

TURING

图灵程序设计丛书

Programming Computer Vision
with Python

Python 计算机视觉

编程



[美] Jan Erik Solem 著

朱文涛 袁勇 译

O'REILLY®



人民邮电出版社
POSTS & TELECOM PRESS



图灵程序设计丛书

Python计算机视觉编程

Programming Computer Vision with Python

[美] Jan Erik Solem 著

朱文涛 袁 勇 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

c.授权人民邮电出版社出版

人 民 邮 电 出 版 社

北 京

图书在版编目 (CIP) 数据

Python计算机视觉编程 / (美) 索利姆
(Solem, J. E.) 著; 朱文涛, 袁勇译. — 北京: 人民邮
电出版社, 2014. 7

(图灵程序设计丛书)

ISBN 978-7-115-35232-3

I. ①P… II. ①索… ②朱… ③袁… III. ①软件工
具—程序设计 IV. ①TP311.56

中国版本图书馆CIP数据核字(2014)第067270号

内 容 提 要

本书是计算机视觉编程的权威实践指南, 依赖 Python 语言讲解了基础理论与算法, 并通过大量示例细致分析了对象识别、基于内容的图像搜索、光学字符识别、光流法、跟踪、三维重建、立体成像、增强现实、姿态估计、全景创建、图像分割、降噪、图像分组等技术。另外, 书中附带的练习还能让读者巩固并学会应用编程知识。

本书适合的读者是: 有一定编程与数学基础, 想要了解计算机视觉的基本理论与算法的学生, 以及计算机科学、信号处理、物理学、应用数学和统计学、神经生理学、认知科学等领域的研究人员和从业者。

-
- ◆ 著 [美] Jan Erik Solem
 - 译 朱文涛 袁 勇
 - 责任编辑 李松峰 毛倩倩
 - 执行编辑 杨 琳
 - 责任印制 焦志炜

- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
- 邮编 100164 电子邮件 315@ptpress.com.cn
- 网址 <http://www.ptpress.com.cn>
- 三河市海波印务有限公司印刷
- ◆ 开本: 800×1000 1/16
- 印张: 17.75
- 字数: 338千字
- 印数: 1—4 000册

2014年7月第1版

2014年7月河北第1次印刷

著作权合同登记号 图字: 01-2014-1134号

定价: 69.00元

读者服务热线: (010)51095186转600 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京崇工商广字第 0021 号



版权声明

© 2012 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2014. Authorized translation of the English edition, 2014 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly Media, Inc. 出版 2012。

简体中文版由人民邮电出版社出版，2014。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未得书面许可，本书的任何部分和全部不得以任何形式重制。



O'Reilly Media, Inc.介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版、在线服务还是面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar 博客有口皆碑。”

——Wired

“O'Reilly 凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference 是聚集关键思想领袖的绝对典范。”

——CRN

“一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim 是位特立独行的商人，他不光放眼于最长远、最广阔的视野，并且切实地按照 Yogi Berra 的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去 Tim 似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

推荐序

计算机视觉是一门实践性很强的课程，虽然已经有了不少教科书，但大多数都只重视其中的理论和算法，很少有实践指导书。因而对于学习者而言，如果希望在实践中学习，往往需要编写大量的程序。在这方面，本书的出版可以算是一个有益的补充，相信本书将成为计算机视觉学习者的一个重要参考。

作为一本面向计算机视觉编程的书，本书涉及了这一学科中相对成熟并且被以往实践验证有效的部分典型算法，因而具有很好的实用性。例如第 2 章描述子部分选择了 Harris 角点检测器和 SIFT 描述子及其实现加以介绍；第 3 章则以全景图的创建为例，给出了 RANSAC 的实现；第 9 章图像分割中讨论了 Graph Cut 的实现等。这些方法大多数具有很好的通用性，因而为读者提供了一种实现范例。

本书的另一个特点是对介绍的单一方法，通过综合运用提升学习者灵活应用这些方法的能力。例如第 4 章给出的增强现实的例子，以及第 8 章给出的图像校正的例子。这些例子能够帮助进一步提升学习者对前述方法的感性认识。

