



普通高等教育“十三五”规划教材  
国家新闻出版改革发展项目库入库项目  
数据科学与大数据技术专业教材丛书

# 计算机视觉

COMPUTER VISION

双 锴◎编著



北京邮电大学出版社  
[www.buptpress.com](http://www.buptpress.com)



普通高等教育“十三五”规划教材  
 国家新闻出版改革发展项目库入库项目  
 数据科学与大数据技术专业教材丛书

# 计算机视觉

双 铭 编著

大数据专业教材编委会

总主编：关 斌

编

编

编

计算机视觉  
 编 双 铭  
 责任编辑：解 强  
 出版发行：北京邮电大学出版社  
 地址：北京信箱 3761号  
 发行部：010-61916000  
 邮购部：010-61916000  
 电子邮箱：bjbupt@pubit.com.cn  
 印刷：北京邮电大学印刷厂  
 开本：787mm×1092mm  
 印张：11.5  
 字数：283千字  
 印次：2020年1月



北京邮电大学出版社  
 www.buptpress.com

010-61916000

ISBN 978-7-309-52948-8

北京邮电大学出版社出版，地址：北京信箱3761号，邮编：100876。

## 内 容 简 介

本书介绍关于计算机视觉的前沿问题,特别关注用深度学习方法解决图像理解方面的任务。在绪论部分,本书回顾了计算机视觉近几十年的发展历程,总览现代计算机视觉的研究内容。本书第2~4章包含计算机视觉的基础知识的讲解,涉及图像处理、机器学习以及深度学习的基础知识,为后续章节的内容提供了必要的先验知识。本书第5~8章讲述图像理解的几大基础任务,包括物体识别、目标检测、语义分割、图片描述以及图片生成,包含相关领域最为经典的案例和计算机视觉领域许多前沿的研究。除讲解计算机视觉的理论知识外,本书还加入各个算法的具体实现思路,书中包含许多可以动手实验的开源代码的入口。使用深度学习方法处理图像理解任务在本书中有了全面而系统的讲述。

## 图书在版编目(CIP)数据

计算机视觉 / 双锴编著. -- 北京:北京邮电大学出版社,2020.1

ISBN 978-7-5635-5946-6

I. ①计… II. ①双… III. ①计算机视觉 IV. ①TP302.7

中国版本图书馆CIP数据核字(2019)第280389号

---

书 名:计算机视觉

作 者:双 锴

责任编辑:刘 颖

出版发行:北京邮电大学出版社

社 址:北京市海淀区西土城路10号(100876)

发 行 部:电话:010-62282185 传真:010-62283578

E-mail:publish@bupt.edu.cn

经 销:各地新华书店

印 刷:保定市中华美凯印刷有限公司

开 本:787 mm×1 092 mm 1/16

印 张:11.5

字 数:283千字

版 次:2020年1月第1版 2020年1月第1次印刷

---

ISBN 978-7-5635-5946-6

定价:39.00元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

# 大数据顾问委员会

宋俊德 王国胤 张云勇 郑宇  
段云峰 田世明 娄瑜 孙少隣  
王柏

## 大数据专业教材编委会

总主编：吴斌

编委：宋美娜 欧中洪 鄂海红 双锴

于艳华 周文安 林荣恒 李静林

袁燕妮 李劼 皮人杰

总策划：姚顺

秘书长：刘纳新

双锴

2019年4月于北京邮电大学

这是一本关于使用深度学习方法处理计算机视觉任务的教材,书中涵盖计算机视觉关于图像理解的各个任务,内容的实时性强,涉及近几年深度学习浪潮下出现的新技术。笔者试图用尽可能少的公式推导和通俗易懂的语言描述,使读者能对现代计算机视觉技术有一个直观的认识。这是一本适合高年级本科生和研究生入门计算机视觉的教材,也可供相关领域的工程技术人员参阅。

