



实用性目标检测与 跟踪算法原理及应用

■ 徐光柱 雷帮军 著

Practical Object Detection
and Tracking Algorithm Theory
and Applications



国防工业出版社
National Defense Industry Press

该书由国家自然科学基金项目(61402259),江苏省三维打印装备与制造重点实验室开放基金项目(L20140713-03),湖北省水电工程智能视觉监测重点实验室开放基金项目(2014KL-A10),三峡库区生态环境教育部工程研究中心开放课题项目(KF2015-10)联合资助。

实用性目标检测与 跟踪算法原理及应用

徐光柱 雷帮军 著

国防工业出版社

·北京·

内 容 简 介

本书从实时与实用两个角度,结合具体实例系统讲述了视觉目标检测与跟踪算法原理及应用。在视觉目标检测部分,本书全面论述了 AdaBoost 算法原理、积分图及 Haar-Like 特征快速计算方法和 Viola-Jones 通用目标检测框架、梯度直方图、SVM 原理、HOG 与 SVM 相结合的目标检测框架,同时给出了这两类算法在人脸检测与行人检测中的应用实例及基于 OpenCV 的实现。在视觉目标跟踪部分,本书系统介绍了单目标跟踪算法的分类与现存问题及性能评价方法;详细论述了以压缩感知理论为背景的压缩跟踪算法原理及改进算法,并对其具体实现给出了细节分析;全面讲述了跟踪、学习及检测相结合的视觉目标跟踪算法的原理,并给出了其性能分析。最后,从具体问题出发,本书详细说明了目标检测与跟踪算法在人眼区域检测与跟踪、多人脸检测与跟踪、交互式鱼体跟踪系统中的应用。

本书内容丰富、案例典型,论述由浅入深、重点突出,理论与实例紧密结合,内容安排合理、可读性强。

本书适用于计算机视觉相关方向的研究生和高年级本科生阅读,同时也可以作为相关领域研究人员和目标检测与跟踪算法工程师的参考资料。

图书在版编目(CIP)数据

实用性目标检测与跟踪算法原理及应用/徐光柱,雷帮军著.

—北京:国防工业出版社,2015.4

ISBN 978-7-118-10220-8

I. ①实… II. ①徐… ②雷… III. ①目标检测—算法
②目标跟踪—算法 IV. ①C931.2 ②TN953

中国版本图书馆 CIP 数据核字(2015)第 108264 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

北京京华虎彩印刷有限公司印刷

新华书店经售

*

开本 710×1000 1/16 印张 10 $\frac{1}{4}$ 字数 201 千字

2015 年 4 月第 1 版第 1 次印刷 印数 1—1800 册 定价 72.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010) 88540777

发行邮购:(010) 88540776

发行传真:(010) 88540755

发行业务:(010) 88540717

前 言

有效且可靠的视频目标自动检测与跟踪技术在视频监控、人机交互、生物特征识别和目标运动行为分析等领域具有广泛的应用,其核心是快速鲁棒的视频目标检测与长时间跟踪算法。

目标检测算法以前景检测技术为基础,通过对变化的前景点进行过滤并配以区块检测算法,来实现前景点集合的融合与分裂,进而实现运动目标的检测。除此以外,基于有监督学习技术的目标检测算法以其抗干扰能力强、智能性高、待检测目标类型可指定等优势,正受到越来越多的关注。鉴于前者在很多专著中均有较为详细的介绍,本书主要侧重于讲述类 Haar 特征与 AdaBoost 相结合的 Viola-Jones 通用目标检测算法和 HOG 与 SVM 相结合的特定目标检测算法的原理及实现。这两类算法的基础是有监督学习中通过弱分类器级联构建强分类器的集成学习技术和统计学习理论框架下的最小结构风险分类器构建技术,是目前应用最为广泛的两类视觉目标检测方法。本书对其原理及 OpenCV 下的实现给出了具体说明,并在后续章节中与目标跟踪算法相结合给出了若干具体应用实例。

