

前 言

本书介绍计算机视觉基础。计算机视觉（也称为机器视觉）的主要目标是基于由各种相机捕获的图像内容来重建和解释自然场景（参见 R. Szeliski[191]）。计算机视觉系统包括诸如卫星测绘、机器人导航系统、智能扫描仪和遥感系统之类的东西。在计算机视觉研究中，重点是从图像中提取有用的信息（参见 S. Prince [162]）。计算机视觉系统通常模仿人类视觉感知。计算机视觉系统中选择的硬件是某种形式的数码相机，被编程为近似视觉感知。因此，计算机视觉、数字图像处理、光学、光度学和光子学之间存在密切联系（参见 E. Stijns 和 H. Thienpont [188]）。

从计算机视觉的角度来看，光子学是捕捉视觉场景中光的科学。图像处理研究数字图像形成（如从模拟光学传感器信号到数字信号的转换）、操作（如图像滤波、去噪、裁剪）、特征提取（如像素强度、梯度方向、梯度幅度、边缘强度）、描述（如图像边缘和纹理）和可视化（如像素强度直方图）。可以参见 B. Jähne [87]和 S.G. Hoggar [82]的图像处理数学框架，并扩展到许多相关人员对图像处理的看法，如 M. Sonka、V. Hlavac 和 R. Boyle [186]，W. Burger 和 M. J. Burge [21]，R. C. 冈萨雷斯和 R. E. 伍兹[58]，R. C. 冈萨雷斯、R. E. 伍兹和 S. L. Eddins [59]，V. Hlavac [81]，C. Solomon 和 T. Breckon [184]。这些有用的信息，即可以检测、分析和分类的图像目标形状和模式，为计算机视觉研究者的关注点提供了基石（如[142]）。实际上，计算机视觉是对数字化图像结构和模式的研究，而这是在图像处理和光子学之上一层的图像分析。计算机视觉将图像处理和光子学包括在其追求图像几何和图像区域模式的技巧集合中。

此外，培养数字图像的智能系统观点可帮助发现隐藏的模式（如图像区域的重复凸轮廓）和嵌入的图像结构（如图像中感兴趣区域的点聚类）。通过量化器可以发现这种结构。量化器将一组（通常是连续的）值限制为离散值。量化器在其计算机视觉里最简单的形式中，观察特定的目标像素强度并在目标邻域中选择最接近的近似值。量化器的输出被 A. Gersho 和 R.M. Gray 称为码本[55, 5.1 节, p.133]（另见 S. Ramakrishnan、K. Rose 和 A. Gersho [164]）。

在图像网格叠加的上下文中，Gersho-Gray 量化器被替换为基于几何的量化器。基于几何的量化器将图像区域限制在其外形轮廓中并在图像中观察特定的目标形状轮廓，与其他具有近似形状的目标轮廓进行比较。在计算机视觉的基础上，基于几何的量化器观察并比较大致相同的图像区域，例如将最大核聚类（MNC）与其他核聚类进行比较。最大核聚类（MNC）是围绕称为核的网格多边形的图像网格多边形的集合（参见 J. F. Peters 和 E. İnan 在 Edelsbrunner 神经中的沃罗诺伊图像镶嵌[150]）。一个图像网格核是一个网格多边形，它是相邻多边形集合的中心。事实上，每个网格多边形都是一个多边形聚类的核。然而，只有一个或多个网格核是最大的。

最大图像网格核是具有最高数量的相邻多边形的网格核。MNC 在计算机视觉中很重

要，因为 MNC 轮廓能近似所考虑图像目标的形状。对图像的沃罗诺伊镶嵌是用多边形对图像的拼贴。图像的沃罗诺伊镶嵌也称为沃罗诺伊网格。对图 0.1 (a) 中音乐家图像的拼贴样本图如图 0.1 (b) 所示。在每个拼贴多边形内的红点●是沃罗诺伊区域(多边形)生成点的示例。更多的信息，参见 1.22.1 小节。这个音乐家网格核是如图 0.2 (b) 所示最大核聚类的中心。这是图 0.1 (b) 中音乐家图像网格中唯一的 MNC。该 MNC 也是沃罗诺伊网状神经的一个例子。对图像 MNC 的研究将我们带到图像几何和图像目标形状检测的入门处。更多内容可见 1.22.2 小节。