作为计算机视觉领域普及类教材,本书从基本的视觉色彩原理到神经网络模型,再到卷积模型与具体应用,对计算机视觉领域的知识进行了较为系统的介绍,对计算机视觉领域的学习具有较大的帮助与指引作用。全书共分为8个章节,前4章介绍了图像处理 and 深度学习的基础知识,建议初学者全面阅读;第5~8章各章节相对独立,内容包括图像分类、目标检测等任务,读者可根据自己的兴趣和需要选择性阅读。

书中每个章节都提供了一些精选的思考题。这些思考题不仅能帮助读者回顾本章的基础知识,更重要的是引导读者对本章的知识点进行思考。对于一般课程,这些习题相对足够。为了给学有余力的读者进一步提升的空间,建议在授课时选择性地辅以编程大作业,把知识融汇贯通到实践中可以加深对知识的理解。

本书在编著时尽可能涵盖计算机视觉图像理解的各个任务,同时在讲解现代计算机视觉技术时尽可能直白,为的是让读者能够快速入门计算机视觉,使读者能够对计算机视觉有一个直观的认识,这就导致笔者无法兼顾一些具体的技术细节,同时计算机视觉的发展日新月异,书中的一些叙述难免有些过时,更多的内容需要读者深入阅读。本书在这些地方用\*号做了标注,为读者提供扩展阅读的材料,便于感兴趣的读者进一步了解。

对于一些英文专业术语,译为中文术语更便于读者理解时,笔者提供了中文翻译,而大部分的英文术语,笔者未做翻译,一是由于意译的结果往往不统一,二是这样可以使读者在日后进一步阅读文献时更加熟悉此类术语。

目前随着人工智能领域的极速发展,计算机视觉也成为研究的热门,大量的研究推动着该领域的进步。作为该领域的研究者之一,由于时间和精力有限,笔者难以对计算机视觉的各个方面都有精深的理解,书中难免有谬误之处,欢迎细心的读者指出。

双 错

2019年4月于北京邮电大学

# 目 录

<b>第1章 绪论</b> .....	1
本章思维导图.....	1
1.1 计算机视觉简史 .....	2
1.2 2012年——计算机视觉发展的新起点 .....	4
1.3 计算机视觉应用 .....	5
1.4 GPU与并行技术——深度学习和计算机视觉发展的加速器.....	7
1.5 基于卷积神经网络的计算机视觉应用 .....	8
1.6 全书章节简介.....	13
本章思考题 .....	13
本章参考文献 .....	14
<b>第2章 图像的表达</b> .....	15
本章思维导图 .....	15
2.1 色彩学基础.....	16
2.1.1 三基色原理.....	16
2.1.2 彩色模型.....	16
2.1.3 小结.....	17
2.2 图像数字化.....	18
2.2.1 采样.....	18
2.2.2 量化.....	19
2.2.3 图像的性质.....	19
2.2.4 像素间的关系.....	19
2.2.5 对比度.....	21
2.2.6 敏锐度.....	21
2.2.7 图像中的噪声.....	22
2.2.8 小结.....	22
2.3 图像预处理.....	22

2.3.1	灰度化	23
2.3.2	几何变换	23
2.3.3	图像增强	25
2.3.4	小结	29
2.4	本章总结	29
	本章思考题	29
	本章参考文献	29
<b>第3章</b>	<b>特征提取</b>	<b>31</b>
	本章思维导图	31
3.1	局部特征点检测概述	32
3.2	角点检测	32
3.2.1	角点介绍	32
3.2.2	Harris 角点	34
3.2.3	Fast 角点	35
* 3.2.4	FAST-ER 角点检测子	36
3.2.5	小结	37
3.3	斑点检测	37
3.3.1	斑点介绍	37
3.3.2	LOG 斑点检测	37
* 3.3.3	DOG 斑点检测	39
* 3.3.4	DOH 斑点检测	39
3.3.5	SIFT 斑点检测	40
3.3.6	SURF 斑点检测	40
3.3.7	小结	41
3.4	特征描述子	41
3.4.1	特征描述子介绍	41
3.4.2	BRIEF 描述子	41
3.4.3	ORB 特征提取算法	42
3.4.4	BRISK 特征提取算法	42
3.4.5	FREAK 特征提取算法	43
3.4.6	小结	43
3.5	边缘检测	43
3.5.1	边缘介绍	43
3.5.2	边缘检测介绍	44
3.5.3	边缘检测的基本方法	44
3.5.4	边缘检测算子的概念	45