视频目标跟踪需要对后续图像序列中特定目标的位置进行连续估计,由于存在目标表观变化、遮挡问题、周边环境的干扰和光照变化等诸多因素的影响,目前还不存在性能优良的通用目标跟踪算法。基于密度估计的 MeanShift 跟踪算法、基于滤波技术的卡尔曼和粒子滤波跟踪算法以及基于多示例学习和增量式子空间学习的跟踪算法均为经典算法,这些算法在很多书籍中均有介绍。本书从实用性、实时性、鲁棒性和长期有效性 4 个方面出发,对目前出现的以压缩感知理论为基础的压缩跟踪算法和将跟踪与学习及检测相结合的 TLD 算法进行了全面的介绍,并给出了一些最新的改进成果。最后,将本书所讲述的目标检测与跟踪算法相结合应用于人眼区域检测与跟踪、多人脸检测与跟踪和交互式鱼体跟踪系统。

整本书共分为 7 章,第 1 章主要对目标检测技术进行综述,常用的目标检测技术包括基于特征的目标检测、基于模板的目标检测、基于运动的目标检测、基于分类器的目标检测;同时详细介绍了基于 Viola-Jones 算法的视频目标检测原理及其在人脸检测中的应用。第 2 章详细介绍了基于 HOG-SVM 算法的行人检测原理及其在 OpenCV 下的实现,对 HOG-SVM 算法进行了深入的分析。主要包括 HOG 特

征提取、SVM 原理、基于 HOG-SVM 算法的行人检测及 OpenCV 下的实现。第 3 章主要对目标跟踪算法进行综述，主要介绍了目标跟踪的应用领域、组成模块，对目标跟踪算法进行了分类和举例，并说明了目标跟踪过程中存在的问题，最后介绍了目标跟踪的测评方法及数据集。第 4 章详细介绍了 CT 算法的原理及其改进的一些方法，包括基于局部区域相似度匹配的改进算法、基于表观模型再匹配的改进算法以及基于加权分块策略应对遮挡问题的目标跟踪改进算法。第 5 章详细介绍了 TLD 算法的原理，首先描述了 TLD 的整体结构，然后对 TLD 的三大组成模块分别介绍，包括跟踪模块、检测模块和学习模块，最后通过实验分析了 TLD 的性能。第 6 章主要介绍了基于 TLD 算法的应用，重点介绍了基于 TLD 算法的人眼检测与跟踪应用和多人脸目标跟踪应用，并对其实现原理和实验结果进行了详细的介绍和分析。第 7 章是 CT 算法的一个具体应用，将 CT 算法用于鱼体运动轨迹提取这一特定应用中，提出了一种交互式的连续跟踪算法来跟踪鱼头所在的位置，并利用鱼头与鱼身相对关系估计出鱼身所在区域，从而抠取出整鱼图像，为完成鱼体游泳活性的判断提供基础数据。

本书适用于计算机视觉相关研究方向的低年级研究生和高年级本科生，也可作为相关领域研究人员和目标检测与跟踪算法工程师的参考资料。

本书得到了国家自然科学基金项目（61402259）、江苏省三维打印装备与制造重点实验室开放基金项目（L2014071303）、湖北省水电工程智能视觉监测重点实验室开放基金（2014KLA10），三峡地区地质灾害与生态环境湖北省协同创新中心项目经费资助！参与本书编写的还有研究生景静、尹潘龙、曹维华、何艳、赵晶晶。在编写过程中还得到了国防工业出版社的大力帮助，在此表示深深的谢意。同时在此要感谢家人的大力支持和理解。

由于作者知识水平和经验的局限性，书中难免有不当和疏漏之处，敬请广大读者批评指正，作者邮箱：xugzhu@163.com。

作者

2015 年 4 月

目 录

第 1 章 视频目标检测综述及 Viola-Jones 算法	1
1.1 目标检测技术综述	2
1.1.1 基于特征的目标检测	2
1.1.2 基于模板的目标检测	3
1.1.3 基于运动的目标检测	4
1.1.4 基于分类器的目标检测	4
1.2 基于 Viola-Jones 算法的视频目标检测原理分析	5
1.2.1 Haar-Like 特征描述	5
1.2.2 积分图像	6
1.2.3 AdaBoost 算法	8
1.2.4 级联分类器	10
1.3 基于 Viola-Jones 算法的人脸检测及其 OpenCV 实现	11
1.3.1 级联分类器的训练	12
1.3.2 实现人脸检测的 OpenCV 代码分析	14
参考文献	18
第 2 章 基于 HOG-SVM 算法的行人检测原理与实现	19
2.1 HOG 特征提取	20
2.1.1 颜色空间标准化	21
2.1.2 计算图像梯度	22
2.1.3 空间和方向上的梯度统计	23
2.1.4 重叠 Block 特征的标准化	26
2.1.5 收集检测空间所有 Block 的 HOG 特征	26
2.2 SVM 原理分析	27
2.2.1 SVM 理论基础	27
2.2.2 线性 SVM	28
2.2.3 非线性 SVM	31

