



MATLAB®&Simulink®
开发实例系列丛书

在线交流，有问有答

(详见封底)



书中所有程序的源代码
可通过扫描此二维码
免费下载。

MATLAB

计算机视觉与

机器认知

主 编 王文峰 阮俊虎 CV-MATH
副主编 刘衍琦 郭裕兰 王海洋

- 集理论、算法、代码于一体
- 快速实现入门到精通的飞跃



北京航空航天大学出版社
BEIHANG UNIVERSITY PRESS

MATLAB® & Simulink® 开发实例系列丛书

MATLAB 计算机视觉 与机器认知

主 编 王文峰 阮俊虎 CV-MATH
副主编 刘衍琦 郭裕兰 王海洋



配套资料(程序源代码+数据)

北京航空航天大学出版社

内 容 简 介

这是一本用 MATLAB 演示计算机视觉原理的基础理论著作,从最初等的视频图像转换入手,层层递进,理论与实战并重但侧重于实战,借助混合编程及图形用户界面(GUI)设计,以简洁的方式展现了有一定挑战性的视频识别、目标跟踪、行为分析等关键视觉技术;同时扩展到机器认知层面,介绍仪器字符识别、机器故障诊断等有趣的应用,使读者可以在最短的时间内完成入门、进阶、精通与实战的跨越。本书主编、副主编均在中国科学院、“985 工程”大学国家重点实验室从事智能算法设计与应用的研究,部分理论功底扎实的优秀研究生也参加了主要章节的编撰。

本书既可作为算法工程师、高校教师和广大科技工作者的参考资料,也可作为高校相关专业的研究生教材和高年级大学生毕业设计的工具书。

图书在版编目(CIP)数据

MATLAB 计算机视觉与机器认知 / 王文峰,阮俊虎,
CV-MATH 主编. -- 北京:北京航空航天大学出版社,
2017.8

ISBN 978-7-5124-2428-9

I. ①M… II. ①王… ②阮… ③C… III. ①Matlab 软
件 IV. ①TP317

中国版本图书馆 CIP 数据核字(2017)第 118646 号

版权所有,侵权必究。

MATLAB 计算机视觉与机器认知

主 编 王文峰 阮俊虎 CV-MATH

副主编 刘衍琦 郭裕兰 王海洋

责任编辑 刘晓明

*

北京航空航天大学出版社出版发行

北京市海淀区学院路 37 号(邮编 100191) <http://www.buaapress.com.cn>

发行部电话:(010)82317024 传真:(010)82328026

读者信箱: goodtextbook@126.com 邮购电话:(010)82316936

北京兴华昌盛印刷有限公司印装 各地书店经销

*

开本:787×1092 1/16 印张:18.25 字数:467 千字

2017 年 8 月第 1 版 2017 年 8 月第 1 次印刷 印数:4 000 册

ISBN 978-7-5124-2428-9 定价:45.00 元

作者简介



王文峰 中国科学院西部之光学者,中国自动化学会认知计算与系统专委会和中国人工智能学会认知系统与信息处理专委会委员,CSIP2016 和 ICISCE2017 委员,SCI 期刊 PJOEs 审稿专家,国家自然科学基金项目评审专家, CV - MATH 发起人。



阮俊虎 西北农林科技大学香江学者,香港理工大学博士后;长期从事数据挖掘、优化理论的应用研究,是有一定国际影响力的物联网与大数据专家,曾受邀担任多个 SCI 期刊的客座编委。



CV - MATH 由计算机视觉和数学研究者构成的研究小组,作为第三主编的成员有(按姓氏拼音顺序)何姣姣(昆明理工大学)、刘帅奇(河北大学)、马海菲(广东科技学院)、邵永胜(西安交通大学)、王平(南京航空航天大学)、伍鹏(长江大学)、余维(湖北科技学院)、曾凡玉(电子科技大学)、张锋(电子科技大学)、邹辉(华侨大学)。



刘衍琦 中国科学院计算所烟台分所工程师,长期从事大规模图像/声纹/视频检索及其大数据应用,对以图搜图、图文识别进行过深入研究及应用,曾主编和参编《MATLAB 计算机视觉与深度学习实战》、《MATLAB 图像与视频处理实用案例详解》等。



郭裕兰 国防科技大学电子科学学院教师,中国人工智能学会优秀博士学位论文获得者,研究方向为三维视觉与模式识别;已在 TPAMI 和 IJCV 等期刊及会议上发表学术论文 50 余篇,合作出版专著 1 部,担任 TPAMI 等 30 余个国际期刊的审稿专家、AAAI 等多个国际会议的程序委员会委员,以及 IEEE TPAMI 期刊客座编辑。



王海洋 中国科学院计算所网络数据科学与技术重点实验室工程师,长期从事数据采集、信息检索、数据挖掘、情报分析等项目的架构和实施,负责和参与天玑团队核心系统等战略项目,开发的网络信息采集软件已被 2 000 多家企业使用。