3.5.5	常见的边缘检测算子	45
3.5.6	梯度算子介绍	46
3.5.7	梯度的衡量方法	46
3.5.8	如何用梯度算子实现边缘检测	47
3.5.9	小结	47
3.6	一阶微分边缘算子	48
3.6.1	一阶微分边缘算子的基本思想	48
3.6.2	Roberts 算子	48
3.6.3	Prewitt 算子	49
3.6.4	Sobel 算子	50
3.6.5	Kirsch 算子	51
3.6.6	小结	51
3.7	二阶微分边缘算子	51
3.7.1	二阶微分边缘算子的基本思想	51
3.7.2	Laplace 算子	52
3.7.3	LOG 算子	53
3.7.4	Canny 算子	53
3.7.5	小结	54
3.8	基于窗口模板的检测方法	54
3.8.1	SUSAN 检测方法介绍	54
3.8.2	小结	56
3.9	新兴的边缘检测算法	56
3.9.1	小波分析	56
3.9.2	模糊算法	56
3.9.3	人工神经网络	56
3.9.4	小结	57
3.10	本章总结	57
	本章思考题	58
	本章参考文献	58
<b>第 4 章</b>	<b>神经网络</b>	<b>60</b>
	本章思维导图	60
4.1	感知器	61
4.1.1	基本概念	61
4.1.2	激活函数	61
4.2	神经网络基础	63

4.3	前向传播与反向传播算法	64
4.3.1	前向传播算法	64
4.3.2	反向传播算法原理	65
4.3.3	反向传播计算过程推导	67
4.4	卷积神经网络概述	70
4.5	卷积神经网络结构	71
4.6	卷积神经网络的组成	72
4.6.1	局部感知	72
4.6.2	空间排列	74
4.6.3	参数共享	75
4.6.4	卷积	75
4.6.5	池化层	76
4.6.6	全连接层	76
4.6.7	卷积神经网络架构	77
4.7	卷积神经网络的应用	77
4.8	循环神经网络概述	77
4.9	循环神经网络与语言模型	78
4.10	循环神经网络结构	79
4.11	循环神经网络的扩展与改进	80
4.11.1	Simple-RNN	80
4.11.2	双向循环神经网络	81
4.11.3	深度循环神经网络	81
4.11.4	长短期记忆网络与门控循环单元网络	82
4.12	本章总结	83
	本章思考题	84
	本章参考文献	84
<b>第5章</b>	<b>物体分类与识别</b>	<b>86</b>
	本章思维导图	86
5.1	从 AlexNet 到 GoogLeNet	87
5.1.1	AlexNet	87
5.1.2	VGGNet	90
5.1.3	GoogLeNet	95
5.1.4	小结	102
5.2	深度残差网络 ResNet	102
5.2.1	平原网络的深度限制	103
5.2.2	ResNet 的提出	104

