



## 第2章 数字图像基础

要想成功，就必须弄清基础问题。

——亚里士多德

### 引言

本章的主要目的是介绍本书所用到的数字图像处理的一些基本概念。2.1 节简述人类视觉系统的机理，包括眼中图像的形成及对亮度的适应和鉴别能力。2.2 节讨论光、电磁波谱的其他分量及它们的成像特点。2.3 节讨论成像传感器及怎样使用它们产生数字图像。2.4 节介绍均匀图像取样及灰度量化的概念。该节中讨论的其他主题包括数字图像表示、图像中取样数和灰度级变化的影响、空间和灰度分辨率的概念，以及图像内插的原理。2.5 节处理像素间的各种基本关系。最后，2.6 节介绍本书用到的主要数学工具。该节的第二个目的是帮助您开始积累一些在各种基本图像处理任务中如何运用这些工具的感觉。这些工具及其应用范围在需要时将在本书的剩余部分再扩展。

### 2.1 视觉感知要素

虽然数字图像处理这一领域建立在数学和概率公式表示的基础之上，但人的直觉和分析在选择一种技术而不选择另一种技术时会起核心作用，这种选择通常是基于主观的视觉判断做出的。因此，将大概了解人类的视觉感知作为学习本书的起点是合适的。鉴于这一主题的复杂性和宽泛性，这里仅涉及人类视觉的最基本方面。特别是，我们的兴趣在于人类形成并感知图像的机理和参数。我们的兴趣还在于通过数字图像处理的一些因素来了解人类视觉的物理限制。因此，人类与电子成像设备相比，在分辨率和对光照变化的适应能力方面不仅有趣，从实践的角度来看也很重要。

#### 2.1.1 人眼的结构

图2.1 显示了人眼的一个简化水平剖面。眼睛的形状近似为一个球体，其平均直径约为 20 mm。有三层薄膜包围着眼睛：角膜与巩膜外壳、脉络膜和视网膜。角膜是一种硬而透明的组织，覆盖着眼睛的前表面。与角膜相连的巩膜是一层包围着眼球其余部分的不透明的膜。

脉络膜位于巩膜的正下方。脉络膜包含有血管网，它是眼睛的重要滋养源。即使是对脉络膜表面并不严重的损害，也有可能严重地损害眼睛，引起限制血液流动的炎症。脉络膜外壳着色很重，因此有助于减少进入眼内的外来光和眼球内反向散射光的数量。脉络膜的的最前面分为睫状体和虹膜。虹膜的收缩和扩张控制着进入眼睛的光量。虹膜中间的开口(瞳孔)的直径是可变的，范围大约在2~8 mm。

虹膜的前面包含有眼睛的可见色素，而后面则包含有黑色色素。

晶状体由同心的纤维细胞层组成，并由附在睫状体上的纤维悬挂着。晶状体包含 60%~70%的水、6%的脂肪和比眼睛中任何其他组织都多的蛋白质。晶状体由稍黄的色素着色，其颜色随着人的年龄的增大而加深。在极端情况下，晶状体会过于混浊，这通常是由白内障等疾病引起的，可能导致低下的彩色辨别能力和视觉清晰度的损失。晶状体吸收大约 8%的可见光谱，对短波长的光有较高的吸收率。在晶状体结构中，蛋白质吸收红外光和紫外光，吸收过量时会伤害眼睛。

眼睛最里面的膜是视网膜，它布满了整个后部的内壁。当眼睛适当地聚焦时，来自眼睛外部物体的光在视网膜上成像。由视网膜表面分布的不连续的光，感受器提供了图案视觉。有两类光感受器：锥状体和杆状体。每只眼睛中的锥状体数量在 600~700 万之间。它们主要位于视网膜的中间部分，称之为中央凹，且对颜色高度敏感。用这些锥状体，人可以充分地分辨图像细节，因为每个锥状体都连接到自身的神经末梢。肌肉控制眼球转动，直到感兴趣的物体图像落到中央凹上。锥状体视觉称为白昼视觉或亮视觉。