前 言

计算机视觉是与机器认知密切联系的一门学科,其研究的目的是给计算机赋予人的视觉特性,使之更好地服务于人类。同时,计算机视觉与机器认知的结合可以为工业机器人提供更多有利于优化产品质量和生产流程的技术支持。此外,机器人技术要在图像视频数据采集分析的基础上才能完成智能判断和决策,这就必须借助计算机视觉搭建高精度的识别系统。目前大多数机器人都被看作是笨拙的、缓慢的、缺乏智能的,只能够用来完成一些非常特定的任务。但是,由于全球城市化进程的推进和机器人时代的来临,未来较长一段时期内,将有大量体力、脑力劳动需要借助机器人智能技术来完成。可以预见,计算机视觉与机器认知作为人工智能应用的关键方向与机器人技术的核心研究领域,即将面临前所未有的机遇和挑战。

美国和欧洲等发达国家和地区已经把计算机视觉与机器认知列为对经济社会和人类发展具有深远意义的重大研究课题。例如,计算机视觉已成为“谷歌大脑”等研究计划中的核心项目,谷歌正试图依托其庞大的数据和语音语义分析与计算架构,创造更强大的认知智能,并且在未来颠覆搜索,创造一个不需要搜索的世界,可以提前预知人的需求并给出解决方案。我国“十三五”规划将大数据、机器人产业作为国家战略,计算机视觉与机器认知作为国家政策重点支持的研究领域,势必也对我国科技自主创新、高科技成果突破产生有益的推进。

计算机视觉与机器认知均为富有挑战性和学科综合性的科学研究领域,其研究方法涉及泛函分析、代数几何、概率统计、微分方程、最优化等多个数学学科。庞大的数学方法体系正是计算机视觉技术的魔力所在,也是机器认知最难跨越的障碍之一。在此背景下,我于2014年开始借助互联网平台宣讲计算机视觉与机器认知相关学科的高级数学理论,目前已有数百名高校教师、算法工程师、博士及硕士研究生参与,我们将编程理论与算法原理体系有机结合起来,展开亦师亦友的在线交流,形成了一个非正式学术组合“CV - MATH”。经过三年的努力,一起互助互学、共同进步,我们对计算机视觉、机器认知与数学方法的关系也有了更深刻的认识,很希望把我们的交流心得分享给大家,这便是撰写本书的初衷。

本书目的

本书将计算机视觉与机器认知作为一个有机整体来研究,改变了计算机视觉研究的方式;基于作者主持和参与的中国科学院西部之光项目(XBBS - 2014 - 16)、国家千人计划项目(Y474161)和深圳市基础研究项目(JCYJ20150630114942260)等研究积累,一层一层、循序渐进地剖析在计算机视觉与机器认知中的数学问题、建模思想、算法原理与编程技巧。

全书理论与实战并重,但仍侧重于实战,探索有一定挑战性的视频识别技术,同时兼顾图像分析处理的常用技术及仪器字符识别、机器故障诊断等有趣的认知问题,提供了价值较高的MATLAB编程及混编实例。通过对数学模型的推导和配套代码的调试,读者可以更加深刻地了解计算机视觉与机器认知的算法流程,轻松领悟各类算法设计的主要思想。

本书力求为计算机视觉与机器认知的研究提供较系统、较便捷的数学方法体系,解决研究算法设计 and 应用过程中较具体的数学问题。这些问题与方法来源于作者在项目研究中发现的

实际问题和积累的实战经验,对计算机视觉与机器认知相关项目研究有一定的启发。同时,考虑到高年级大学生毕业设计和其他对计算机视觉感兴趣的入门读者的特殊需求,本书尽可能地与一些有趣的大众应用紧密结合,以便读者可以有针对性地进行算法调试、运行和仿真,快速入门,更有效地进行科学实验或完成毕业设计。

本书特点

作者阵容强大,互补优势明显

本书的主编、副主编大都在中国科学院、“985 工程”大学国家重点实验室负责认知计算、系统建模、大数据分析、机器人技术应用软件的研发工作,部分理论功底扎实、有实战经验的优秀博士及硕士研究生也参加了主要章节的编撰,做出了卓有成效的贡献,并形成了强大的作者梯队和明显的互补优势。无论是有一定编程基础的高校教师、科技工作者、算法工程师,还是相关学科专业的博士及硕士研究生、高年级大学生,都能通过阅读本书有所收获。