5.2.3 残差学习突破深度限制 .....	105
5.2.4 小结 .....	108
5.3 迁移学习图像分类 .....	108
5.3.1 迁移学习简介 .....	108
5.3.2 迁移学习图像分类策略 .....	109
5.3.3 小结 .....	111
5.4 本章总结 .....	111
本章思考题 .....	111
本章参考文献 .....	112
<b>第 6 章 目标检测与语义分割</b> .....	<b>113</b>
本章思维导图 .....	113
6.1 从 RCNN 到 Faster R-CNN .....	114
6.1.1 R-CNN 网络结构 .....	114
6.1.2 交并比 .....	115
6.1.3 边框回归算法 .....	115
6.1.4 非极大值抑制 .....	116
6.1.5 SPP-Net 网络结构 .....	117
6.1.6 Fast R-CNN 网络结构 .....	118
6.1.7 Faster R-CNN 网络结构 .....	120
6.1.8 小结 .....	123
6.2 端到端方法:YOLO、SSD .....	123
6.2.1 One Stage 和 Two Stage 方法比较 .....	123
6.2.2 YOLO 网络结构 .....	123
6.2.3 SSD 网络结构 .....	125
6.2.4 小结 .....	127
6.3 从 FCN 到 Mask R-CNN .....	127
6.3.1 FCN 网络结构 .....	128
6.3.2 DeepLab 网络结构 .....	130
6.3.3 Mask R-CNN 网络结构 .....	131
6.3.4 小结 .....	132
6.4 本章总结 .....	133
本章思考题 .....	133
本章参考文献 .....	133
<b>第 7 章 图片描述与关系识别</b> .....	<b>135</b>
本章思维导图 .....	135

7.1	单词、句子在深度学习模型中的表示	136
7.1.1	One-Hot 表示	136
7.1.2	词嵌入表示	136
7.1.3	小结	139
7.2	Encoder-Decoder 模型	139
7.2.1	Encoder-Decoder 基本结构	139
7.2.2	Attention 机制	140
7.2.3	小结	145
7.3	基于 Encoder-Decoder 的图片描述与关系识别模型	145
7.3.1	NIC 网络模型	146
7.3.2	基于 Attention 的图片描述	146
7.3.3	小结	148
7.4	本章总结	148
	本章思考题	148
	本章参考文献	149
<b>第 8 章</b>	<b>生成对抗网络</b>	<b>150</b>
	本章思维导图	150
8.1	GANs 模型介绍	151
8.1.1	生成模型与判别模型	151
8.1.2	对抗网络思想	151
8.1.3	详细实现过程	151
8.1.4	小结	153
8.2	GANs 的简单理论介绍	154
8.2.1	GANs 的理论灵感	154
8.2.2	GANs 的理论证明	155
8.2.3	小结	157
8.3	GANs 的应用	157
8.3.1	文本转图像——CGAN	157
8.3.2	照片风格转化——CycleGAN	160
8.3.3	局部变脸术——StarGAN	163
8.3.4	定制图片生成——InfoGAN	165
8.3.5	小结	168
8.4	本章总结	169
	本章思考题	169
	本章参考文献	170

# 第1章

## 绪论

### 本章思维导图

作为人类,我们可以轻松感知周围的三维世界。相比之下,不管近年来计算机视觉已经取得多么令人瞩目的成果,要让计算机能像人类那样理解和解释图像,却仍然是一个遥远的梦想。为什么计算机视觉会成为如此富有挑战性的难题?它的发展历史与现状又是怎样的?本章将对计算机视觉的发展简史及应用现状进行介绍。



## 1.1 计算机视觉简史

都说“眼睛是心灵的窗口”，乍一听觉得只是一个很好听的比喻，但仔细想想，视觉对于人类的重要性是不言而喻的。现代的科学的研究也表明，人类的学习和认知活动有 80%~85% 都是通过视觉完成的。也就是说，视觉是人类感受和理解这个世界的最主要的手段。在当前机器学习成为热门学科的背景下，人工智能领域自然也少不了视觉的相关研究，这即是本书将要介绍的计算机视觉。