杆状体的数量更多：约有 7500~15 000 万个杆状体分布在视网膜表面。由于分布面积较大而且几个杆状体连接到一个神经末梢，因此减少了这些感受器感知细节的数量。杆状体用来给出视野内的一般的总体图像。它们没有彩色感觉，而对低照明度敏感。例如，在白天呈现鲜明色彩的物体，在月光下都没有颜色，因为此时只有杆状体受到刺激。这种现象称为暗视觉或微光视觉。

图 2.2 显示了右眼中通过眼睛光神经出现区的剖面的杆状体和锥状体密度。在这一区域由于没有感受器而导致了所谓的盲点(见图 2.1)。除了这一区域，感受器的分布是关于中央凹径向对称的。感受器密度根据距中央凹的度数来度量(即离开视轴的度数，它由视轴和通过晶状体中心并相交于视网膜的一条直线的夹角度量)。注意，图 2.2 中的锥状体在视网膜的中心最密(在中央凹的中心区域)。还应注意到，从该中心向外到偏离视轴大约  $20^\circ$  处，杆状体的密度逐渐增大，然后向外到视网膜的极限边缘处，密度逐渐下降。

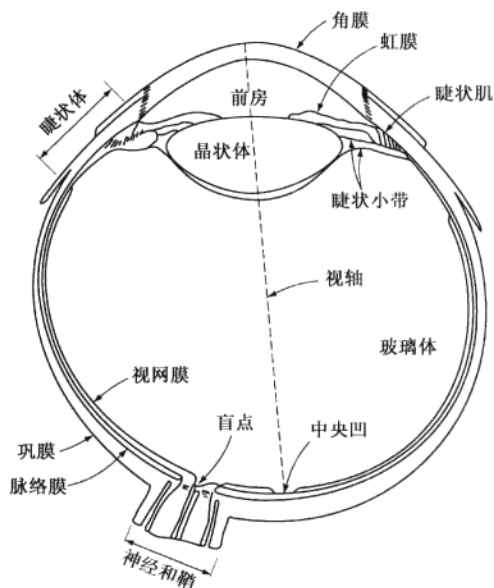


图 2.1 人眼剖面简图

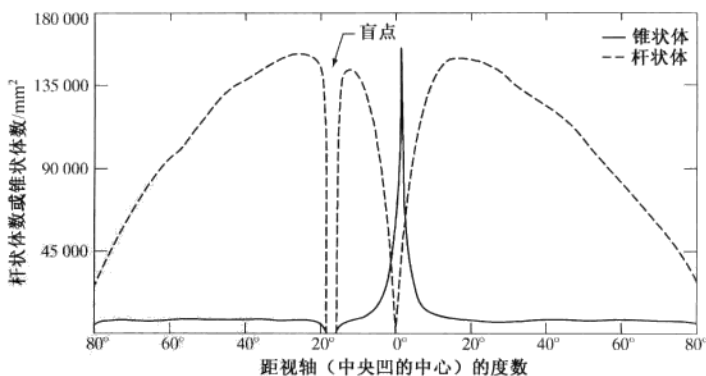


图 2.2 视网膜中杆状体和锥状体的分布

中央凹本身是视网膜中直径约为 1.5 mm 的圆形凹坑。然而,为便于后续的讨论,我们把它近似为方形或矩形敏感元素的阵列会更有用。这样,为解释的灵活性,我们可以把中央凹看成是大小为  $1.5 \text{ mm} \times 1.5 \text{ mm}$  的方形传感器阵列。在视网膜这一区域中,锥状体的密度大约为  $150\,000 \text{ 个}/\text{mm}^2$ 。基于这一近似,眼睛中最高敏感区域的锥状体数量约为 337 000 个。从自然分辨能力的角度看,恰好与一个中等分辨率的电荷耦合元件(CCD)成像芯片具有的元素数量相当,接收器阵列不大于  $5 \text{ mm} \times 5 \text{ mm}$ 。尽管人类整体的智慧和视觉经验的能力使得这种比较有些肤浅,但记住,人眼分辨细节能力的进一步讨论与当前电子成像传感器当然是可比较的。