技术来源高端,阅读价值较高

不同于其他算法类书籍,本书算法原理与编程技巧并重,同时也集成了中国科学院西部之光项目(XBBS-2014-16)、国家千人计划项目(Y474161)和深圳市基础科研项目(JCYJ20150630114942260)等省部级、国家级研发项目部分可公开的技术成果,这些技术成果涉及较高端的 MATLAB 混合编程技术;不同于其他 MATLAB 书籍,本书讲解的计算机视觉与机器认知算法模块均结合了图形用户界面(GUI)设计,提供了直观的界面演示,这使得初学者入门更快,开发应用也更容易上手;最后,不同于计算机视觉领域其他经典理论著作,本书文字力求通俗易懂,一层一层、循序渐进地帮读者搭建思维大楼,阅读价值较高。

程序代码实用,原创性、可拓展性强

本书借助 MATLAB 提供了实用价值较高的混编代码实例,这些实例是基于作者在中国科学院西部之光项目(XBBS-2014-16)、国家千人计划项目(Y474161)和深圳市基础科研项目(JCYJ20150630114942260)等的研究过程进行归纳总结的结果,大多数实例本身就是一个独立的技术模块,具有一定的原创性和可拓展性。作者在此无私分享了在上述省部级、国家级项目研究过程中所积累的算法设计思想和部分可公开的技术源代码,这些程序代码兼具实用性和原创性,开发者可以针对其工作中的定制化需求,在现有工作基础上完成二次开发。

算法分层递进,夯实数学基础

计算机视觉与机器认知均为富有挑战性和学科综合性的科学研究领域,其研究方法涉及泛函分析、代数几何、概率统计、微分方程、最优化等多个数学学科。庞大的数学方法体系正是计算机视觉技术的魔力所在,也是机器认知最难跨越的障碍之一。本书以视频识别跟踪的技术流程为主线,算法分层递进,读者通过对书中理论与模型的阅读和推导,可以更加全面地了解这些关键视觉应用技术的相关数学问题、算法设计思想与数学建模过程。本书一层一层、循序渐进地解释图像分析处理、仪器字符识别、机器故障诊断、视频识别跟踪、目标理解描述等关键视觉技术的算法原理和算法实现流程中的数学思想。

内容架构

全书分为 9 章,以计算机视觉为主,并延伸和扩展到机器认知,理论与案例实践并重,在有限的篇幅内尽可能对各技术模块的算法思想、建模过程、编程实现及 GUI 设计等进行比较系统的介绍,具体内容架构如下:

若您对此书内容有任何疑问,可以凭在线交流卡登录 MATLAB 中文论坛与作者交流。

【第1章】 视频图像采集及读取

本章介绍视频图像采集与读取等基础知识。数字视频图像处理是计算机视觉与机器认知的基础。通常图像数据有两个来源,一是由视频转换得到的图像序列,二是直接采集的图像数据。不管是哪一种图像数据来源, MATLAB 都可将其以矩阵的形式读入工作空间,从而将对视频图像的处理简化为对矩阵数据的处理。完成处理之后,又可以通过图像显示这一环节,将矩阵数据还原成图像,当然也可以将处理后的图像序列还原成视频。在 MATLAB 视频图像处理工具箱中就有一些相关函数可以直接使用,本章也提供了这些函数的主要用法及实例。考虑到本章所有函数和技术模块是全书的基础,所有代码均给出了详尽的注释。值得一提的是,本章介绍的视频图像转换同时也包含了用 MATLAB 设计 GUI 的主要思想及核心代码,后续各章节的 GUI 程序设计都以此为基础,继续集成常用算法并完成衔接和扩展。

【第2章】 视频图像变换及融合

视频变换需要先把视频转换为图像,然后进行每一帧的图像变换,完成后再将图像转换为视频,因此这里只需要介绍图像变换。图像变换是指用数学建模的方法来描述图像位置、大小和形状等变化的方法。实际应用中,如果一幅图像的画面过大或过小,均需要对其进行缩小或放大处理。若拍摄时,景物与摄像头不成相互平衡的关系,则图像会发生几何畸形,比如会将一个圆形拍摄成一个椭圆形。出现这样的情况后,需要对图像进行畸变校正。对目标物进行匹配时,需要对图像进行缩放、旋转、平移等变换。所以,图像变换作为图像处理和分析的基础,是本章的重点内容,而图像融合作为一个补充内容。两个视频的融合,需要先把两个视频分别转换为图像,然后进行每一个对应帧的图像融合,完成后再将融合的图像转换为视频,因此这里只需要介绍图像融合。图像融合是把不同传感器的同一目标或同一场景的两幅或多幅图像进行综合,通过多幅图像间互补信息的利用来提高图像的可用性、清晰度和可识别性,从而获得对同一场景目标更为准确、可靠、全面的图像描述,以便更适合人眼的观察及做进一步的技术处理。考虑读者基础的差异,本章代码注释仍然较为详细。