计算机视觉(Computer Vision)顾名思义是一门“教”会计算机如何去“看”世界的学科。计算机视觉与自然语言处理(Natural Language Process, NLP)及语音识别(Speech Recognition)并列为机器学习方向的三大热点方向。而计算机视觉也由诸如梯度方向直方图(Histogram of Gradient, HOG)以及尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)等传统的手动提取特征(Hand-Crafted Feature)与浅层模型的组合(如图 1-1 所示)逐渐转向了以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度学习模型。然而计算机视觉正式成为一门学科,则要追溯到 1963 年美国计算机科学家拉里·罗伯茨在麻省理工大学的博士毕业论文“Machine Perception of Three-Dimensional Solids”<sup>[1]</sup>。加拿大科学家大卫·休伯尔(David Hubel)和瑞典科学家托斯坦·维厄瑟尔(Torsten Wiesel)从 1958 年起通过对猫视觉皮层的研究,提出在计算机的模式识别中,和生物的识别类似,边缘是用来描述物体形状的关键信息。拉里在论文中根据上述研究,通过对输入图像进行梯度操作,进一步提取边缘,然后在 3D 模型中提取出简单形状结构,之后利用这些结构像搭积木一样去描述场景中物体的关系,最后获得从另一角度看图像物体的渲染图。在拉里的论文中,从二维图像恢复图像中物体的三维模型的尝试,正是计算机视觉和传统图像处理学科思想上最大的不同:计算机视觉的目的是让计算机理解图像的内容。所以这算是与计算机视觉相关的最早的研究。



图 1-1 传统的手动提取特征与浅层模型的组合

20 世纪 70 年代:从有了计算机视觉的相关研究开始,一直到 20 世纪 70 年代,人们关心的热点都偏向图像内容的建模,如三维建模、立体视觉等。比较有代表性的弹簧模型<sup>[2]</sup>(Pictorial Structure, 如图 1-2 所示)和广义圆柱体模型(Generalized Cylinder, 如图 1-2 所示)就是在这个时期被提出来的。那个时期将视觉信息处理分为三个层次:计算理论、表达和算法、硬件实现。在如今看来,或许有些不合理,但是却将计算机视觉作为了一门正式学科的研究。而且其方法论到今天仍然是表达和解决问题的好向导。

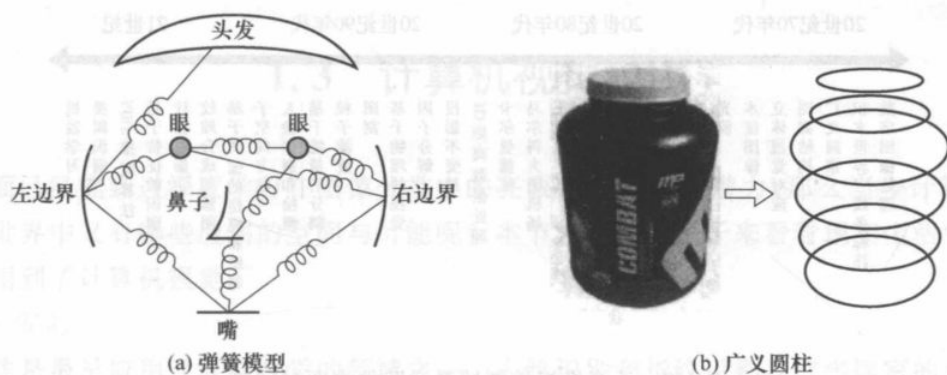


图 1-2 弹簧模型与广义圆柱

20 世纪 80 年代:在视觉计算理论提出后,计算机视觉在 20 世纪 80 年代进入了最蓬勃发展的一个时期。主动视觉理论和定性视觉理论等都在这个时期被提出,这些理论认为人类的视觉重建过程并不是马尔理论中那样直接,而是主动的、有目的性和选择性的。同时从 20 世纪 80 年代起,这个学科开始慢慢脱胎于神经科学,更多偏重计算和数学的方法开始发展起来,相关的应用也变得更加丰富。著名的图像金字塔和 Canny 边缘检测算法在这个时期被提出,图像分割和立体视觉的研究在这个时期也蓬勃发展,当然还有和本书更紧密的基于人工神经网络的计算机视觉研究,尤其是模式识别的研究也伴随着人工神经网络的第一次复兴变得红火起来。