与早期计算机视觉领域多数程序都是由 C/C++ 写就的情形不同。随着计算机硬件速度越来越快，研究者在考虑选择实现算法语言的时候会更多地考虑编写代码的效率和易用性，而不是像早年那样把算法的执行效率放在首位。这直接导致近年来越来越多的研究者选择 Python 来实现算法。与此同时，Python 的开放性使不同领域的研究者能够有机会在 Python 中加入他们需要的特性，甚至可以纳入 Python 的标准库，这也大大吸引了众多研究者对 Python 的参与。

本书的第三个特点是提供了与 OpenCV 接口的介绍。这为利用 OpenCV 中的资源提供了方便的途径。

今天在计算机视觉领域，越来越多的研究者使用 Python 开展研究。本书中文版的出版一方面能够鼓励更多的研究者采用这一语言，另一方面则为 Python 的学习者提供了一种尝试不同领域算法的机会。

A handwritten signature in black ink, appearing to read '陈乾' (Chen Qian), written in a cursive style.

2014 年 4 月 21 日

前言

今天，图像和视频无处不在，在线照片分享网站和社交网络上的图像有数十亿之多。几乎对于任意可能的查询图像，搜索引擎都会给用户返回检索的图像。实际上，几乎所有手机和计算机都有内置的摄像头，所以在人们的设备中，有几 G 的图像和视频是一件很寻常的事。

计算机视觉就是用计算机编程，并设计算法来理解在这些图像中有什么。计算机视觉的有力应用有图像搜索、机器人导航、医学图像分析、照片管理等。

本书旨在为计算机视觉实战提供一个简单的切入点，让学生、研究者和爱好者充分理解其基础理论和算法。本书中的编程语言是 Python，Python 自带了很多可以免费获取的强大而便捷的图像处理、数学计算和数据挖掘模块，可以免费获取。

写作本书的时候，我遵循了以下原则。

- 鼓励探究式学习，让读者在阅读本书的时候，在计算机上跟着书中示例进行练习。
- 推广和使用免费且开源的软件，设立较低的学习门槛。显然，我们选择了 Python。
- 保持内容完整性和独立性。本书没有介绍计算机视觉的全部内容，而是完整呈现并解释所有代码。你应该能够重现这些示例，并可以直接在它们之上构建其他应用。
- 内容追求广泛而非详细，且相对于理论更注重鼓舞和激励。

总之，如果你对计算机视觉编程感兴趣，希望它能给你带来启发。

先决条件和概述

本书主要针对各种应用和问题探讨理论及算法，下面简单概括一下。

读者须知

- 基本的编程经验。你需要会使用编辑器，能够运行脚本，知道如何构建代码以及基本数据类型。熟悉 Python，或诸如 Ruby、Matlab 等其他脚本语言，这也会对你理解本书有所帮助。
- 数学基础。如果你知道矩阵、向量、矩阵乘法、标准数学函数以及导数和梯度等概念，这对于充分利用其中示例非常有益。对于一些较高级的数学例子，你可以轻松跳过。

本书内容

- 用 Python 对图像进行实战编程。
- 现实世界中各种应用背后的计算机视觉技术。
- 一些基本算法，以及如何实现并应用这些算法。

本书中的代码示例会向你展示物体识别、基于内容的图像检索、图像搜索、光学字符识别、光流、跟踪、三维重建、立体成像、增强现实、姿态估计、全景创建、图像分割、降噪、图像分组等内容。

