目录

1

第一章　介绍

图 1-1　人类视觉系统可以毫无问题地解释这张照片中半透明和阴影的细微变化，并从背景中正确分割出物体。

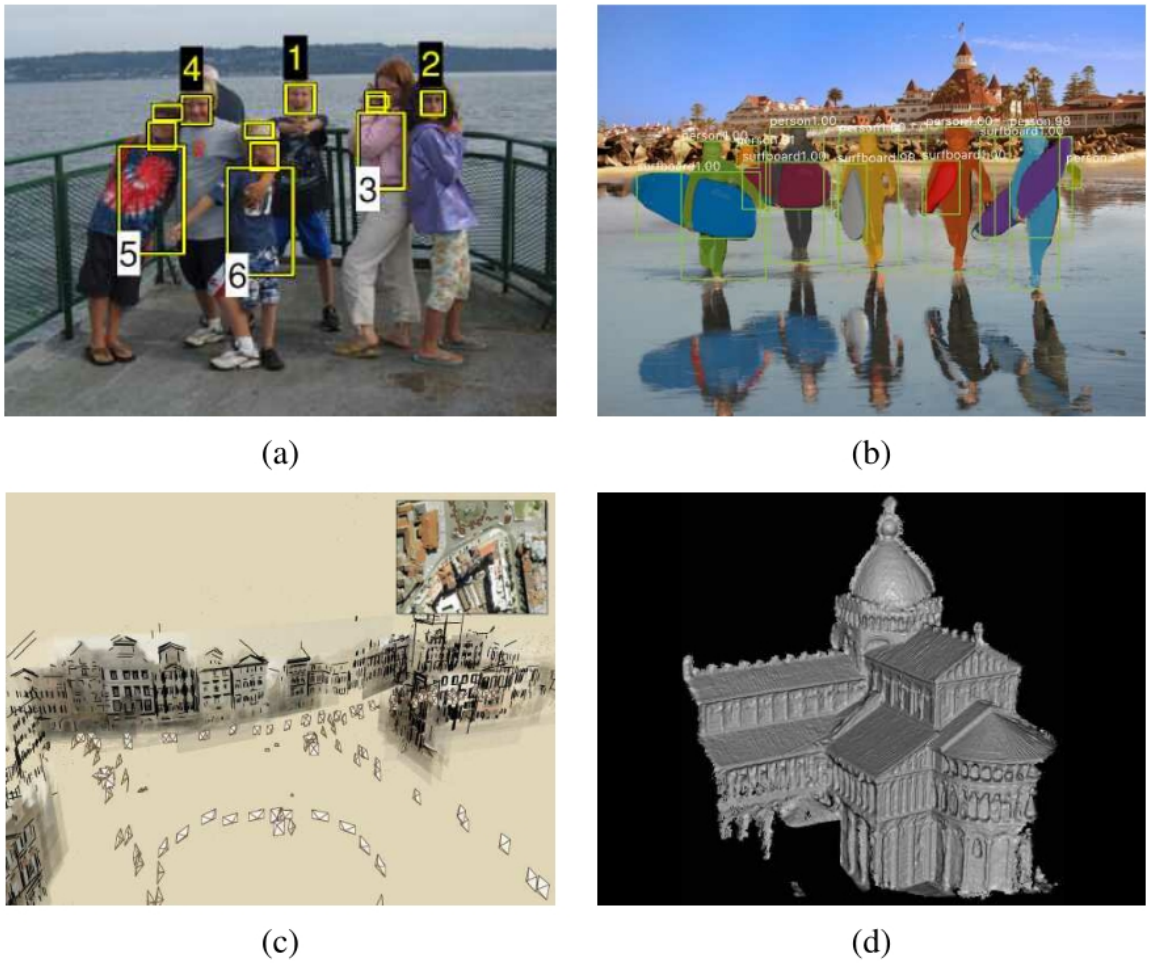


图 1-2 计算机视觉算法与应用的一些例子。(a) 人脸检测算法，结合基于颜色的服装和头发检测算法，可以定位和识别该图像中的个人 (西维奇，齐特尼克，和谢利斯基 2006) ©2006 施普林格。(b) 对象实例分割可以描绘复杂场景中的每个人和对象 (何，吉卡萨里等 2017) 2017 IEEE。(c) 运动算法的结构可以从数百张部分重叠的照片中重建大型复杂场景的稀疏 3D 点模型。(斯内夫利，塞兹，和谢利斯基 2006) © 2006 ACM。(d) 立体匹配算法可以从互联网上拍摄的数百张不同曝光的照片中建立建筑物正面的详细 3D 模型。(格塞勒，斯内夫利等 2007) © 2007 IEEE。

1.1　什么是计算机视觉?

作为人类，我们很容易感知周围世界的三维结构。想想看，当你看着旁边桌子上的一瓶花时，三维感觉是多么生动。你可以通过花瓣表面微妙的光线和阴影模式来判断每个花瓣的形状和半透明性，并毫不费力地从场景背景中分割出每个花朵 (图 1.1)。看着一张裱好的团体照，你可以很容易地叫出照片中所有人的名字，甚至可以从他们的面部表情猜测他们的情绪。(图 1.2a)。感知心理学家花了几十年时间试图理解视觉系统是如何工作的，尽管他们可以设计光学图像[①]来梳理其一些原理 (图 1.3)，但这个难题的完整解决方案仍然难以捉摸 (马尔 1982; 万德尔 1995; 帕尔默 1999; 利文斯通 2008; 弗里斯比和斯通 2010)。

计算机视觉领域的研究人员一直在并行开发数学技术，用于恢复图像中物体的三维形状和外观。在这方面，过去二十年的进展一直很快。我们现在有了可靠的技术，可以从成千上万张部分重叠的照片中精确计算出环境的 3D 模型 (图 1.2c)。给定特定对象或立面的足够大的视图集，我们可以使用立体匹配创建精确的密集 3D 表面模型 (图 1.2d)。我们甚至可以适度成功地描绘出照片中的大多数人和物体 (图 1.2a)。虽然取得了现有的这些进步，但是让一台计算机像一个两岁的孩子一样用同样的细节和因果关系解释一幅图像的梦想仍然难以实现。

视觉为什么这么难？在某种程度上，这是因为它是一个反问题，在这个问题中，我们试图在给定信息不足以完全指定解决方案的情况下恢复一些未知量。因此，我们必须借助基于物理的概率模型，或者从大量例子中进行机器学习，来消除在解决方案之间潜存的歧义。然而，用视觉世界丰富的复杂性来建模要比对产生声音的声道建模困难得多。

我们在计算机视觉中使用的前向模型通常是在物理 (辐射测量、光学和传感器设计) 和计算机图形学中开发的。这两个领域都模拟了物体如何运动和活动，光线如何从物体表面反射，如何被大气散射

[①] 一些有明显视错觉的有趣页面，包括：https://michaelbach.de/ot,https://www.illusionsindex.org/, 和http://www.ritsumei.ac.jp/~akitaoka/index-e.html

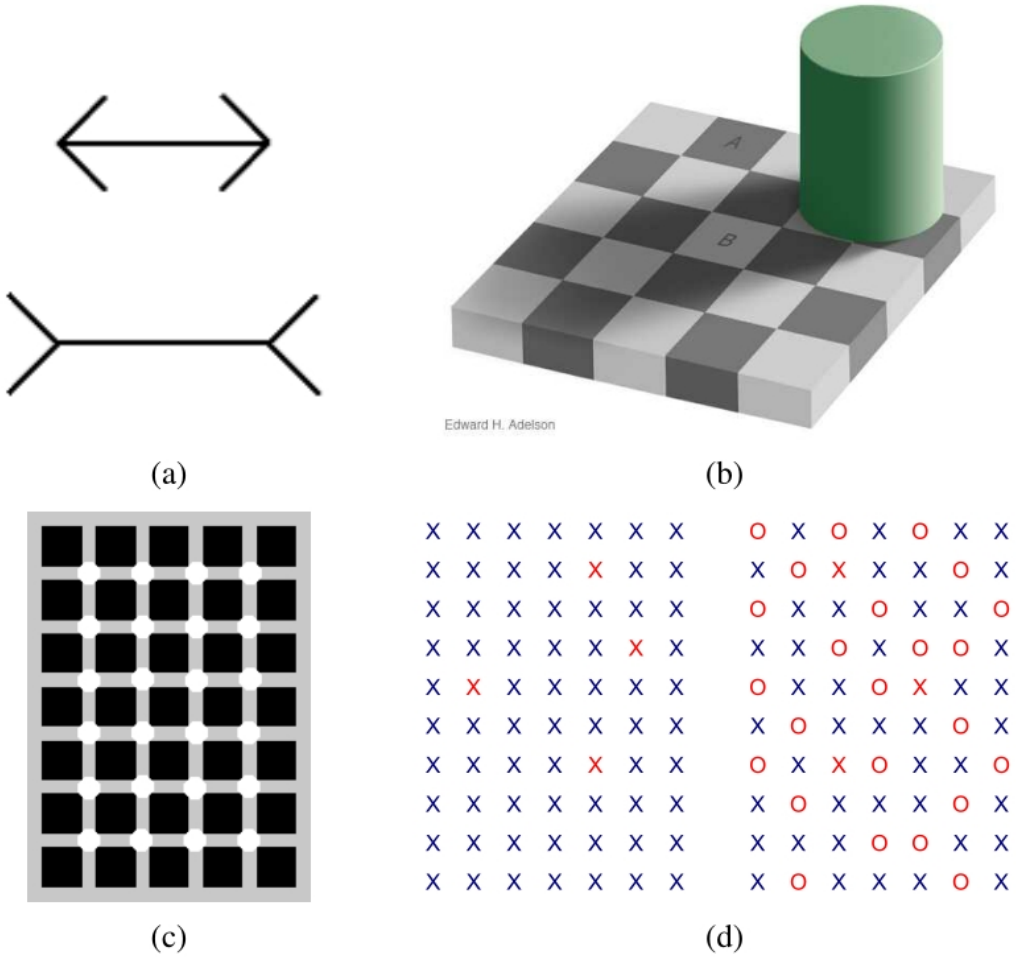


图 1-3 一些常见的视错觉以及它们能够告诉我们的视觉系统:(a) 经典的穆勒-莱尔错觉，其中两条水平线的长度看起来不同，可能是由于想象中的透视效果。(b) 阴影中的“白色”方块 B 和光线中的“黑色”方块 A 实际上具有相同的绝对强度值。这种感觉是由于亮度恒常性，视觉系统在解释颜色时试图忽略照明。图片由泰德·埃德森提供，http://persci.mit.edu/gallery/checkershadow。(c) 赫尔曼网格错觉的变体，由哈尼·法里德提供。当你把眼睛移过这个图形时，灰色的斑点出现在交叉点上。(d) 统计图左半部分红色的 X。现在在右边一半统计他们。是不是明显更难？这种解释与一种突出效应有关 (特雷斯曼 1985)，这种效应告诉我们大脑中平行感知和整合路径的运作。

30 通过照相机镜头 (或人眼) 折射，最后投射到平面 (或曲面) 图像平面上。虽然计算机图形学还不
31 完善，但在许多领域，比如渲染由日常物体组成的静止场景，或者制作恐龙等灭绝生物的动画，错