### 2.1.2 眼睛中图像的形成

在普通照相机中,镜头有固定的焦距,各种距离的聚焦是通过改变镜头和成像平面间的距离实现的,胶片放置在成像平面上(数码相机情况下是成像芯片)。在人眼中,则与此相反;晶状体和成像区域(视网膜)之间的距离是固定的,实现正确聚焦的焦距是通过改变晶状体的形状来得到的。睫状体中的纤维可实现这一功能,在远离或接近目标物时纤维会分别变扁或加厚晶状体。晶状体中心和视网膜沿视轴的距离大约是 17 mm。焦距约为 14~17 mm,在眼睛放松且聚焦距离大于 3 m 时,焦距约为 17 mm。

图 2.3 中的几何关系说明了如何得到一幅在视网膜上形成的图像的尺度。例如,假设一个人正在观看距其 100 m 处的高为 15 m 的一棵树。令  $h$  表示视网膜图像中该物体的高度,由图 2.3 的几何形状可以看出  $15/100 = h/17$  或  $h = 2.55 \text{ mm}$ 。正如 2.1.1 节指出的那样,视网膜图像主要聚焦在中央凹区域。然后,光接收器的相对刺激作用产生感知,把辐射能转变为电脉冲,最后由大脑解码。

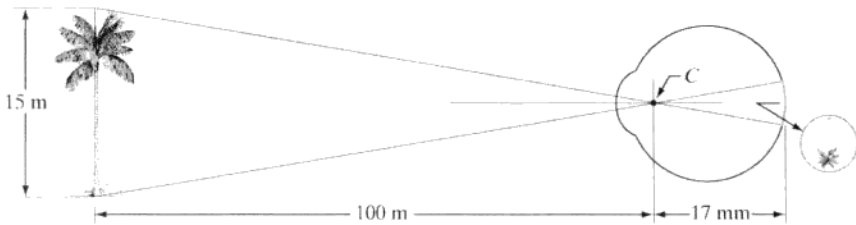


图 2.3 人眼观看一棵棕榈树的图解,点 C 是晶状体的光心

### 2.1.3 亮度适应和辨别

因为数字图像作为离散的灰度集来显示,所以眼睛对不同亮度级别之间的辨别能力在显示图像处理结果中是一个重要的考虑因素。人的视觉系统能够适应的光强度级别范围是很宽的——从暗阈值到强闪光约有  $10^{10}$  个量级。实验数据指出,主观亮度(即由人的视觉系统感知的亮度)是进入人眼的光强的对数函数。图 2.4 中画出的光强度与主观亮度的关系曲线说明了这一特性。长实线代表视觉系统能适应的光强范围。在亮视觉中,该范围大约是  $10^6$ 。由暗视觉逐渐过渡到亮视觉的近似范围约为  $0.001 \sim 0.1 \text{ mL}$ (朗伯)(在对数坐标中为  $-3 \sim -1 \text{ mL}$ ),图中画出了该范围内这一适应曲线的两个分支。

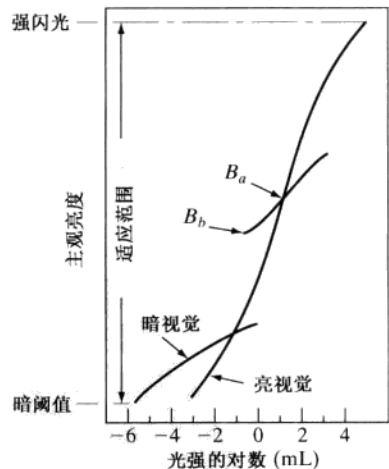


图 2.4 显示了特殊适应级别的主观亮度感知范围

解释图 2.4 中描述的令人印象深刻的动态范围的基本要点是视觉系统不能同时在一个范围内工作。确切地说，它是通过改变其整个灵敏度来完成这一较大的变动的，这就是周知的亮度适应现象。眼睛可以同时辨别不同强度级别的总范围与整个适应范围相比是很小的。对于任何给定的条件集合，视觉系统的当前灵敏度级别称为亮度适应级别，例如，它可能对应于图 2.4 中的亮度  $B_a$ 。较短的交叉线表示当眼睛适应这一强度级别时人眼所能感知的主观亮度范围。注意，这一范围是有一定限制的，级别  $B_b$  或低于级别  $B_b$  的刺激都被感知为不可辨别的黑色。该曲线的上部实际上不受限制，但如果延伸太远也会失去意义，因为高得多的强度将会把适应能力提高到比  $B_a$  更高的水平。

