

HZ BOOKS  
华章IT

RAV 丛书

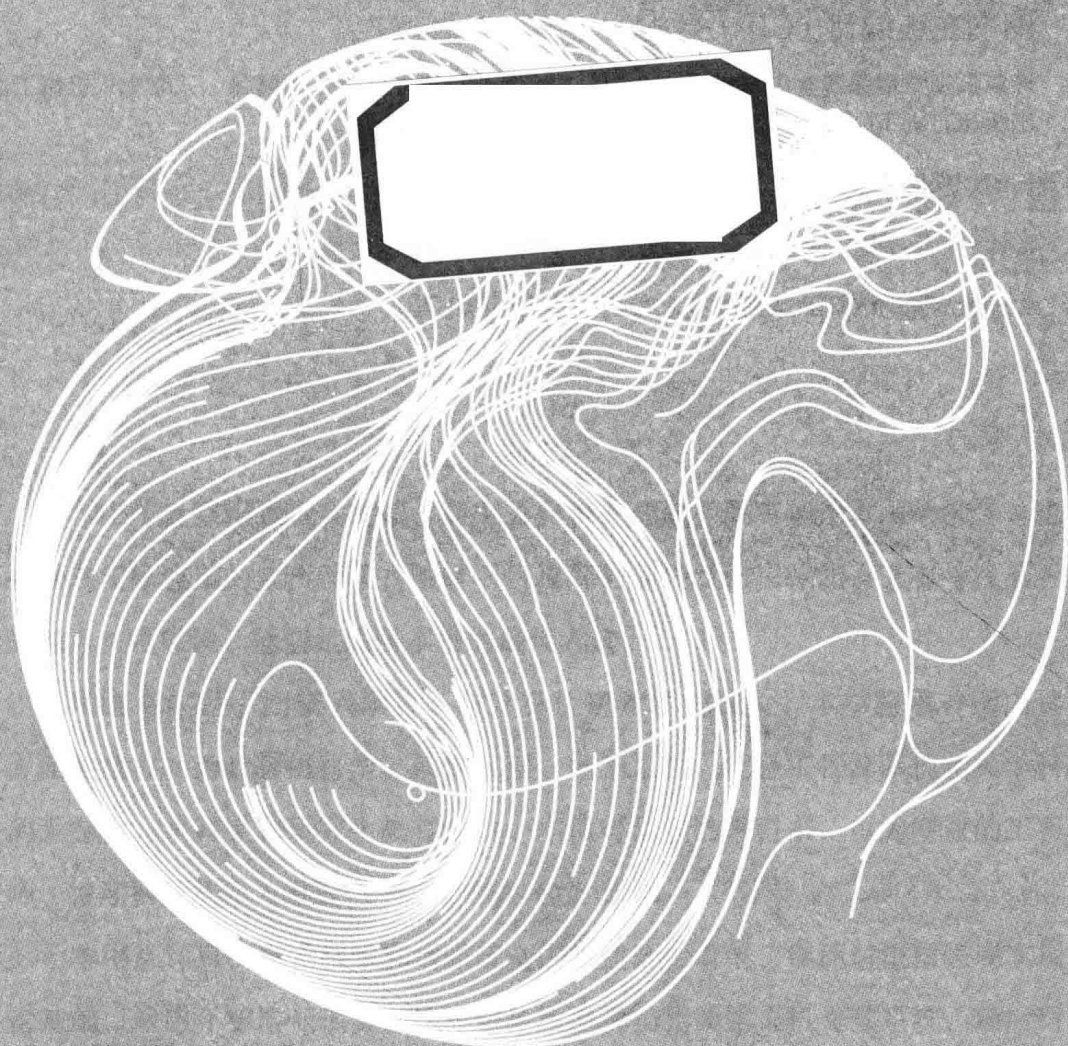
站在技术的前沿，紧跟视觉技术发展趋势

# 计算机视觉增强现实 应用概论

深圳中科呼图信息技术有限公司 ©编著



机械工业出版社  
China Machine Press



# 计算机视觉增强现实 应用概论

深圳中科呼图信息技术有限公司 ©编著



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

计算机视觉增强现实应用概论 / 深圳中科呼图信息技术有限公司编著. —北京: 机械工业出版社, 2017.7

ISBN 978-7-111-57689-1

I. 计… II. 深… III. 计算机视觉—研究 IV. TP302.7

中国版本图书馆 CIP 数据核字 (2017) 第 186001 号

## 计算机视觉增强现实应用概论

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 曲 熠

责任校对: 李秋荣

印 刷: 北京文昌阁彩色印刷有限责任公司

版 次: 2017 年 9 月第 1 版第 1 次印刷

开 本: 185mm×260mm 1/16

印 张: 12.25

书 号: ISBN 978-7-111-57689-1

定 价: 39.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379426 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东



