

Computer Vision Metrics
Survey, Taxonomy, and Analysis

异步图书
www.epubit.com.cn

Apress®

计算机视觉度量 深入解析

[美] Scott Krig 著
刘波 靳小波 于俊伟 译

- 来自计算机视觉专家的声音
- 英特尔推荐阅读书目



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

Computer Vision Metrics
Survey, Taxonomy, and Analysis

TP3
1126

计算机视觉度量 深入解析

[美] Scott Krig 著
刘波 靳小波 于俊伟 译

人民邮电出版社
北京

图书在版编目 (C I P) 数据

计算机视觉度量深入解析 / (美) 斯科特·克里格
(Scott Krig) 著 ; 刘波, 靳小波, 于俊伟译. — 北京 :
人民邮电出版社, 2016. 10
ISBN 978-7-115-43304-6

I. ①计… II. ①斯… ②刘… ③靳… ④于… III.
①计算机视觉—研究 IV. ①TP302.7

中国版本图书馆CIP数据核字(2016)第200096号

版 权 声 明

Computer Vision Metrics: Survey, Taxonomy, and Analysis

By Scott Krig, ISBN: 978-1-4302-5929-9

Original English language edition published by Apress Media.

Copyright © 2014 by Apress Media

Simplified Chinese-language edition copyright © 2016 by Post & Telecom Press.

All rights reserved.

本书中文简体字版由 **Apress Media** 授权人民邮电出版社独家出版。未经出版者书面许可, 不得以任何方式复制本书的内容。

版权所有, 侵权必究。

-
- ◆ 著 [美]Scott Krig
 - 译 刘 波 靳小波 于俊伟
 - 责任编辑 王峰松
 - 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>
 - 固安县铭成印刷有限公司印刷
 - ◆ 开本: 700×1000 1/16
 - 印张: 27
 - 字数: 535 千字 2016 年 10 月第 1 版
 - 印数: 1—2 500 册 2016 年 10 月河北第 1 次印刷

著作权合同登记号 图字: 01-2015-2076 号

定价: 89.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315

内容提要

计算机视觉作为人工智能的一个重要分支，目前已经广泛应用于智能驾驶、机器人、生物识别等众多领域。本书对计算机视觉特征描述子的性质进行了分类，并给出了计算机视觉处理流程的一般性框架。同时，本书也对目前较新的计算机视觉技术进行了介绍，这些技术包括 3D 深度感知方法、稀疏编码、卷积神经网络、深度学习等。

本书针对的读者为从事计算机视觉的工程技术人员、研究人员等。读者可根据不同的应用，利用本书提供的知识来选择合适的特征描述子。本书将按各种鲁棒性属性来理解各类计算机视觉的特征描述子，读者在阅读本书时最好具备一定的图像处理的基础知识。

作者简介



Scott Krig 是计算机成像学、计算机视觉和图形可视化方面的先驱。他在 1988 年成立了 Krig Research 公司 (krigresearch.com)，该公司提供了世界上第一个基于高性能工程工作站、超级计算机和专有成像硬件的成像和视觉系统，并为来自全球 25 个国家的客户提供服务。Scott 为全球客户提供成像和视觉方面的解决方案，并且和多个行业，包括航空、军事、情报部门、执法机关、政府研究部门、学术组织等密切合作过。

近年来，Scott 主要为大型公司和服务于商业市场的初创公司提供服务，帮助它们解决计算机视觉、图形成像、图像学、可视化、机器人、过程控制、工业自动化、计算机安全、密码学，以及成像学和机器视觉在电子消费品（如 PC 机、笔记本电脑、手机和平板电脑）方面的应用问题。最近，Scott 在将深度感知和计算机视觉方法用于嵌入式系统和移动平台的相关领域为英特尔公司提供技术性指导。

Scott 也是全球范围的许多专利应用的发明人，其涉及的范围包括嵌入式系统、成像学、计算机视觉、DRM 和计算机安全，他也曾在斯坦福大学做过研究。

同时，Scott 也很喜欢木吉他设计和弦乐器制作，特别是 12 弦木吉他，还喜欢为木吉他作曲并演奏。

译者简介

1. 刘波，博士，重庆工商大学计算机科学与信息工程学院教师，主要从事机器学习理论、计算机视觉和最优化技术研究，同时爱好 Hadoop 和 Spark 平台上的大数据分析，也对 Linux 平台的编程和 Oracle 数据库感兴趣。