在任何特定的适应级别上，人眼辨别光强变化的能力也是值得考虑的因素。用于确定人类视觉系统亮度辨别能力的一个著名实验，由一个注视对象和大到足以使其占有全部视野的均匀发光区组成。这一区域是一个典型的漫反射体，譬如不透明玻璃，使用强度  $I$  可变的一个光源从后面照向该漫反射体。在该视野中增加一个照射分量  $\Delta I$ ，形成一个持续时间很短的闪烁，该闪烁以均匀光场中央的圆形方式出现，如图 2.5 所示。

如果  $\Delta I$  不够亮，则目标不变，表明没有可觉察的变化。当  $\Delta I$  逐渐加强时，目标会给出肯定的响应“是”，指出一个可觉察的变化。最后，当  $\Delta I$  足够强时，目标将始终给出肯定的响应“是”。 $\Delta I_c / I$  称为韦伯比，其中  $\Delta I_c$  是在背景照明为  $I$  时可辨别照明增量的 50%。 $\Delta I_c / I$  值较小意味着可辨别强度较小的百分比变化，这表示亮度辨别能力“较好”。相反， $\Delta I_c / I$  较大意味着要求有较大强度的百分比变化，这表示亮度辨别能力“较差”。

作为  $\log I$  的函数， $\log \Delta I_c / I$  曲线具有图 2.6 所示的一般形状。这一曲线表明，在低照明级别，亮度辨别较差(韦伯比大)，且它会随着背景照明的增加而明显改善(韦伯比降低)。曲线中的两个分支反映了这样一个事实，即在低照明水平情况下，视觉由杆状体执行，在高照明水平情况下(表示较好的辨别能力)，视觉由锥状体执行。

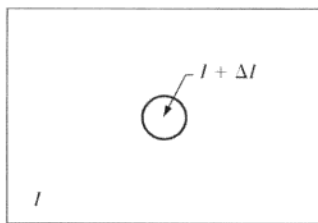


图 2.5 用于表征亮度辨别的基本实验设置

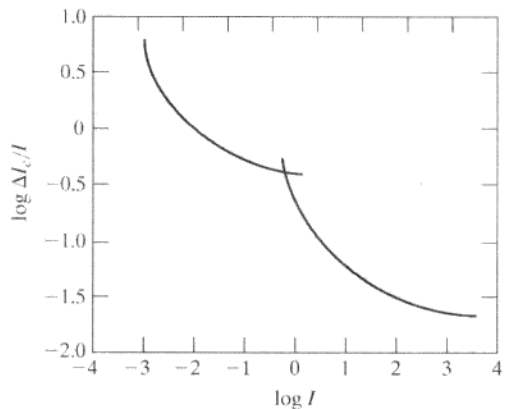


图 2.6 作为强度函数的典型韦伯比

如果背景照明保持恒定，并且代替闪光的其他光源的强度现在允许从不能觉察逐渐变化到总可以被觉察，典型的观察者可以辨别总共 12 级到 24 级的不同强度变化。粗略地看，该结果与一个人观看单色图像中的任意一点时所觉察到的不同强度的数量相关。这个结果并不意味着一幅图像可以由这样少的强度值来表示，因为当眼睛扫视图像时，平均背景在变化，从而允许在每个新的适应级别上检测一组不同的增量变化。最终结果是眼睛能够辨别更宽范围的整个强度。事实上，我们将在 2.4.3 节说明，眼睛能够检测到用不足 24 级强度表示的单色图像中的令人讨厌的轮廓效应。

两种现象清楚地表明感知亮度不是简单的强度的函数。第一种现象基于这样一个事实，即视觉

系统往往会在不同强度区域的边界处出现“下冲”或“上冲”现象。图 2.7(a) 显示了这种现象的一个典型例子。虽然条带的强度恒定，但在靠近边界处我们实际上感知到了带有毛边的亮度模式 [见图 2.7(c)]。这些看起来带有毛边的带称为马赫带，厄恩斯特·马赫于 1865 年首次描述了这一现象。