2.3 基于 HOG-SVM 算法的行人检测及其 OpenCV 实现	32
2.3.1 行人检测器的训练	33
2.3.2 实现行人检测的 OpenCV 代码剖析	34
2.3.3 实验结果与分析	36
参考文献	38
第 3 章 视频目标跟踪技术综述	39
3.1 视频目标跟踪的应用领域	40
3.2 视频目标跟踪系统模块组成	41
3.2.1 目标描述	41
3.2.2 目标特征的选择	43
3.3 跟踪算法分类与举例	44
3.4 目标跟踪所面临的挑战问题	46
3.4.1 目前跟踪方法存在的问题	46
3.4.2 长时间目标跟踪	47
3.4.3 TLD 算法	48
3.5 目标跟踪评测方法与数据集	49
3.5.1 评测方法	49
3.5.2 评测数据集	50
参考文献	54
第 4 章 CT 算法的原理与实现及改进	56
4.1 相关理论基础介绍	57
4.1.1 朴素贝叶斯分类器	57
4.1.2 随机 Haar-Like 特征	58
4.1.3 压缩感知理论	59
4.1.4 目标跟踪算法性能评价指标	62
4.2 CT 算法原理及实现细节分析	62
4.2.1 CT 算法的原理	62
4.2.2 CT 算法实现步骤	63
4.2.3 CT 算法的细节分析	66
4.2.4 CT 算法的 Matlab 实现难点	68
4.3 基于局部区域相似度匹配的改进算法	72
4.3.1 区域选择	73
4.3.2 相似度计算	73

4.3.3	实验结果与分析	74
4.4	基于表观模型再匹配的改进算法	78
4.4.1	朴素贝叶斯分类器值的剖析	79
4.4.2	改进算法的实现	79
4.4.3	实验结果与分析	80
4.5	利用加权分块策略应对遮挡问题的目标跟踪改进算法	86
4.5.1	改进算法的实现	87
4.5.2	实验结果与分析	91
4.6	小结	98
	参考文献	99
第 5 章	跟踪与检测及学习相结合的目标跟踪算法	102
5.1	TLD 算法介绍	102
5.2	TLD 中的跟踪算法	106
5.2.1	光流法简介	106
5.2.2	牛顿迭代法	107
5.2.3	经典光流法	110
5.2.4	金字塔 LK 光流法	111
5.2.5	光流法的应用	112
5.3	TLD 中的目标检测算法	114
5.3.1	方差分类器	114
5.3.2	级联分类器	116
5.3.3	KNN 分类器	119
5.4	实验结果及性能分析	120
5.4.1	TLD 性能分析	120
5.4.2	P-N 学习的性能分析	124
	参考文献	124
第 6 章	目标检测与 TLD 算法的应用举例	126
6.1	人眼检测与跟踪	126
6.1.1	基于 AdaBoost 的人眼区域检测	127
6.1.2	基于尺度均衡策略的快速人脸检测	129
6.1.3	基于 TLD 的人眼跟踪系统	130
6.1.4	实验结果与分析	131
6.2	多人脸目标跟踪	133

6.2.1	人脸目标检测	134
6.2.2	多人脸的标记与分离	135
6.2.3	多人脸跟踪	138
6.2.4	实验结果与分析	139
	参考文献	140
第 7 章	基于压缩跟踪算法的交互式鱼体跟踪	142
7.1	计算鱼类游泳速度方法的选择	142
7.2	交互式视频跟踪	146
7.2.1	交互式视频跟踪算法的实现	146
7.2.2	目标样本再选择方法介绍	148
7.2.3	目标样本再选择方法实验结果分析	149
7.2.4	算法流程	151
7.2.5	交互式跟踪得到鱼头位置的数据	153
7.2.6	交互式视频跟踪技术的 Matlab 实现	155
7.3	整鱼图像抠取算法	156
7.3.1	抠取整鱼图像的方法	156
7.3.2	抠取整鱼图像的实验结果	160
7.3.3	算法流程	161
7.3.4	抠取整鱼图像算法的 Matlab 实现	162
	参考文献	163