觉的真实性在本质上是存在的。

在计算机视觉中，我们试图做相反的事情，即描述我们在一幅或多幅图像中看到的世界，并重建其属性，如形状、照明和颜色分布。令人惊讶的是，人类和动物毫不费力地做到了这一点，而计算机视觉算法却很容易出错。没有在这个领域工作过的人经常低估这个问题的难度。这种认为视觉应该很容易的误解可以追溯到人工智能的早期 (见 1.2 节)，当时人们最初片面地认为智能的认知 (逻辑证明和规划) 本质上比感知部分更难 (博登 2006)。好消息是，计算机视觉如今被广泛应用于现实世界，包括:

光学字符识别 (OCR): 读取信件上的手写邮政编码 (图 1.4a) 和自动车牌识别 (ANPR)；

机器检查: 快速零件检查以保证质量，使用立体视觉和专用照明测量飞机机翼或汽车车身零件的公差 (图 1.4b)，或使用 x 光寻找钢铸件中的缺陷；

零售: 自动结账通道和全自动商店的对象识别 (温菲尔德 2019)；

仓库提货： 包裹的自主递送和托盘-携带“驱动器”(圭佐 2008 年)；奥布莱恩 2019 年)；

医学影像: 记录手术前和手术中的影像 (图 1.4d) 或对人随着年龄增长的大脑形态进行长期研究；

自驾车: 能够在城市之间点对点行驶 (图 1.4e 米勒，坎贝尔等人，2008 年；蒙特默罗，贝克尔等人，2008 年；乌姆森，安哈尔特等人，2008 年；贾奈，居内伊等人，2020 年) 以及自主飞行 (考夫曼，格赫里希等人，2019 年)；

3D 模型构建 (摄影测量): 从航拍和无人机照片全自动构建 3D 模型 (图 1.4f)；

匹配移动: 通过跟踪源视频中的特征点来估计 3D 相机运动和形状环境，从而将计算机生成的图像 (CGI) 与实时动作镜头合并，这样的手法在好莱坞被广泛使用 (如在《侏罗纪公园》等电影中)(罗布莱 1999；罗布莱和扎夫拉 2009)；它们还需要使用精确的抠图在前景和背景元素之间插入新元素 (庄，阿加瓦拉等人，2002)。

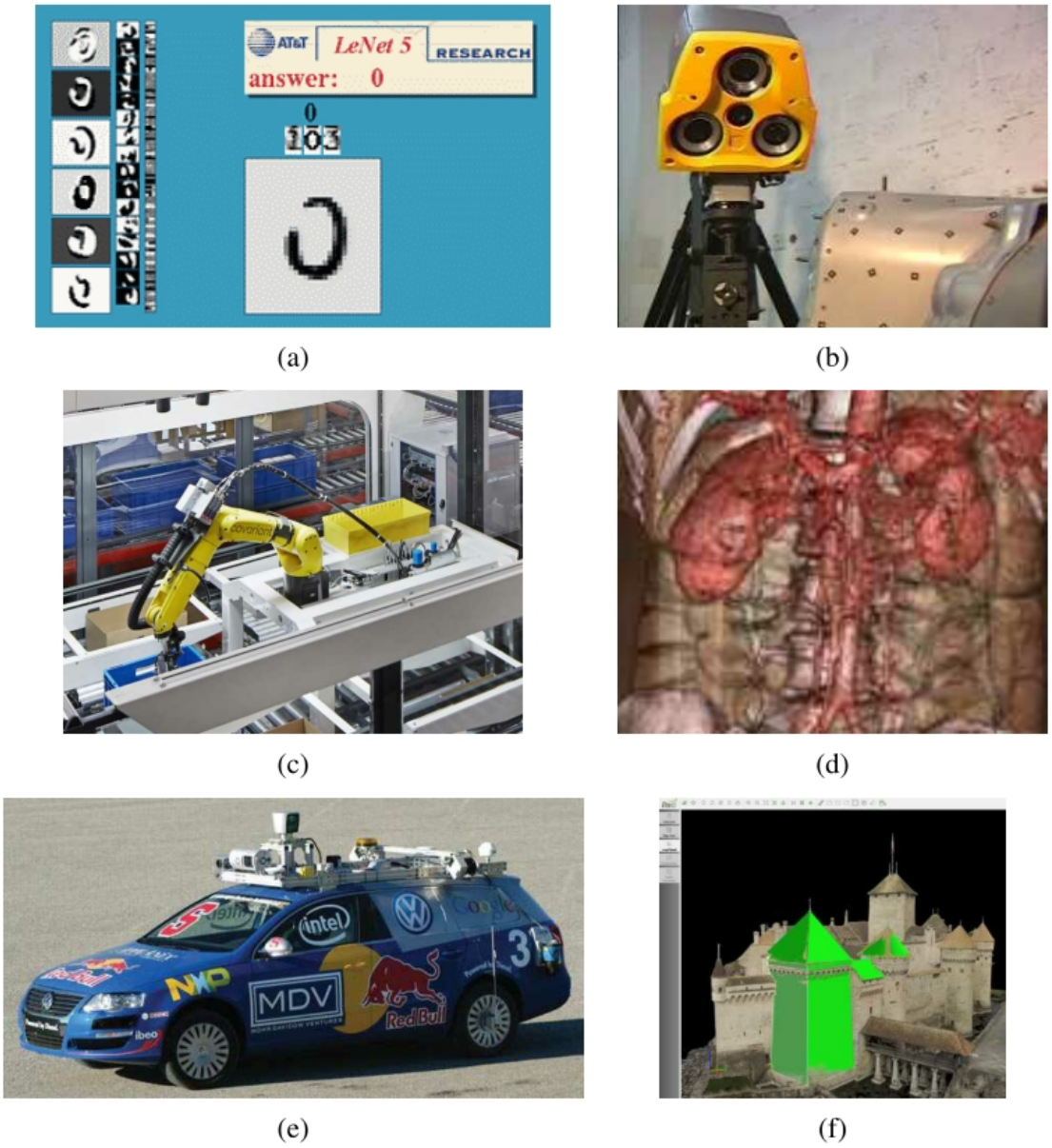


图 1-4 计算机视觉的一些工业应用：(a) 光学字符识别 (OCR)，http://yann.lecun.com/exdb/lenet/; (b) 机械检验，http://www.cognitens.com/;(c) 仓库提货 https://covariant.ai/;(d) 医疗成像 http://www.clarontech.com/;(e) 自动驾驶汽车，(蒙特梅洛，贝克尔等人，2008 年)2008 年威利；(f) 基于无人机的摄影测量，https://www.pix4d.com/blog/mapping-chillon-castle-with-drone

62 **动作捕捉 (mocap):** 使用从多个摄像机或其他基于视觉的技术观察到的回射标记来捕捉演

员，用于计算机动画；

监测: 监测溺水受害者，分析道路交通和监测溺水受害者 (例如，https://swimeye . com/)；

指纹识别和生物识别: 用于自动访问认证和取证应用。

大卫・劳的工业视觉应用网站 (http://www.cs.ubc.ca/spider/lowe/・vision.html) 列出了计算机视觉的许多其他有趣的工业应用。虽然上述应用都非常重要，但它们大多属于相当专业的图像种类和狭窄的领域。

除了所有这些工业应用程序之外，还有无数的消费者级应用程序，例如您可以用自己的个人照片和视频做的事情。其中包括:

拼接: 将重叠的照片变成单一无缝拼接的全景图 (图 1.5a)，如 8.2 节所述；

曝光包围: 如第 10.2 节所述，将在具有挑战性的光照条件 (强烈的阳光和阴影) 下拍摄的多重曝光的图像合成一张完美曝光的图像 (图 1.5b)；

变形: 使用无缝的变形过渡，将你的一个朋友的照片变成另一个朋友的照片 (图 1.5c)；

3D 建模: 将一个或多个快照转换为您正在拍摄的对象或人的 3D 模型 (图 1.5d)，如第 13.6 节所述

视频匹配移动和稳定: 通过自动跟踪附近的参考点 (参见第 11.4.4 节)[②]或使用运动估计来消除视频中的抖动 (参见第 9.2.1 节)，将 2D 图片或 3D 模型插入到您的视频中；

基于照片的漫游: 通过在不同的 3D 照片之间飞行来浏览大量的照片，例如你房子的内部 (见第 14.1.2 和 14.5.5 节)；

人脸检测: 为了改善相机聚焦以及更相关的图像搜索 (见第 6.3.1 节)；

视觉认证: 当家庭成员坐在网络摄像头前时，自动将他们登录到您的家用电脑上 (参见第 6.2.4 节)。

这些应用的伟大之处在于，它们已经为大多数学生所熟悉；它们至少是学生可以立即欣赏并在他们自己的个人媒体上使用的技术。由于计算机视觉是一个具有挑战性的课题，考虑到所涉及的数学范围很广[③]以及所要解决的问题本质上很难，因此有乐趣和相关的问题去研究会非常激励人心。

[②] 关于这个主题的有趣的学生项目，请参阅 http://www.cc.gatech.edu/dvfx/videos/dvfx2005.html 的“摄影书”项目。

[③] 这些技术包括物理、欧几里得和射影几何、统计学和最优化。它们使计算机视觉成为一个引人入胜的研究领域，也是学习广泛应用于其他领域的技术的好方法。

这本书之所以强烈关注应用程序的另一个主要原因是，它们可以用来描述和约束视觉中潜在的开放性问题。因此，与其抓住你可能听说过的第一个技巧，不如从手头的问题回想合适的技巧。这种从问题到解决方案的工作是视觉研究的典型工程方法，反映了我自己在这个领域的背景。

首先，我想出一个详细的问题定义，并决定问题的约束和规范。然后，我试着找出哪些技术是已知有效的，实现其中的一些，评估它们的性能，最后做出选择。为了使这一过程发挥作用，重要的是要有真实的测试数据，既有可用于验证正确性和分析噪声敏感度的合成数据，也有系统最终使用方式的真实数据。如果使用机器学习，更重要的是要有足够数量的代表性无偏训练数据，以便在真实世界的输入上获得良好的结果。

然而，这本书不仅仅是一个工程文本 (食谱的来源)。它还采用科学的方法来解决基本的视力问题。在这里，我试图提出手头系统物理的最佳模型: 场景是如何创建的，光如何与场景和大气效应相互作用，以及传感器如何工作，包括噪声和不确定性的来源。接下来的任务是试图颠倒采集过程，以给出场景的最佳描述。

这本书经常使用统计学的方法来阐述和解决计算机视觉问题。在适当的情况下，使用概率分布来模拟场景和有噪声的图像采集过程。先验分布与未知的关联通常被称为贝叶斯建模 (附录 B)。有可能将风险或损失函数与错误估计答案联系起来 (第二节)，并设置您的推理算法以最小化预期风险。(考虑一个机器人试图估计到障碍物的距离: 低估通常比高估更安全。) 通过统计技术，收集大量训练数据来学习概率模型通常会有所帮助。最后，统计方法使您能够使用成熟的推理技术来估计最佳答案 (或答案的分布)，并量化结果估计中的不确定性。

因为计算机视觉的很多内容都涉及到反问题的求解或者未知量的估计，所以我的书也非常强调算法，尤其是那些已知在实践中效果很好的算法。对于许多视觉问题，很容易得出问题的数学描述，要么不符合现实世界的条件，要么不适合对未知量的稳定估计。我们需要的是对噪声和模型偏差都很稳健的算法，以及在运行时资源和空间方面相当高效的算法。在这本书里，我详细讨论了这些问题，在适用的情况下，使用贝叶斯技术来确保鲁棒性，以及有效的搜索、最小化和线性系统求解算法来确保效率。[④]这本书里描述的大多数算法都是高水平的，大多是学生或通过阅读其他地方更详细的描述来填写的步骤列表。事实上，许多算法都是在练习中概述的。