【第3章】 视频图像噪声及处理

视频和图像作为最常见的视觉信息的基本载体在人们的日常生活和工作中所起的作用越来越显著,这是其他信息所不能取代的。然而,在视频图像产生、传输、量化等过程中,自然图像常常会被引入各种噪声,进而导致视频和图像质量下降,大大增加了信息解译的复杂性;同时,降低了视频图像的可编译性,影响了认知效果以及信息的可用性,给视频图像增强、复原、重建、识别等后续工作带来了严重的影响。视频噪声处理需要先将视频转换为图像,处理完成后再将减少噪声的图像转换为视频,因此这里仅需介绍图像去噪。本章主要介绍了图像去噪的算法设计思想及在工程应用中涉及到的一系列算法(包括均值滤波、中值滤波、小波阈值滤波、轮廓波变换和非局部均值滤波等),并借助 MATLAB 加以实现。本章开始接触到一些真实的工程应用,读者可以更好地理解算法原理,对于如何进行算法设计、数学建模,以及如何将算法表达为代码将有更加深刻的认识。

【第4章】 视频图像阈值及分割

图像分割(Image Segmentation)是把图像分割成若干个特定的、具有独特性质的区域并提取出感兴趣的区域的技术和过程。在对图像的研究和应用中,人们往往仅对图像的某些部分(目标或背景)感兴趣,它们一般对应图像中的特定的、具有独特性质的区域。为了分析和识别目标,需要将它们分割并提取出来。图像分割是由图像处理转到图像分析的关键。一方面,它是目标图像表达的基础,对特征测量有重要的影响;另一方面,图像分割和分割的目标表达、

特征提取及参数测量等将原始图像转化为数学表达形式,使得利用计算机进行图像分析和理解成为可能。本章主要介绍图像分割的基本概念和分割所用的主要方法,具体包括边缘检测、灰度阈值分割、全局阈值分割、动态阈值分割和区域生长及分割等一系列常见的图像分割方法;在代码注释上,侧重于前几章没有注释过的技术模块,其他基础部分的代码注释被省略,甚至部分基础代码不予展示,以便适应读者进阶的过程与需求。

【第 5 章】 图像特征计算及应用

本章以人脸检测、行人检测等应用为例,主要对图像处理中的几个比较常用的特征提取方法及应用实例进行介绍,这些常用的特征包括 Haar-like 特征、Hog 特征、LBP 特征以及 SIFT 特征。这些特征都各有优缺点,在使用的时候应该根据实际的场景进行选择,这样才能够在识别性能和运算性能之间找到平衡点。显而易见,本章已经开始接触一些相对高级的工程应用了,因此相应地对于每个技术模块,都补充了相关基础定义、基础函数和基础模块的说明。在介绍 Haar-like 特征人脸检测之前,说明了 Haar-like 特征的种类、计算和个数;在介绍 Hog 特征行人检测之前,介绍了一些必要的准备工作,包括图像的预处理、梯度和滑动。为了适应读者进阶的需求,在本章的最后两节,基础介绍开始精减,重心转移到算法思想上发,在介绍利用 LBP 特征进行纹理分类之前,重点介绍了 LBP 特征的演化;而在介绍利用 SIFT 特征进行模板匹配之前,侧重说明了 SIFT 算法的特点、设计框架与设计过程。

【第 6 章】 运动目标检测及跟踪

运动目标检测是计算机视觉与机器认知的一个重要环节,在计算机视觉的多个应用领域,诸如运动编码、行为分析及场景理解中,运动目标检测一直是研究的热点问题。运动目标检测的主要目的是将图像中的前景和背景分割开来,提取感兴趣的运动目标,并获取描述每一个运动目标特征的相关参数,为解决视觉问题提供必要的信息。本章介绍运动目标检测及跟踪的算法设计思想、数学建模过程,并提供算法编程实现、拓展与优化。这一章开始接触到更实用、更高级的知识层面,读者经过前面几章的学习已经具备了计算机视觉研究的基础,相应地,本章算法的介绍更加趋于具体化、系统化和数学化。本章介绍了运动目标检测及跟踪的算法设计思想、数学建模过程和算法编程实现、拓展与优化,其中数学建模过程具体介绍了帧间差分法建模和改进过程,其后的算法代码实现不仅包括了帧间差分法和改进的帧间差分法,而且考虑了一些特殊问题与对应的算法,并结合 MATLAB 编程实现。

【第 7 章】 目标定位及字符识别

这一章读者将开始接触到机器认知层面的知识,具体算法是计算机视觉中的图像目标提取和识别。同时,有了前面 6 章的基础,这里不仅仅介绍图像目标提取和识别,同时也将深入地介绍视频识别与跟踪。除此之外,对于图像目标定位、提取和识别的算法思想也做了详尽的描述,而模型方面的介绍也趋于简单化,这就意味着从这一章开始,计算机视觉研究者就要开始尝试机器认知层面的应用,需要学会基于算法的设计思想自行完成与机器认知相关的数学建模了。视频轨迹分析是目标识别与跟踪的延伸,为了帮助读者顺利完成计算机视觉到机器认知的过渡与衔接,本章从内容架构上分为图像目标提取与识别和视频识别与轨迹分析,其中图像目标提取与识别具体分为图像采集与目标检测、图像目标归一化处理及图像字符提取与识别,而视频识别与轨迹分析则分为问题描述与建模思想、算法功能与技术流程,最后提供了 MATLAB 进行视频人脸识别与跟踪的核心代码。