## Forward 序 言

本书作者在增强现实的诸多研究方向都有建树，既有理论方面的准备，也有开发实际应用的宝贵经验；既有项目统筹的领导及组织才能，又有一线编程的能力。可以说，他们是编著本书的最佳人选。

增强现实是一门新兴学科，与计算机视觉紧密相关，本书将填补国内这方面的空白，为计算机科学、软件工程、电子工程、信息工程及其他相关学科的大专生、本科生和研究生教育提供一本值得信赖的教科书，也为专业的工程师、程序员及从业人员提供一本自学的有益读本。

本书从计算机视觉技术和增强现实的概述开始着笔，继而讨论增强现实理论基础，然后介绍增强现实系统，包括增强现实相关的人机交互系统，最后讨论增强现实行业应用及发展趋势。

本书的特色是在计算机视觉这个当前最有活力的领域展开系统性的讨论，既有理论基础，也有算法实现，以便读者融会贯通，方便他们在开发计算机视觉系统和增强现实系统时实现和应用书中介绍的算法。本书还介绍了现在市场上主流增强现实系统的硬件配置及主流增强现实引擎。在增强现实相关的人机交互系统方面，本书介绍了手势识别、语音识别及眼动追踪等交互系统，这些都是目前人工智能研究的热点。本书在增强现实的应用方面也着墨较多，讨论了增强现实在教育与技能培训、游戏娱乐、市场营销、文化旅游、工业与医疗及军事领域的应用。在展望未来增强现实的发展趋势时，作者特别讨论了深度学习和大数据分析在增强现实的研究和应用中所扮演的角色。

本书章节安排合理，内容组织紧贴计算机视觉和增强现实的最新研究动

向，知识深度适中，适合不同领域的读者阅读，也适合各大专院校用作教材。  
我相信通过学习本书，读者们一定受益匪浅。

池哲儒

香港理工大学电子及资讯工程系

2017年6月

## Preface 前言

近年来,随着增强现实(AR)技术的火热,越来越多的人开始关注相关领域的动态及发展,尤其是相当一部分软件开发者及技术爱好者已经开始计划从事AR行业相关的研究或开发工作。由于增强现实技术近几年才进入公众的视野,且涉及众多计算机视觉的复杂概念,相关知识又有着相当的深度及广度,因此目前市面上还没有一本系统介绍增强现实相关理论基础、软硬件系统及应用案例的专业书籍。

本书从增强现实的基本概念出发,系统介绍AR相关的理论、系统设计、行业应用及发展趋势,可以作为计算机科学、电子工程及其他相关院系的计算机视觉或增强现实基础课程的参考教材,也可以作为有意从事AR相关行业的从业人员的业务参考书。同时,对增强现实技术感兴趣的读者,也可以通过本书对AR的不同方面有一个系统的认识,为之后进一步的研究打下良好的基础。

本书部分章节涉及一些数学公式及程序实例,若想透彻理解这些内容,需要读者有良好的高等数学基础并有一定的实用基础软件编程能力,例如C++或C#。考虑到大多数读者的需要,本书第2章对增强现实相关的计算机视觉基本概念进行了简单描述。建议一般读者按照章节的顺序逐章进行阅读,有计算机视觉基础的专业读者则可按照自身情况选择相关内容进行阅读。对于非技术人员或者对AR应用及发展感兴趣的读者,可重点阅读第5~7章。在本书的编写过程中,我们得到了众多友人的支持。感谢浙江大学周泓教授为本书的修改多次提出宝贵意见,感谢香港理工大学池哲儒教授在百忙之中为本书写序。最后,感谢机械工业出版社的编辑为本书的出版付出的辛勤劳动。