第 1 章 视频目标检测综述及 Viola-Jones 算法

随着计算机技术的发展和计算机视觉的广泛应用，利用计算机图像与视频处理技术对目标进行实时检测的研究日趋走热。目标动态实时检测在车辆监控、行人检测、人脸识别、军事目标检测、入侵检测、行为识别、虚拟现实、医学诊断及人机交互等方面具有广泛的应用价值。目标检测是计算机视觉方向备受关注的根本问题，也是众多视觉应用的基础，其性能直接决定了后续应用的可行性与稳健性。

过去 20 年以来，研究人员在目标检测领域做了大量的研究，目标检测研究涵盖了多个学科领域，涉及图像处理、机器学习、线性代数、优化、统计学、拓扑学等领域。目标检测算法主要分为静态目标检测和运动目标检测。静态目标检测是指对单个静态图像（数码照片、扫描图像等）中的感兴趣目标进行检测，如检测一副数码照片中的人脸；运动目标检测是指对视频中运动的目标进行检测，如检测交通监控视频中的运动车辆、行人等。相对于静态目标检测而言，运动目标检测从视频序列图像中将变化的区域从背景中分割出来。

对于目标检测，目前还没有一种通用的方法，一般都只是针对特定的应用环境来研究和设计检测算法。近年来，目标检测研究方向有了多样化的发展，目标检测算法大致可以分为四大类：基于特征的目标检测、基于模板的目标检测、基于运动的目标检测、基于分类器的目标检测。其中，对基于分类器的目标检测算法的研究尤为活跃，它将分类决策理论引入视觉目标检测领域，并取得了不错的进展。例如：2001 年 Viola 与 Jones^[1]提出了一种基于类 Haar 特征和 AdaBoost 学习算法的目标检测方法，该算法可以实时处理并给出了很好的物体检出率，在人脸识别应用中显示出很好的效果；2005 年 Dalal N^[2]等实现了一种基于 HOG 特征和 SVM 学习算法的行人检测系统，该方法是目前行人检测领域检测精度较好且应用最为广泛的一种算法。

本章将详细介绍目标检测技术和基于 Viola-Jones 算法的视频目标检测算法原理及其 OpenCV 实现。

1.1 目标检测技术综述

现有的常用目标检测算法大致可以分为 4 大类：基于特征的目标检测、基于模板的目标检测、基于运动的目标检测、基于分类器的目标检测，如表 1-1 所列。

表 1-1 目标检测技术分类

类别	主要方法
基于特征的目标检测	基于形状的方法
	基于颜色的方法
基于模板的目标检测	基于固定模板的方法
	基于变形模板的方法
基于运动的目标检测	帧差法
	光流法
	背景差分法
基于分类器的目标检测	基于 Haar 特征和 AdaBoost 算法的检测方法
	基于 HOG 特征和 SVM 学习算法的检测方法

1.1.1 基于特征的目标检测

在基于特征的目标检测算法中，选取的图像特征是很重要的。最初的目标检测是提取一个或多个特征用来对感兴趣的目标进行建模，然后在目标模型的基础上进行模板匹配，这些特征可以是物体的形状、大小、纹理和颜色等。其中，比较典型的方法有基于形状特征的主动形状模型^[5]（Active Shape Models, ASM）和基于颜色特征的 CamShift 算法^[6]。

主动形状模型是 Tim Coote 于 1995 年提出来的一种基于形状特征的目标检测方法，已广泛应用于各种与图像相关的领域中。主动形状模型利用形状特征对目标进行描述和建模，其基本思想是：使用手工标定物体的轮廓边界点，构成训练集；对这些样本的形状进行配准，使用主成分分析的方法对配准后的形状向量建立统计形状模型；建立局部的灰度模型，得到初始形状，根据已经训练完成的模型对输入的图像加以匹配和搜索，对参数不断进行最优化调整，迭代搜索出与物体最佳匹配的目标轮廓，从而得到要检测的目标。