2. 靳小波，博士，副教授，硕士生导师，2009 年 7 月从中国科学院自动化研究所模式识别国家重点实验室博士毕业。2010 年 5 月入河南工业大学信息科学与工程学院参加工作至今。近年来，在国际顶级杂志和顶级会议上发表论文多篇（其中，Pattern Recognition 一篇，Neurocomputing 一篇，ICPR 会议 3 篇），申请专利一项。多次参与互联网作弊检测挑战赛获第一名，曾主持青年科学基金一项（61103138），并参与两项青年科学基金。开发了一款开源的 Java 机器学习库 JMLP。主要研究兴趣为机器学习、互联网挖掘和计算机视觉。

3. 于俊伟，博士，副教授，2009 年 12 月从中国科学院自动化所毕业，随后在河南工业大学信息科学与工程学院工作至今，在国内外学术期刊及国际会议上发表论文 10 余篇，当前主持一项青年科学基金（61300123），主要从事计算机视觉、模式识别和智能信息处理等方向的研究。

致谢

若没有与 ITSEEZ 的 Vadim Pizarevsky 进行早期技术反馈、交流和对粗糙原始数据的观察，本书内容就不会有这样全面。Vadim 也是成立 OpenCV 的主要力量，它具有多个计算机视觉应用领域的广阔、丰富的专业知识。感谢 Vadim。

还要特别感谢 Intel 出版社的 Stuart Douglas 约我写作本书，是他将我引荐给 Apress 出版社的编辑。此外，特别感谢与本书出版相关的主要编辑，包括 Melissa Maldonado、Mark Powers、Jeffrey Pepper、Steve Weiss、Robert Hutchinson、James Markham 和 Carole Berglie，他们的编辑处理为本书增色不少。

感谢我的妻子 Janie，也感谢我的家庭和父母，他们是我生命的重要部分。

前言

本书从特征描述的度量出发，或者说从如何描述、计算和设计宏特征（macro-features）和微特征（micro-features）（它们用于构造图像中的更大对象）的角度来着眼于一个细分领域，即计算机视觉度量。它关注的是视觉流程中像素层面的内容，而不是后端的训练、分类、机器学习和匹配等阶段的内容。本书可以用作计算机视觉方面的参考书、高级课程或者自学教材，它针对那些已经了解计算机视觉和图像处理的读者。但这个领域的新手，也可通过本书丰富的插图、汇总表，从较高层面来获取一些重要概念。

我将计算机视觉看作是一个数学的艺术形式，而它的研究者和实践者们就是艺术家。因此，本书更像是一个艺术画廊，而不是一本技术或科学的专著。它提供了一些观测现象；给出了令人感兴趣的问题；并提出了一套视觉分类方法，通过该方式勾画了该领域的一幅蓝图。本书试图绘制一幅以特征度量为中心的地图，这幅图可能不那么准确、清晰，但希望能抛砖引玉，启发大家以自己的方式丰富其细节，以便比我一个人甚至几个人完成得更好。我要是发现市面上有某本类似的书涵盖了这个主题的这一特定细分领域，我就不会承担写作这本书的任务了。

本书不包含什么

读者在本书中不会找到如何实现计算机视觉的例子和相应的源代码、讨论指南、性能分析以及一些捷径，因为这些可以通过阅读广受好评的 OpenCV 库资源（<http://opencv.org>）得到，它含有大量的优秀图书、在线资源、源代码示例和几个博客。对着手实际应用的开发者而言，没有比 OpenCV 更合适的了。因此，本书会避开与 OpenCV 社区和其他地方重复的“如何做”的材料，转而提供与之对应的讨论，包括综述、方法分类学和分析。另外，本书不会试图对包含计算机视觉的所有主题进行讨论和性能分析，因为其他资料非常全面地提供了这方面的材料，例如，由于本书关注的是特征度量，因而机器学习、训练和分类方法在这里只是简单地介绍一下。