20 世纪 90 年代:进入 20 世纪 90 年代,伴随着各种机器学习算法的全面开花,机器学习开始成为计算机视觉,尤其是识别、检测和分类等应用中一个不可分割的重要工具。各种识别和检测算法迎来了大发展。尤其是人脸识别在这个时期迎来了一个研究的小高潮。各种用来描述图像特征的算子也不停地被发明出来,如耳熟能详的 SIFT<sup>[3]</sup>算法就是在 20 世纪 90 年代末被提出的。另外伴随着计算机视觉在交通和医疗等工业领域的应用越来越多,其他一些的基础视觉研究方向,如跟踪算法、图像分割等,在这个时期也有了一定的发展。

21 世纪:进入 21 世纪之后,计算机视觉已经俨然成为计算机领域的一个大学科。国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR) 和 ICCV 等会议已经是人工智能领域,甚至是整个计算机领域内的大型盛会,甚至出现了一些新的子方向,如计算摄影学 (Computational Photography) 等。在传统的方向上基于特征的图像识别成了一个大热门,斯坦福大学的李飞飞教授牵头创立了一个非常庞大的图像数据库 ImageNet。ImageNet 里包含 1 400 万张图像,超过 20 000 个类别。基于这个数据库,自 2010 年开始,每年举办一次的大规模视觉识别挑战比赛 (ImageNet Large Scale Visual Recognition Challenge, ILSVRC),采用了 ImageNet 里 1 000 个子类目的超过 120 万张图片作为数据,参赛者来自世界各国的大学、研究机构和公司,成为了计算机视觉领域最受关注的事件之一。图 1-3 为计算机视觉领域最活跃的主题时间轴。

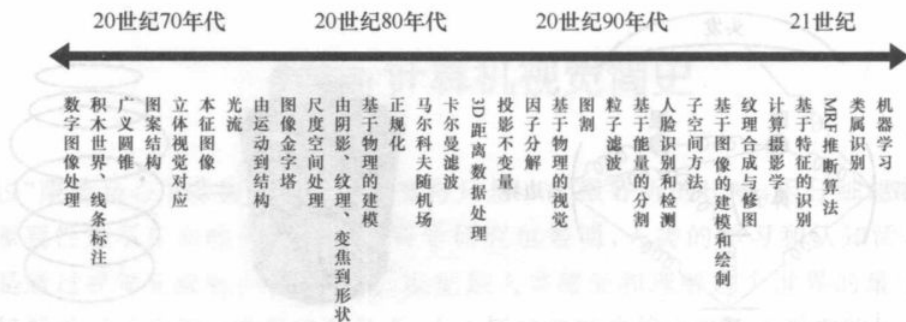


图 1-3 计算机视觉领域最活跃的主题时间轴

若想详细了解论文“Machine Perception of Three-Dimensional Solids”，请扫描书右侧的二维码。



## 1.2 2012年——计算机视觉发展的新起点 论文“Machine Perception of Three-Dimensional Solids”

ILSVRC 举办的前两年，各种“手工设计特征+编码+SVM”框架下的算法一直是该项比赛的前几名。ILSVRC 的分类错误率的标准是让算法选出最有可能的  $S$  个预测，如果有一个是正确的，则算通过，如果都没有预测对，则算错误。2010 年 ILSVRC 的冠军是 NEC 的余凯带领的研究组，错误率达到了 28%。2011 年施乐欧洲研究中心的小组将这个成绩提高到了 25.7%。