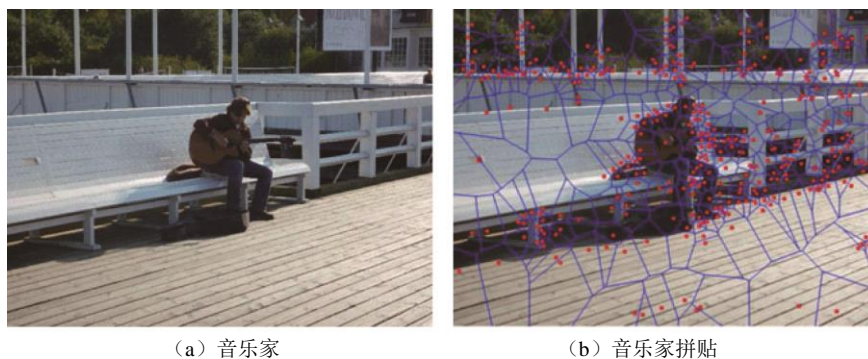


图 0.1 音乐家图像的沃罗诺伊镶嵌

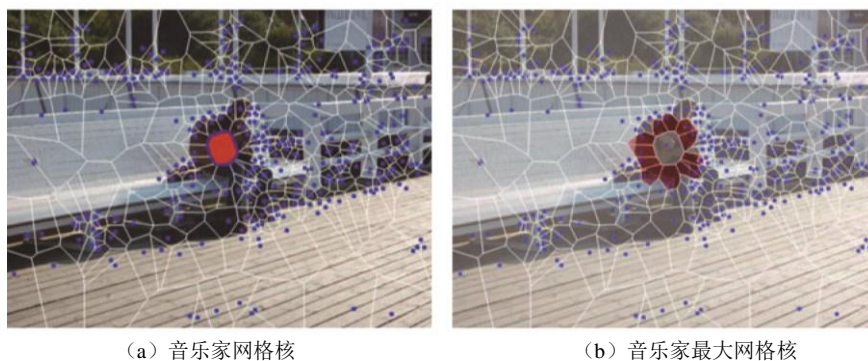


图 0.2 音乐家图像的最大核聚类

每个**图像拼贴多边形**是内部和顶点像素的**凸包**。一组图像点的凸包是该组点的最小凸集。一组图像点 A 是**凸集**，只要集合 A 中任意两点之间每条直线段上的所有点都包含在集合中。换句话说，知识发现是计算机视觉的核心。知识和对数字图像的理解可用于计算机视觉系统的设计。在视觉系统设计中，需要理解数字图像的组成和结构以及用于分析捕获图像的方法。

本书的重点是对光栅图像的研究。本书的续集将集中在由点(矢量)，线和曲线组成的矢量图像。每幅光栅图像基本由像素(典型的像素称为网点或网格生成点)，边缘(共同的、平行的、交叉的、凸的、凹的、直的、弯曲的、连通的、不连通的)，角度(矢量角度、矢量之间的角度、像素角度)，图像几何(沃罗诺伊区域[141]、德劳内三角剖分[140])，颜色，形状和纹理组成。许多计算机视觉和场景分析中的问题可通过找到某些隐藏或未观察到的图像变量核结构的最可能值来解决(参见 P. Kohi 和 P. H. S. Torr [96])。这样的结构和变量

包括像素的拓扑邻域、像素集合的凸包、图像结构的接近度（和分离度）、像素梯度分布以及描述捕获场景元素的特征矢量。

其他计算机视觉问题包括图像匹配、特征选择、最佳分类器设计、图像区域测量、兴趣点识别、轮廓组合、分割、配准、匹配、识别、图像聚类、模式聚类，见 F. Escolono、P. Suau、B. Bonev [45]和 N. Paragios、Y. Chen、O. Faugeras [138]，地标点形状匹配、图像变形、形状梯度[138]，假着色、像素标记、边缘检测、几何结构检测、拓扑邻域检测、目标识别和图像模式识别。

在计算机视觉中，重点是检测数字图像中常见的几何结构和物体形状。这导致了对图像处理和分析基础知识以及对矢量空间和图像计算几何观点的研究。图像处理的基础包括彩色空间、滤波、边缘检测、空间描述和图像纹理。数字图像是欧几里得空间（2-D 和 3-D）的示例。因此，数字图像的矢量空间观点是其基本特征的自然结果。数字图像结构本质上是几何或视觉拓扑结构。图像结构的示例包括图像区域、线段、生成点（例如，洛韦关键点）、像素组、像素的邻域、半空间、像素凸集和图像像素集合的凸包。例如，这种结构可以根据最接近选定点的图像区域或具有指定直径范围的图像区域来查看。一个图像区域是数字图像内部的一组图像点（像素）。任何图像区域的直径是该区域中一对点之间的最大距离。这样的结构也可以在由 2-D 和 3-D 图像中选定之间连接的线段所形成的三角形区域中找到。

这种结构也常见于 2-D 和 3-D 图像中由封闭的半空间所形成的点集凸包里、或如 G. M. 齐格勒所称的多面体[221]里。图像半空间是一条线上方或下方的所有点的集合。在所有这三种情况下都能获得对数字化图像的区域视图。更多有关多面体的信息参见附录 B.15。

每个图像区域都有一个形状。一些区域的形状比其他区域的更有趣。有趣的图像区域形状包含感兴趣的目标。这些图像的区域视图导致各种形式的图像分割，它们在识别图像中的目标时具有实用价值。此外，检测感兴趣图像区域的形状可帮助发现图像模式，它们超越了图像处理中对纹理元的研究。纹理元是一个由像素组表示的区域。更多有关形状的信息，参见附录 B.18 关于形状和形状边界的介绍。