【第 8 章】 机器故障认知及检测

作为前几章的延伸,本章将进一步开阔机器视觉的研究和应用的思路,主要介绍如何用声

音信息进行机器故障诊断。读者会惊喜地发现,这部分扩展应用不会占用太多的研究时间,可作为算法跨领域的扩展应用,它与图像滤波去噪算法有很多的相似之处。这一章将一个实际应用案例作为一个专门模块进行系统、全面的介绍,包括机器故障诊断的问题描述与建模思想、分析过程与编程实现。其中,分析过程与编程实现又具体到该问题的不同方面,包括齿轮振动信号的数字滤波处理、频谱分析,波形分析中常用的有量纲指标、无量纲指标的计算,以及齿轮振动信号的相关分析等。

【第9章】深度学习及人脸识别

作为前几章的升华,本章介绍的深度学习(Deep Learning, DL),旨在弥补计算机视觉与机器认知经典算法的不足。以图像识别为例,首先要获取图像数据,然后经过预处理、特征提取、特征选择,最后进行推理、预测,以实现图像识别。通过推理、预测来实现图像识别的这一阶段就是用机器学习算法来实现的,而数据的预处理、特征提取以及特征选择这一阶段也至关重要,这一阶段称为特征表达。良好的特征表达对最终算法的有效性影响非常大,而特征表达一般都是靠人工完成的,包括人工提取特征、人工选择特征,能否获取良好的特征很大程度上需要靠经验和运气,并且费时。DL就可以很好地解决这一问题。由于涉及到MATLAB与Caffe的混合编程,所以对MATLAB混合编程、Caffe语言以及DL工具箱也进行了比较系统的介绍,主要包括算法形成背景、算法基本思想、算法数学原理及人脸识别应用等方面。其中,为了更好地帮助读者理解深度学习算法,将算法数学原理解释作为本章的重点,具体地分为权值共享、CNN结构、算法设计框架、混合编程思想和算法应用详解等方面。

作者贡献

第1章主要由余维、王文峰编写,第2章主要由曾凡玉、马海菲、何姣姣编写,第3章主要由刘帅奇、王文峰编写,第4章主要由王平、王文峰编写,第5章主要由张锋、王文峰编写,第6章主要由伍鹏、王文峰编写,第7章主要由王文峰、刘衍琦编写,第8章主要由邵永胜、王文峰编写,第9章主要由邹辉、王文峰编写。在全书各章节撰写及审校过程中,郭裕兰对算法思想、建模过程、程序代码进行了修改、补充与完善。各章节配套的GUI程序主要由刘衍琦提供,阮俊虎、王海洋也参与了部分章节的GUI程序设计。

如何使用本书

计算机视觉与机器认知的算法思想和建模过程涉及较复杂的数学理论体系。本书各章节配套了GUI程序,正是希望借助图形用户界面,让读者在最短的时间内完成对各技术模块算法框架的理解。此外,计算机视觉与机器认知属于较大规模的编程,为确保读者阅读的完整性和连贯性,部分章节难免出现大段程序代码。考虑到这一特殊情况,我们尽可能对每一行代码都添加了必要的注释,在最大程度上提高读者的阅读效率。北京航空航天大学出版社在本书扉页也提供了二维码,读者可扫码下载,轻松获取这些代码,并在短时间内完成调试。

为了及时解答读者阅读过程中的疑难问题和满足读者进一步的需求,本书作者还将联合MATLAB中文论坛为广大读者提供优质配套资源及网络互动答疑服务,我们会尽量每周登录网站2~3次,集中解答读者的疑难问题,并给出提示,根据读者进一步的需求,上传和更新书籍辅助资源。本书已在MATLAB中文论坛申请和设立了专版,用于在线答疑和源码分享:

书码验证:<http://www.ilovematlab.cn/book.php>;

在线答疑:<http://www.ilovematlab.cn/forum-258-1.html>。

特别致谢

在本书的编写过程中,参考了 MATLAB 中文论坛大量的 MATLAB 帮助文档、MATLAB 书籍及其他相关资源。MATLAB 中文论坛注册会员超过 100 万人,是全球最大的 MATLAB 中文门户。MATLAB 中文论坛不仅是 MATLAB 编程类书籍的 Facebook,而且也是 MATLAB 开发商 MathWorks 公司的战略合作伙伴。论坛运营已超过 10 年,采取线上、线下交流互助模式,培养了数百万名 MATLAB 爱好者。本书所有作者正是在论坛开始接触并逐渐喜欢上 MATLAB 的。不仅如此,论坛创始人张延亮作为 MathWorks 机器人工具箱 (Robotics System Toolbox) 首任产品负责人,亲自担任了本书的统审专家,做出了指导性的贡献。

