



经典译丛

WILEY

# 数字图像目标 检测与识别 ——理论与实践

Object Detection and Recognition in Digital Images  
Theory and Practice

【波兰】Bogusław Cyganek 著

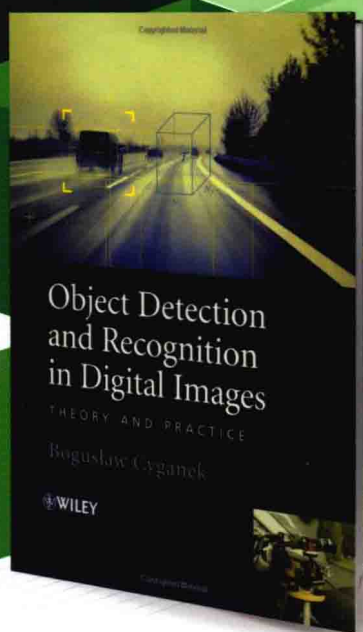
宋晓炜 杨 蕾 瞿博阳 译

李 锵 审校

信息与通信  
技术

Object Detection and Recognition in Digital Images:  
Theory and Practice

.73



中国工信出版集团



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY  
<http://www.phei.com.cn>

经典译丛·信息与通信技术

# 数字图像目标检测与识别

## ——理论与实践

Object Detection and Recognition in Digital Images  
Theory and Practice

[波兰] Bogusław Cyganek 著

宋晓炜 杨 蕾 瞿博阳 译

李 锵 审校



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

计算机视觉和机器模式识别是当前热门的研究领域,目标检测和识别是其中的关键技术。本书以作者自身丰富的项目实践经验为基础,提供了一些优选的目标检测和识别方法,特别是基于统计和基于张量的目标检测与识别方法。本书力求理论与实践密切结合,不仅以简洁明了的方式提供了这些方法的理论综述和必要的数学背景,还提供了以C++编程语言为平台的可用于指导或直接移植的实现代码,读者可基于文中及网站给出的代码开发自己工作中所需的方法。本书的实践领域主要涉及汽车应用,包括用于路标识别或驾驶监控的视觉系统。

本书可作为高等院校、科研院所相关专业的研究生、高年级本科生的教材或者教学参考书,也可以作为图像、计算机视觉和机器模式识别方向的科研和工程技术人员的参考用书。

Object Detection and Recognition in Digital Images: Theory and Practice, 9780470976371, Bogusław Cyganek.

Copyright © 2013 John Wiley & Sons, Ltd.

All rights reserved. This translation published under license.

AUTHORIZED TRANSLATION OF THE EDITION PUBLISHED BY JOHN WILEY & SONS, Ltd.

No part of this book may be reproduced in any form without the written permission of John Wiley & Sons, Ltd.

本书简体中文字版专有翻译出版权由英国 John Wiley & Sons, Ltd. 授予电子工业出版社。

未经许可,不得以任何手段和形式复制或抄袭本书内容。

版权贸易合同登记号 图字:01-2013-9185

## 图书在版编目(CIP)数据

数字图像目标检测与识别:理论与实践/(波)齐加尼克(Cyganek,B.)著;宋晓炜,杨蕾,瞿博阳译.

北京:电子工业出版社,2016.5

(经典译丛·信息与通信技术)

书名原文: Object Detection and Recognition in Digital Images: Theory and Practice

ISBN 978-7-121-28680-3

I. ①数… II. ①齐… ②宋… ③杨… ④瞿… III. ①数字图像处理-高等学校-教材 IV. ①TN911.73

中国版本图书馆CIP数据核字(2016)第092193号

策划编辑:马 岚

责任编辑:李秦华

印 刷:三河市华成印务有限公司

装 订:三河市华成印务有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

开 本:787×1092 1/16 印张:28 字数:716.8千字 彩插:6

版 次:2016年5月第1版

印 次:2016年5月第1次印刷

定 价:89.00元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888,88258888。

质量投诉请发邮件至 zlt@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: classic-series-info@phei.com.cn。

# 译 者 序

俗话说，眼见为实。对于人和动物来说，在场景中对目标进行检测与识别是一种很轻松自然的生存能力。然而，对于诸如计算机这样的机器来说，怎样才能像人和动物一样观察目标，甚至比人和动物观察得更好、检测识别得更快，这成为计算机视觉和模式识别领域的主要热点问题。对于从事该领域研究的教师、学生以及科研人员而言，一本全面翔实、富有实践价值的参考书是十分必要的。