图像分析侧重于各种数字图像测量（例如，像素大小、像素邻接、像素特征值，像素邻域，像素梯度，图像邻域的接近度）。图像分析中三种基于区域的标准方法是等值线阈值（二值化图像）、分水岭分割（使用从前景像素到背景区域的距离图计算）和非最大抑制（通过抑制所有比周围像素更不可能的像素来找到局部最大值）[212]。

在图像分析中，前景和背景像素具有不同的邻接（邻域），见 T. Abera [3]。有三种基本的邻域类型：即罗森菲尔德邻接邻域[101, 169]，豪斯多夫邻域[73, 74]，以及 J. F. Peters [142]和 C. J. Henry [75, 76]的描述性邻域。使用不同的几何形状，像素的邻接邻域由与给定像素相邻的像素限定。一个像素 p 的罗森菲尔德邻接邻域是一组与 p 邻接的像素。邻接邻域常用于数字图像中的边缘检测。

一个像素 p 的豪斯多夫邻域由与 p 的距离小于一个正数 r （称为邻域半径）的所有像素构成。一个像素 p 的描述性邻域（由 $N(\text{img}(x, y), r)$ 表示）由一组具有特征矢量的像素构成，这些特征矢量与描述 $\text{img}(x, y)$ 的特征矢量匹配或相似，且在规定的 r 半径内。

与邻接邻域不同，描述性邻域可以拥有其中的孔，即具有与邻域中心不匹配和不属于邻域的特征矢量的像素。其他类型的描述性邻域在[142, 1.16 节, pp.29-34]中介绍。



本书的章节源于过去若干年中自己教授过的本科生计算机视觉课程的笔记。本书中的许多主题源于自己与一些研究生和其他人员的讨论和交流,尤其是 S. Ramanna(很多形状,特别是在水晶中)、Anna Di Concilio(邻近性、区域自由几何和如图 0.3 的海景形状)、Clara Guadagni(花神经结构)、Arturo Tozzi(Borsuk-Ulam 定理的见解和 Gibson 形状、Avenarius 形状)、Romy Tozzi(记得 $8, \infty$)、Zdzisław Pawlak(波兰乡村绘画中的形状)、Lech Polkowski(那些纯粹的,拓扑的和粗糙集结构)、Piotr Artiemjew(蜻蜓翅膀)、Giangiacomo Gerla(那些 UNISA 庭院三角形和空间区域的指示(点))、Gerald Beer(Som Naimpally's 生活中的时刻)、Guiseppe Di Maio(关于邻近性的见解)、Somashekhhar (Som) A. Naimpally(拓扑结构)、Chris Henry(色彩空间,颜色形状集)、Macek Borkowski(3-D 空间观点)、Homa Fashandi、Dan Lockery、Irakli Dochviri、Ebubekir İnan(接近关系和接近组)、Mehmet Ali Öztürk(优美的代数结构)、Mustafa Uçkun、Nick Friesen(住宅的形状)、Özlem Umdu、Doungrat Chitcharoen、Çenker Sandoz(德劳内三角化)、Surabi Tiwari(许多类别)、Kyle Fedoruk(计算机视觉应用:斯巴鲁 EyeSight®)、Amir H. Meghdadi、Shabnam Shahfar、Andrew Skowron(Banacha 的接近性)、Alexander Yurkin、Marcin Wolksi(束或滑轮)、Piotr Wasilewski、Leon Schilmoeler、Jerzy W. Grzymala-Busse(关于粗糙集和 LATEX 提示的见解)、Zbigniew Suraj(Petri 网)、Jarosław Stepaniuk、Witold Pedrycz、Robert Thomas(倾斜的形状)、Marković G. oko(多形体)、Miroslaw Pawlak、Pradeepa Yahampath、Gabriel Thomas、Anthony (Tony) Szturm、Sankar K. Pal、Dean McNeill、Guiseppe (Joe) Lo Vetri、Witold Kinsner、Ken Ferens、David Schmidt(集合论)、William Hankley(基于时间的规范)、Jack Lange(黑板拓扑涂鸦)、Irving Sussman(定理和证明中的金块)和 Brian Peters(在墙壁上稍纵即逝的几何形状一瞥)。



图 0.3 沿意大利维耶特里海岸线的海景形状

我们系的技术人员非常乐于助人,尤其是 Mount-First Ng, Ken Biegun, Guy Jonatschick 和 Sinisa Janjic。

我的很多学生提出了有关本书主题的重要建议,特别是 Drew Barclay、Braden Cross、Binglin Li、Randima Hettiarachchi、Enoch A-iyeh、Chidoteremndu(Chido)Chinonyelum Uchime、D. Villar、K. Marcynuk、Muhammad Zubair Ahmad 和 Armina Ebrahimi。

各章的问题已被分了类。以  标记的问题可以迅速地回答。以  标记的问题有可能需要用喝一杯茶或咖啡的时间来完成。解决剩下的问题所需的时间各不相同。

詹姆斯 F.彼得斯
温尼伯，加拿大