第二种现象称为同时对比，它与这样一个事实有关，即感知区域的亮度并不简单地取决于其强度，如图 2.8 所示。所有的中心方块都有完全相同的强度。然而，随着背景变得更亮，它们在眼睛里会变得更暗。一个更熟悉的例子是一张纸，一张纸放在桌子上时，看上去似乎比较白，而当我们使用它遮住眼睛来直视明亮的天空时，它看起来会总是呈现黑色。

人类感知现象的另一一些例子就是错觉，在错觉中，眼睛填充了不存在的信息或者错误地感知了物体的几何特点。图 2.9 给出了一些例子。在图 2.9(a) 中，正方形的轮廓看起来很清楚，尽管图像中并没有定义这样一个图形的直线。在图 2.9(b) 中可以看到相同的效果，只是这次是一个圆；注意，仅有几条直线就足以导致一个完整的圆的错觉。图 2.9(c) 中的两条水平线段的长度相同，但看起来一条显得比另一条短。最后，图 2.9(d) 中 45° 方向的所有直线都是等间距的平行线，然而画有交叉影线就产生了错觉，觉得这些线不再平行。

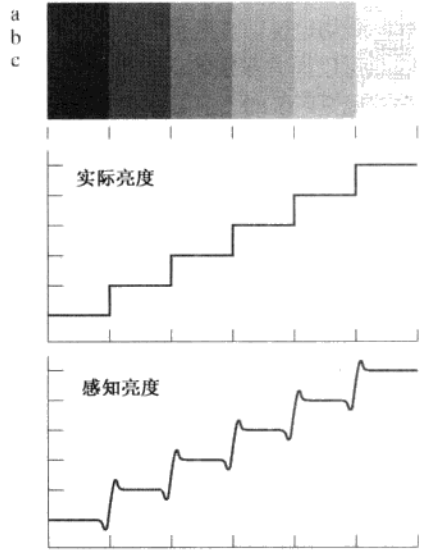


图 2.7 马赫带效应图解

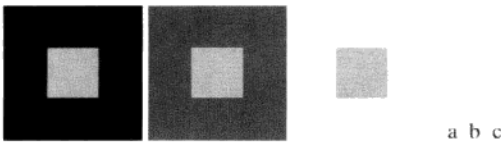


图 2.8 同时对比的几个例子。所有的中心方块都有相同的强度，但随着背景的变亮，它们看起来逐渐变暗

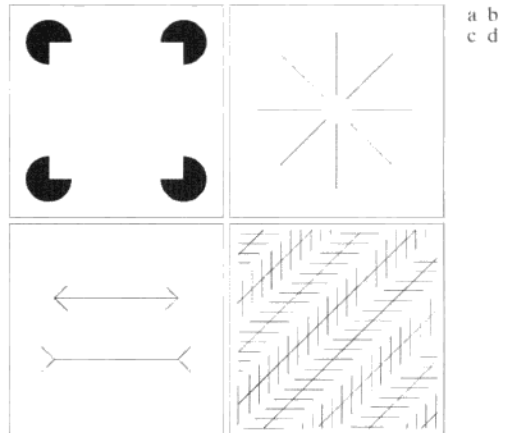


图 2.9 一些典型的错觉

## 2.2 光和电磁波谱

电磁波谱在 1.3 节中已做过介绍，现在我们更详细地研究这一主题。1666 年，艾萨克·牛顿发现，当一束太阳光通过一个玻璃棱镜时，显示的光束不再是白光，而是由一端为紫色而另一端为红色的连续色谱组成。如图 2.10 所示，我们感受到的可见光的彩色范围只占电磁波的一小部分。在波谱的一端是无线电波，其波长是可见光波长的几十亿倍。波谱的另一端是伽马射线，其波长比可见光小几百万倍。电磁波谱可用波长、频率或能量来描述。波长( $\lambda$ )和频率( $\nu$ )的关系可用下式描述：