总而言之，本书主要讨论特征度量，展示开发者使用的是“什么”方法，同时

详细观察和分析这些方法“为什么”起作用，侧重通过观察提出问题，而不是给出过多的答案。我非常喜欢这些问题，因为好的问题能激起好的回答，每个回答常常孕育更多好的问题。

本书的目的是在综述层面进行分类和分析，所以没有单个用例的详细例子，也没有涉及方法之间的比较。若想获取更多细节，可参考本书后面的 540 多个参考文献。另外，附录 C 提供了一些“如何做”和“动手实践”的资源。少量与本书有关的简单源代码可以在线获取。附录 A 包含了兴趣点检测子的评估，在第 7 章会引入这些合成的兴趣点字母。附录 D 包含扩展的 SDM 度量（在第 3 章会介绍）。

本书包含什么

第 1 章是开场白，它描述了二维图像的构造和三维深度成像。第 2 章进一步介绍智能图像预处理技术，这些技术会被用来增强特征的描述。第 3 章到第 6 章构成了特征描述的核心部分，特别强调了局部特征。第 3 章包含了全局和区域度量，第 4 章介绍了特征描述子的概念，第 5 章介绍了视觉分类，第 6 章包含了局部特征描述子。第 7 章引入了基准数据，第 8 章从工程的角度讨论了假设视觉流程和假设优化过程，并给出了一系列的练习题目，由此把各种视觉概念综合在一起，与实际的系统发生关联（第 8 章的练习题目用于实现和改进这些假设的例子）。第 7 章研究一系列合成的兴趣点字母，并用 10 个常用的描述子与这些字母做对比，其结果可在附录 A 中找到。在图像处理和计算机视觉中，很难对所有的主题做一个清晰的划分，因此各章之间存在重叠。在实际中也会这样混合使用，因此，在第 5 章的视觉分类会出现重叠，并且创新总是在以出人意料的新方式使用旧方法的过程中形成的。但分类学是一个起点，有助于阐述清楚本书的结构。

因此，本书的主要目标是研究和理解特征描述方法的使用范围，而不是判定方法的优劣。本书还有一个目的，就是通过展示一些发展历程来介绍为什么要研究这种方法，探讨它的不变性和性能分别是什么，但不会对所介绍的内容进行评价，这会留待其他研究者来做。因为每种方法如何实现决定了其性能和精度，而每种方法使用基准数据测试的是什么则表明了其他内容。如果我们能够从其他人的研究工作中获得好的想法，这就是他们研究工作很成功的一个标志。

范围

为确保全书的简洁，我对一些与计算机视觉本身关系不大的主题未做深入探讨；这种方式有点与众不同，因为讨论计算机视觉一般会涉及广泛的议题。具体来说，这里讨论的主题不包含统计机器学习、分类和训练、特征数据库的构造和优化，以

及搜索和排序。本书讨论了距离函数，因为它与特征度量之间有直接关系（本书将来的某个版本可能会涵盖与计算机视觉相关的另一个领域——统计机器学习，但不是现在）。

术语说明

有时当描述类似的概念时，文献上的术语并不一致。因此本书采用了一些各独立研究机构未做标准化的术语。实际上，这里引入了一些新的非标准术语，也许是因为本书作者没有意识到有更好的术语存在（也许这里引入的一些术语在将来会变成标准术语）。术语的不一致在与数学（比如聚类、回归、分组距离、误差最小化）和计算机视觉（关键点、兴趣点、锚点等）相关的主题方面表现得最为明显。由于人们学到的太多概念都是基于术语，因而作者也认识到改变这些术语可能会有些勉强。我想起一位叫 Homer Mead 的朋友，他在波音公司任月球车和预警机雷达方面的首席工程师，他就下意识地用旧词 *condenser* 而不用新词 *capacitor*（两者都表示电容器）。

作者的灵感有几个来源，主要是有开疆拓土的机遇。任何新领域的疆界在扩展时，总会遇到某种程度上的界限不明、缺少结构和缺乏组织的问题，所以在这一广阔的领域，探索的机遇让人无法抗拒：确定这门知识的结构和路径，让他人沿着这条路找寻新的研究领域，设置更明确的路标并将知识之路扩展得更远。