既然我已经描述了这本书的目标和我使用的框架，我将在本章的剩余部分讨论另外两个主题。第 1.2 节是计算机视觉历史的简要概述。对于那些想接触到本书新材料的“肉”，而不太关心谁在什么时候发明了什么的人来说，这很容易被忽略。第二部分是对这本书内容的概述，第 1.3 节，这对于每个打算研究这个主题的人来说都是有用的阅读材料 (或者中途跳过去，因为它描述了章节的相互依赖性)。这份大纲对于希望围绕这一主题构建一门或多门课程的教师也很有用，因为它提供了基于这本书内容的示例课程。

[④] 在某些情况下，深度神经网络也被证明是加速以前依赖于迭代的算法的有效方法 (陈，徐和科尔屯 2017)。

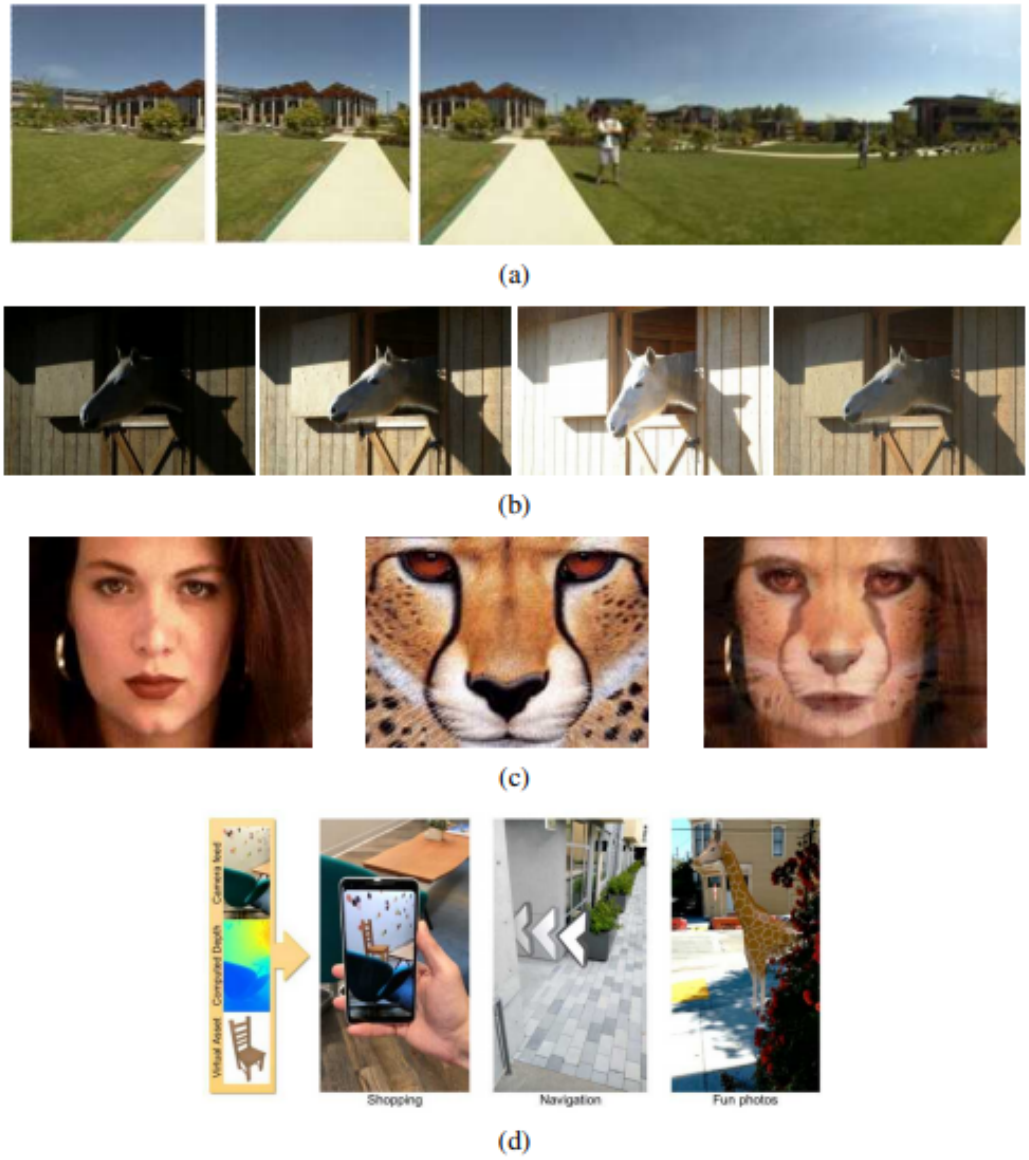


图 1-5　计算机视觉的一些消费者应用:(a) 图像拼接: 融合不同的视图 (Szeliski 和 Shum 1997)1997 ACM；(b) 风险包围: 合并不同的风险；变形: 两张照片之间的混合 (Gomes，Darsa 等人，1999 年)1999 年摩根・考夫曼；(d) 显示实时深度遮挡效果的智能手机增强现实 (Valentin，Kowdle 等人，2018 年)2018 年 ACM。

1.2 简史

在这一部分，我提供了过去五十年来计算机视觉主要发展的简要个人简介 (图 1.6)；至少，那些我个人觉得有趣的，而且似乎经得起时间考验的。对各种不同想法的出处以及这个领域的发展不感兴趣的读者应该跳到第 1.3 节的书籍概述。

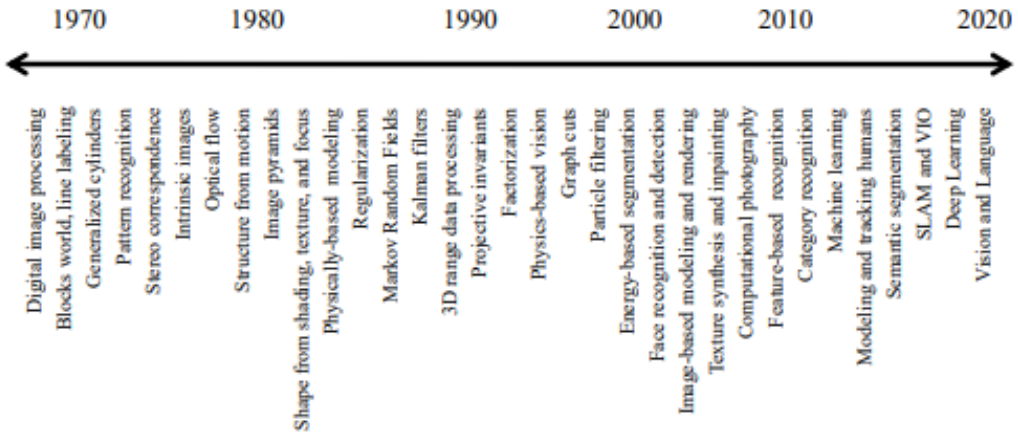


图 1-6 计算机视觉中一些最活跃的研究主题的大致时间表

1970s. 当计算机视觉在 20 世纪 70 年代初首次出现时，它被视为模拟人类智能并赋予机器人智能行为的雄心勃勃的议程的视觉感知组成部分。当时，一些人工智能和机器人学的早期先驱 (在麻省理工学院、斯坦福和 CMU 等地) 认为，解决“视觉输入”问题将是解决更高层次推理和规划等更困难问题的简单步骤。根据一个广为人知的故事，1966 年，麻省理工学院的马文・明斯基让他的本科生杰拉德・让伊・萨斯曼“花一个夏天的时间把照相机和计算机连接起来，让计算机来描述它所看到的东西”(博登 2006，第 781 页)。[5]我们现在知道这个问题比那个稍微难一点。[6]

什么使计算机视觉区别于已经存在的数字图像处理领域 (罗森菲尔德和普法茨 1966；罗森菲尔德和卡卡 (1976) 渴望从图像中恢复世界的三维结构，并以此作为充分理解场景的垫脚石。温斯顿 (1975) 和汉森和里斯曼 (1978) 提供早期的两篇经典论文。

场景理解的早期尝试包括提取边缘，然后从 2D 线的拓扑结构推断出物体或“块世界”的 3D 结构 (罗伯特 1965)。当时开发了几种线标注算法 (图 1.7a)(Huffman 1971；Clowes 1971 华尔兹 1975；罗森菲尔德，胡梅尔和扎克 1976；Kanade 1980)。纳尔瓦 (1993) 给出了这个领域的一个很好的评论。边缘检测也是一个活跃的研究领域；在 (戴维斯 1975) 中可以找到对同时代作品的一个很好的调查。

[5] 博登 (2006) 引用 (克雷维尔 1993) 作为原始来源。真正的愿景备忘录是由西蒙・派珀特 (1966 年) 撰写的，涉及到一整群学生。

[6] 要了解机器人视觉在过去 60 年里取得了多大的进步，请看一下波士顿动力公司 https://www.bostondynamics.com/,・斯凯迪奥・https://www.skydio.com/, 和 https://covariant.ai/协变公司网站上的一些视频。

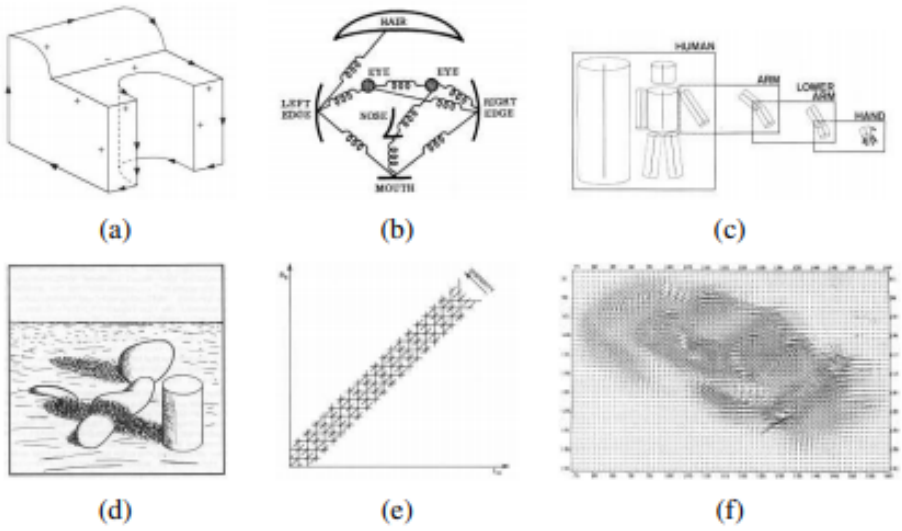


图 1-7　计算机视觉算法的一些早期 (20 世纪 70 年代) 例子:(a) 线条标记 (Nalva 1993)1993 Addison-Wesley, (b) 图像结构 (FischLerandelschlager 1973)1973 IEEE, (c) 关节体模型 (Marr 1982) 1982 David Marr，(d) 内在图像 (Barrow 和 Tenebaum 1981)1973 IEEE，(e) 立体对应 (Marr 1982) 1982 David Marr,(f) 光流 (Nagel 和 Enkelmann 1986) 1986 年 IEEE。