63  
~  
65

$$\lambda = c / \nu \quad (2.2-1)$$

其中  $c$  是光速 ( $2.998 \times 10^8$  m/s)。电磁波谱的各个分量的能量由下式给出:

$$E = h\nu \quad (2.2-2)$$

其中  $h$  是普朗克常数。波长的单位是米 (m)，最常用的单位是微米 (表示为  $\mu\text{m}$ ,  $1 \mu\text{m} = 10^{-6}$  m) 和纳米 (表示为 nm,  $1 \text{ nm} = 10^{-9}$  m)。频率用赫兹 (Hz) 来度量, 1 Hz 表示正弦波每秒 1 个周期。常用的能量单位是电子伏特。

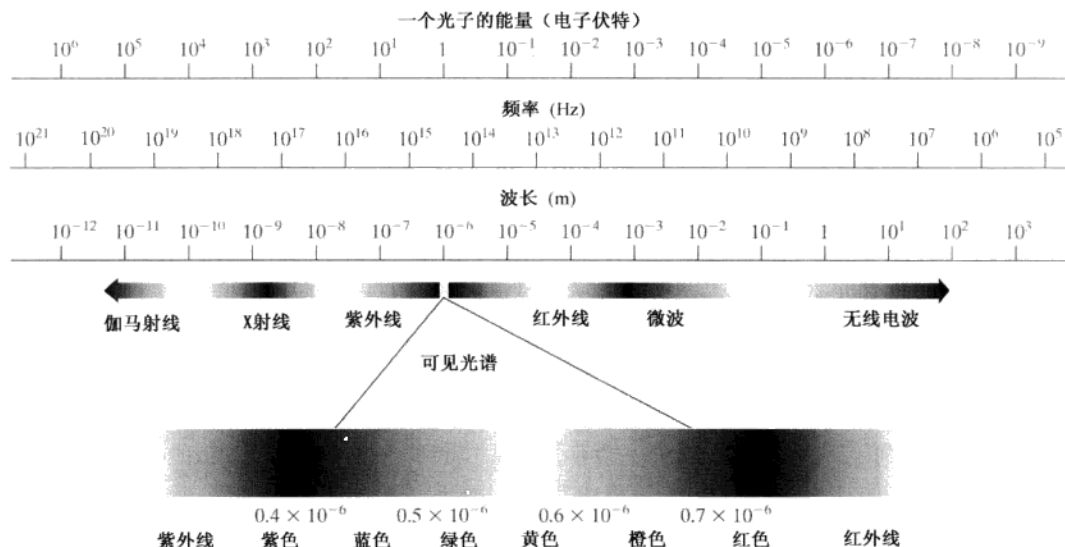


图 2.10 电磁波谱。为便于解释, 可见光谱已被放大, 但请注意, 可见光谱是电磁波谱中相当窄的一部分

电磁波可以看成是以波长  $\lambda$  传播的正弦波 (见图 2.11), 或者可以看成是没有质量的粒子流, 每个粒子以波的模式以光速传播和移动。每个无质量的粒子包含一定的 (一束) 能量, 每束能量称为一个光子。我们从式 (2.2-2) 可以看出能量与频率成正比, 因此更高频率 (更短波长) 的电磁现象的每个光子携带有更多的能量。这样, 无线电波有低能量光子, 微波比无线电波具有较多的能量, 红外波还要多, 然后是可见光、紫外线、X 射线和最后的伽马射线, 其中伽马射线的能量最高, 这就是伽马射线对活体组织危害很大的原因。

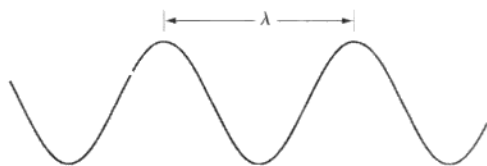


图 2.11 一个波长的图形表示

光是一种特殊的电磁辐射, 它可以被人眼感知。为便于讨论, 把可见光谱 (彩色) 段扩展后示于图 2.10 (第 6 章将更详细地讨论彩色)。电磁波谱的可见光波段的跨越范围约为  $0.43 \mu\text{m}$  (紫色)  $\sim 0.79 \mu\text{m}$  (红色)。为方便起见, 彩色谱分为 6 个主要区域: 紫色、蓝色、绿色、黄色、橘黄色和红色。每种颜色 (或电磁波谱的其他分量) 不是突然终止的, 而是混合平滑地过渡到另一种颜色的, 如图 2.10 所示。