本书还得到了 MathWorks 公司总部机器人产品部和市场部、松山湖国际机器人研究院、中国科学院新疆生态与地理研究所、新疆维吾尔自治区科技厅、中国科学院计算所烟台分所、西北农林科技大学等单位领导和同事的大力支持,在此对他们表示衷心的感谢。本书写作之初还得到了北京航空航天大学出版社陈守平编辑的鼓励和支持,在此深表谢意。

感谢中国科学院西部之光项目(XBBS-2014-16)、国家千人计划项目(Y474161)和深圳市基础研究项目(JCYJ20150630114942260)等的支持。感谢我的授业恩师陈曦研究员,陈老师一直给予我正确的引导和前行的勇气与力量!本书在审校过程中还得到中国自动化学会认知计算与系统专委会、中国人工智能学会认知系统与信息处理专委会的支持,在此一并致谢!最后,必须感谢我的父母、妻子和所有家人的默默支持!感谢女儿和即将出生的小宝贝,是你们赋予了我神奇的力量,让爸爸突破了时间和精力的局限,最终完成这本书!

最后,感谢与我并肩作战、共同撰写此书的其他主编、副主编。郑重感谢我的团队 CV-MATH,尤其是作为第三主编参与本书撰写的 CV-MATH 成员,感谢他们始终如一的支持与鼓励!值此书稿完成之际,我想说:你们都是好样的!非常荣幸能与你们一起工作!

由于时间仓促,加之作者水平和经验有限,书中疏漏甚至错误在所难免,希望广大读者批评指正,您的建议将是我们创作和研究的最大动力与源泉。

王文峰

2017年6月



配套资料(程序源代码+数据)

本书所有程序的源代码均可通过 QQ 浏览器扫描二维码免费下载。读者也可以通过以下网址下载全部资料:<http://www.buaapress.com.cn/upload/download/matlabsjrz.rar>。

配套资料下载或与本书相关的其他问题,请咨询北京航空航天大学出版社理工图书分社,电话(010)82317036,邮箱:goodtextbook@126.com。

目 录

第 1 章 视频图像采集及读取	1
1.1 视频图像采集	1
1.1.1 视频生成技术	1
1.1.2 智能采集技术	1
1.2 视频图像读取	8
1.2.1 视频读取函数	8
1.2.2 图像读取函数	10
1.2.3 视频图像转换	23
第 2 章 视频图像变换及融合	32
2.1 视频图像变换	32
2.1.1 平移变换	32
2.1.2 镜像变换	34
2.1.3 裁剪操作	36
2.1.4 缩放操作	38
2.1.5 旋转变换	39
2.1.6 几何变换	40
2.1.7 邻域操作	42
2.2 视频图像融合	57
2.2.1 算法思想	57
2.2.2 算法实例	58
第 3 章 视频图像噪声及处理	63
3.1 算法基本思想	63
3.2 噪声均值滤波	64
3.2.1 数学建模	64
3.2.2 编程实现	65
3.3 噪声中值滤波	65
3.3.1 数学建模	65
3.3.2 编程实现	66
3.4 小波阈值滤波	67
3.4.1 算法思想	67
3.4.2 数学建模	68
3.4.3 编程实现	69
3.5 轮廓波变换	70
3.5.1 算法设计	70
3.5.2 算法实现	72

3.5.3	代码实例	73
3.6	非局部均值滤波	75
3.6.1	算法设计	75
3.6.2	算法实现	77
第 4 章	视频图像阈值及分割	93
4.1	图像分割的基本概念	93
4.1.1	图像分割的定义	93
4.1.2	图像分割算法分类	94
4.2	边缘检测	94
4.2.1	边缘检测概述	94
4.2.2	边缘检测梯度算法	95
4.2.3	拉普拉斯算子	97
4.2.4	LoG 算子	98
4.2.5	Canny 算子	99
4.3	阈值分割	101
4.3.1	阈值分割介绍	101
4.3.2	全局阈值分割	103
4.3.3	动态阈值分割	109
4.4	区域生长及分割	114
4.4.1	区域生长的基本原理、步骤及流程图	114
4.4.2	生长准则和过程	115
4.4.3	区域分裂与合并	119
第 5 章	图像特征计算及应用	123
5.1	Haar-like 特征	123
5.1.1	Haar-like 特征的种类	123
5.1.2	Haar-like 特征的计算	124
5.1.3	Haar-like 特征的个数	125
5.1.4	Haar-like 特征人脸检测	125
5.2	Hog 特征	128
5.2.1	图像的预处理	128
5.2.2	图像的梯度	128
5.2.3	图像的滑动	129
5.2.4	Hog 特征行人检测	129
5.3	LBP 特征	132
5.3.1	LBP 特征的演化	132
5.3.2	LBP 特征纹理分类	134
5.4	SIFT 特征	142
5.4.1	SIFT 算法的特点	142
5.4.2	SIFT 算法设计框架	143
5.4.3	SIFT 算法设计过程	143

