

作者前言

FOREWORD

本书通过采用程序实现的实际操作方法介绍了媒体处理的概念和原理及其在模式识别中的应用。本书的主要目的是让读者了解可以使用数据分析和可视化工具 MATLAB 来读取、修改和写入图像、音频和视频文件的工具和技术。本书是为学习图像处理、语音和语言处理、信号处理、视频目标检测和跟踪以及相关多媒体技术的毕业班大学生和研究生而写的,重点是使用编程结构和技能开发的实际实现。本书的特点是介绍了与媒体处理相关的技术及其各种应用的简明软件解释,以及相关的理论背景和实际实施步骤,同时避免了冗长的理论论述。本书也适用于模式识别、计算机视觉和基于内容的检索领域的研究人员,以及学习有关媒体处理、统计分析和数据可视化的 MATLAB 课程的学生。

MathWorks 公司的 MATLAB 是一种数据分析和可视化工具,适用于数值计算、算法开发和仿真应用。与其他同时期编程工具相比, MATLAB 的一个优势是它存储了大量现成的函数,可用于广泛的媒体处理任务,这些任务可以直接包含在自定义应用程序和问题解决任务中,而无需任何额外的编程工作。因此,未来的程序员和学生有必要了解这些函数,以便在必要时使用它们,从而可以快速开发应用程序并减少工作量和时间。MATLAB 函数一般分为两大类:基本集和扩展集。本书中称为基本 MATLAB(BM)函数的基本集具有基本处理功能,如数据类型转换、算术关系运算和逻辑运算、代数和三角运算、插值和傅里叶分析、各种图形绘图和注释、文件 I/O 操作,以及不同类型的基本编程结构。由于已假设读者具备 MATLAB 的基本知识,并熟悉矩阵代数和三角学等基本数学运算,因此,本书并不打算向初学者讲授初级 MATLAB,尽管在本书中首次使用函数的地方已经提供了函数的简明描述。

第二组更专业的函数称为“工具箱”,它们扩展了为特定领域和应用场合定制的基本功能集。本书讨论了许多工具箱中的函数,用于说明各种媒体处理任务,分为图像处理、音频处理和视频处理 3 章。每个函数都使用示例进行了说明,同时将程序输出以图形的方式显示,以便于可视化。第 4 章讨论了前 3 章讨论的媒体处理任务在解决模式识别问题中的应用。在此过程中,本书深入研究了以下工具箱中一些常用函数的选定子集:

- 图像处理工具箱(IPT): 为图像处理、分析、可视化和算法开发提供一整套参考标准算法和应用工作流程。
- 音频系统工具箱(AST): 为音频处理系统的设计、仿真和桌面原型设计提供算法和工具。
- 计算机视觉系统工具箱(CVST): 提供用于设计和模拟计算机视觉和视频处理系统的算法、功能和应用程序。

- 数字信号处理系统工具箱(DSPST): 提供设计、模拟和分析信号处理系统的算法和应用程序。
- 统计和机器学习工具箱(SMLT): 提供用于描述、分析和建模数据的函数和应用程序。
- 神经网络工具箱(NNT): 提供创建、训练、可视化和模拟神经网络的算法、预训练模型和应用程序。
- 信号处理工具箱(SPT): 提供从均匀和非均匀采样信号中分析、预处理和提取特征的函数和应用程序。
- 小波工具箱(WT): 提供使用基于小波的建模以分析和合成信号与图像的函数和应用程序。

因为本书的重点是让读者熟悉 MATLAB 函数,因此对理论概念介绍不多,只介绍与所讨论函数相关的概念。这与其他媒体处理书籍不同,这些书籍更多地关注于解释理论概念并将其中一些概念通过示例加以说明。此外,本书主要讨论用于执行各种媒体处理任务的 MATLAB 软件包的功能,并在必要时讨论了简短的相关理论部分,以理解基本概念。这使得本书能够以紧凑的方式呈现思想,读者无须阅读长篇大论就能快速理解。另外,本书通过解决具体数值问题的实例说明了这些特点,这有助于未来的学生和读者快速熟悉所有相关函数,并利用这些功能解决定制问题。本书将代码以简单的复制和执行格式提供,以方便初学者理解。代码的输出以图形和绘图的形式显示,以帮助学习者将结果可视化,以便快速吸收知识。根据相关工具箱中 MATLAB 函数的层次结构,对每章中的主题进行了排序。以下是每章各节涵盖的主要主题列表,从每章的第 3 节开始,大部分主题都用至少一个编码示例以及显示程序输出的可视化绘图进行了说明。