人感受一个物体的颜色由物体反射光的性质决定。以所有可见波长相对平衡地反射光的物体, 对观察者而言是白色的。然而, 一个物体在可见光谱的有限范围内反射时会呈现各种颜色色调。例如, 绿色物体反射波长范围为  $500 \sim 570 \text{ nm}$  的光, 而吸收其他波长的大部分能量。

没有颜色的光称为单色光或无色光。单色光的唯一属性是它的强度或大小。因为感知单色光的强度从黑色到灰色变化, 最后到白色, 灰度级一词通常用来表示单色光的强度。在后续讨论中, 我们

将交替地使用强度和灰度级这两个术语。从黑到白的单色光的度量值范围通常称为灰度级,而单色图像常称为灰度图像。

正如前面指出的那样,彩色光的电磁能谱跨越  $0.43\ \mu\text{m}$ (紫色)~ $0.79\ \mu\text{m}$ (红色)的范围。除了频率之外,有三个基本量用于描述彩色光源的质量:发光强度、光通量和亮度。发光强度是从光源流出能量的总量,通常用瓦特(W)来度量。用流明数(lm)度量的光通量给出观察者从光源感受到的能量。例如,从远红外光谱范围的光源发射出的光具有实际意义的能量,但观察者却很难感知到它。它的光通量几乎是零。最后,如 2.1 节所讨论的那样,亮度是光感知的主观描绘子,它实际上不能度量。它具体体现了强度的无色概念,是描述彩色感觉的参数之一。

继续图 2.10 的讨论,我们注意到电磁波谱的短波长的一端是伽马射线和 X 射线。正如 1.3.1 节讨论的那样,伽马辐射对医学和天文学成像很重要,对核环境中的辐射成像也很重要。硬(高能)X 射线用于工业应用中。胸透和牙科使用的 X 射线位于 X 射线频段的低能(软)端。软 X 射线波段过渡到远紫外线的区域,而远紫外区域又与较长波长的可见光谱混合。逐步移向更高的波长,我们会遇到红外波段,红外波段会辐射热,这使得它在依靠“热特性”成像的应用中非常有用。红外波段靠近可见光谱的部分称为近红外区域,而该波段的另一端称为远红外区域。远红外区域的后部与微波频段混合。这个波段是众所周知的微波炉的能量,但它还有许多其他用途,包括通信与雷达。最后,无线电波段包括电视、调幅收音机和调频收音机。在更高的能级,从某些天体发出的无线电信号在天文观测中很有用。1.3 节给出了刚才讨论的多数波段成像的例子。

原理上,如果可以开发出一种传感器来检测由一种电磁波谱发射的能量,那么我们就可以在该波段上对感兴趣的事件成像。但要注意的一点是,要求“看到”一个物体的电磁波的波长必须小于等于物体的尺寸。例如,水分子直径的量级是  $10^{-10}\ \text{m}$ ,若要研究该分子,则需要一个能在远紫外或软 X 射线范围发射的光源。这个限制与传感材料的物理特性一起确立了成像传感器功能的基本限制,譬如今天所用的可见光、红外线和其他传感器。

虽然成像主要以电磁波发射的能量为基础,但这并不是生成图像的唯一方法。例如,如 1.3.7 节讨论的物体反射的声波也可用于形成超声波图像。其他的主要数字图像源是电子显微镜的电子束和用于图形与可视化的合成图像。

## 2.3 图像感知和获取

我们感兴趣的多数图像都是由“照射”源和形成图像的“场景”元素对光能的反射或吸收而产生的。把“照射”和“场景”加上引号是为了强调这样一个事实,即比我们所熟悉的一个可见光源每天照射普通的三维场景情况更一般。例如,照射可能由电磁能源引起,如雷达、红外线或 X 射线系统。但是,正如前面指出的那样,照射也可以由非传统光源(如超声波)甚至由计算机产生的照射模式产生。类似地,场景元素可能是熟悉的物体,但它们也可能是分子、沉积岩或人类的大脑。依赖光源的特性,照射被物体反射或透射。第一类例子是从平坦表面反射。第二类例子是为了产生一幅 X 射线照片,让 X 射线透过病人的身体。在某些应用中,反射能或透射能可聚焦到一个光转换器上(如荧光屏),光转换器再把能量转换为可见光。电子显微镜和某些伽马成像应用就使用这种方法。