2012 年，Hinton 的小组也参加了竞赛，主力选手是 Hinton 的一名研究生 Alex Krizhevsky。在这一年的竞赛上，Alex 提出了一个 5 卷积层+2 全连接层的卷积神经网络 AlexNet<sup>[4]</sup>，并利用 CUDA 给出了实现，这个算法将前 5 类错误率从 25.7% 降到了 15.3%。在之前的 ImageNet 竞赛中，哪怕只有一个百分点的提升都是很不错的成绩，而深度学习第一次正式应用在图像分类竞赛就取得了 10 个百分点的改进，并且完胜第二名(26.2%)。这在当时对传统计算机视觉分类算法的冲击是不言而喻的。现在概括起来，当时的改进主要有以下 3 点：更深的网络结构、校正线性单元(Rectified Linear Unit, ReLU)、Dropout 等方法的应用、GPU 训练网络。

尽管在当年许多传统计算机视觉的学者仍然对 AlexNet 抱有种种质疑，如算法难以解释、参数过多(实际上比许多基于 SVM 的办法参数少)等，但自从 2012 年后，ImageNet 的参赛者几乎全体转向了基于卷积神经网络的深度学习算法，或者说计算机视觉领域全体转向了深度学习。基于深度学习的检测和识别、基于深度学习的图像分割、基于深度学习的立体视觉等如雨后春笋般发展起来。深度学习，尤其是卷积神经网络，就像一把万能的大杀器，在计算机视觉的各个领域开始发挥作用。

若想详细了解 ImageNet 竞赛及数据集相关信息，请扫描书右侧的二维码。

ImageNet 竞赛  
及数据集相关信息



## 1.3 计算机视觉应用

前面已经提到过深度学习在图像分类中的亮眼表现与应用潜力,那么当今计算机视觉在人类世界中又有哪些应用的空间与可能呢?本节将举几类例子来看看现实生活中都有哪些地方用到了计算机视觉。

### (1) 安防

安防是最早应用计算机视觉的领域之一。人脸识别和指纹识别在许多国家的公共安全系统里都有应用,因为公共安全部门拥有真正意义上最大的人脸库和指纹库。常见的应用有利用人脸库和公共摄像头对犯罪嫌疑人进行识别和布控。例如,利用公共摄像头捕捉到的画面,在其中查找可能出现的犯罪嫌疑人,用超分辨率技术对图像进行修复,并自动或辅助人工进行识别以追踪犯罪嫌疑人的踪迹;将犯罪嫌疑人照片在身份库中进行检索以确定犯罪嫌疑人身份也是常见的应用之一;移动检测也是计算机视觉在安防中的重要应用,利用摄像头监控画面移动用于防盗或者劳教和监狱的监控。

### (2) 交通

提到交通方面的应用,一些开车的朋友们一定立刻就想到了违章拍照,利用计算机视觉技术对违章车辆的照片进行分析提取车牌号码并记录在案,这是大家都熟知的一项应用。此外很多停车场和收费站也用到车牌识别。除车牌识别外,还有利用摄像头分析交通拥堵状况或进行隧道桥梁监控等技术,但应用并没有那么广泛。前面说的是道路应用,针对汽车和驾驶的计算机视觉技术也有很多,如行人识别、路牌识别、车辆识别、车距识别,还有更进一步的也即是近两年突然火起来的无人驾驶等。计算机视觉技术在交通领域虽然有很多研究,但因为算法性能或实施成本等因素,目前真正能在实际应用中发挥作用的仍然不多,交通领域仍是一个有着巨大空间的领域。

### (3) 工业生产

工业领域也是最早应用计算机视觉技术的领域之一。例如,利用摄像头拍摄的图片对部件长度进行非精密测量;利用识别技术识别工业部件上的缺陷和划痕等;对生产线上的产品进行自动识别和分类用来筛选不合格产品;通过不同角度的照片重建零部件三维模型。