本书是波兰 AGH 科技大学电子学系研究员及讲师 Bogusław Cyganek 博士在 2009 年 3 月出版其知名著作《三维计算机视觉技术及算法导论》(An Introduction to 3D Computer Vision Techniques and Algorithms)之后于 2013 年 8 月出版的又一力作。本书出版后，因其内容系统全面，理论体系严谨，讲解深入浅出，理论与实践结合紧密，获得了读者的普遍好评。本书具有以下几个特色。

首先，本书内容全面、重点突出。本书从张量的基本定义开始，介绍了应用于计算机视觉和模式识别任务的张量方法，并概述了常见的各种分类方法，还基于波兰道路场景中的路标识别应用阐明了目标检测与跟踪、目标识别的关键问题。全书以清晰简洁的语言，涵盖了计算机视觉的几个关键技术：分类、检测、跟踪与识别。

其次，本书的适用面广。它既适合电子工程、计算机科学、计算机工程等专业的本科生作为教材使用，也适合图像、视频信号处理，模式识别和计算机视觉方向的硕士、博士研究生使用，还可以作为相关专业的科研工作者参考用书。

第三，基于作者多年从事软件开发的实践经验，本书提供了具有很高参考价值的代码和伪代码。特别是提供了一些 C++ 语言的完整实现和用于矩阵和张量处理的 MATLAB 函数，附录中还给出用于将串行代码转换为并行版本的 OpenMP 库，原书相关网站还提供了完整的软件平台、彩色图像以及其他有用的链接，这都便于工程应用时参考使用，尤其适合高校的硕士、博士研究生进行算法研究时参考使用。

当然，目标检测、跟踪与识别是图像模式识别和计算机视觉领域的研究热点，各种新算法和技术层出不穷，本书只提供了迄今为止现代计算机视觉方法和算法的一个深入阐述。尽管如此，译者仍然认为本书是一本不可多得的优秀著作。

全书由中原工学院的宋晓炜、杨蕾和瞿博阳分工翻译。其中，前言、第 1 章和第 2 章由宋晓炜翻译，第 3 章和第 4 章由杨蕾翻译，第 5 章和附录由瞿博阳翻译。全书由天津大学李锵教授审校。另外，研究生吴源昭、牛林林、刘清丽、杨满意、蔡文静等也参与了书稿的部分整理准备工作。

感谢 Bogusław Cyganek 博士对中文译书出版方面给予的合作。感谢国家自然科学基金项目(60902063, 61440031, 61305080)在本书翻译过程中给予的支持。感谢电子工业出版社对翻译工作的大力支持，正是他们的严谨细致和辛勤付出，本书才得以顺利出版。

由于译者学识所限，疏漏乃至错误在所难免，恳请广大读者及专家不吝赐教，提出修改意见，我们将不胜感激。

译 者

2015 年 11 月于郑州

# 前 言

我们生活在一个技术革命的时代，在这个时代中，某一领域的进展常常会引发另一领域的突破。类似于 19 世纪的工业革命，近几十年可以称为计算机革命的新纪元。多年来，我们已经见证了微芯片技术的快速发展，该技术造成了成本逐年降低的、持续增长的计算能力。图形处理单元和现场可编程门阵列的并行计算系统的最新发展使之得以加强。所有这些硬件方面的成就也开启了寻求使计算机看得见并能理解它所见内容的新应用领域和可能性——这是计算机视觉领域的首要目标。然而，虽然快速计算机在这方面有很大帮助，但是真正产生影响的是新的、更好的处理方法及其实现。

本书给出了一些优选的目标检测和识别方法，特别强调了统计的以及对该领域较新的基于张量的方法。然而，有趣且重要方法的数量正在快速增长，这使得在一本书中很难提供对这些方法的全面涵盖。因此，本书的目标略有不同，即这里所选择的这些方法是我自己和我的同事在很多项目中使用过的、并且在实践中被证明是有用的方法。我们的主要领域涉及汽车应用，其中我们尝试开发用于道路标志识别或者驾驶员监控的视觉系统。当开始写这本书的时候，我的主要目的是：不仅给出这些方法的综述，还提供虽然简明但是必要的数学背景。然而，同样重要的是所讨论方法的实现。我深信详尽理论与其实现的联系是深入理解该主题的先决条件。就这方面来说，对实现平台的选择也不足为奇。贯穿本书并且在所附软件库中使用的 C++ 编程语言是世界性的工业标准。这并不意味着不可以使用不同的编程平台来完成实现，可以将所提供的代码示例用做指南或者用来直接移植。本书附有一个配套网址：[www.wiley.com/go/cyganekobject](http://www.wiley.com/go/cyganekobject)，其中包含了代码和彩图，以及 PPT、勘误表和其他有用的链接。