若您对此书内容有任何疑问，可以凭在线交流卡登录 MATLAB 中文论坛与作者交流。

5.4.4 SIFT 特征模板匹配	150
第 6 章 运动目标检测及跟踪	156
6.1 算法设计思想	156
6.2 数学建模过程	157
6.2.1 帧间差分法建模	157
6.2.2 帧间差分法改进	158
6.3 算法编程实现	158
6.3.1 帧间差分法	158
6.3.2 改进的帧间差分法	161
6.4 算法拓展与优化	165
6.4.1 特殊问题与算法	165
6.4.2 MATLAB 编程实现	166
第 7 章 目标定位及字符识别	188
7.1 图像目标提取与识别	188
7.1.1 图像采集与目标检测	189
7.1.2 图像目标归一化处理	194
7.1.3 图像字符提取与识别	197
7.2 视频识别与轨迹分析	203
7.2.1 问题描述与建模思想	204
7.2.2 算法功能与技术流程	204
7.2.3 MATLAB 核心代码	204
第 8 章 机器故障认知及检测	211
8.1 问题描述与建模思想	211
8.2 分析过程与编程实现	214
8.2.1 齿轮振动信号的数字滤波处理	214
8.2.2 齿轮振动信号的频谱分析	226
8.2.3 波形分析中常用的指标计算	232
8.2.4 齿轮振动信号的相关分析	235
第 9 章 深度学习及人脸识别	243
9.1 算法形成背景	243
9.2 算法基本思想	243
9.3 算法数学原理	244
9.3.1 权值共享	244
9.3.2 CNN 结构	245
9.3.3 算法设计框架	246
9.3.4 混合编程思想	249
9.3.5 算法应用详解	249
9.4 基于深度学习的人脸识别	266
参考文献	276

1.1 视频图像采集

1.1.1 视频生成技术

视频图像处理是计算机视觉与机器认知的重要内容。视频生成技术是利用人类的“视觉滞留”原理,将多幅图像以超过 24 帧每秒的速度播放,形成平滑连续的画面,即视频。因此,视频实质上是图像时间序列,而视频处理就是对一系列静态图像信号的采集、标记、保存、传输和重现及其他智能化的处理和操作。既然视频可以看作是图像在时间轴上扩展所得到的图像序列,那么视频图像的采集与读取就是视频图像处理的基础。

MATLAB 包含了一个强大的视频及图像处理工具箱,具备视频图像处理和图形展示等多元化功能,可以轻易地实现视频读取、图像采集、视频信息获取、视频图像转换(帧图像序列提取)、视频播放、软件截屏等基本功能和操作。采用 MATLAB 研究计算机视觉与机器认知,易上手,进阶快。

作为记录多媒体信息的重要载体,视频同时包含图像、声音、备注信息(例如字幕)等内容, MATLAB 以数字形式进行存储和记录视频(称为数字视频),可以方便地对不同应用场景的智能算法进行编程,实现计算机视觉与机器认知的各类技术模块,包括仪器字符识别、机器齿轮声音分析等辅助模块。

1.1.2 智能采集技术

机器认知与计算机视觉的算法原理相似,但不同的是,计算机视觉(例如,较经典的智能监控应用)主要是在后端(监控中心)进行视频图像分析和处理,而机器认知除了车间工作可以通过后台的机器视觉系统实现控制之外,更多时候需要在前端运行智能算法,例如,机器人导航等高端工业技术的实现。机器认知的实现需要将计算机视觉中的视频图像分析处理算法移植到 ROS(Robot Operator System),例如,智能相机主要用于实现视频和图像的智能采集甚至直接进行智能分析与处理。机器视觉系统采用的工业相机主要分为三大类,具体包括 FA 镜头(智能相机)、远心镜头和工业显微镜。其中远心镜头主要是为纠正传统工业镜头的视差而设计的,它可以在一定的物距范围内,使得到的图像放大倍率不会变化,这对被测物不在同一物面上的情况是非常重要的应用。工业显微镜主要适用于金属、半导体行业的超精细检测。我们将在其他章节介绍视频图像智能处理算法时作对应的、更详细的介绍,这里主要介绍视频图像的智能采集技术。