各章概览

第 1 章“基本的图像操作和处理”介绍用来处理图像的基本工具及本书用到的核心 Python 模块，同时涵盖了很多贯穿全书的基础示例。

第 2 章“局部图像描述子”讲解检测图像兴趣点的方法，以及怎样使用它们在图像间寻找相应点和区域。

第 3 章“图像到图像的映射”描述图像间基本的变换及其计算方法。涵盖从图像扭曲到创建全景图像的示例。

第 4 章“照相机模型与增强现实”介绍如何对照相机建模、生成从三维空间到图像特征的图像投影，并估计照相机视点。

第 5 章“多视图几何”讲解如何对具有相同场景、多视图几何基本面的图像进行处理，以及怎样从图像计算三维重建。

第 6 章“图像聚类”介绍一些聚类方法，并展示如何基于相似性或内容对图像进行分组和组织。

第 7 章“图像搜索”展示如何建立有效的图像检索技术，以便能够存储图像并表示，并基于图像的视觉内容搜索图像。

第 8 章“图像内容分类”描述了图像内容分类算法，以及怎样使用它们识别图像中的物体。

第 9 章“图像分割”介绍了通过聚类、用户交互或图像模型，将图像分割成有意义区域的不同技术。

第 10 章“OpenCV”展示怎样使用常用的 OpenCV 计算机视觉库 Python 接口，以及如何处理视频及摄像头的输入。

本书结尾有参考文献。文献条目的引用用方括号表示，如 [20]。

计算机视觉简介

计算机视觉是一门对图像中信息进行自动提取的学科。信息的内容相当广泛，包括三维模型、照相机位置、目标检测与识别，以及图像内容的分组与搜索等。本书中，我们使用广义的计算机视觉概念，包括图像扭曲、降噪和增强现实等^①。

计算机视觉有时试图模拟人类视觉，有时使用数据和统计方法，而有时几何是解决问题的关键。在本书中，我们试图对此进行全面介绍。

实用计算机视觉混合了编程、建模和数学技巧，有时很难掌握。我本着“力求简单，又不影响理解”的精神，有意用最少的理论来展示这些内容。书中对于数学知识的介绍是为了帮助读者理解算法，有些章（主要是第 4 章和第 5 章）无法避免地涉及很多数学理论。只要读者愿意，可以跳过这些数学理论，直接使用示例代码。

Python和NumPy

Python 是一门编程语言，其使用贯穿了全书的示例代码。Python 是一种简洁明了的语言，对于输入/输出、数字、图像及绘图都具有良好的支持。这门语言的一些特性需要我们逐渐适应，比如缩进和紧凑语法。要运行代码示例，你需要安装 Python 2.6 或之后的版本，因为只有这些版本才提供本书中用到的很多工具包。Python 3.x 版本与 2.x 版本有很多语法差异，并且不兼容 2.x 版本，也不兼容我们所需的工具包。

熟悉一些基本 Python 操作会更容易理解这些内容。对于 Python 初学者，我建议读一下 Mark Lutz 的书 *Learning Python*[20] 和 <http://www.python.org/> 上的在线文档。

在进行计算机视觉编程的时候，我们需要在向量、矩阵的表示上进行操作，这可

注 1：这些例子生成新的图像，并且比实际地从图像中提取信息需要更多的图像处理。

以通过 Python 的 NumPy 模块处理；在该模块中，向量和矩阵是用 `array` 类型表示的。对于图像，我们也将采用这种类型的表示。Travis Oliphant 的免费电子书 *Guide to NumPy*[24] 是一本不错的 NumPy 参考手册；<http://numpy.scipy.org/> 上的文档对于刚接触 NumPy 的读者来说是一个很好的起点。对于结果的可视化，我们会用到 Matplotlib 模块；而对于更高级的数学，我们会用到 SciPy 模块。这些就是你会用到的核心模块，详见第 1 章。

除了这些核心模块，对于某些特殊目的，比如读取 JSON 或 XML、载入并保存数据、生成图、图形编程、Web 演示、分类器等，我们还会用到很多其他免费模块。这些模块只有在特殊的应用和演示中才需要，如果你对于某种应用不感兴趣，可以跳过。

这里有必要提一下 IPython，它是一个交互式 Python 壳，使调试和实验变得更简单。对应文档及下载地址见 <http://ipython.org/>。

排版约定

代码如下：

```
# 一些点
x = [100,100,400,400]
y = [200,500,200,500]

# 绘制这些点
plot(x,y)
```

本书中的字体约定如下。

- 楷体
用于定义。
- 等宽字体 (Constant width)
用于函数、Python 模块及代码示例，也用于控制台打印输出。

数学公式为内联式（如 $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$ ），或者单独居中：

$$f(\mathbf{x}) = \sum_i \mathbf{w}_i x_i + b$$

只有需要参考的时候我们才对公式进行编号。

在介绍数学知识的部分，标量使用小写字母 ($s, r, \lambda, \theta \dots$)，矩阵（包括图像数组 $\mathbf{I}$ ）使用大写加粗 ($\mathbf{A}, \mathbf{V}, \mathbf{H}, \dots$)，向量则小写加粗 ($\mathbf{t}, \mathbf{c}, \dots$)，二维（图像）和三维中的点分别用 $\mathbf{x}=[x, y]$ 和 $\mathbf{X}=[X, Y, Z]$ 表示。