非多面体物体的三维建模也正在研究中 (鲍姆加特 1974；贝克 1977)。一种流行的方法使用广义圆柱体，即旋转体和扫掠闭合曲线 (阿金和宾福特 1976；内瓦蒂亚和宾福特，1977)，经常被安排成部分关系[7](辛顿，1977；Marr 1982)(图 1.7c)。菲舍尔和埃尔施拉格 (1973) 称这种弹性排列的零件为绘画结构 (图 1.7b)。

巴罗和特南鲍姆 (1981) 在他们关于内在图像的论文 (图 7 1.7d) 中, 连同马尔 (1982) 的相关 21/2-D 草图思想, 提出了一种定性的方法来理解强度和阴影变化，并通过图像形成现象 (如表面方向和阴影) 的影响来解释它们。这种方法经历了周期性的复兴，例如在塔彭、弗里曼和埃德森 (2005) 以及巴伦和马利克 (2012) 的工作中。

当时还开发了更多的计算机视觉定量方法，包括许多基于特征的立体对应算法中的第一个 (图 1.7e)(Dev 1974；Marr 和 Poggio 1976，1979；巴纳德和菲舍尔 1982；Ohta 和 Kanade 1985 格里姆森 1985；波拉德、梅休和弗里斯比 1985) 和基于强度的光流算法 (图 1.7f)(霍恩和舒克 1981；黄 1981；卢卡斯和卡纳德 1981；Nagel 1986)。同时恢复 3D 结构和相机运动的早期工作 (见第 11 章) 也是在这个时候开始的 (乌尔曼 1979；朗格特-希金斯 1981)。

[7] 在机器人学和计算机动画中，这些相连的零件图通常被称为运动链。

大卫・马尔 (1982) 的书总结了当时视觉是如何工作的许多哲学。[8]特别是，马尔介绍了他对 (视觉) 信息处理系统的三个描述层次的概念。根据我自己的解释，这三个层次是:

计算理论: 计算 (任务) 的目标是什么，已知的或能对问题产生影响的约束是什么?

表示和算法: 如何表示输入、输出和中间信息，以及使用哪些算法来计算期望的结果?

硬件实现: 如何将表示和算法映射到实际的硬件上，例如生物视觉系统或一块专用的硅? 反过来，如何用硬件约束来指导表示和算法的选择? 随着图形芯片 (GPU) 和多核架构在计算机视觉中的日益使用 (见第二节)，这个问题再次变得非常相关。

正如我在前面的介绍中提到的，我坚信，对图像形成和先验 (科学和统计方法) 的问题规范和已知约束的仔细分析必须与高效和鲁棒的算法 (工程方法) 相结合，以设计成功的视觉算法。因此，马尔的哲学似乎和 25 年前一样，可以很好地指导我们构建和解决当今领域的问题。

1980s. 80 年代在 20 世纪 80 年代，许多注意力集中在用于执行定量图像和场景分析的更复杂的数学技术上。

图像金字塔 (见 3.5 节) 开始被广泛用于执行任务，如图像混合 (图 1.8a) 和粗到细的对应搜索 (罗森菲尔德 1980; 罗森菲尔德 1984; Quam 1984Anandan 1989)。使用尺度空间处理概念的金字塔的连续版本也被开发出来 (维特金 1983; Witkin, Terzopoulos, 和 Kass 1986Lindeberg 1990)。在 20 世纪 80 年代后期，小波 (见第 3.5.4 节) 开始在一些应用中取代或增加规则的图像金字塔 (Mallat 1989Simoncelli 和 Adelson 1990aSimoncelli，Freeman 等人，1992 年)。

立体声作为定量形状线索的使用被广泛的各种形状 from- x 技术所扩展，包括阴影形状 (图 1.8b)(见章节 13.1.1 和 Horn)1975;Pentland 1984; 布莱克、齐塞尔曼和诺尔斯，1985 年; 霍恩和布鲁克斯 1986 年，1989 年)，光度立体声 (见 13.1.1 节和 Woodham 1981)，纹理形状 (见 13.1.2 节和 Witkin 1981;Pentland 1984; 马利克和罗森霍尔茨，1997 年) 和 shape from 重点 (见第 13.1.3 节和 Nayar, Watanabe 和 Noguchi 1995)。霍恩 (1986) 有一个漂亮的讨论这些技术中的大多数。

在这一时期，对更好的边缘和轮廓检测 (图 1.8c)(见 7.2 节) 的研究也很活跃 (Canny 1986;Nalwa 和 Binford 1986 年)，包括引入动态进化的轮廓跟踪器 (第 7.3.1 节)，如蛇 (Kass, Witkin, and Terzopoulos 1988 年)，以及基于物理的三维模型 (1.8d)(Terzopoulos,Witkin,Kass, 1987;Kass, Witkin 和 Terzopoulos 1988;Terzopoulos 和 Fleischer1988)。

研究人员注意到，很多立体声、流、X 形变和边缘检测算法可能是统一的，如果他们冒充变分优化问题他们至少是使用相同的数学框架，使得更稳定更正规化（ 图 1.8 e)(参见 4.2 节和 Terzopoulos 1983; 博吉奥，托瑞和科赫 1985 年;Terzopoulos 1986 b;Blake 和 Zisserman 1987; 伯泰罗、博乔和托瑞 1988;Terzopoulos 1988)。大约在同一时间，german 和 german(1984) 指出，这些问题同样可以用离散马尔科夫随机场 (MRF) 模型 (见 4.3 节) 来很好地表述，这使得可以使用更好的 (全局的) 搜索和优化算法，如模拟热处理。

在线变异的 MRF 算法中，使用卡尔曼滤波的建模和更新不确定性是稍后引入的 (Dickmanns and

[8] 视觉感知理论的最新发展包括在 (万德尔 1995; Palmer 1999Livingstone 2008 弗里斯比和斯通 2010)。

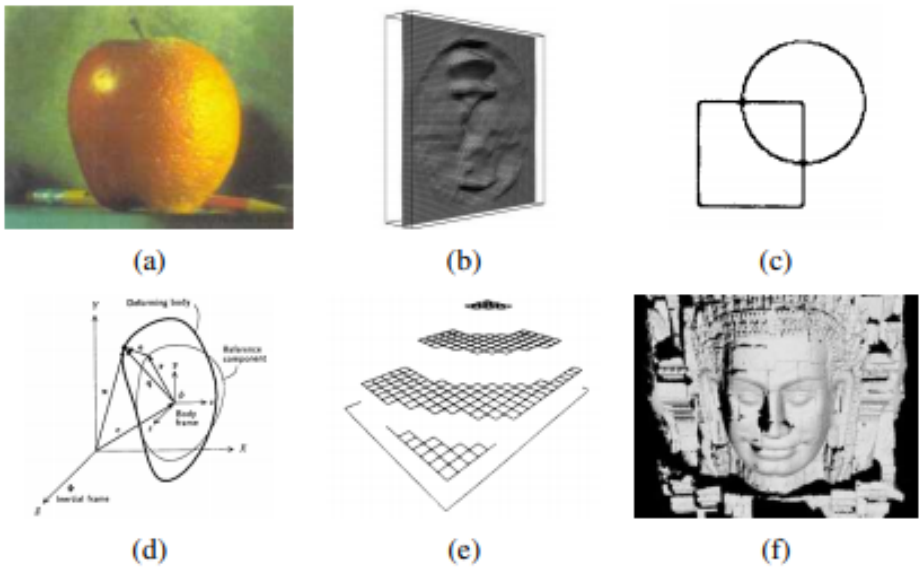


图 1-8　20 世纪 80 年代计算机视觉算法的例子:(a) 金字塔混合 (伯特和阿德尔森 1983b) 1983 ACM，(b) 阴影形状 (弗里曼和阿德尔森 1991) 1991 IEEE，(c) 边缘检测 (弗里曼和阿德尔森 1991) 1991 IEEE，(d) 基于物理的模型 (Terzopoulos 和 Witkin 1988) 1988 IEEE，(e) 基于正则化的表面重建 (Terzopoulos 1988) 1988 IEEE，(e)1988。

Graefe 1988;Matthies, Kanade 和 Szeliski 1989 年;Szeliski 1989)。还尝试绘制正规化和并行硬件上的 MRF 算法。(Poggio and Koch 1985;Poggio, Little 等 1988;Fischler, Firschein 等人 1989 年)。Fischler 和 Firschein(1987) 所著的书包含了一系列聚焦于所有这些主题 (立体声、流、正规化、MRFs，甚至更高层次的视觉) 的文章。

三维距离数据处理 (采集、合并、建模和识别；(见图 1.8f) 在这十年中继续积极探索 (Agin 和 Binford)1976;1985 年，Besl 和 Jain;Faugeras 和 Hebert 1987;Curless 和 Levoy 1996 年)。Kanade(1987) 的汇编包含了许多这方面有趣的论文。

1990s. 虽然继续探讨前面提到的许多问题，但其中有几个问题变得更加活跃。

使用投影不变量进行识别的活动 (Mundy 和 Zisserman1992 年) 演变为从运动问题解决结构问题的共同努力 (见第 11 章)。很多最初的活动都是针对投影重建的并且是不需要相机校准知识的 (Faugeras 1992; Hartley, Gupta, and Chang1992; Hartley 1994a; Faugeras and Luong 2001; Hartley and Zisserman 2004)。与此同时，因式分解技术 (第 11.4.1 节) 被发展来有效地解决了适用于正投影相机近似法的问题 (图 1.9a) (Tomasi andKanade 1992; Poelman and Kanade 1997; Anandan and Irani 2002), 然后延长视角情况 (Christy and Horaud 1996; Triggs 1996)。最终，该领域开始使用全局优化，(见 11.4.2 节 and Taylor, Kriegman, and Anandan 1991;Szeliski and Kang 1994; Azarbayejani and Pentland 1995) 此种优化

在后来被人们认为与传统摄影测量中使用的束式平差技术相同 (Triggs, McLauchlan et al. 1999)。使用此技术建立了全自动（稀疏）3D 建模系统 (Beardsley, Torr, and Zisserman 1996; Schaffalitzky and Zisserman2002; Brown and Lowe 2003; Snavely, Seitz, and Szeliski 2006; Agarwal, Furukawa et al.2011; Frahm, Fite-Georgel et al. 2010)。

工作开始于 20 世纪 80 年代，使用详细的测量颜色和强度结合，通过精确的物理模型，辐射传输和彩色图像形成自己的创造子领域被称为基于物理的视觉。关于这个题目的三卷本合集可以找到关于这个领域的一份很好的概览 c (Wolff, Shafer, and Healey 1992a; Healey and Shafer 1992;Shafer, Healey, and Wolff 1992)。