第 1 章讨论图像处理,共分为 11 节。1.1 节介绍了基本概念,包括像素、数字化、二值图像、灰度图像、彩色图像、图像采集和输出设备、图像转换、图像调整、彩色模型、压缩方案和文件格式。1.2 节列出了该章涵盖的基本函数和工具箱函数,并提供了一个包含五组约 115 个基本 MATLAB 函数的列表;以及一个包含 116 个属于图像处理工具箱(IPT)函数的列表。IPT 函数分为五类:导入导出和转换、显示和探索、变换和配准、滤波和增强、分割和分析。1.3 节涉及图像的导入、导出和转换,涵盖了诸如读取和写入图像数据,图像类型转换,二值化阈值和大津方法,图像量化和灰度,索引图像和抖动,使用无符号 8 位整数和 64 位双精度的图像表示,图像彩色表示、三刺激和色度值,sRGB 和 AdobeRGB 彩色空间,RGB、CMY、XYZ、HSV、 $L^* a^* b^*$ 表示之间的彩色转换,棋盘格和幻影等合成图像,图像噪声表示和高斯函数等主题。1.4 节讨论了图像的显示和探索,涵盖了基本显示技术、图像融合和蒙太奇、图像序列和扭曲表面,以及交互式探索工具等主题。1.5 节讨论了几何变换和图像配准,包括常见几何变换、仿射和投影变换、图像配准、协方差和相关性,以及缩放和插值等主题。1.6 节涉及图像滤波和增强,涵盖了核、卷积、图像模糊、噪声滤波器、排序统计滤波器、盖伯滤波器、边缘检测算子、图像梯度和图像偏导数、对比度调整和伽马曲线、直方图均衡化、形态学运算、感兴趣区域和块处理、算术和逻辑运算、点扩展函数、反卷积、逆滤波器、维纳反卷积、露西·理查森(Lucy-Richardson)反卷积和盲反卷积等主题。1.7 节讨论了图像分割和分析,包括图像分割、目标分析、哈夫变换、二叉树分解、提取区域特性、像素连通性、纹理分析、灰度共生矩阵、图像质量、信噪比、均方误差、结构相似性指数、图像变换、离散

傅里叶变换、离散余弦变换和离散小波变换等主题。1.8节讨论了频域处理,涵盖卷积定理、理想和 高斯低通滤波器以及理想和 高斯高通滤波器等主题。1.9节讨论了使用 Simulink 进行图像处理,并涵盖了用于图像类型转换、彩色转换、色调调整、边缘检测、几何变换、形态学操作和团点分析等任务的 Simulink 模型的开发等主题。1.10节和 1.11节分别讨论了各种二维和三维绘图功能的句法、选项和参数,以及可视化数据分布及其定制内容。

第2章介绍音频处理,共分为11节。2.1节介绍了基本概念,包括声波特性(如振幅和频率)及其感知表示、即响度和音调,用于操纵环境声音的设备,如麦克风放大器和扬声器,立体声和单声道,音频数字化,采样和奈奎斯特采样定理,声卡组件,CD质量数字音频的特性,音频滤波,合成器和 MIDI 协议,压缩方案和文件格式。2.2节列出了该章涵盖的基本函数和工具箱函数,并提供了一个包含52个基本 MATLAB 函数的列表,这些函数分为五组。还有一个包含29个属于音频系统工具箱(AST)的函数的列表,这些函数分为五类:音频 I/O 和波形生成,音频处理算法设计,测量和特征提取,仿真、调整和可视化,以及乐器数字接口(MIDI)。除了 AST 函数外,还列出了其他两个工具箱,即 DSP 系统工具箱(DSPST)和信号处理工具箱(SPT)。2.3节讨论了声波的研究和表征,涵盖了波形、相位、采样频率、混叠、正弦音调、复合音符、音频信号的傅里叶域表示、系数和基函数等主题。2.4节讨论了音频 I/O 和波形生成,包括读取和写入数字音频文件、绘制音频波形、记录和回放数字音频、示波器显示和波形表合成器等主题。2.5节讨论了音频处理算法设计,涵盖混响、噪声门、动态范围压缩器和扩展器以及交叉滤波器等主题。2.6节讨论了测量和特征提取,涵盖了基音、语音活动检测(VAD)、瞬时响度、短期响度和综合响度以及梅尔(Mel)频率倒谱系数(MFCC)等主题。2.7节介绍了仿真、调整和可视化,涵盖了时间范围、正弦波发生器、频谱分析仪和阵列图等主题。2.8节介绍了 MIDI,并涵盖一些主题,如使用各种乐器声音播放 MIDI 音符。2.9节讨论了时间滤波器,涵盖了有限脉冲滤波器(FIR)、无限脉冲滤波器(IIR)和窗口函数等主题。2.10节涉及频谱滤波器,涵盖了频率表示、频谱图、变频信号、低通和高通 FIR 滤波器、低通和高通 IIR 滤波器、带阻和带通 FIR 滤波器以及带阻和带通 IIR 滤波器等主题。2.11节介绍了使用 Simulink 进行音频处理,并涵盖了诸如为混响、噪声门、交叉滤波器、VAD、响度测量仪和频谱等任务开发 Simulink 模型等主题。