主动轮廓模型能很好地适应形状复杂的目标检测，由于它是参数化的模型，可以通过改变参数来限定特定约束范围内的可容许的形状样本，保持了形状的特异性，同时它还充分利用物体的纹理信息建立局部纹理模型，结合全局形状模型使目标检测更加精确。但是，主动形状模型也有许多不足之处：即：①它需要对训练集

中的图像样本进行手工标定，因此必然会存在误差，当需要大量样本时会使工作非常繁琐；②模型中的参数数量比较多；③它对目标初始化位置和形状的依赖性比较大，当初始位置和形状与目标位置和形状相差较大时，就有可能匹配失败。

CamShift 算法建立在 MeanShift 算法^[7]的基础上，主要通过视频图像中运动物体的颜色特征达到目标检测与跟踪的目的，是一种动态变化的非参数密度函数梯度估计方法。MeanShift 算法根据单一图像中目标和背景区域的分布概率不变的特性进行目标的检测，而 CamShift 算法在此基础上进行了扩展。首先，用户手动选择一个感兴趣的目标区域，由于红、蓝、绿（RGB）颜色空间对光照亮度变化比较敏感，为了减少光照变化的影响，采用 HSV 颜色模型代替 RGB 颜色模型；然后，计算选中目标区域对应的颜色直方图得到其颜色概率分布，建立相应的目标模型，并以此为特征利用 MeanShift 算法在整个图像中进行搜寻，从而检测并跟踪目标。CamShift 算法在简单背景下具有较佳的检测跟踪效果，但在复杂的背景下，尤其是在背景干扰以及跟踪目标高速运动、遮挡情况下，常会导致检测跟踪失败。

1.1.2 基于模板的目标检测

如果一个模板描述一个特殊物体是有效的，那么目标检测就转化成了模板与图像序列之间的特征匹配问题。模板匹配是基于图像相似性度量，在实时获取的图像序列中寻找最接近目标模板的图像区域的一种目标检测方法。目标模板匹配方法有两种：基于固定模板的匹配和基于变形模板的匹配。

在物体形状相对于摄像机视角不变的情况下（即刚性物体），基于固定模板匹配^[8]的方法是非常有效的。基于固定模板匹配的方法就是计算出固定的目标模板和候选图像区域之间的差异性，对该差异性结果阈值化后判断是否匹配。这种方法具有简单、直观的特点。当被检测目标的模式过于复杂而不能用单一模板建模时，可使用了几个相关的目标局部模板来检测目标。

变形模板（Deformable Template）匹配^[9]的目标检测方法适用于会发生变形的非刚性物体，由于在大多数视频中要检测的目标是可变形的非刚性物体，因此变形模板匹配的目标检测方法广泛应用于各种检测跟踪任务中。它实质上是由参数化的简单图元模型与相应的能量函数构成，具体采用何种图元视待提取特征的目标物体形状而定，能量函数的设计同样也必须以目标物体的知识为依据。变形模板是利用先验概率分布知识建立起来的参数化集合模型，模型中的参数按照一定的优化准则迭代运算使模板最大限度地与目标图像相匹配，最终达到精确检测目标的目的。其整个处理流程为：在提取特征轮廓时，图元模型在图像上一定范围内移动，并动态地调整模型的参数；当图元模型的位置与参数达到最佳拟合时，能量函数应达到极小，此时就认为检测到了目标。

1.1.3 基于运动的目标检测

基于运动的目标检测算法是指从图像序列中将运动的目标区域检测出来。通常情况下，目标分类、跟踪和行为理解等后处理过程仅仅考虑图像中对应的运动目标像素区域，因此运动目标的正确检测对后期处理非常重要。由于场景的动态变化，如天气、光照、阴影及杂乱背景干扰等，使得基于运动的目标检测的精度与可靠性受场景条件影响大。目前常用的基于运动的目标检测方法有帧差法^[10]、光流法^[11]、背景差法^[12, 13]等。