本书的完成是在我与合作作者 J. Paul Siebert 致力于三维图像处理的前一本书之后，我对现代计算机视觉方法着迷的结果。因此，虽然这两本书都可以单独阅读，但在某种意义上可以将本书视为我们前一本书的延续。

因此，本书可供与计算机视觉和机器模式识别相关的所有科学家和行业从业者使用，也可以用做对此快速发展领域感兴趣的学生的教程。

Bogusław Cyganek

波兰

## 致 谢

写一本书是一个巨大的任务。如果没有朋友、同事、合作伙伴以及许多其他人的帮助，那将是不可能的，有时我甚至不知道这些人的名字，但我知道他们做出了杰出的工作来使本书诞生。

我尤其感谢位于波兰克拉科夫 (Kraków) 的 AGH 科技大学以及学术计算机中心 Cyfronet 的许多同事。特别感谢 Ryszard Tadeusiewicz 教授和 Kazimierz Wiatr 教授持续的鼓励和支持。

我还想要表达对来自德国柏林洪堡大学和德国航空航天中心的 Ralf Reulke 教授以及他所在团队的所有同事的感谢，感谢在感兴趣的科学领域中我们富有成效的倾力合作。

我非常感谢 Wiley 团队，他们的帮助使这本书的出版成为可能。我想要表达对 Richard Davies、Alex King、Nicky Skinner、Simone Taylor、Liz Wingett 和 Nur Wahidah Binte Abdul Wahid、Shubham Dixit、Caroline McPherson 的特别感谢，以及对我不知道他们名字但我知道他们为促使本书诞生做了杰出工作的所有其他人的特别感谢。再次非常感谢！

我还要非常感谢世界各地的许多同事，并且特别是我前一本关于三维计算机视觉的书的读者，感谢他们的电子邮件、提问、建议、缺陷报告以及我们所进行的所有讨论。所有这些都帮助我开发更好的文本和软件。在现在和将来我还要请求他们的支持。

我想要好好表达对波兰共和国国家科学中心 CNC 的感激之情，感谢他们对在 2007 - 2009 年以及 2011 - 2013 年在合同号 no. DEC-2011/01/B/ST6/01994 下进行的科研项目的经费支持，这对本书做出了很大贡献。我还想表达对 AGH 科技大学出版社授权使用我前一本书的部分内容的感谢。

最后，我想要感谢我的家人：我的妻子 Magda、我的孩子 Nadia 和 Kamil，以及我的母亲，感谢他们在我完成本书的所有日子里的容忍、支持和鼓励。

# 符号和缩写

$B$	基矩阵	$\{x_i\}$	对于给定范围指标 $i$ 的向量 $x_i$ 的集合
$C$	数据类的数目	$x_i^{(k)}$	对于矩阵 $X_i$ 的第 $k$ 列向量
$C$	系数矩阵	$\hat{x}$	归一化列向量
$C_x$	数据集 $\{x_i\}$ 的相关矩阵	$\bar{x}$	均值向量
$D$	数据矩阵	$\bar{x}$	正交残差向量
$D$	距离函数	$x_i$	向量 $x$ 的第 $i$ 个分量
$E$	统计期望	$\Sigma_x$	数据集 $\{x_i\}$ 的协方差矩阵
$i, j, k, m, n$	自由坐标, 矩阵指标	$\rho$	直方图中的屈数目
$\mathbf{1}_n$	所有元素设置为 1 的 $n \times n$ 维矩阵	$\Delta$	直方图中的屈宽度
$I_n$	$n \times n$ 维单位矩阵	$\Omega$	分类标记的集合
$I$	图像, 一幅图像的强度信号	$\odot$	Khatri-Rao 积
$I_x, I_y$	图像 $I$ 在 $x, y$ 方向的空间导数	$\otimes$	克罗内克 (Kronecker) 积
$J$	一系列中成分的数目	$\circledast$	按元素的乘法 [阿达马 (Hadamard) 积]
$K$	核矩阵	$\oslash$	按元素的除法
$L$	向量中分量的数目, 空间维度	$\circ$	向量的外积
$M$	聚类的数目, 图像通道的数目	$\vee$	最大积
$N$	(数据)点的数目	$\wedge$	最小积
$P$	概率质量函数	$\times$	形态外积
$p$	概率密度函数	$\forall$	任意
$P, Q, C$	张量中指标的数目 (张量维度)	AD	各向异性扩散
$p, q$	张量的协变和逆变度	ALS	交替最小二乘
$R$	主成分的数目	AMI	仿射不变矩
$\Re$	实数集	AWG	自适应窗口增长
$T$	张量	CANDECOMP	(张量的) 标准分解
$T_{(k)}$	张量 $T$ 的第 $k$ 个展平模式	CID	彩色图像辨别
$T_C$	紧凑结构张量	CNMF	约束 NMF
$T_E$	扩展结构张量	CP	标准分解 (CANDECOMP)/平行因子分解 (PARAFAC)
$t$	时间坐标	CST	紧凑结构张量
$W$	向量空间	CST	卷积标准化变换
$W^*$	双向量空间	CV	计算机视觉
$X$	矩阵	CVS	计算机视觉系统
$X^T$	$X$ 的转置矩阵	DAS	驾驶辅助系统
$X_i$	(一系列矩阵中) 第 $i$ 个矩阵	DFFS	距特征空间的距离
$x, y$	空间坐标		
$x$	列向量		
$x_i$	(一系列向量中) 第 $i$ 个向量		