第3章讨论视频处理,共分为8节。3.1节介绍了视频帧、帧速率、运动错觉、光栅扫描、隔行扫描、分量视频信号、合成视频信号、亮度和色度、RGB 到 YC 信号格式的转换以及色度亚采样等基本概念。3.2节列出了本章涵盖的基本函数和工具箱函数,并提供了一个包含约21个基本 MATLAB 函数的列表,这些函数分为五组。还提供了一个属于计算机视觉系统工具箱(CVST)的20个函数的列表,这些函数分为三类:输入、输出和图形,对象检测和识别,目标跟踪和运动估计。3.3节涉及视频输入-输出和播放,包括读取和写入视频文件、视频帧子集的显示和播放以及从图像集合创建电影等主题。3.4节涉及视频帧的处理,包括创建用于存储视频帧的4-D结构,将图像和视频帧互相转换,在视频帧的特定位置插入文本,以指定帧速率选择性播放帧,将视频帧从彩色转换为灰度和二进制版本,以及对视频帧应用图像滤波器。3.5节涉及视频彩色空间,包括 RGB 与 $YCbCr$ 彩色空间的互相转换,RGB 与 NTSC 彩色空间的互相转换,以及 RGB 与 PAL 彩色空间的互相转换。3.6节介绍了目标检测,包括团点检测器、前景检测器、人体检测器、人脸检测器和光学文字识别等

主题。3.7节讨论了运动跟踪,涵盖了基于直方图的跟踪器、光流、点跟踪器、卡尔曼滤波器和块匹配器等主题。3.8节介绍了使用 Simulink 的视频处理,包括视频彩色空间转换、几何变换、彩色到灰度和二进制转换,以及将图像滤波器应用于视频帧等主题。

第4章讨论模式识别,共分为8节。4.1节介绍了基本概念,包括聚类、分类、监督学习、无监督学习、训练阶段、测试阶段、特征向量、特征空间和相似性度量。4.2节列出了本章涵盖的基本函数和工具箱函数,并列出了属于计算机视觉系统工具箱(CVST)的10个函数和属于统计和机器学习工具箱(SMLT)的26个函数。4.3节涉及数据采集,包括将工作空间变量保存到MAT文件、将变量从MAT文件加载到工作空间、使用 Fisher-Iris 数据集、从任意文件夹读取多个媒体文件以及使用图像数据存储等主题。4.4节讨论了预处理,包括媒体类型转换、彩色转换、几何变换、色调校正、噪声滤波、边缘检测、形态学操作、目标分割以及时间和光谱滤波等主题。4.5节介绍了特征提取方法,包括最小本征值法、哈里斯角点检测器、加速分段测试特征(FAST)算法、最大稳定极值区域(MSER)算法、加速鲁棒特征(SURF)算法、KAZE算法、二进制鲁棒不变可伸缩关键点(BRISK)算法、局部二进制模式(LBP)算法和梯度方向直方图(HOG)算法。4.6节讨论了聚类,涵盖了相似性测度、 k -均值聚类算法、 k -中心点聚类算法、分层聚类算法和基于GMM的聚类算法等主题。4.7节涉及分类,涵盖了 k -NN分类器、人工神经网络(ANN)分类器、决策树分类器、鉴别分析分类器、朴素贝叶斯分类器、支持向量机(SVM)分类器和分类学习器应用程序等主题。4.8节涉及性能评估,包括剪影值、卡林斯基-哈拉巴斯指数和混淆矩阵等主题。

由于图像处理、音频处理和视频处理在模式识别、计算机视觉、目标检测、人工智能、语音和说话人识别、语音激活、视频监控、人脸识别、车辆跟踪、运动估计等的广泛应用,相关课程在全世界都有需求,本书对大多数人来说都是有用的。本书末尾包含一个函数汇总,提供了本书中讨论的大约400个MATLAB函数,并按字母顺序排序,同时给出了它们的原始工具箱并对每一个函数给出了一行描述,以便于读者参考。本书提供了大约75本参考书和研究论文,供读者进一步阅读本书中讨论的各种算法、标准和方法。每章结尾都有一组用于自我评估的复习问题。本书包含大约250个已解示例及其相应的MATLAB代码。虽然这些代码是在2018版MATLAB中进行的测试,但其中大多数代码能在2015版及其后的版本中正确执行。粗体字用于突出每个章节中讨论的重要理论/概念术语。本书中有100多个这样的术语。红色粗体文本用于表示首次出现的MATLAB函数名。本书讨论了400多个MATLAB函数。读者可使用MATLAB帮助工具获取有关所讨论函数的更多信息。本书提供了超过300幅彩图,以帮助读者直观地可视化程序输出。

鼓励所有读者提供有关本书内容以及任何遗漏或打字错误的反馈。

Ranjan Parekh
Jadavpur University