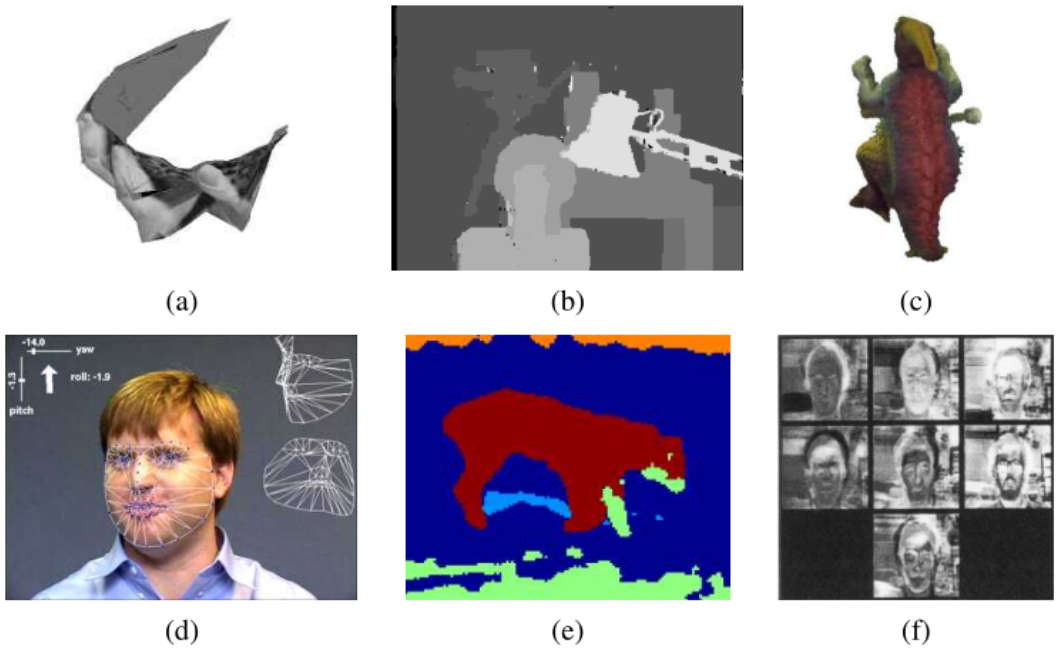


图 1-9 20 世纪 90 年代计算机视觉算法的例子。(a) 基于运动的分解结构 (Tomasi and Kanade 1992) ©1992 年施普林格；(b) 立体密集匹配 (Boykov, Veksler, and Zabih 2001)；(c) 多视角重建 (Seitz 和 Dyer1999)©1999 施普林格；(d) 人脸跟踪 (Matthews, Xiao, and Baker 2007)；(e) 图像分割 (Belongie, Fowlkes et al. 2002) ©2002 年施普林格；(f) 人脸识别 (Turk and Pentland 1991)。

通过不断地研究光流方法 (见第 9 章) 不断改进 (Nagel and Enkelmann1986; Bolles, Baker, and Marimont 1987; Horn and Weldon Jr. 1988; Anandan 1989; Bergen,Anandan et al. 1992; Black and Anandan 1996; Bruhn, Weickert, and Schnörr 2005; Papenberg, Bruhn et al. 2006) (Nagel 1986; Barron, Fleet, and Beauchemin 1994; Baker,Scharstein et al. 2011)，同样，密集立体对应算法也取得了很大的进展（见

12 章，OkutomiandKanade(1993,1994);Boykov, Veksler, and Zabih (1998); Birchfield and Tomasi (1999); Boykov, Veksler, and Zabih (2001)，以及 Scharstein 和 Szeliski(2002) 的调查和比较，其中最大的突破可能是使用图割技术的全局优化 (图 1.9b)。(Boykov,Veksler, and Zabih 2001)。

多视角立体视觉算法（图 1.9c）产生完整的 3D 表面（见第二节）也是一个活跃的研究课题 (Seitz and Dyer 1999; Kutulakos and Seitz2000)，且一直持续到今天 (Seitz, Curless et al. 2006; Schops, Schonberger et al. 2017; Knapitsch, Park et al. 2017)。制作三围体积描述的技术从二元轮廓（见 12.7.3 节）继续发展 (Potmesil 1987; Srivasan, Liang, and Hackwood 1990; Szeliski 1993; Laurentini 1994),，与基于跟踪和重建平滑一样 (见 12.2.1 节 and Cipolla and Blake 1992; Vaillant and Faugeras 1992; Zheng 1994; Boyer and Berger 1997;Szeliski and Weiss 1998; Cipolla and Giblin 2000)。

跟踪算法也得到了很大的改进，包括使用活动轮廓的轮廓跟踪（见 7.3 节），如蛇 (Kass, Witkin, and Terzopoulos 1988)，粒子过滤器 (Blake and Isard 1998)，水平集 (Malladi, Sethian, andVemuri 1995)，以及基于强度的（直接）技术 (Lucas and Kanade 1981; Shi and Tomasi 1994; Rehg and Kanade 1994)，通常用于人脸（图 1.9d）(Lanitis, Taylor, and Cootes 1997; Matthews and Baker 2004; Matthews, Xiao, and Baker 2007), 和整个身体 (Sidenbladh, Black, and Fleet 2000; Hilton, Fua, and Ronfard 2006; Moeslund, Hilton, and Krüger 2006)。

图像分割（见 7.5 节）（图 1.9e）, 一个一直活跃的主题 (Brice and Fennema1970;Horowitz and Pavlidis 1976;Riseman and Arbib 1977; Rosenfeld and Davis 1979; Haralick and Shapiro 1985; Pavlidis and Liow 1990)，最早的计算机视觉，也是一个活跃的研究主题，产生的技术基于最小能量 (Mumford and Shah 1989) 和最小描述长度 (Leclerc 1989) 归一化切割 (Shi and Malik 2000)，以及平均移位 (Comaniciu and Meer 2002)。

统计学习技术开始出现，首先是将主成分特征脸分析应用于人脸识别（图 1.9f）（见 5.2.3 节 Turk and Pentland 1991）和线性动力系统用于曲线跟踪（见 7.3.1 节 Blake and Isard 1998)。

也许在这十年中，计算机视觉最显著的发展是计算机图形学的相互交叉 (Seitz and Szeliski 1999)，特别是在基于图像的建模和渲染的交叉学科领域（见 14 章）。操纵的想法真是图像直接创建新的动画第一次来突出图像变形技术（图 1.5c）（见 3.6.3 节 Beier and Neely 1992），后来应用于视图插值（Chen and Williams 1993; Seitz and Dyer 1996），全景图像缝合（图 1.5a）（见 8.2 节 Mann and Picard 1994; Chen1995; Szeliski 1996; Szeliski and Shum 1997; Szeliski 2006a）和全光场呈现（图 1.10a）（见 14.3 节 Gortler, Grzeszczuk et al. 1996; Levoy and Hanrahan1996; Shade, Gortler et al. 1998）。同时，基于图像的建模技术（图 1.10b）自动创建逼真的 3D 模型集合的图像也被引入（(Beardsley, Torr, and Zisserman 1996; Debevec, Taylor, and Malik 1996;Taylor, Debevec, and Malik 1996)。

2000s. 这十年继续加深了视觉和图形之间的相互作用领域，但更重要的是，将数据驱动和学习方法作为愿景的核心组成部分。许多主题介绍都是基于图像的渲染，例如图像拼接（见第 8.2 节），光场捕获和渲染（见 14.3 节）以及通过曝光支架进行的高动态范围（HDR）图像捕获（图 5b）（见 10.2 节以及 Mann 和 Picard 1995 年；;Debevec 和 Malik 1997 年），被重命名为计算机摄影（见第 10 章），以承认这种技术在日常数码摄影中越来越多人使用。例如，快速采用曝光包围法来创建高

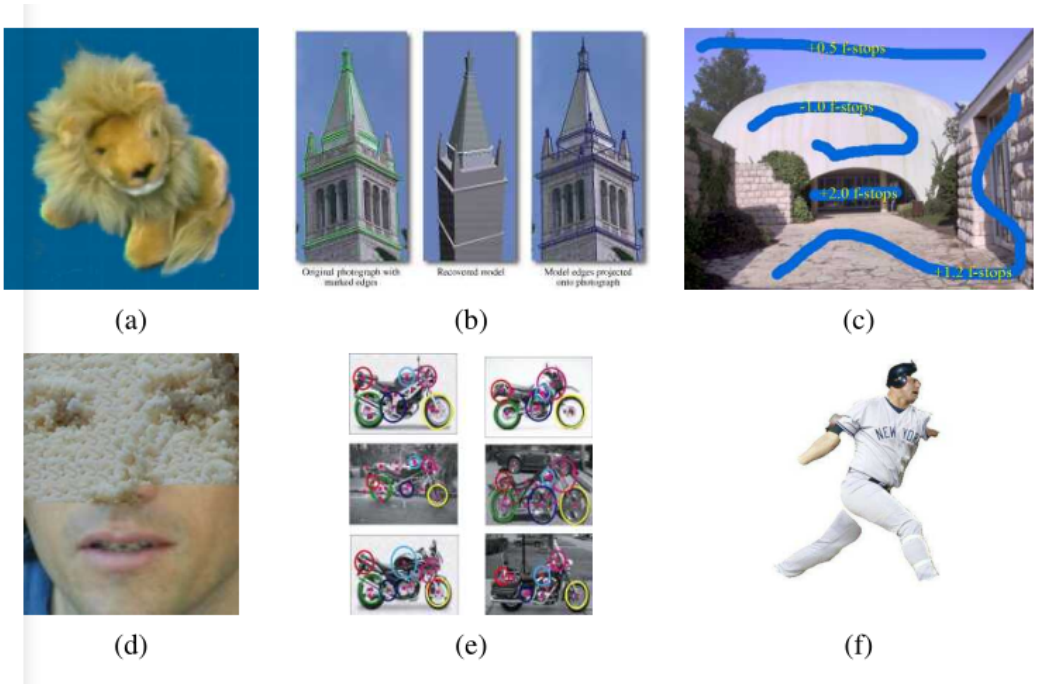


图 1-10 21 世纪初计算机视觉算法的例子:(a)基于图像的渲染 (Gortler, Grzeszczuk et al. 1996),(b)基于图像的建模 (Debevec, Taylor, and Malik 1996) © 1996 ACM, (c)交互式色调映射 (Lischinski, Farbman et al. 2006)(d)纹理合成 (Efros and Freeman 2001), (e)特征识别 (Fergus, Perona,and Zisserman 2007),(f)区域识别 (Mori, Ren et al. 2004) © 2004 IEEE。

动态范围的图像，就需要开发色调映射算法（图 1.10c）（见第 10.2.1 节）来将这些图像转换回可显示的结果 (Fattal, Lischinski, and Werman 2002; Durand and Dorsey 2002; Reinhard, Stark et al. 2002; Lischinski, Farbman et al. 2006)。除了合并多次曝光外，还开发了将闪光图像与非闪光图像合并的技术 (Eisemann and Durand 2004; Petschnigg, Agrawala et al. 2004)，并从重叠图像中交互式或自动选择不同区域 (Agarwala, Dontcheva et al. 2004)。

纹理合成（图 1.10d）（见 10.5 节），行缝 (EfrosandLeung1999;Efros and Freeman 2001; Kwatra, Schödl et al. 2003) 和修复 (Bertalmio, Sapiro et al. 2000;Bertalmio, Vese et al. 2003; Criminisi, Pérez, and Toyama 2004) 是可以归类为计算摄影技术的额外主题，因为他们重新组合输入图像样本来产生新的照片。

这十年中第二个值得注意的趋势是基于特征的技术的出现（结合学习）进行物体识别（见 6.1 节和 Ponce, Hebert et al.2006）。该领域一些著名论文包括 Fergus,Perona 和 Zisserman（2007）的星座模型（图 1.10e）和 Felzenszwalb 和 Huttenlocher（2005）的图形结构。基于特征的技术也主导了其他识别任务，如场景识别（Zhang, Marszalek et al. 2007）、全景和位置识别（Brown and Lowe 2007;