使用代码示例

本书旨在帮助你完成工作。通常，你可以在程序或文档中使用本书的代码。你不必联系我们请求许可，除非你要复制本书的大量代码。例如，用本书的几段代码编写程序不需要获得许可；售卖或再分发 O'Reilly 的图书示例光盘需要获得许可；引用本书和示例代码回答问题不需要获得许可；将本书中的大量示例代码纳入产品文档中需要获得许可。

我们对你在使用时声明引用信息表示感谢，但不强制要求。引用信息通常包括标题、作者、出版商和 ISBN，例如“*Programming Computer Vision with Python* by Jan Erik Solem (O'Reilly). Copyright © 2012 Jan Erik Solem, 978-1-449-31654-9.”

如果你觉得使用的代码示例超出合理引用或上述许可范围，请随时和我们联系：permissions@oreilly.com。

Safari® Books Online



Safari Books Online (www.safaribooksonline.com) 是应需而变的数字图书馆。它同时以图书和视频的形式出版世界顶级技术和商务作家的专业作品。

Safari Books Online 是技术专家、软件开发人员、Web 设计师、商务人士和创意人士开展调研、解决问题、学习和认证培训的第一手资料。

对于组织团体、政府机构和个人，Safari Books Online 提供各种产品组合和灵活的定价策略。用户可通过一个功能完备的数据库检索系统访问 O'Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones & Bartlett、Course Technology 以及其他几十家出版社的上千种图书、培训视频和正式出版之前的书稿。要了解 Safari Books Online 的更多信息，我们网上见。

联系我们

请把对本书的评价和问题发给出版社。

美国：

O'Reilly Media, Inc.

1005 Gravenstein Highway North
Sebastopol, CA 95472

中国:

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室 (100035)
奥莱利技术咨询(北京)有限公司

O'Reilly 的每一本书都有专属网页, 你可以在那儿找到本书的相关信息, 包括勘误表、示例代码以及其他信息。本书的网站地址是:

<http://shop.oreilly.com/product/0636920022923.do>

对于本书的评论和技术性问题, 请发送电子邮件到:

bookquestions@oreilly.com

要了解更多 O'Reilly 图书、培训课程、会议和新闻的信息, 请访问以下网站:

<http://www.oreilly.com>

我们在 Facebook 的地址如下:

<http://facebook.com/oreilly>

请关注我们的 Twitter 动态:

<http://twitter.com/oreillymedia>

我们的 YouTube 视频地址如下:

<http://www.youtube.com/oreillymedia>

致谢

感谢参与本书写作及出版的每一个人, 感谢整个 O'Reilly 团队给予的帮助。特别要感谢 Andy Oram (O'Reilly) 的编辑工作, 以及 Paul Anagnostopoulos (Windfall Software) 的高效出版工作。

很多人对于我分享在网上的书稿发表了评论。感谢 Klas Josephson 和 Håkan Ardö 就本书细致给出意见及反馈, 而 Fredrik Kahl 和 Pau Gargallo 帮助进行了核实。感谢所有读者的鼓励, 以及对本书文本和代码示例的改进。很多陌生人通过邮件分享了对书稿的看法, 这对我而言是一个巨大的动力。

最后, 感谢朋友和家人对我夜以继日写作的理解和支持。特别要感谢长久以来一直支持着我的妻子 Sara。

目录

推荐序	XI
前言	XIII
第 1 章 基本的图像操作和处理	1
1.1 PIL: Python 图像处理类库	1
1.1.1 转换图像格式	2
1.1.2 创建缩略图	3
1.1.3 复制和粘贴图像区域	3
1.1.4 调整尺寸和旋转	3
1.2 Matplotlib	4
1.2.1 绘制图像、点和线	4
1.2.2 图像轮廓和直方图	6
1.2.3 交互式标注	7
1.3 NumPy	8
1.3.1 图像数组表示	8
1.3.2 灰度变换	9
1.3.3 图像缩放	11
1.3.4 直方图均衡化	11
1.3.5 图像平均	13
1.3.6 图像的主成分分析 (PCA)	14
1.3.7 使用 pickle 模块	16
1.4 SciPy	17
1.4.1 图像模糊	18
1.4.2 图像导数	19
1.4.3 形态学: 对象计数	22