虽然本书经过了多次校对，但难免会有疏漏之处。欢迎读者指出其中的错误，以便我们及时更正。

最后，愿广大读者朋友在本书的帮助下，都能很好地理解增强现实技术的各方面内容。愿本书能为增强现实在国内的普及以及在世界范围内的发展献上绵薄之力。

## 推荐阅读



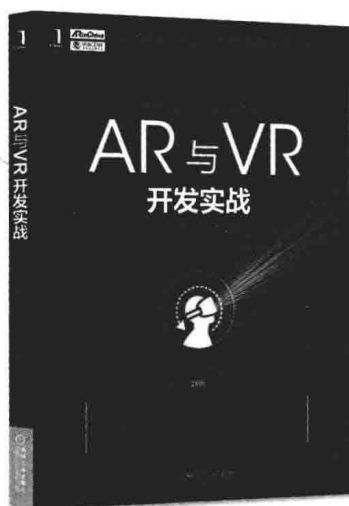
### 计算机视觉：模型、学习和推理

作者：Simon J. D. Prince 译者：苗启广等 ISBN：978-7-111-51682-8 定价：119.00元



### 计算机与机器视觉：理论、算法与实践（英文版·第4版）

作者：E. R. Davies ISBN：978-7-111-41232-8 定价：128.00元



### AR与VR开发实战

作者：张克发等 ISBN：978-7-111-55330-4 定价：69.00元



### VR/AR/MR开发实战——基于Unity与UE4引擎

作者：刘向群等 ISBN：978-7-111-56326-6 定价：129.00元



## Contents 目 录

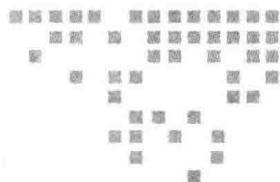
序言  
前言

|                      |    |
|----------------------|----|
| 第 1 章 计算机视觉与增强现实应用概论 | 1  |
| 1.1 计算机视觉概述          | 1  |
| 1.1.1 计算机视觉发展历程      | 2  |
| 1.1.2 计算机视觉常见模块      | 5  |
| 1.1.3 计算机视觉应用领域      | 8  |
| 小结                   | 9  |
| 1.2 增强现实概述           | 9  |
| 1.2.1 增强现实发展历程       | 11 |
| 1.2.2 增强现实表现形式       | 14 |
| 1.2.3 增强现实应用领域       | 19 |
| 小结                   | 21 |
| 第 2 章 增强现实理论基础       | 23 |
| 2.1 增强现实理论基础简介       | 23 |
| 2.2 增强现实的摄像机空间理论     | 25 |
| 2.2.1 摄像机透视模型        | 25 |
| 2.2.2 摄像机参数矩阵        | 27 |
| 2.2.3 双目摄像机立体视觉系统    | 32 |
| 小结                   | 36 |
| 2.3 图像处理基本原理         | 37 |
| 2.3.1 数字图像的色彩模型      | 37 |
| 2.3.2 图像的截取与缩放       | 40 |

|                       |                 |            |
|-----------------------|-----------------|------------|
| 2.3.3                 | 线性滤波器的运用        | 42         |
|                       | 小结              | 46         |
| 2.4                   | 图像局部特征的提取与匹配    | 46         |
| 2.4.1                 | SIFT 特征描述子      | 47         |
| 2.4.2                 | ORB 特征描述子       | 56         |
| 2.4.3                 | 模板匹配            | 64         |
|                       | 小结              | 66         |
| 2.5                   | 图像的边缘检测         | 67         |
| 2.5.1                 | 一阶导数边缘检测        | 67         |
| 2.5.2                 | 二阶导数边缘检测        | 69         |
| 2.5.3                 | Canny 算子边缘检测    | 73         |
| 2.5.4                 | 基于二值图像的模板匹配     | 74         |
|                       | 小结              | 80         |
| 2.6                   | 位姿估计            | 80         |
| 2.7                   | 用于运动状态预测的滤波器    | 83         |
| 2.7.1                 | 贝叶斯滤波器          | 83         |
| 2.7.2                 | 粒子滤波器           | 85         |
| 2.7.3                 | 卡尔曼滤波器          | 90         |
|                       | 小结              | 95         |
| 2.8                   | 即时定位与地图构建系统     | 95         |
| 2.8.1                 | 单目 LSD-SLAM     | 97         |
| 2.8.2                 | 单目 ORB-SLAM     | 104        |
|                       | 小结              | 107        |
| <b>第 3 章 增强现实系统简介</b> |                 | <b>108</b> |
| 3.1                   | 增强现实软件系统        | 108        |
| 3.1.1                 | 增强现实软件系统概述      | 108        |
| 3.1.2                 | 基于平面图像识别的增强现实系统 | 110        |