Schindler, Brown, and Szeliski 2007）。虽然兴趣点（块拼接）特征在当前研究中占主导地位，但一些群体正在追求基于轮廓（Belongie, Malik, and Puzicha 2002）和区域分割（图 1.10f）的识别（Mori, Ren et al. 2004）。

这十年的另一个重要趋势是开发更有效的算法来解决复杂的全局优化问题（见第四章和附录 B.5 以及 Szeliski, Zabih et al. 2008; Blake, Kohli, and Rother 2011b）。虽然这一趋势始于对图像分割的研究（Boykov, Veksler, and Zabih 2001; Kohli and Torr 2007），但消息传递算法方面也取得了很多进展，如环路信念传播（LBP）（Yedidia, Freeman, and Weiss 2001; Kumar and Torr 2006）。

这十年来最显著的趋势是将复杂的机器学习技术应用于计算机视觉问题（见第五章和第六章），到目前为止，它已经完全接管了视觉识别和计算机视觉的大多数其他方面。这一趋势与互联网上大量部分标记数据的可用性增加以及计算能力的显著提高相吻合，这使得在不使用仔细人类监督的情况下学习对象类别变得更加可行。

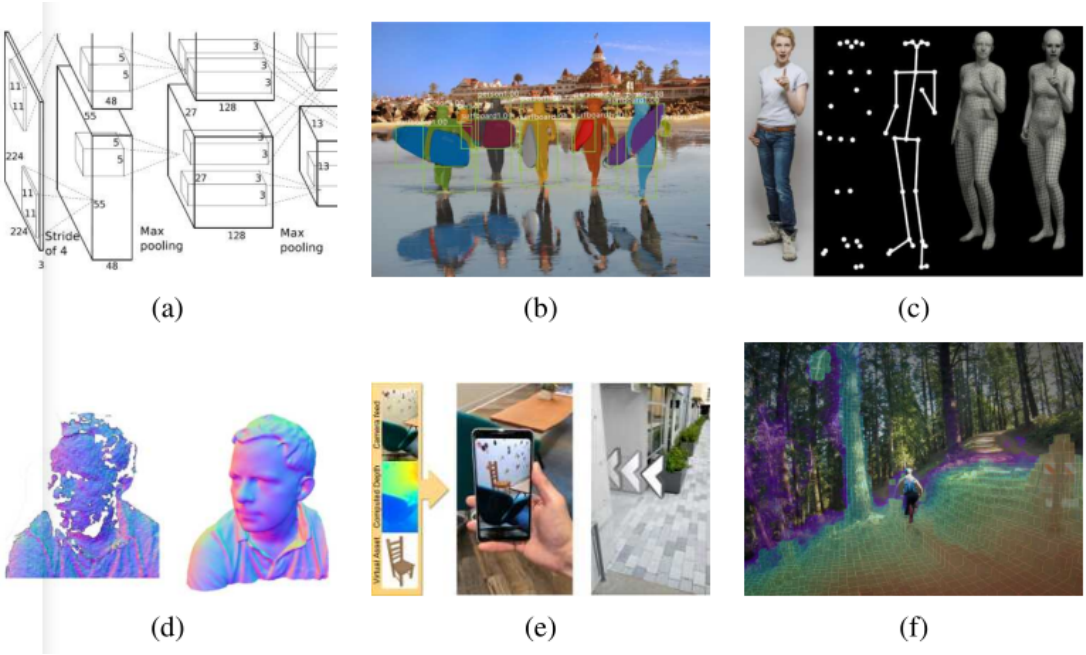


图 1-11　2010 年代计算机视觉算法的例子：（a）监督深度神经网络 ©Krizhevsky,Sutskever, and Hinton (2012)；（b）对象实例分割 (He, Gkioxari 等. 2017) © 2017 IEEE；（c）整个身体，表情和手势都符合一个单一的图像 (Pavlakos, Choutas 等 2019) © 2019 IEEE（d）使用 KinectFusion 实时系统融合多色深度图像 (Newcombe, Izadi 等 2011) © 2011 ACM；（e）智能手机增强现实与实时深度遮挡效果 (Valentin, Kowdle 等 2018) © 2018 ACM；（f）在全自动无人机上实时计算 3D 地图（跨越 2019 年）

2010s.【新文本】使用大型标记（以及自我监督）数据集来开发机器学习算法的趋势成为了一

场浪潮，彻底改革了图像识别算法以及其他应用的发展，如去噪和光流，这些以前使用贝叶斯和全局优化技术。

这一趋势是由 ImageNet（Deng, Dong 等 2009;Russakovsky, Deng 等 2015）、MicrosoCOCO（上下文的普通对象）和 LVIS（Gupta, Dollár, and Girshick 2019）等高质量大规模注释数据集的发展所推动的。这些数据集不仅为跟踪识别和语义分割算法的进展提供了可靠的指标，更重要的是，有足够多的标记数据来开发基于机器学习的完整解决方案。

另一个主要趋势是图形处理单元（GPGPU）上的通用（数据并行）算法的开发极大地提高了计算能力。突破监督（“AlexNet”）深度神经网络（图 1.11a；Krizhevsky, Sutskever, and Hinton 2012），是第一个赢得年度 ImageNet 大规模视觉识别挑战的神经网络，凭借 GPU 训练，以及一些技术进步，其戏剧性的性能。在本文发表后，深度卷积架构的使用进展显著加快，目前已成为识别和语义分割任务中唯一考虑的架构（图 1.11b），也是许多其他视觉任务的首选架构。(第五章; LeCun, Bengio, and Hinton 2015)，包括视觉流 (Sun, Yang 等 2018))，去噪和单目深度推断（Li, Dekel 等 2019）。

大型数据集和 GPU 架构，加上快速传播的思想通过及时的出版物在 arXiv 以及发展的语言深度学习和神经网络模型的开源，导致这个地区在飞速发展和功能以及大量出版物和研究人员现在讨论这些议题的一个爆炸性的增长。他们还使图像识别方法扩展到视频理解任务，如动作识别 (Feichtenhofer, Fan 等 2019) 以及结构化回归任务，如实时多人身体姿态估 (Cao,Simon 等 2017)。

专门用于计算机视觉任务的传感器和硬件也在继续发展。微软于 2010 年发布的 Kinect 深度相机，在迅速成为许多 3D 建模（图 1.11d）和个人跟踪 (Shotton, Fitzgibbon 等 2011) 系统的重要组成部分。在过去的十年中，3D 体型建模和跟踪系统不断发展，现在可以从单个图像中通过手势和表情推断出一个人的 3D 模型（图 1.11c）。

虽然深度传感器还没有普及（除了高端手机上的安全应用），但计算机摄影算法在今天的所有智能手机上都能运行。在计算机视觉领域引入的创新，如全景图像拼接和括号内的高动态范围图像合并现在是标准特征，多图像弱光去噪算法也变得普遍 (Liba, Murthy 等 2019)。光场成像算法，允许创建柔和的景深效果，现在也越来越可用 (Garg, Wadhwa 等 2019)。最后，使用特征跟踪和惯性测量组合执行实时姿态估计和环境增强的移动增强现实应用程序是常见的，目前正在扩展到包括像素精确的深度遮挡效果（图 1.11e）。

在自动驾驶汽车和无人机等高端平台上，强大的实时 SLAM（同时定位和绘图）和 VIO（视觉惯性测程）算法 (Engel, Schöps, and Cremers 2014；Forster, Zhang 等 2017；Engel, Koltun, and Cremers 2018) 可以构建精确的 3D 地图，从而实现自主飞行，通过具有挑战的场景，如森林（图 1.11f）。

总之，在过去的十年里，计算机视觉算法在性能和可靠性方面取得了令人难以置信的进步，这在一定程度上是由于转向了在非常大的显示世界数据集上进行机器学习和训练。它也看到了视觉算法的应用在无数的商业和消费场景中。

1.3　本书概述/书籍

在本书导论的最后部分，我简要介绍了这本书的内容，并提供了一些有关符号的注释和一些其他一般参考资料。由于计算机视觉是一个广阔的领域，因此可以研究它的某些方面，例如几何图像形成和 3D 结构恢复，而无需其他部分，如反射率和阴影建模。本书中的某些章节仅与其他章节松散地结合在一起，因此没有必要严格地按顺序阅读所有的材料。

图 1.12 显示了本书内容的粗略布局。由于计算机视觉涉及从图像到语义理解以及场景的 3D 结构描述，因此我根据内容在频谱中的位置水平放置了章节，并根据它们的依赖关系垂直放置了章节。[9]

在本书中穿插的是示例应用程序，这些示例程序将在各章中介绍的算法和数学材料与有用的实际应用程序相关联。在练习部分中还将介绍许多这些应用程序，以便学生自己编写。

在每个部分的结尾，我提供了一组练习，学生可以用来实现，测试和完善每个部分中介绍的算法和技术。其中一些练习适合作为书面家庭作业，其他练习适合作为为期一周的较短项目，而另一些则适合作为挑战性最终项目的开放式研究问题。到本书的结尾，完成这些练习的有动机的学生将拥有一个计算机视觉软件库，该库可用于各种有趣的任务和项目。

如果学生或课程对编程语言的偏爱不强，那么 Python，NumPy 科学和数组算术库以及 OpenCV 视觉库，是开发算法和学习视觉的理想环境。学生不仅将学习如何使用数组/张量表示法和线性/矩阵代数（这是以后使用 PyTorch 进行深度学习的良好基础）进行编程的方法，还可以使用 Jupyter 笔记本准备课堂作业，使您可以选择将描述性教程，示例代码和要扩展/修改的代码结合在一个方便的位置。[10]

作为一本参考书，我将尽可能地讨论哪些技术和算法在实践中行之有效，并提供指向我所涵盖领域的最新研究成果的最新指针。这些练习可用于建立你自己的经过自我测试和验证的视觉算法的个人库，从长远来看（假设您有时间），这比简单地将算法从您并不真正了解的库中抽出更有价值。

该书从第 2 章开始，回顾了图像形成过程，这些过程创建了我们看到并捕获的图像。如果您想采用科学的（基于模型的）计算机视觉方法，那么了解此过程至关重要。急于开始实施算法（或时间有限的课程）的学生可以跳至下一章，稍后再阅读该材料。在第二章中，我们将图像形成分为三个主要部分。几何图像形成（第 2.1 节）处理点，线和平面，以及如何使用投影几何和其他模型（包括径向透镜畸变）将它们映射到图像上。光度成像法（第 2.2 节）涵盖了辐射度测量法，它描述了光如何与世界上的各个表面相互作用，而光学器件则将光投射到传感器平面上。最后，第 2.3 节介绍了传感器的工作方式，包括诸如采样和混叠，颜色感测以及相机内压缩等主题。

第 3 章介绍了图像处理，几乎所有计算机视觉应用程序都需要图像处理。其中包括线性和非线性滤波（第 3.3 节），傅立叶变换（第 3.4 节），图像金字塔和小波（第 3.5 节）以及几何变换（例如

[9] 为了与人类视觉系统的已知情况进行有趣的比较，例如在很大程度上平行的什么和在哪里的路径 (Goodale 和 Milner 1992)，参见一些关于人类感知的教科书 (Palmer 1999Livingstone 2008 弗里斯比和斯通 2010)。