本书的灵感也来自于多年来很多研究人员之间的对话。具体从哪里开始的呢？它开始于 20 世纪 80 年代初的波音公司，当时我还在上大学。我在高级发展研究实验室（Advanced Development Research）工作时接触到了计算机图形学，其中航天飞机的第一个计算机三维渲染是以光栅的形式出现的。在当时，主要是使用矢量图机器，如 Evans & Sutherland 图形系统，最后，实验室添加了 BARCO 帧缓冲设备，Jeff Lane 和他的研究组以及 Loren Carpenter，从图形学模型发展了阴影图的高级光栅计算机渲染。实验室发明了几个方法，包括分形学（Fractals）、NURBS 和 A-buffer 技术，刚开始计算机图形学涉及的数学知识，如双三次样条（bi-cubic patches）和双五次样条（bi-quintic patches）让我望而却步。但是随后我被 BARCO 帧缓冲里面的单像素深深地吸引，因为它们看起来非常直观和明显，一次只处理一个像素、一行或一帧。最初我研究成像学和计算机视觉，而不是所有的计算机图形学以及与之有关的数学。然而，事实证明，计算机视觉和图像处理的数学更加变化多样，无论怎样，至少和前者（计算机图像学及其相关的数学）一样复杂。从那以后，我在计算机图像学上也花费了相当多的时间。到 20 世纪 80 年代中期，我的老板 Don Snow 先是与人合作成立了太平洋西部系统（Pacific Western Systems）公司，并成为负责研究工作的副总裁，后来去了应用精度（Applied Precision）公司，他让我分析用于模式识别的 View-PRB 固定功能硬件单元，主要将它用在自动晶圆（wafer）探测中

(以防我们需要自己构造一个类似它的东西) 以便定位晶片上的模式和调校探测仪。它使用相关性进行模式匹配并通过一个我们称之为“超像素”的尺度步长(scale-space)的方法进行搜索。在 NTSC 上其亚像素精度匹配的速度是每秒 4 个 32×32 的分块, 我计算位置、旋转和偏移量以便调整晶圆探测阶段, 为晶片探测做准备。这被称为自动调整。我设计了一个模式识别伺服系统, 它可以以几个微弧度的旋转精度和几分之一微米的位置精度对模式进行定位。在 20 世纪 80 年代末期, 我转到 Mentor Graphics 工作, 几年后, 我向总裁 Gerry Langelier 提出辞职, 离开了公司的 R&D 组, 然后成立了一家名为 Krig Research 的研究公司, 主要关注基于现在已经很少见的工作站(如 SGI、Apollo、Sun, 现在所有的都没有了), 并且是面向高端的军方和以研究为主的客户提供计算机视觉和成像学方面的服务, 一直到现在, 我对它们仍然有浓厚的兴趣。现在的工业相比过去发生了很多变化, 软件看起来几乎是免费的, 硬件或 SOC 也几乎是免费的, 所以我不清楚现在别人是怎么赚钱的。

就在最近, 好多同人给了我一些灵感。感谢 Paul Rosin 提供的合成图像和一些架构上的想法。感谢 Yann LeCun 提供了关于深度学习和卷积网络方面的重要参考文献。感谢 Shree Nayar 准许我在书中使用了他的几幅插图, 并且他仍在通过 Cave 研究项目为计算机视觉社区提供更多的灵感。感谢 Luciano Oviedo 帮我完善了关于工业界的行为和策略以及未来发展趋势方面的大量内容, 并且我还和他进行了充满活力的讨论。