DIFS	特征空间中的距离	MRI	磁共振成像
DSP	数字信号处理	MSE	均方误差
DT	距离变换	NIR	近红外
EMML	期望最大化最大似然	NMF	非负矩阵因子分解
EMD	土堆转移距离	NTF	非负张量因子分解
EST	扩展结构张量	NUMA	非均匀内存访问
FCM	模糊 $c$ 均值	OC-SVM	一类支持向量机
FIR	远红外	PARAFAC	(张量)的平行因子
FN	假阴性	PCA	主成分分析
FP	假阳性	PDE	偏微分方程
GHT	广义霍夫(Hough)变换	PDF	概率密度函数
GLOH	梯度位置和方向直方图(图像描述符)	PERCLOSE	闭眼百分比
GP	高斯过程/遗传程序设计	PR	模式识别
GPU	图形处理单元(显卡)	PSNR	峰值信噪比
HDR	高动态范围(图像)	R1NTF	秩 1 非负张量因子分解
HI	高光谱图像	RANSAC	随机采样一致性
HNN	汉明神经网络	RBF	径向基函数
HOG	梯度直方图	RLA	Richardson-Lucy 算法
HOOI	高阶正交迭代	RMSE	均方根误差
HOSVD	高阶奇异值分解	ROC	受试者工作特征
ICA	独立成分分析	ROI	感兴趣区域
IMED	图像欧几里得距离	RRE	相对重构误差
IP	图像处理	SAD	绝对差值和
ISM	隐式形状模型	SIMCA	类模拟软独立建模
ISRA	图像空间重构算法	SIFT	尺度不变特征变换(图像描述符)
KFCM	核模糊 $c$ 均值	SLAM	同步定位和映射
$k$ -NN	$k$ 最近邻	SMO	序列最小优化
KPCA	核主成分分析	SNR	信噪比
K-R	Khatri-Rao 积	SOM	自组织映射
LP	对数极坐标	SPD	显著点检测器
LN	像素局部邻域	SSD	差的平方和
LSH	位置敏感哈希(Hashing)	ST	结构张量
LSQE	采用二次等式约束的最小二乘问题	SURF	快速鲁棒特征(图像描述符)
LQE	线性二次型估计	SVM	支持向量机
MAP	最大后验概率分类	TDCS	张量辨别彩色空间
MICA	多线性 ICA	TIR	热红外
ML	最大似然	TN	真阴性
MNN	形态神经网络	TP	真阳性
MoG	混合高斯	WOC-SVM	加权一类支持向量机
MPI	消息传递接口	WTA	胜者通吃



# 目 录

第1章 引言 .....	1
1.1 计算机视觉的一个例子 .....	2
1.2 全书内容概览 .....	5
参考文献 .....	7
第2章 计算机视觉中的张量方法 .....	8
2.1 摘要 .....	8
2.2 张量——一个数学对象 .....	8
2.2.1 线性空间的主要属性 .....	9
2.2.2 张量的概念 .....	9
2.3 张量——数据对象 .....	11
2.4 张量的基本属性 .....	13
2.4.1 张量指标和分量的符号 .....	14
2.4.2 张量积 .....	16
2.5 张量距离测量 .....	17
2.5.1 张量距离概述 .....	19
2.5.2 欧几里得图像距离和标准化变换 .....	24
2.6 张量场的滤波 .....	27
2.6.1 张量数据的顺序统计滤波 .....	27
2.6.2 各向异性扩散滤波 .....	30
2.6.3 扩散过程的实现 .....	32
2.7 采用结构张量观察图像 .....	37
2.7.1 二维图像空间中的结构张量 .....	39
2.7.2 空时结构张量 .....	41
2.7.3 多通道和尺度空间结构张量 .....	43
2.7.4 扩展结构张量 .....	45
2.8 采用惯性张量和矩的目标表示 .....	52
2.9 张量的特征分解和表示 .....	57
2.10 张量不变量 .....	60
2.11 多视点几何：多焦点张量 .....	60
2.12 多线性张量方法 .....	63
2.12.1 多线性代数的基本概念 .....	65
2.12.2 高阶奇异值分解(HOSVD) .....	94