[10] 您也可以使用谷歌 Colab 服务在 https://Colab.research.Google.com/上运行您的笔记本和训练您的模型。

图像变形）（第 3.6 节）等主题。第 3 章还介绍了诸如无缝图像融合和图像变形之类的应用程序。

【新案文：概述部分的其余部分已更新】第 4 章以关于数据拟合和插值的新部分开始，该部分为全局优化技术提供了一个概念框架，如正则化和马尔可夫随机场 (MRFs) 以及机器学习，我们将在下一章中讨论。第 4.2 节涵盖了经典的正则化技术，即使用快速迭代线性系统求解器实现的分段连续平滑样条（又称变分技术），这在移动增强现实等时间关键应用中仍然是首选的方法。下一节（4.3）提出了马尔可夫随机场的相关主题，该主题也是对贝叶斯推理技术的介绍，附录 B 在更抽象的层次上介绍了贝叶斯推理技术。本章还讨论了在交互式着色和分段中的应用。

第五章是一个全新的章节，涵盖了机器学习、深度学习、深度学习和深度神经网络。它从第 5.1 节开始，回顾了经典的监督机器学习方法，这些方法旨在基于中级特征对图像（或回归值）进行分类。第 5.2 节介绍了无监督学习，这对理解无标签的训练数据和提供真实分布的模型都很有用。第 5.3 节介绍了前馈神经网络的基本要素，包括权值、层、激活函数，以及网络训练的方法。第 5.4 节详细介绍了卷积网络及其在识别和图像处理中的应用。本章的最后一节讨论了更复杂的网络，包括三维网络、时空网络、循环网络和生成网络。

第 6 章涉及识别问题。在本书的第一版中，本章是最后一章，因为它建立在早期的方法，如分割和特征匹配。随着深度网络的出现，这些中间表示中的许多不再是必要的，因为网络可以作为培训过程的一部分学习它们。由于计算机视觉研究现在致力于各种识别主题，我决定把这一章移到前面，以便学生在课程中更早地了解它。

本章从实例识别的经典问题开始，即在杂乱的场景中找到已知的 3D 对象的实例。第 6.2 节涵盖了传统和深层网络方法对整个图像分类，即过去称为类别识别的方法。还讨论了面部识别的特殊情况。第 6.3 节介绍了对象检测的算法（绘制识别对象周围的包围框），并简要回顾了以前的人脸和行人检测方法。第 6.4 节涵盖了各种类型的语义分割（生成每像素标签），包括实例分割 (描述单独的对象)、姿态估计 (用身体部分标记像素) 和泛光分割 (同时标记事物和事物）。在第 6.5 节中，我们简要地回顾了一些在视频理解和动作识别方面最近的论文，而在第 6.6 节中，我们提到了在图像字幕和视觉问答（视觉和语言）方面的一些最新工作。本章的最后一部分回顾了一些广泛使用的识别数据集和基准测试。

在第七章中，我们介绍了特征检测和匹配。当前的许多三维重建和识别技术都是建立在提取和匹配特征点之上的（第 7.1 节），因此这是随后的许多章（第 8 章和第 11 章）所需要的基本技术，甚至例如重新认知（第 6.1 节）。我们还涵盖了第 7.2 节和第 7.4 节中的边缘和直线检测，第 7.3 节中的轮廓跟踪和第 7.5 节中的低级分割技术。

第 8 章中使用特征检测和匹配来执行图像对齐（或配准）和图像拼接。我们介绍了基于特征的对齐的基本技术，并展示了如何使用线性或非线性最小二乘来解决这个问题，这取决于所涉及的运动。我们还引入了额外的概念，如不确定性加权和鲁棒回归，这对使现实系统工作至关重要。然后将基于特征的对准用作二维应用块，如图像拼接（第 8.2 节）和计算摄影（第 10 章），以及三维几何对准任务，如姿态估计和运动结构（第 11 章）。

第 8 章的第二部分是图像拼接，即大全景图和复合材料的构建。虽然缝纫只是计算摄影的一个

例子（见第 10 章），但这里有足够的深度来保证一个单独的部分。我们首先讨论各种可能的运动模型（8.2.1 节），包括平面运动和纯摄像机旋转。然后，我们讨论全局对齐（8.3 节），这是一个特殊的(简化的）情况下的一般束调整，然后提出全景识别，即自动发现哪些图像实际上形成重叠全景图的技术。最后，我们讨论了图像合成和混合的主题（8.4 节），其中包括选择哪些像素用于图像，并将它们混合在一起，以掩盖曝光差异。

图像拼接是一个很好的应用程序，将本书前面部分所涵盖的大部分材料连接在一起。它也是一个很好的中期课程项目，可以建立在以前开发的技术，如图像扭曲和特征检测和匹配。第 8.2-8.4 节还提供了更专门的拼接变体，如白板和文档扫描、视频总结、面板析、全 360° 球形全景以及用于将重复动作镜头的交互式照片混合在一起。

在第 9 章中，我们推广了基于特征的图像对齐的概念，以涵盖基于密集强度的运动估计，即光流。我们从最简单的运动模型开始，平移运动（第 9.1 节），并涵盖了分层 (粗到细）运动估计、基于傅里叶的技术和迭代细化等主题。然后，我们提出了参数运动模型，可以用来补偿相机的旋转和缩放，以及仿射或平面透视运动（9.2 节）。然后将其推广到基于样条的运动模型（9.3 节），最后推广到一般的每像素光流 (9.4 节）。我们结束了第 9.5 节的章节，讨论了分层和学习的运动模型以及视频对象的分割和跟踪。运动估计技术的应用包括自动变形、视频去噪和帧内插（慢运动）。

第 10 章介绍了计算摄影的额外例子，这是从一个或多个输入照片中创建新图像的过程，通常是基于对图像形成过程的仔细建模和校准（第 10.1 节）。计算摄影技术包括合并多个曝光以创建高动态范围图像（10.2 节），通过模糊去除和超分辨率提高图像分辨率 (10.3 节），以及图像编辑和合成操作 (10.4 节）。我们还涵盖了 10.5 节中的纹理分析、合成和修补（孔填充）主题，以及非真实感渲染和风格转移。

从第 11 章开始，我们深入研究了从图像中重建三维模型的技术。我们首先在第 11.1 节和第 11.2 节中介绍了本征摄像机校准的方法，即外部校准。这些部分还描述了建筑模型的单视图重建和三维位置识别的应用，然后我们讨论了三角剖分的主题（11.2.4 节），即当摄像机位置已知时，从匹配特征中重建点的三维重建。

然后，第 11 章继续讨论运动中的结构主题，该主题涉及从跟踪的 2D 特征集合中同时恢复 3D 摄像机运动和 3D 场景结构。我们从运动的两帧结构开始（第 11.3 节），对此存在代数技术，以及鲁棒的采样技术（如 RANSAC）可以消除错误的特征匹配。然后，我们将介绍运动中的多帧结构技术，包括因式分解（第 11.4.1 节），束调整（第 11.4.2 节）以及受约束的运动和结构模型（第 11.4.8 节）。我们为大型（例如互联网）照片集展示了视觉效果（比赛移动）和稀疏 3D 模型构建中的应用。本章的最后部分（第 11.5 节）增加了有关同时定位和地图绘制（SLAM）及其在自主导航和移动增强现实（AR）中的应用的新章节。

在第 12 章中，我们转向立体声对应的主题，这可以看作是运动估计的一个特殊情况，其中相机的位置已经知道（第 12.1 节）。这种额外的知识使立体声算法能够在更小的对应空间中进行搜索，以使用各种匹配标准、优化算法和/或深度网络的组合来产生密集的深度估计（第 12.3-12.6 节）。我们还涵盖了多视图立体算法，它构建了一个真正的三维表面表示，而不是单个深度图（第 12.7 节），

以及单眼深度推理算法，从一个图像产生深度图的幻觉 (第 12.8 节)。立体声匹配的应用包括头部和注视跟踪，以及基于深度的背景替换（ Z 键控)。

第 13 章涉及额外的三维形状和外观建模技术。这些包括经典的形状从 X 技术，如形状从阴影，形状从纹理，形状从焦点（ 第 13.1 节)。除了所有这些被动的计算机视觉技术之外，另一种方法是使用主动测距（ 第 13.2 节)，即将图案光投射到场景上，并通过三角剖分恢复三维几何。处理所有这些三维表示通常涉及插值或简化几何（ 第 13.3 节)，或使用替代表示，如表面点集 (第 13.4 节) 或隐式函数 (第 13.5 节)。

从一个或多个图像到部分或完整 3D 模型的技术集合通常称为基于图像的建模或 3D 摄影。第 13.6 节研究了另外三个专门的应用领域（ 体系结构，面部和人体)，它们可以使用基于模型的重构来将参数化模型拟合到感测到的数据。第 13.7 节探讨了外观建模的主题，即估算纹理贴图，反照率甚至有时是描述 3D 表面外观的完整双向反射分布函数（ BRDFs ）的技术。

在第 14 章中，我们讨论了在过去三十年中发展起来的大量基于图像的渲染技术，包括更简单的技术，如视图 1.4 示例大纲 29 透视 (第 14.1 节)、分层深度图像 (第 14.2 节)、精灵和层 (第 14.2.1 节)，以及更一般的光场和 Lumigraphs 框架 (第 14.3 节) 和更高阶的场，如环境遮罩 (第 14.4 节)。这些技术的应用包括使用照片旅游导航 3D 照片集合。接下来，我们讨论基于视频的渲染，这是基于图像的渲染的时间扩展。我们涵盖的主题包括基于视频的动画 (第 14.5.1 节)、将周期性视频转化为视频纹理 (第 14.5.2 节) 以及从多个视频流构建的 3D 视频 (第 14.5.4 节)。这些技术的应用包括动画静态图像和创建基于 360♀ 视频的家庭旅游。我们完成的这一章是对神经渲染这一新兴领域的概述。为了支持这本书作为教科书的使用，附录和相关网站包含更详细的数学主题和附加材料。附录 A 涵盖了线性代数和数值技术，包括矩阵代数、最小二乘法和迭代技术。附录 B 涵盖贝叶斯估计理论，包括最大似然估计、稳健统计、马尔可夫随机场和不确定性建模。附录 C 描述了可用于补充本书的补充材料，包括图像和数据集、软件指针和课程幻灯片。

1.4 教学大纲样本

在四分之一或学期的课程中教授这本书所涵盖的所有材料是一项艰巨的任务，很可能不值得尝试。最好简单地选择与讲师首选重点相关的主题，并根据为学生设想的迷你项目进行定制[11]。史蒂夫・塞茨和我已经成功地使用了为期 10 周的类似于表 1.1 的教学大纲，作为计算机视觉的本科和研究生水平的课程。本科课程[12]，倾向于数学更轻，需要更多的时间回顾基础，而研究生级课程[13]，掌握它们在 Linux 操作系统中的使用方法。更深入技术，假设学生在视觉或相关数学技术方面已经有了一个像样的基础。此外，我们还教授了有关 3D 摄影和计算摄影的相关课程。附录 C.3 和本书的

[11]一些大学，如斯坦福大学 (二硫化碳 31A231N)、伯克利大学 (CS194-294-26-26280) 和密歇根大学 (EECS498/598442)，现在将材料分为两门课程。