数字视频图像处理是计算机视觉与机器认知的基础。通常图像数据有两个来源,一个是由视频转换得到的图像序列,另一个就是直接采集的图像数据。智能相机(smart camera)是

一种高度集成化的微型机器视觉系统。它将图像的采集、处理与通信功能集成于单一相机内,从而提供了具有多功能、模块化、高可靠性、易于实现的机器视觉解决方案。同时,由于应用了最新的 DSP、FPGA 及大容量存储技术,其智能化程度不断提高,可满足多种机器认知的应用需求。

不管是哪一种图像数据来源, MATLAB 都可以将其以矩阵的形式读入工作空间,从而把对视频图像的处理简化为对矩阵数据的处理。完成处理之后,又可以通过图像显示这一环节将矩阵数据还原成图像,当然也可以将处理后的图像序列还原成视频。视频图像转换我们将在 1.2.3 小节进行介绍,这里先介绍利用 MATLAB 进行视频和图像采集的方法。其大致分为如下几个步骤:

(1) 安装和配置视频图像采集设备,确保摄像头处于连接和正常运转状态。

(2) 获取能够唯一标识这个视频图像采集设备的硬件信息(在创建数字视频、输入对象时,要用此类信息),包括 Adaptor name 软件适配器名、Device ID 硬件设备的 ID、Video format(视频格式)等。利用 MATLAB 图像视频处理工具箱里的 `imaqhwinfo` 函数可以检测到安装了一个适配器 `winvideo`,用此适配器名可以连接视频图像采集设备,获取该设备的硬件信息。

初次使用 `imaqhwinfo` 时,可不设定输入参数:

```
info = imaqhwinfo
```

运行结果如下:

```
info =
    InstalledAdaptors: {}
    MATLABVersion:    '8.6 (R2015b)'
    ToolboxName:      'Image Acquisition Toolbox'
    ToolboxVersion:   '4.10 (R2015b)'
```

此时返回的是一个结构体,它包含了系统中存在的适配器和 MATLAB 相关的版本信息。

若要获取指定适配器名的适配器的相关信息,使用命令格式如下:

```
info = imaqhwinfo('winvideo')
```

运行结果如下:

```
info =
    AdaptorDllName:
    'C:\MATLAB\SupportPackages\R2014a\genericvideo\adaptor\win64\mwwinvideoimg.dll'
    AdaptorDllVersion: '4.7 (R2014a)'
    AdaptorName:      'winvideo'
    DeviceIDs:        {[1]}
    DeviceInfo:       [1x1 struct]
```

此时返回的是适配器 `winvideo` 的具体参数,包括适配器 dll 文件绝对路径、适配器 dll 文件版本、适配器名称、设备 ID 及图像采集设备的相关参数。

若想获得更多关于硬件设备的信息,可使用命令:


```
dev_info = imaqhwinfo('winvideo', 1)
```

运行结果如下：

```
dev_info =
    DefaultFormat: 'MJPG_1280x720'
    DeviceFileSupported: 0
    DeviceName: 'Integrated Camera'
    DeviceID: 1
    VideoInputConstructor: 'videoinput('winvideo', 1)
    VideoDeviceConstructor: 'imaq.VideoDevice('winvideo', 1)
    SupportedFormats: {1x15 cell}
```

当然,如果计算机连接了两个摄像头,还可以使用命令:

```
dev_info2 = imaqhwinfo('winvideo', 2)
```

(3) 创建视频输入对象。利用第(2)步中获取的适配器名、设备 ID 和设定的视频格式,使用 videoinput 函数创建视频输入对象,可以只用适配器名作为 videoinput 函数的输入参数,videoinput 函数会使用默认的设备 ID 和视频格式。创建视频输入对象的命令格式如下:

```
vid = videoinput('winvideo')
```

运行结果如下:

```
Summary of Video Input Object Using 'Integrated Camera'.
Acquisition Source(s): input1 is available.
Acquisition Parameters: 'input1' is the current selected source.
                        10 frames per trigger using the selected source.
                        'MJPG_1280x720' video data to be logged upon START.
                        Grabbing first of every 1 frame(s).
                        Log data to 'memory' on trigger.
Trigger Parameters: 1 'immediate' trigger(s) on START.
Status: Waiting for START.
                0 frames acquired since starting.
                0 frames available for GETDATA.
```

(4) 预览视频流。为了测试得到的视频是否满足实际需求,可以先对其进行预览。根据预览的情况对摄像设备的位置、光照、对焦等影响视频效果的参数进行调节。预览视频流的命令格式如下:

```
preview(vid)
```

此时弹出了视频预览窗口,里面是实时监控图像。

运行结果如图 1-1 所示。

若要关闭预览窗口,可使用如下命令格式:

```
closepreview(vid)
```

(5) 配置视频对象的属性。工具箱中与视频设备的连接由两种类型对象表示: