学金

2012 年 中国科学院大学 “国家奖学金”

已发表文章目录

- **Jixiang Liang,** Qixiang Ye, Jie Chen, Jianbian Jiao, "Evaluation of Local Feature Descriptors and Their Combination for Pedestrian Representation," IEEE *International Conference of Pattern Recognition,* pp.2496-2499, 2012. (EI)
- Qixiang Ye, **Jixiang Liang**, Jianbin Jiao, "Pedestrian Detection in Video Images via Error Correcting Output Code Classification of Manifold Subclasses," IEEE *Transactions on Intelligent Transportation Systems, vol.1,* pp.1-10, 2011. (SCI 3.452)
- Bo Wu, **Jixiang Liang**, Qixiang Ye, Zhenjun Han, Jianbin Jiao, "Fast Pedestrian Detection with Laser and Image Data Fusion," IEEE *International Conference on Image and Graphics,* pp. 605-608, 2011. (EI)

致　谢

在中国科学院大学攻读硕士学位三年的学习生活中，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，由衷地感谢这三年来曾经给予我无数帮助的老师、同学、朋友和家人。

本课题的研究工作是在焦建彬教授和叶齐祥副教授的悉心指导下完成的。首先，我要感谢导师叶齐祥老师在我攻读硕士学位期间从学习和生活各个方面给予的无微不至的关怀与指导，以及在我论文的撰写和修改中倾注的心血。感谢焦建彬教授在学习中对我的每一点进步的指导和鼓励。两位恩师在科研上精益求精，在学术上认真严谨，他们的科学精神令人敬佩；两位恩师在生活上关心爱护学生，和蔼可亲、平易近人，他们春风化雨的教诲让人感动。

其次，感谢韩振军老师在理论学习和课题研究过程中给我提供的耐心的引导和帮助。感谢陈孝罡、张立国、彭艺、李策以及毕业的各位师兄、师姐，在我研究生的三年中给予的学习上的耐心引导与生活中的种种帮助。

感谢我的同届好友高山、武利军、邹佳凌、杨威以及张晓丹等同学，三年的科研生活中大家相互鼓励、献计献策、相互提点，一起渡过了三年快乐的日子。这些快乐的时光在我记忆中永远不会褪色。

感谢我的父母，他们给了我巨大的、无私的爱和永远无条件的支持，永远是我最坚强的后盾和避风的港湾，愿他们永远平安健康。

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

最后，感谢参加论文评审和答辩的各位老师。

梁吉祥

2013 年 4 月