1.4.4 一些有用的 SciPy 模块	23
1.5 高级示例：图像去噪	24
练习	28
代码示例约定	29
第 2 章 局部图像描述子	31
2.1 Harris 角点检测器	31
2.2 SIFT（尺度不变特征变换）	39
2.2.1 兴趣点	39
2.2.2 描述子	39
2.2.3 检测兴趣点	40
2.2.4 匹配描述子	43
2.3 匹配地理标记图像	47
2.3.1 从 Panoramio 下载地理标记图像	47
2.3.2 使用局部描述子匹配	50
2.3.3 可视化连接的图像	52
练习	54
第 3 章 图像到图像的映射	57
3.1 单应性变换	57
3.1.1 直接线性变换算法	59
3.1.2 仿射变换	60
3.2 图像扭曲	61
3.2.1 图像中的图像	63
3.2.2 分段仿射扭曲	67
3.2.3 图像配准	70
3.3 创建全景图	76
3.3.1 RANSAC	77
3.3.2 稳健的单应性矩阵估计	78
3.3.3 拼接图像	81
练习	84
第 4 章 照相机模型与增强现实	85
4.1 针孔照相机模型	85
4.1.1 照相机矩阵	86
4.1.2 三维点的投影	87
4.1.3 照相机矩阵的分解	89
4.1.4 计算照相机中心	90
4.2 照相机标定	91

4.3	以平面和标记物进行姿态估计	93
4.4	增强现实	97
4.4.1	PyGame 和 PyOpenGL	97
4.4.2	从照相机矩阵到 OpenGL 格式	98
4.4.3	在图像中放置虚拟物体	100
4.4.4	综合集成	102
4.4.5	载入模型	104
	练习	106
第 5 章	多视图几何	107
5.1	外极几何	107
5.1.1	一个简单的数据集	109
5.1.2	用 Matplotlib 绘制三维数据	111
5.1.3	计算 F : 八点法	112
5.1.4	外极点和外极线	113
5.2	照相机和三维结构的计算	116
5.2.1	三角剖分	116
5.2.2	由三维点计算照相机矩阵	118
5.2.3	由基础矩阵计算照相机矩阵	120
5.3	多视图重建	122
5.3.1	稳健估计基础矩阵	123
5.3.2	三维重建示例	125
5.3.3	多视图的扩展示例	129
5.4	立体图像	130
	练习	135
第 6 章	图像聚类	137
6.1	K-means 聚类	137
6.1.1	SciPy 聚类包	138
6.1.2	图像聚类	139
6.1.3	在主成分上可视化图像	140
6.1.4	像素聚类	142
6.2	层次聚类	144
6.3	谱聚类	152
	练习	157
第 7 章	图像搜索	159
7.1	基于内容的图像检索	159
7.2	视觉单词	160

7.3 图像索引	164
7.3.1 建立数据库	164
7.3.2 添加图像	165
7.4 在数据库中搜索图像	167
7.4.1 利用索引获取候选图像	168
7.4.2 用一幅图像进行查询	169
7.4.3 确定对比基准并绘制结果	171
7.5 使用几何特性对结果排序	172
7.6 建立演示程序及 Web 应用	176
7.6.1 用 CherryPy 创建 Web 应用	176
7.6.2 图像搜索演示程序	176
练习	179
第 8 章 图像内容分类	181
8.1 K 邻近分类法 (KNN)	181
8.1.1 一个简单的二维示例	182
8.1.2 用稠密 SIFT 作为图像特征	185
8.1.3 图像分类: 手势识别	187
8.2 贝叶斯分类器	190
8.3 支持向量机	195
8.3.1 使用 LibSVM	196
8.3.2 再论手势识别	198
8.4 光学字符识别	199
8.4.1 训练分类器	200
8.4.2 选取特征	200
8.4.3 多类支持向量机	201
8.4.4 提取单元格并识别字符	202
8.4.5 图像校正	205
练习	206
第 9 章 图像分割	209
9.1 图割 (Graph Cut)	209
9.1.1 从图像创建图	211
9.1.2 用户交互式分割	216
9.2 利用聚类进行分割	218
9.3 变分法	224
练习	226
第 10 章 OpenCV	227
10.1 OpenCV 的 Python 接口	227