|                                    |            |
|------------------------------------|------------|
| 3.1.3 基于实物识别的增强现实系统 .....          | 113        |
| 3.1.4 基于 SLAM 的增强现实系统 .....        | 116        |
| 小结 .....                           | 117        |
| 3.2 增强现实硬件系统 .....                 | 117        |
| 3.2.1 基于高清摄像机的舞台增强现实系统 .....       | 117        |
| 3.2.2 基于普通智能手机的增强现实系统 .....        | 119        |
| 3.2.3 基于包含深度摄像模块的手持设备增强现实系统 .....  | 119        |
| 3.2.4 基于单目智能眼镜的增强现实系统 .....        | 122        |
| 3.2.5 基于双目可穿透式智能眼镜的增强现实系统 .....    | 124        |
| 3.2.6 基于投影的增强现实系统 .....            | 125        |
| 小结 .....                           | 127        |
| 3.3 国际主流增强现实引擎简介 .....             | 127        |
| <b>第 4 章 增强现实相关的人机交互系统简介 .....</b> | <b>129</b> |
| 4.1 手势识别交互系统 .....                 | 129        |
| 4.1.1 2D 静态手势识别 .....              | 129        |
| 4.1.2 2D 动态手势识别 .....              | 130        |
| 4.1.3 3D 手势识别 .....                | 130        |
| 4.2 语音识别交互系统 .....                 | 132        |
| 4.3 眼动追踪交互系统 .....                 | 134        |
| 小结 .....                           | 135        |
| <b>第 5 章 增强现实行业应用 .....</b>        | <b>136</b> |
| 5.1 增强现实在教育和技能培训领域的应用 .....        | 136        |
| 5.1.1 早教类增强现实应用 .....              | 137        |
| 5.1.2 学科教学类增强现实应用 .....            | 138        |
| 5.1.3 工业培训类增强现实应用 .....            | 140        |
| 5.2 增强现实在游戏娱乐领域的应用 .....           | 141        |

|                            |            |
|----------------------------|------------|
| 5.3 市场营销类增强现实应用·····       | 143        |
| 5.4 文化旅游类增强现实应用·····       | 146        |
| 5.5 工业和医疗类增强现实应用·····      | 147        |
| 5.6 军事类增强现实应用·····         | 149        |
| 小结·····                    | 150        |
| <b>第6章 增强现实系统发展趋势·····</b> | <b>151</b> |
| 6.1 增强现实软件发展趋势·····        | 151        |
| 6.1.1 深度学习与增强现实的结合·····    | 151        |
| 6.1.2 云计算技术与增强现实的结合·····   | 155        |
| 6.1.3 大范围三维重建技术·····       | 156        |
| 小结·····                    | 160        |
| 6.2 增强现实硬件发展趋势·····        | 160        |
| 6.2.1 基于光场成像技术的增强现实系统····· | 160        |
| 6.2.2 基于半导体光学的增强现实系统·····  | 169        |
| 小结·····                    | 172        |
| 6.3 增强现实人机交互技术发展趋势·····    | 172        |
| 6.3.1 新形态用户界面·····         | 172        |
| 6.3.2 新形态交互技术·····         | 181        |
| 小结·····                    | 182        |
| <b>第7章 结语·····</b>         | <b>183</b> |



# 计算机视觉与 增强现实应用概论

## 1.1 计算机视觉概述

通过视觉可以获取环境中物体和事件的信息，进而从物体发射或者反射出来的光中提取信息。人的视觉系统处理过程包括：眼球接收光信号的输入，信息的逐级提取，大脑皮层的处理，最后传递到大脑的不同功能区以做出反馈。

计算机视觉是计算机对生物的一种模拟，是教机器如何“看”的科学，也是人工智能的重要组成部分。它的任务就是通过对采集的图像视频流进行处理以获得相应场景的三维信息，就像人类在非睡眠时间所做的那样。在计算机视觉中，计算机用摄像机代替了人的眼睛，收集外部世界的图像；用算法代替了神经，对输入的图像视频流进行运算；用处理器（CPU 或者 GPU）代替了大脑，进行信息的存储和运算。