还有很多需要感谢的人, 他们为我提供了帮助, 无法在此一一列出。尽管我和他们的交谈有时可能非常简短, 有时甚至只是通过虚拟的电子邮件或 SKYPE 网络电话, 他们的研究工作和想法的影响力仍然存在。对下列同人我要特别感谢, 感谢他们给本书的草稿或大纲提出的有意义的评论、插图, 或是可能他们自己都没有意识到的灵感。感谢 Rahul Suthankar 以及 Alexandre Alahi 和我讨论了图像的使用; 感谢 Steve Seitz、Bryan Russel、Liefeng Bo 以及 Xiaofeng Ren 同我就 RGB-D 计算机视觉和其他研究主题所做的深入讨论; 感谢 Gutemberg Guerra-filho、Harsha Viswana、Dale Hitt、Joshua Gleason、Noah Snavely、Daniel Scharstein、Thomas Salmon、Richard Baraniuk、Carl Vondrick、Hervé Jégou 和 Andrew Richardson; 也感谢英特尔公司的几个同人, 包括 Ofri Weschler、Hong Jiang、Andy Kuzma、Michael Jeronimo、Eli Turiel, 还有许多其他我没有提到的人员, 我和他们就计算机视觉的主题做了很多令人感兴趣的讨论。

总结

总的来说, 我的目标是考察研究人员用于特征描述的一些方法, 即生成的重要度量, 并且使得人们在实际中理解这些方法相对容易一些, 同时考察如何使用计算机视觉分类学和鲁棒性准则评估这些方法以便获得他们所需要的结果, 发现能改进

的最领先的技术和方法领域。我希望在得到对本书第一版的所有反馈之后，能把第二版做得更好。

Scott Krig
Anno Domini, 2014

目录

■ 第 1 章	图像的获取和表示	1
1.1	图像传感器技术	1
1.1.1	传感器材料	2
1.1.2	传感器光电二极管元件	3
1.1.3	传感器配置：马赛克、Faveon 和 BSI	3
1.1.4	动态范围和噪声	5
1.1.5	传感器处理	5
1.1.6	去马赛克	6
1.1.7	坏像素的校正	6
1.1.8	颜色和照明校正	6
1.1.9	几何校正	7
1.2	摄像机和计算成像	7
1.2.1	计算成像概述	7
1.2.2	单像素的摄像头计算	8
1.2.3	二维可计算摄像机	9
1.2.4	三维深度的摄像机系统	10
1.3	三维深度处理	21
1.3.1	方法概述	21
1.3.2	深度感知和处理中存在的问题	22
1.3.3	单目深度处理	27
1.4	三维表示：体元、深度图、网格和点云	31
1.5	总结	32
■ 第 2 章	图像预处理	33
2.1	图像处理概述	33
2.2	图像预处理要解决的问题	34

2.2.1	计算机视觉的流程和图像预处理	34
2.2.2	图像校正	36
2.2.3	图像增强	36
2.2.4	为特征提取准备图像	37
2.3	图像处理方法分类	41
2.3.1	点运算	42
2.3.2	直线运算	42
2.3.3	区域运算	42
2.3.4	算法	42
2.3.5	数据转换	43
2.4	色度学	43
2.4.1	色彩管理系统概述	44
2.4.2	光源、白点、黑点和中性轴	44
2.4.3	设备色彩模型	45
2.4.4	颜色空间与色彩感知	45
2.4.5	色域映射与渲染目的	46
2.4.6	色彩增强的实际考虑	47
2.4.7	色彩的准确度与精度	48
2.5	空间滤波	48
2.5.1	卷积滤波与检测	48
2.5.2	核滤波与形状选择	50
2.5.3	点滤波	51
2.5.4	噪声与伪像滤波	52
2.5.5	积分图与盒式滤波器	53
2.6	边缘检测器	54
2.6.1	核集合: Sobel, Scharr, Prewitt, Roberts, Kirsch, Robinson 和 Frei-Chen	54
2.6.2	Canny 检测器	55
2.7	变换滤波、Fourier 变换及其他	56
2.7.1	Fourier 变换	56
2.7.2	其他变换	58
2.8	形态学与分割	59
2.8.1	二值形态学	59
2.8.2	灰度和彩色形态学	61
2.8.3	形态学优化和改进	61
2.8.4	欧氏距离映射	61

2.8.5	超像素分割	62
2.8.6	深度图分割	63
2.8.7	色彩分割	64
2.9	阈值化	64
2.9.1	全局阈值化	65
2.9.2	局部阈值化	68
2.10	总结	69