### (4) 在线购物

例如,淘宝和京东的拍照购物功能。事实上计算机视觉在电商领域的应用还有更多。图片信息在电商商品列表中扮演着信息传播最重要的角色,尤其是在手机上。当我们打开购物 App 时,最先最快看到的信息一定是图片。而为了让每一位用户都能看到最干净、有效、赏心悦目的图片,电商背后的计算机视觉就成了非常重要的技术。几乎所有的电商都有违规图片检测的算法,用于检测一些带有违规信息的图片。在移动网络主导的时代,一个手机 App 的一个页面能展示图片数量非常有限,如果搜索一个商品返回的结果里有重复图片出现,则是对展示画面的巨大浪费,于是重复图片检测算法发挥了重要的作用。对于第三方商家,一些商家在商品页面发布违规或是虚假宣传的文字很容易被检测,这个时候文字识别(Optical Character Recognition, OCR)就成了保护消费者利益的防火墙。除保护消费者利

益外,计算机视觉技术也在电商领域里保护着一些名人的利益,一些精通 Photoshop 的商家常常把明星的脸放到自己的商品广告中,人脸识别便成了打击这些行为的一把利剑。

#### (5) 信息检索

搜索引擎可以利用文字描述返回用户想要的信息,图片也可以作为输入来进行信息的检索。最早做图片搜索的是一家老牌网站 TinEye,上传图片就能返回相同或相似的结果。后来随着深度学习在计算机视觉领域的崛起,Google 和百度等公司也推出了自己的图片搜索引擎,只要上传自己拍摄的照片,就能从返回的结果中找到相关的信息。

#### (6) 游戏娱乐

在游戏娱乐领域,计算机视觉的主要应用是在体感游戏,如 Kinect、Wii 和 PS4 等。在这些游戏设备上会用到一种特殊的深度摄像头,用于返回场景到摄像头距离的信息,从而用于三维重建或辅助识别,这种办法比常见的双目视觉技术更加可靠实用。此外就是手势识别、人脸识别、人体姿态识别等技术,用来接收玩家指令或和玩家互动。

#### (7) 摄影摄像

数码相机诞生后,计算机视觉技术就开始应用于消费电子领域的照相机和摄像机上。最常见的就是人脸,尤其是笑脸识别,不需要再喊“茄子”,只要露出微笑就会捕捉下美好的瞬间。新手照相也不用担心对焦不准,相机会自动识别出人脸并对焦。手抖的问题也在机械技术和视觉技术结合的手段下,得到了一定程度上的控制。近些年一个新的计算机视觉子学科——计算摄影学的崛起,也给消费电子领域带来了新玩意——“光场相机”。有了光场相机甚至不需要对焦,拍完之后回家慢慢选对焦点,聚焦到任何一个距离上的画面都能一次捕捉到。除图像获取外,图像后期处理也有很多计算机视觉技术的应用,如 Photoshop 中的图像分割技术和抠图技术,高动态范围(High Dynamic Range, HDR)技术用于美化照片,利用图像拼接算法创建全景照片等。

#### (8) 机器人/无人机

机器人和无人机中主要利用计算机视觉和环境发生互动,如教育或玩具机器人利用人脸识别和物体识别对用户和场景做出相应的反应。无人机也是近年来火热的一个领域。用于测量勘探的无人机可以在很低成本下采集海量的图片用于三维地形重建;用于自动物流的无人机利用计算机视觉识别降落地点,或者辅助进行路线规划;用于拍摄的无人机,目标追踪技术和距离判断等可以辅助飞行控制系统做出精确的动作,用于跟踪拍摄或自拍等。

#### (9) 体育

高速摄像系统已经普遍用于竞技体育中。球类运动中结合时间数据和计算机视觉的进球判断、落点判断、出界判断等。基于视觉技术对人体动作进行捕捉和分析也是一个活跃的研究方向。

#### (10) 医疗

医学影像是医疗领域中一个非常活跃的研究方向,各种影像和视觉技术在这个领域中至关重要。计算断层成像(Computed Tomography, CT)和磁共振成像(Magnetic Resoiance Imaging, MRI)中重建三维图像,并进行一些三维表面渲染都有涉及一些计算机视觉的基础手段。细胞识别和肿瘤识别用于辅助诊断,一些细胞或者体液中小型颗粒物的