2.12.3	HOSVD 的计算	96
2.12.4	HOSVD 诱导基	102
2.12.5	张量最佳秩 1 近似	103
2.12.6	张量的秩 1 分解	105
2.12.7	最佳秩( $R_1, R_2, \dots, R_p$ )近似	110
2.12.8	最佳秩( $R_1, R_2, \dots, R_p$ )近似的计算	112
2.12.9	子空间数据表示	123
2.12.10	非负矩阵因子分解	126
2.12.11	非负矩阵因子分解的计算	129
2.12.12	采用 NMF 的图像表示	133
2.12.13	非负矩阵因子分解的实现	135
2.12.14	非负张量因子分解	141
2.12.15	目标识别的多线性方法	144
2.13	结束语	149
2.13.1	本章小结	149
2.13.2	延伸阅读	150
	习题	151
	参考文献	152
<b>第 3 章</b>	<b>分类方法和算法</b>	<b>158</b>
3.1	摘要	158
3.2	分类框架	158
3.3	用于目标识别的子空间方法	162
3.3.1	主成分分析	162
3.3.2	子空间模式分类	180
3.4	目标识别的统计公式	184
3.4.1	参数化和非参数化方法	184
3.4.2	概率框架	184
3.4.3	贝叶斯决策规则	185
3.4.4	最大后验分类方案	186
3.4.5	二元分类问题	187
3.5	参数化方法——混合高斯	188
3.6	卡尔曼滤波器	192
3.7	非参数化方法	195
3.7.1	基于直方图的技术	195
3.7.2	比较直方图	197
3.7.3	多维直方图的实现	201
3.7.4	Parzen 方法	203

3.8	均值移位方法	207
3.8.1	均值移位简介	208
3.8.2	连续自适应均值移位方法	212
3.8.3	均值移位跟踪的算法方面	214
3.8.4	CamShift 方法的实现	217
3.9	神经网络	220
3.9.1	概率神经网络	220
3.9.2	概率神经网络的实现	222
3.9.3	汉明神经网络	226
3.9.4	汉明神经网络的实现	229
3.9.5	形态神经网络	233
3.10	视觉模式识别中的核	241
3.10.1	核函数	245
3.10.2	核的实现	249
3.11	数据聚类	253
3.11.1	$k$ 均值方法	255
3.11.2	模糊 $c$ 均值	257
3.11.3	核模糊 $c$ 均值	259
3.11.4	聚类质量的测量	261
3.11.5	实现问题	262
3.12	支持向量域描述	271
3.12.1	支持向量机的实现	276
3.12.2	一类分类器集成的体系结构	276
3.13	本章附录——用于模式分类的 MATLAB 和其他软件包	278
3.14	结束语	279
3.14.1	本章小结	279
3.14.2	延伸阅读	279
	习题	280
	参考文献	281
<b>第4章</b>	<b>目标检测和跟踪</b>	<b>288</b>
4.1	简介	288
4.2	直接像素分类	288
4.2.1	基准数据采集	289
4.2.2	实例研究——人类皮肤检测	289
4.2.3	实例研究——基于像素的路标检测	293
4.2.4	采用分类器集成的基于像素的图像分割	300
4.3	基本形状检测	303

4.3.1	线段的检测	304
4.3.2	凸形状的 UpWrite 检测	305
4.4	图形检测	307
4.4.1	从特征点进行的规则形状检测	308
4.4.2	显著点的聚类	311
4.4.3	自适应窗生长方法	312
4.4.4	图形验证	313
4.4.5	实例研究——路标检测系统	315
4.5	实例研究——路标跟踪和识别	319
4.6	实例研究——用于目标跟踪的框架	323
4.7	行人检测	329
4.8	结束语	334
4.8.1	本章小结	334
4.8.2	延伸阅读	334
	习题	334
	参考文献	335
<b>第 5 章</b>	<b>目标识别</b>	<b>339</b>
5.1	摘要	339
5.2	从张量相位直方图和形态尺度空间进行的识别	339
5.2.1	在形态尺度中张量相位直方图的计算	341
5.2.2	张量相位直方图的匹配	343
5.2.3	实例研究——在形态尺度空间中采用张量相位直方图进行的目标识别	344
5.3	基于不变量的识别	349
5.3.1	实例研究——采用仿射不变矩的象形图识别	349
5.4	基于模板的识别	352
5.4.1	用于路标识别的模板匹配	353
5.4.2	用于模板匹配的专用距离	355
5.4.3	采用对数极坐标和尺度空间进行的识别	356
5.5	从可变形模型进行的识别	362
5.6	分类器集成	363
5.7	实例研究——用于从变形原型中进行路标识别的分类器集成	365
5.7.1	路标识别系统的体系结构	368
5.7.2	用于警告标志识别的模块	371
5.7.3	仲裁单元	375
5.8	基于张量分解的识别	376
5.8.1	在由模式张量 HOSVD 分解所张成的子空间中进行的模式识别	376
5.8.2	实例研究——基于采用可变形模式原型的张量分解的路标识别系统	378