计算机视觉研究有两个终极目标：一是让计算机按照人类的方法去看待这个世界，代替人完成各种复杂的任务，甚至是人类根本无法完成的工作；二是实现图像理解系统，即利用输入的图像数据自动构造场景。当然这不是那么容易实现的，毕竟到目前为止人类对自己大脑的了解程度还是个未知数。人类视

觉系统是迄今为止人们所知道的功能最强大和完善的视觉系统。同时,生物的神经系统对计算机视觉也非常重要,很多计算机视觉的处理和算法都是模拟生物视觉系统设计的。比如,人类大脑最强大的功能之一便是慢慢为所学到的东西建立知识关联库,随着知识的扩充与完善,大脑的网络越来越复杂。计算机视觉中也模拟了这样的结构,发展出机器学习和神经网络,使算法本身具备学习能力,随着数据的样本量增多、代表性增强、涵盖面变广,运算结果能够更精确,也更符合人类的预期。

本章将先从计算机视觉的发展历史出发,介绍计算机视觉研究方向的演化以及现状。之后将介绍计算机视觉所涉及的不同领域的理论发展和典型应用,让读者对计算机视觉这门学科有更全面和充分的认识。

### 1.1.1 计算机视觉发展历程

计算机视觉领域的突出特点是其多样性和不完善性。

计算机视觉的历史要追溯到 20 世纪 50 年代,那时候的计算机视觉研究主要集中在二维图像的分析 and 识别上,比如工件表面、显微图片和航空图片的分析和解释。到了 60 年代,模式识别发展成为一门独立的学科,也就是当今最流行的深度学习和机器学习的基础。同时 MIT 的 Roberts 的研究工作开创了以理解三维场景为目的的三维计算机视觉的研究,他实现了从数字图像中提取立方体、棱柱体等多面体的三维结构,并对物体形状及物体的空间关系进行描述。Roberts 对积木世界创造性的研究给人们以极大的启发,许多人相信由简单的“立体积木”创造的三维世界可以推广到更复杂的三维场景。计算机视觉正是从这时逐渐开始步入蓬勃发展时期的。

到了 70 年代,当计算机的性能提高到足以处理图像这样的大规模数据时,一些视觉应用系统开始出现,可惜这些方面的应用都很有局限性,如脸孔、指纹、文字等,因而无法被广泛应用在更多场景中。在 70 年代另外一件标志性



事件是 MIT 人工智能实验室的 David Marr 教授提出了一套计算机视觉理论, 该理论立足于计算机科学, 系统地概括了心理生理学和神经心理学等方面取得的重要成果。在该理论中, Marr 认为可将视觉看作从三个层次对信息进行处理, 主要包括以下三个阶段。

第一阶段是对输入的原始图像进行处理, 抽取图像中的角点、边缘、纹理等特征, 这些基本的灰度变动以线条勾画出的草图形式出现, 称为基元图。

第二阶段是在以观测者为中心的坐标系中, 由输入图像和基元图恢复场景部分的深度和轮廓等, 这些信息包含了深度信息但又不是三维物体的全面表示, 称为二维半图。

第三阶段是在以物体为中心的坐标系中, 由输入图像、基元图和二维半图来对物体形状进行全面而清晰的描述。

有人认为 Marr 的理论是迄今为止最为完善的视觉理论, 同时也为计算机视觉领域创造了许多新的起点。进入 80 年代, 计算机视觉的研究又进入了新的阶段, 从实验室走向了实际应用的发展。90 年代, 局部特征描述子开始崛起, David Lowe 提出尺度不变特征变换 (Scale Invariant Feature Transform, SIFT), 计算机视觉研究的世界开始发生革命性的改变。SIFT 是一种对图像块进行比较和匹配的鲁棒解决方案。SIFT 对计算机视觉 (包括立体视觉、运动重构等) 几何方面的研究有着重要影响, 之后还成为用于对象识别的词袋模型的基础。

20 世纪末, 随着互联网科技的成熟, 基于更大数据集的计算机视觉研究成为研究的主流。随着加州理工学院的 Caltech-101 数据集得到普及, 分类的研究在不断增多。Caltech-101 也是更现代的 ImageNet 数据库的鼻祖。