■ 第 3 章 全局特征和区域特征

3.1	视觉特征的历史概述	70
3.1.1	核心思想：全局、区域和局部	71
3.1.2	纹理分析	73
3.1.3	统计方法	76
3.2	纹理区域度量	77
3.2.1	边缘度量	77
3.2.2	互相关和自相关	79
3.2.3	Fourier 频谱、小波和基签名	79
3.2.4	共生矩阵和 Haralick 特征	80
3.2.5	Laws 纹理度量	89
3.2.6	LBP 局部二值模式	90
3.2.7	动态纹理	91
3.3	统计区域度量	91
3.3.1	图像矩特征	92
3.3.2	点度量特征	92
3.3.3	全局直方图	94
3.3.4	局部区域直方图	94
3.3.5	散点图和 3D 直方图	95
3.3.6	多分辨率和多尺度直方图	97
3.3.7	径向直方图	98
3.3.8	轮廓或边缘直方图	99
3.4	基空间度量	99
3.4.1	Fourier 描述	101
3.4.2	Walsh-Hadamard 变换	102
3.4.3	HAAR 变换	103
3.4.4	斜变换	103
3.4.5	Zernike 多项式	103

3.4.6	导向滤波器	104
3.4.7	Karhunen-Loeve 变换与 Hotelling 变换	104
3.4.8	小波变换和 Gabor 滤波器	105
3.4.9	Hough 变换与 Radon 变换	106
3.5	总结	108
■ 第 4 章	局部特征设计、分类和学习	109
4.1	局部特征	109
4.1.1	检测器、兴趣点、关键点、锚点、标注	110
4.1.2	描述子、特征描述、特征提取	110
4.1.3	稀疏局部模式方法	111
4.2	局部特征属性	111
4.2.1	选择特征描述子和兴趣点	111
4.2.2	特征描述子和特征匹配	112
4.2.3	好特征的标准	112
4.2.4	可重复性，相对于困难的查找算容易	113
4.2.5	判别性与非判别性	114
4.2.6	相对和绝对位置	114
4.2.7	匹配代价和一致性	114
4.3	距离函数	115
4.3.1	关于距离函数的早期研究成果	115
4.3.2	欧氏或笛卡儿距离度量	116
4.3.3	网格距离度量	118
4.3.4	基于统计学的差异性度量	119
4.3.5	二值或布尔距离度量	120
4.4	描述子的表示	121
4.4.1	坐标空间和复数空间	121
4.4.2	笛卡儿坐标	121
4.4.3	极坐标和对数极坐标	121
4.4.4	径向坐标	122
4.4.5	球面坐标	122
4.4.6	Gauge 坐标	122
4.4.7	多元空间和多模数据	122
4.4.8	特征金字塔	123
4.5	描述子的密度	123
4.5.1	丢弃兴趣点和描述子	124

4.5.2	稠密与稀疏特征描述	124
4.6	描述子形状拓扑	125
4.6.1	关联性模板	125
4.6.2	块和形状	125
4.6.3	对象多边形	127
4.7	局部二值描述与点对模式	128
4.7.1	FREAK 视网膜模式	129
4.7.2	Brisk 模式	130
4.7.3	ORB 和 BRIEF 模式	131
4.8	描述子判别性	131
4.8.1	谱的判别性	132
4.8.2	区域、形状和模式的判别性	133
4.8.3	几何判别因素	133
4.8.4	通过特征可视化来评价判别性	134
4.8.5	精度与可跟踪	136
4.8.6	精度优化、子区域重叠、Gaussian 权重和池化	138
4.8.7	亚像素精度	138
4.9	搜索策略与优化	139
4.9.1	密集搜索	139
4.9.2	网格搜索	139
4.9.3	多尺度金字塔搜索	140
4.9.4	尺度空间和图像金字塔	140
4.9.5	特征金字塔	142
4.9.6	稀疏预测搜索与跟踪	142
4.9.7	跟踪区域限制搜寻	143
4.9.8	分割限制搜索	143
4.9.9	深度或 Z 限制搜索	143
4.10	计算机视觉、模型和结构	144
4.10.1	特征空间	144
4.10.2	对象模型	145
4.10.3	约束	146
4.10.4	选择检测器和特征	146
4.10.5	训练概述	147
4.10.6	特征和对象的分类	148
4.10.7	特征学习、稀疏编码和卷积网络	154
4.11	总结	158