5.8.3 实例研究——采用张量分解方法进行的手写数字识别 .....	384
5.8.4 张量子空间分类器的实现 .....	387
5.9 用于驾驶员状态监控的人眼识别 .....	391
5.10 目标分类识别 .....	396
5.10.1 基于部分的目标识别 .....	397
5.10.2 采用视觉词袋的识别 .....	397
5.11 结束语 .....	400
5.11.1 本章小结 .....	400
5.11.2 延伸阅读 .....	401
习题 .....	402
参考文献 .....	402
附录 A .....	406

# 第1章 引言

向里面看，不要忽略任何一件东西的特质或价值。

——马可·奥里利乌斯《沉思录》，公元170-180  
(Meric Casaubon 1634年译)

本书给出了在计算机视觉中所选择的目标检测和识别方法，结合了理论、实现以及应用。大多数所选择的方法在实际汽车视觉系统中使用。然而，区分出两组方法。第一组方法包含基于张量的方法，这些方法在过去10年中已经在图像处理和模式分析领域开拓了新前沿。第二组方法建立在数学统计上。在许多情况下，从这两组方法中提取目标检测和识别方法。正如标题中所示，解释这些方法的主要概念并且给出这些方法的数学推导与它们在实际应用中的实现和使用同等重要。虽然目标检测和识别是密切关联的，在某种程度上可以将这两个领域视为模式分类，并且检测通常在识别之前，我们对二者进行了区分。在定义中，目标检测主要涉及回答关于给定类型的目的是否存在于图像中的问题。有时，它们当前的外观和位置也是重要的。另一方面，目标识别的目的是辨别它的特定类型。例如，我们可以检测一张脸，或者在此之后对一个具体的人进行识别。同样，在路标识别系统中，对于诸如“让行”等某些标志，它们的检测无异议地揭示了它们的类别。然而，对于它们中的大多数，首先检测它们的特征形状，随后识别它们的特定类型，诸如“40 km/h 限速”等。

在所观察场景中对目标进行检测与识别是一种自然生物能力。在日常生活中，人和动物可很轻松地行使这种能力，以便在移动中不发生碰撞、寻找食物、避免威胁等。然而，用于场景分析的类似的计算机方法和算法尽管有极大的进展，但是它们并不那么直接。不过，在密切观察和分析之后，生物系统为它们的机器实现提供了某些提示。在这里，一个很好的例子是人工神经网络，在多样性方面，类似于生物神经元系统，并且在软件实现方面，计算机经常使用它们进行目标识别。这就是计算机科学的分支、即所谓的计算机视觉(CV)如何发展的情况。它的主要目标是使计算机像人一样观看，或者甚至比人更好。有时这成为可能。

由于技术突破，目标检测与识别领域不断变化，以至于为该领域大多数重要的主题准备一个多卷出版物似乎是不可能的。每个月都会发表具有新的想法、定理、算法等数百篇新的论文。另一方面，最快最丰富的信息源是因特网。你可以在诸如维基百科(Wikipedia)的大量网页上查找几乎所有的主题。因此，必须声明现在写一本关于计算机视觉的书的目的与几年前有所不同。当面对一个新的技术问题并且我们的任务是解决它或者设计一个为解决该问题的系统时，丰富的信息集与知识和经验之间的差异开始变得尤为重要。在这种情况下，需要一种思考的方式帮助人们理解自然状况，并且也需要一种方法带领人们更接近可能的解决方案。本书就是关于图像目标识别中不