帧差法是使用最多也是最简单的一种算法。基本原理就是将前后两帧图像对应的像素值相减，在环境亮度变化不大的情况下，如果对应的像素值相差很小，就认为此处是背景像素点，如果对应的像素值相差很大，则可以认为这是由图像中运动的物体引起的，标记该处为前景像素点，这样就可以求出运动目标在图像中的位置和大小。帧差法对于动态的环境具有较强的适应性，鲁棒性较好，能够适应各种动态环境，但一般不能完全提取与所有目标相关的像素点，容易在运动目标内部形成空洞现象。

光流法是常用的基于运动的目标检测方法之一。光流是空间运动物体被观测面上的像素点运动产生瞬时速度场（光流场），包含了物体表面结构和动态行为的重要信息。一般情况下，光流由相机运动、场景中目标运动或两者共同运动产生。通常有基于全局光流场和特征点光流场两种方法。光流法的优点是不需要预先得知场景的任何信息，就能检测到运动的目标，且可以处理动态背景的情况，但由于噪声、多光源、阴影和遮挡等因素，使得计算出的光流分布不可靠，而且光流法计算复杂，很难达到实时处理。

背景差法是一种常用的运动目标检测算法。基本思想是利用背景的参数模型来近似背景图像的像素值，将输入图像与背景图像进行对比，通过判定灰度等特征的变化或用直方图等统计信息的变化来区分前景和背景。背景差法要求背景图像必须是随着光照或外部环境的变化而实时更新，因此背景差法的关键是背景建模及其更新。针对如何建立对于不同场景的动态变化均具有自适应性的背景模型，研究人员已提出了许多背景建模算法。例如，基于卡尔曼滤波的自适应背景模型^[12]可以适应天气和光照的变化；Stauffer 与 Grimson 提出的自适应的混合高斯背景模型^[13]，可以在线更新背景模型，很好地解决了光照变化、背景混乱变化的干扰。与帧差法相比，背景差分法可以检测视频中停止运动的物体，其不足之处在于背景的实时更新增加了算法的计算量，实时性降低。

1.1.4 基于分类器的目标检测

前面的几种检测方法都是基于对要检测的目标有比较清晰的刻画，如基于特征的方法、基于模板的方法等，但是这些方法总体的泛化能力不够强。如果检测目标

比较固定且对时间要求不高，则可以考虑这类方法，但是考虑到有些应用无法提供模板，或者需要应对更大程度或更多样的形变（即需要检测算法有很好的泛化能力），这时前面的方法就无法适应了。为此，基于分类器的目标检测算法被提出来，它把目标检测问题看作图像中背景和待检测目标的分类问题。

基于分类器的目标检测算法，也称作基于统计理论的方法，它把目标检测这一图像处理问题上升到统计理论的高度，目前最常用的基于分类器的目标检测算法有神经网络方法、AdaBoost 算法^[14]、支持向量机（SVM）学习算法^[15]等。这些方法通常包括以下 3 个步骤：①获取大量用于训练分类器的正样本（待检测的目标）和负样本（图像背景），分别计算其特征；②利用各种分类算法（神经网络、AdaBoost、SVM 等）对正、负样本的特征进行训练得到分类器；③根据训练得到的分类器来判断图像中的检测窗口是否包含待检测的目标。大量研究表明，基于分类器的目标检测算法具有很好的适应性，能很好地解决多尺度、旋转变形等问题，泛化能力很强。基于分类器的目标检测算法在实际应用中得到了广泛的使用。例如，2001 年 Viola 与 Jones^[1]提出了一种基于 Haar 特征和 AdaBoost 学习算法的目标检测方法，2005 年 Dalal N^[2]等人实现了一种基于 HOG 特征和 SVM 学习算法的行人检测系统，这些方法都取得了不错的检测效果。

1.2 基于 Viola-Jones 算法的视频目标检测原理分析

基于 Viola-Jones 算法的视频目标检测框架是 Viola 和 Jones 在 2001 年提出的，是第一个可以实时处理并有较高检测率的物体检测方法，提出该方法的论文在 2011 年的 CVPR 会议上荣获“Longuet-Higgins 奖”。它可以被训练来检测多种物体，但主要应用于解决人脸检测方面的问题。该算法框架主要由三个部分组成：Haar-Like 特征、AdaBoost 算法和级联分类器架构。