[12]http://www.cs.washington.edu/education/courses/576/

[13]http://www.cs.washington.edu/education/courses/455/

网站列出了使用本书教授类似课程的其他课程。当史蒂夫和我教这门课程时，我们更喜欢在课程的早期给学生几个小的编程作业，而不是专注于书面作业或测验。

表 1-1　一学期 13 周课程的教学大纲样本。一个为期 10 周的季度可能会进入较低的深度或省略一些主题。

周次	章节	主题
1.	第 1-2 章	引言和图像形成
2.	第 3 章	图像处理
3.	第 4-5 章	优化和学习
4.	第 5 章	深度学习
5.	第 6 章	识别
6.	第 7 章	特征检测与匹配
7.	第 8 章	图像拼接
8.	第 9 章	稠密运动估计
9.	第 10 章	计算摄影学
10.	第 11 章	由运动到结构
11.	第 12 章	深度估计
12.	第 13 章	3D 重建
13.	第 14 章	基于图像的绘制

有了合适的主题选择，这些项目就有可能相互构建。例如，早期引入特征匹配可以在第二个任务中使用来进行图像对齐和缝合。或者，可以使用直接（光流）技术来进行对齐，更多的焦点可以关注图形切缝选择或多分辨率混合技术。以前，我们还要求学生们在课程中途提出期末项目（我们为需要想法的人提供一组建议主题），并在课程的最后一周供学生演示。有时，其中一些项目实际上已经变成了会议提交！无论你决定如何构建课程或如何选择使用这本书，我鼓励你至少尝试一些小的编程任务来了解视觉技术是如何工作的以及它们是如何失败的。更好的是，选择那些有趣的、可以用在你自己的照片上的主题，并努力推动你的创意边界，想出令人惊讶的结果。

1.5　标记法说明

无论好坏，计算机视觉和多视图几何教科书中的符号往往在地图上有所不同（(Faugeras1993；Hartley 和 Zisserman2004；Girod，Greiner，1.6 附加阅读 31 和 Niemann2000；Faugeras 和 Luong2001；

Forsyth 和 Ponce2003)。在这本书中，我使用了我在高中物理课上第一次学到的惯例（以及后来的多元演算和计算机图形课程），即向量 v 是小写粗体，矩阵 M 是大写粗体，标量 (T，s) 是混合写斜体。除非另有说明，向量作为列向量操作，即它们的乘后矩阵 Mv，尽管它们有时被写成逗号分隔的圆括号列表 x=(x，y)，而不是括号内的列向量 $\mathbf{x} = \begin{bmatrix} x & y \end{bmatrix}^T$。一些常用的矩阵是 R 表示旋转，K 表示校准矩阵，I 表示恒等矩阵。齐次坐标（第 2.1 节）在向量上用一个 tilde 表示，例如 $\tilde{\mathbf{x}} = (\bar{x}, \tilde{y}, \tilde{w}) = \tilde{w}(x, y, 1) = \tilde{w}\bar{\mathbf{x}}$ 在 $\mathcal{P}^2$ 中。矩阵形式的交叉乘积算子用 $[]_\times$ 表示。

1.6 附加阅读说明

这本书试图是独立的，这样学生就可以实现这里描述的基本作业和算法，而不需要外部参考。然而，它确实预先预先熟悉线性代数和数值技术的基本概念，在附录 A 中回顾，以及在第 3 章中回顾的图像处理。想要深入研究这些主题的学生可以在 (Golub 和 VanLoan，1996) 中寻找矩阵代数和 (Strang，1988) 中寻找线性代数。在图像处理方面，有一些流行的教科书，包括 (Crane1997；Gomes 和 Velho1997；J¨ahne1997；Pratt2007；Russ2007；汉堡和汉堡 2008；Gonzalez 和 Woods2017)。对于计算机图形学，流行的文本包括 (Hughes，vanDam 等。2013 年；Marschner 和 Shirley，2015 年)，(Glassner，1995 年) 提供了更深入的图像形成和渲染。对于统计和机器学习，ChrisBishop（2006）的书是一个精彩而全面的介绍，有丰富的练习，而 (Murphy2012) 提供了一个更新的领域和 (Hastie，Tibshirani 和 Friedman2009) 一个更经典的治疗。深入学习的一个很好的介绍性文本是 (Glassner2018)，而 Goodfellow、Bengio 和 Cour ville（2016）和 Zhang、Lipton 等人。（2020 年）提供更全面的治疗。学生也可能希望在其他关于计算机视觉的教科书中寻找我们在这里不涉及的材料，以及其他项目想法（纳尔瓦 1993 年；1998 年哈特利和齐塞曼 2004 年；福赛斯 2011 年庞塞和庞塞；2012 年王子；戴维斯 2017 年）。

然而，这里没有任何东西可以代替阅读最新的研究文献，无论是最新的思想和技术，还是相关文献的最新参考文献[14]。在本书中，我试图引用每个领域的最新作品，以便学生可以直接阅读它们，并将其用作自己工作的灵感。从主要的视觉、图形和机器学习会议上浏览过去几年的会议记录，如 CVPR、ECCV、ICCV、SIGGRAPH 和神经科，以及关注 arXiv 上的最新出版物，将提供丰富的新想法。这些会议上提供的教程也是一种宝贵的资源，其中的幻灯片或笔记通常可以在网上获得。

[14]对于计算机视觉研究的综合书目和分类，KeithPrice 的注释计算机视觉书目 http://www.visionbib.com/bibliography/contents.html 是一个宝贵的资源

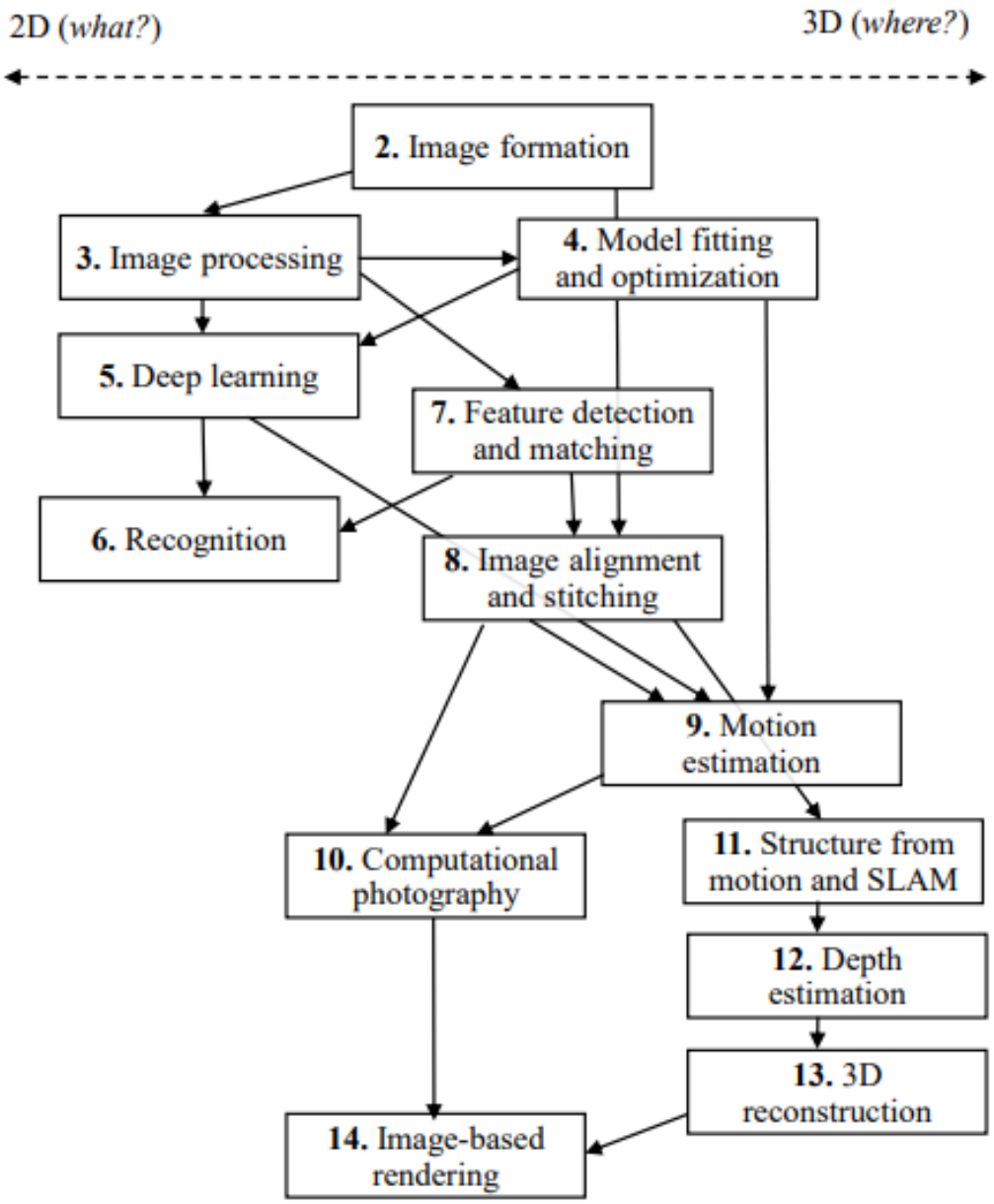
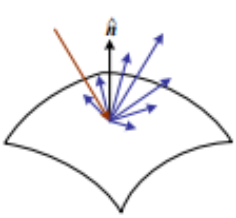


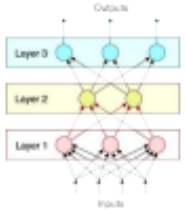
图 1-12　本书涵盖的主题的分类，显示了不同章节之间的（粗略）依赖关系，这些依赖关系沿左右轴大致定位，具体取决于它们与图像（左）还是与 3D 几何关系更紧密（正确）表示。沿顶部轴的“何处”是对视觉系统中单独的视觉路径方式的引用（Goodale 和 Milner 1992），但不应太当真。



2. Image formation

3. Image processing

4. Optimization

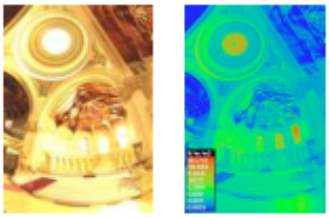


5. Deep learning

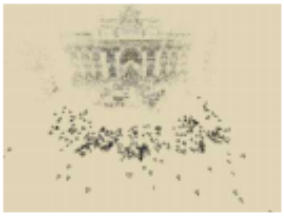
6. Recognition

7–8. Features & alignment

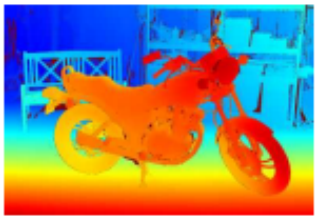
9. Motion estimation



10. Computational Photography

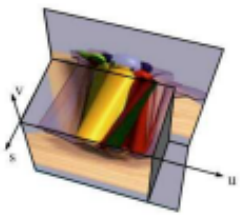


11. Structure from motion



12. Depth estimation

13. 3D reconstruction



14. Image-based Rendering

图 1-13 本章内容的图形摘要。资料来源：Burt and Adelson（1983b）；Agarwala，Dontcheva 等（2004）；格拉斯纳（2018）；他，Gkioxari 等（2017）；Brown，Szeliski 和 Winder（2005）；Butler，Wulff 等（2012）；Debevec 和 Malik（1997）；Snavely，Seitz 和 Szeliski（2006）；Scharstein，Hirschmuller 等（2014）；Curless 和 Levoy（1996）；Gortler，Grzeszczuk 等（1996）-有关版权信息，请参见各章中的图。