同项目的工作。为了能够应用一种给定的方法，我们需要首先理解它。在这个阶段，不仅是用一个最终的公式来总结一种方法，而且它的详细数学背景也很有用。另一方面，仅仅是公式还不能解决问题，我们需要它们的实现。这是第二个阶段，有时需要比前一阶段付出更多的时间和工作。本书的主要目标之一是将两个领域结合在一组选定的有用目标检测和识别方法上。在这方面，我希望本书对自学以及在处理具体问题作为参考的人都有实际用途。然而，我们不能贯穿全部方法的所有阶段，但是我希望本书至少能为在这个引人入胜并且不断变化领域内的深入学习和开发提供一个良好的开端。

正如标题所示，我的目标之一是将理论和实践相结合，经验是这种结合会导致对该主题的深入理解。这可以通过目标检测和识别汽车应用的实例研究获得进一步巩固。因此，本书的章节可以分组如下：

- 方法的提出、它们的主要概念以及数学背景。
- 包括C++代码列表的方法实现(这类章节用单词实现指示)。
- 特定应用分析(它们的名字以实例研究开始)。

除此之外，还有一些专用词目，其包括采用示例对某些数学概念进行的简要说明，旨在帮助理解附近章节中的数学推导。

对代码示例的评论。在这样的一本书里不应该浪费篇幅介绍C、C++或者计算机科学的其他基本原理。原因至少有两个：第一个原因是对于计算机科学，有许多很好的书籍可用，在书中提供了这些书籍作为参考文献。第二个原因是为了不使读者的注意力偏离本书的主要目的，这个目的是深入阐述现代计算机视觉方法和算法。另一方面，不熟悉C++的读者可以跳过详细代码说明并且关注在其他平台中的实现。然而，没有比通过在实际应用中实际测试和使用更好的学习该方法的方式。

本书以我在很多科研项目工作中积累的经验为基础。我也在一些会议和期刊文章中发表了这些工作的结果。在这方面，有两本我以前写的书值得关注。第一本书是与J. Paul Siebert合著的《三维计算机视觉技术及算法导论》(An Introduction to 3D Computer Vision Techniques and Algorithms)<sup>[1]</sup>，由Wiley出版社于2009年出版。第二本书是我的教授资格论文<sup>[2]</sup>，由位于波兰克拉科夫(Kraków)的AGH科技大学出版社于2009年出版。第二本书的扩展部分包含在了本书的不同章节中，并获得了AGH大学出版社的授权许可。

## 1.1 计算机视觉的一个例子

在本节中，将简要介绍计算机视觉在驾驶监控系统以及场景分析系统中的一些应用。这两个系统都属于车载驾驶辅助系统，旨在如通过告知驾驶员即将到来的道路标志并且防止由于驾驶员睡着引起的交通事故。

图1.1描绘了一个安装在测试车辆上的摄像机系统。这些摄像机可以观察驾驶员，并允许该系统监控他的状态。摄像机也可以观察车辆的前方，以便进行行人检测或路标识别，在这种情况下，它们可以发送一幅如图1.2中所示的图像。

我们可以从这幅图像中提取什么类型的信息呢？这当然取决于我们的目标。在所描绘的实际交通情况下，人们主要感兴趣的是安全驾驶车辆、避让行人以及其他运动或停放的车辆，以及对交通信号和标志进行辨认并做出反应。然而，在某人发送给我们这幅图像的情况下，可能有兴趣找出街道的名称。计算机视觉可以为人们做什么？某种程度上所有上述内容都可以做到，而且很快计算机视觉就可以驾驶一辆汽车，至少在某些特定条件下可以做到。让我们看看通过计算机视觉方法进行处理的一些阶段，在下面各章节中讨论这些阶段的细节。



图 1.1 安装在汽车中的摄像机系统。这些摄像机可以观察驾驶员以便监控他的状态。这些摄像机也可以观察车辆的前方以便进行行人检测或路标识别。这种视觉模块将可能很快作为汽车车载驾驶员辅助系统的一部分并成为标准装置