基于 Viola-Jones 算法的视频目标检测的大体过程可以描述如下：首先，由积分图像快速地计算出大量简单的 Haar-Like 特征；然后，用 AdaBoost 学习算法从大量简单特征中挑选出一些重要的特征，经过训练得到最佳弱分类器，多个弱分类器通过线性组合，组合成一个强分类器；最后，将训练得到的多个强分类器按照复杂性的顺序组成一种级联结构，该级联架构大大地提高了检测速度，使检测系统具有实时的性能。本节将对基于 Viola-Jones 算法的视频目标检测原理进行详细的分析和介绍。

1.2.1 Haar-Like 特征描述

Viola-Jones 算法是基于简单 Haar-Like 特征对待检测的目标进行分类的。和其他检测算法一样，Viola-Jones 算法选择使用特征而不是直接使用像素，主要是因为

特征可以解决特定领域知识很难学会使用有限训练样本的问题，并且基于特征的检测系统运行速度远比基于像素的检测系统快。

Haar-Like 特征最早是由 Papageorgiou 等提出的，主要用于人脸表示，因为与 Haar 小波变换极为相似而得名。一个 Haar-Like 特征定义为矩形中几个区域的像素和的差值，可以具有任意的位置和尺寸。Haar-Like 特征值反映了图像的灰度变化情况。例如，脸部的一些特征可以由 Haar-Like 特征简单描述，如眼镜比脸颊颜色要深、鼻梁两侧比鼻梁颜色要深、嘴巴比周围颜色要深等。Viola 和 Jones 在此基础上使用 3 种类型 4 种形式的 Haar-Like 特征，3 种类型包括 2 矩形特征、3 矩形特征和 4 矩形特征，4 种形式如图 1-1 所示。其中：2 矩形特征的值定义为两个矩形区域里像素和的差，区域具有相同的大小和尺寸，并且水平或者垂直相邻；3 矩形特征的值等于两个外侧矩形的像素和减去中间矩形的和的值；4 矩形特征的值是计算两组对角线矩形的像素和差值得到的。从图 1-1 可以看出，Haar-Like 特征能提供水平、垂直和对角线等特征，提供了丰富的图像表征，支持有效的学习。Haar-Like 特征的数量取决于训练样本图像的分辨率，当训练样本分辨率设定为 24×24 时，Haar-Like 特征的总数量将超过 180000。

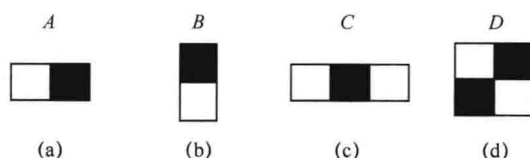


图 1-1 Haar-Like 特征的 4 种形式

Rainer Lienhart 等于 2002 年提出了倾斜 45° 的 Haar-Like 特征，对原有的 Haar-Like 特征库做了进一步的扩展。这种对特征的扩充是为了提升对物体的检测，这些特征对一些物体的描述更为适合，如一个倾斜的特征可以描述一个倾斜 45° 的边缘。另外，针对这种扩展后的 Haar-Like 特征，他们还提出了倾斜的积分图用于快速的特征值计算。图 1-2 显示了扩展后的 Haar-Like 特征，从图 1-2 中可以看出扩展后的 Haar-Like 特征大致可以分为 4 种类型，即边缘特征、线性特征、中心环绕特征和对角线特征等。

1.2.2 积分图像

由于一幅图像中的 Haar-Like 特征的数量远远超过图像中的像素值。在目标检测过程中，通常需要计算不同尺度下的大量 Haar-Like 特征值，如果不能解决大量矩形特征快速计算的问题，检测速度会非常缓慢。Viola-Jones 算法的一个重要贡献就是使用积分图^[16]计算图像的 Haar-Like 特征，解决了快速计算 Haar-Like 特征的问题，大大提高了图像特征的计算效率。