进入 21 世纪后, 研究人员逐步开始深入研究目标识别问题, 涌现出词袋、空间金字塔、矢量量化等各种机器学习工具。原生 SIFT 算法对于一对一的匹配问题解决得很好, 但无法对不同的视觉对象类别进行分类。由 Josef Sivic 和 Andrew Zisserman 在 2003 年提出的视觉单词是一种从大规模的文本匹配中借

鉴的方法。视觉字典可以通过执行对 SIFT 的无监督学习，将 128 维的 SIFT 描述子映射为整数。使用这些视觉单词的直方图来表示图像是一种相当可靠的方法。至今词袋模型的变种仍然被大量使用在视觉研究中。

到了 2010 年之后，数据集越来越大，人工智能和深度学习开始飞速发展。深度学习其实是基于神经网络发展起来的，因为如今计算机的运算速度比 90 年代快了几个数量级，使很多之前难以实现的大型运算成为可能。可以将深度学习系统看作应用线性算子的多阶段过程，并通过非线性激活函数放到一个管道中进行处理。与以核函数为基础的学习系统相比，深度学习更类似于一种将多个线性支持向量机进行巧妙组合的系统。深度学习系统的构架需要手工设计，但不用为图像设计具体的特征描述子。

目前，计算机视觉技术被广泛地应用于计算几何、计算机图像学、图像处理、机器人学等多个领域中。这也让计算机视觉与其他许多知识领域产生了紧密关联，比如图像处理、模式识别、投影几何、统计学习和机器学习等。以下是计算机视觉与其他领域的关系图（图 1-1）。

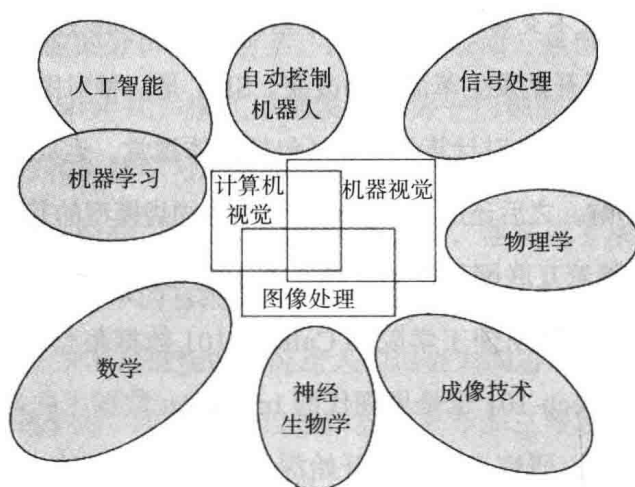


图 1-1 计算机视觉与其他领域的关系

接下来将分别介绍计算机视觉的常见模块及其在各个领域中的应用。

### 1.1.2 计算机视觉常见模块

众所周知,计算机视觉的快速发展是由于现实生活中有许多问题需要使用相关方案解决。典型的可以利用计算机视觉解决的问题有图像处理、图像分类、目标追踪和场景建模四个模块。每个模块对计算机视觉算法的设计思路和应用都不尽相同。一些更复杂的解决方案往往需要几个模块的结合,或用一种解决方案解决一系列相同的问题。

#### 图像处理

这部分包含了所有基于二维平面图像的计算机视觉处理算法,尤其针对像素级的操作,其感兴趣的往往是整张图片或局部区域像素的强度、梯度等,与图像的具体内容无关。

首先是平面图像的预处理,即在对图像实施具体的计算机视觉算法前,需要一些预处理来使图像满足后续要求。例如,二次采样保证图像的坐标正确;平滑去噪去除设备收集图像时引入的噪声;提高对比度来增强图像的反差效果;将彩色图像转成灰度或二值图像以满足后续处理的要求;进行边缘检测来提取图像中感兴趣的区域,比如在一张含有人脸的图片中把人脸区域直接提取出来;调整图像的尺度空间使图像结构适合局部运动等。

其次是对图像的特征提取,特征可以是线和面,也可以是图像的强度、梯度和纹理等,还可以是比较复杂的角点和局部特征描述点,具体的提取方法在下一章中详细介绍。这些提取的特征点可以用于进行图像或单一物体的识别等。这一步更大的利用价值是为更高级的处理提供图像的特征信息,在下面会介绍。

#### 图像分类

图像分类就属于高级的图像识别,不仅仅是图与图之间的匹配,而且涉及基于内容的图像识别。比如在庞大的图像库中寻找包含某指定内容的所有图