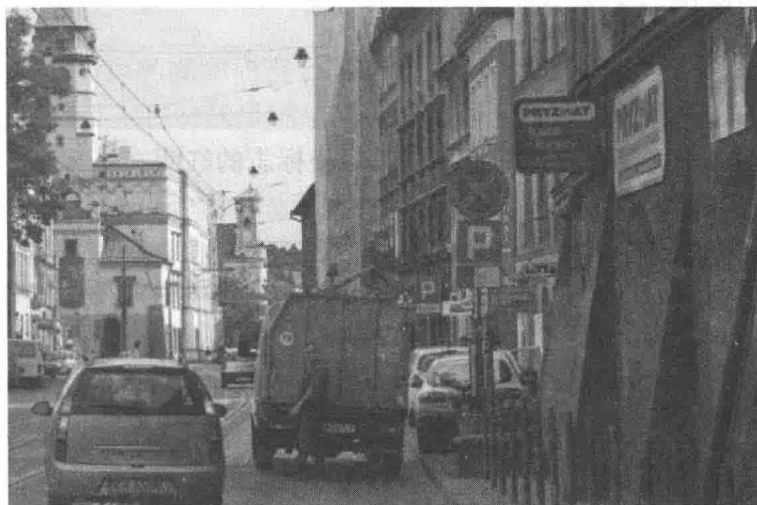


图 1.2 一个交通场景。具有摄像机的车载计算机可以提供关于道路场景的信息，以帮助安全驾驶。但是计算机视觉也可以帮助识别这幅照片是在什么地方拍摄的



如图 1.3(a)所示,即便是单独一幅彩色图像也具有三个维度。在多幅图像或者视频流的情况下,维度将会进一步增长。这样,需要工具对这种结构进行分析。张量在这方面提供了新的可能性。同时,最新开发的分析方法可深刻理解多维结构中的信息,并且深刻理解它们的压缩或者特征提取用于进一步分类。很多对计算机视觉和模式识别的研究是关于特征检测和属性的。在这方面,对这种变换进行了研究,其将原始亮度或彩色像素变为某些新的表示,这些表示提供了关于图像内容的某些知识,或者更适合于找到特定目标。在图 1.3(b)中示出了将结构张量应用于图 1.2 中的图像以便对具有较强局部结构的区域进行检测的一个例子。对找到的结构用颜色进行编码——用不同的颜色代表其方向,而用色饱和度代表其强度。我们观察到,无显著结构的区域对该滤波器没有显示出响应。在图 1.3(b)中是黑色的。如将要示出的,证明这种表示对在图像中找到诸如行人、汽车或者路标等特定图形是非常有用的。

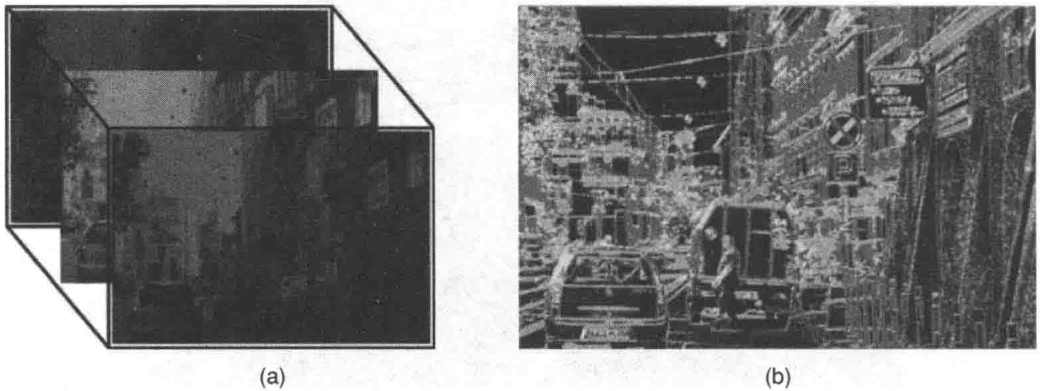


图 1.3 (a)可以将彩色图像视为三维结构;(b)可以采用张量来分析这种多维信号的内部属性。可以采用结构张量来检测局部结构。在这里,不同的颜色对具有强信号变化区域的方向进行编码,如边缘。具有弱纹理的区域是黑色的。这些特征可以用于检测行人、汽车、路标以及其他目标(对于该图的彩色版本,参见彩插)

现在,让我们简要示出在图 1.2 的图像中进行路标检测的可能步骤。在该方法中,首先通过将特定彩色特性快速分割到不同的预期标志组对标志进行检测。例如,使用红色分割辨认全红目标,这些全红目标也可以是禁止标志的红圈,对于所有感兴趣的颜色,以此类推。

图 1.4 示出了图 1.2 中的图像分别经过红色和蓝色分割之后获得的二值图。有很多分割方法,在本书中将进行详细讨论。在这种情况下,我们使用手工采集的彩色样本用于训练支持向量分类器。

我们需要从图 1.4 中找到一种选择目标的方式,这些目标的形状和大小可能对应于正在寻找的路标。这可以通过依靠显著点检测的特定方法以及定义候选目标的可能形状和大小的模糊逻辑规则来完成。

图 1.5 示出了这些标志的检测区域。现在需要将它们供给将要提供最终响应的下一个分类器,该最终响应是:首先我们是否真的观察到一个标志而不是一个交通灯,然后它是哪一类特定标志。然而,所观察的标志可以是任何大小,并且也可以旋转。可以处