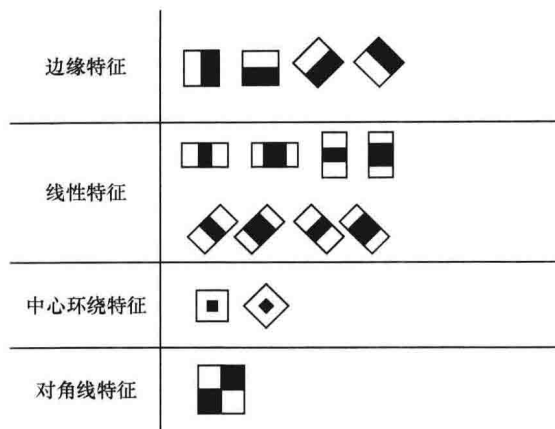


图 1-2 扩展后的 Haar-Like 特征

积分图是一种只需遍历一次图像，就可以求出图像在对应位置 (x, y) 左上角所有像素和的快速算法。积分图像可表示为

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

式中： $ii(x, y)$ 为积分图像； $i(x, y)$ 为原始图像。图 1-3 中黑色矩形区域显示了点 (x, y) 处的积分图像。

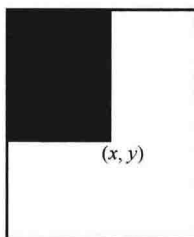


图 1-3 点 (x, y) 处的积分图像

可以通过以下公式循环遍历一次原始图像求得积分图像，即

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

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

式中： $s(x, y) = \sum_{y_1 \leq y} i(x, y_1)$ 为原图像 (x, y) 点所在列纵坐标不超过该点的所有像素值之和。

积分图的主要思想是将图像在对应位置左上角所有像素之和作为一个数组的元素保存在内存中，当要计算某个矩形区域的像素和时，只需对数组中的元素检索 4 次，不用重新计算就可以得到这个矩形区域的像素和，从而加快了计算。如图 1-4

所示，矩形区域内的像素和可以通过检索 4 个数组进行计算得到。位置 1 的积分图像的值就是矩形 A 中的像素之和，位置 2 的积分图像的值是 $A+B$ ，位置 3 的积分图像的值是 $A+C$ ，位置 4 的积分图像的值是 $A+B+C+D$ ，所以矩形 D 中的像素和就是 $4+1-(2+3)$ 位置处的积分图像的值。

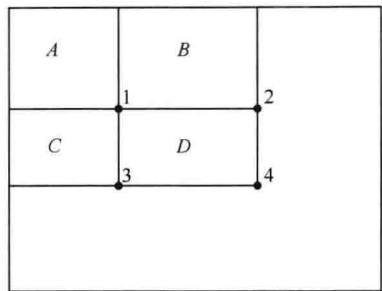


图 1-4 利用积分图求矩形区域 D 的像素值之和

Haar-like 特征值定义为几个矩形区域像素和的差，所以 Haar-Like 特征的特征值计算只与此特征的矩形区域端点的积分图有关。显然，两个矩形像素和之差需要检索 8 个数组。因为 2 矩形特征的定义是两个相邻的矩形像素和之差，所以只需检索 6 个数组就可以计算出特征值。同理，3 矩形特征的特征值计算需要检索 8 个数组，4 矩形特征的特征值计算需要检索 9 个数组。不管此特征矩形的尺度如何变换，特征值的计算所消耗的时间都是常量，这样只要遍历图像一次就可以求得所有子窗口的特征值。积分图能够在多种尺度下，使用相同的时间来计算不同的特征，因此大大提高了检测速度。

1.2.3 AdaBoost 算法

如果已知一个特征集和一个包含正样本和负样本的训练集，那么任何机器学习的方法都可以学习分类功能，常用的方法有神经网络、SVM、AdaBoost 等。基于 Viola-Jones 算法的视频目标检测采用自适应提升 (Adaptive Boosting, AdaBoost) 来训练分类器，它可以对弱学习得到的弱分类器的错误进行自适应性调整，将迭代得到的 N 个弱分类器按一定权重叠加起来，构成最终的强分类器。研究表明，经由该算法训练得到的分类器具有很好的泛化能力，并且识别准确率较高。

AdaBoost 由 Boosting 算法发展而来。Boosting 算法^[17]的思想起源于 Valiant 于 1984 年提出的 (Probably Approximately Correct PAC) 学习模型概率近似正确^[18]，在该模型中 Valiant 提出了弱学习和强学习的概念。如果一个学习算法识别准确率仅比随机猜测略高，则称该算法为弱学习算法；相反，如果学习算法能达到理想的识别准确率，并能在多项式时间内完成，则称该算法为强学习算法。在实际应用中，人们根据实际的经验可以很容易地找到弱学习的算法，而多数情况下强学习的方法