图 2.12 显示了用于将照射能量变换为数字图像的三种主要传感器配置。原理很简单:通过将输入电能和对特殊类型检测能源敏感的传感器材料相组合,把输入能源转变为电压。输出电压波形是传感器的响应,通过把传感器响应数字化,从每一个传感器得到一个数字量。在这一节中,我们将关注图像感知和生成的主要方式。图像数字化将在 2.4 节中讨论。

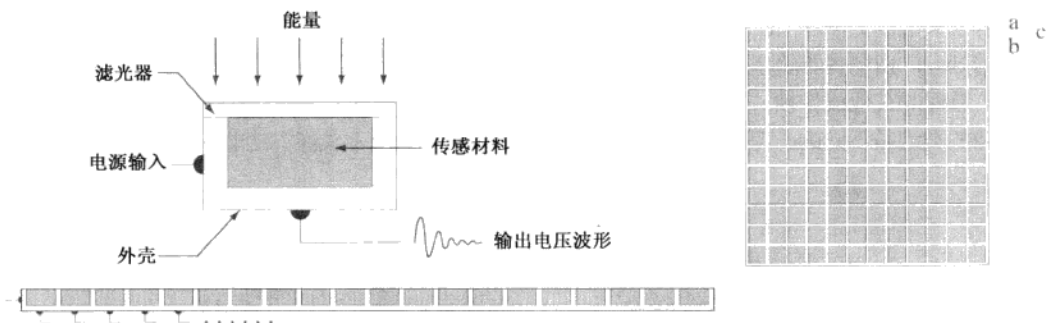


图 2.12 (a) 单个成像传感器; (b) 条带传感器; (c) 阵列传感器

### 2.3.1 使用单个传感器获取图像

图 2.12(a) 显示了单个传感器的部件。也许最熟悉的这类传感器是光二极管，它由硅材料构成，并且其输出电压波形与入射光成正比。在传感器前面用一个滤光器改善选择性。例如，光传感器前面的绿色(通过)滤光器有利于彩色谱的绿波段光通过。因此，传感器输出的绿光比可见光谱中的其他分量要强。

为使用单一传感器产生二维图像，在传感器和成像区域之间必须有  $x$  方向和  $y$  方向的相对位移。图 2.13 示出了一个用于高精度扫描的配置，其中底片安装在一个鼓上，鼓的机械转动提供了一个维度的位移。单个传感器安装在引导螺杆上，它提供与转动相垂直的方向上的移动。因为机械运动可高精度地控制，所以这一方法是得到高分辨率图像的一种廉价方法(但速度较慢)。另一种类似的机械配置使用一个平面床，传感器则在两个方向线性移动。这些类型的机械数字化仪有时称为微密度计。

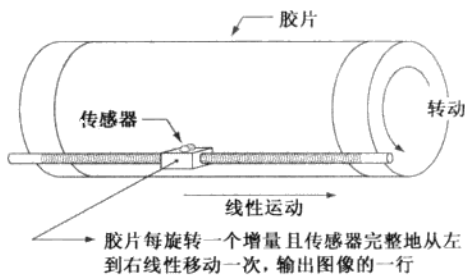


图 2.13 单个传感器通过运动来生成二维图像

使用单个传感器成像的另一个例子是将一个激光源和传感器放在一起。使用镜子来控制扫描模式的发射光束，并将反射的激光信号引导到传感器。这种配置也可使用条形或矩形传感器获取图像，这种配置将在后续两节中加以讨论。

### 2.3.2 使用条带传感器获取图像

比单个传感器更常用的几何结构是由内嵌传感器形式组成的传感器带，如图 2.12(b) 所示。该传感器带在一个方向上提供成像单元。垂直于传感器带的运动在另一方向上成像，如图 2.14(a) 所示。这是大多数平板扫描仪中使用的排列方式。感知设备可能内嵌有 4000 个或更多的传感器。内嵌传感器常用于航空成像应用中，在这种应用中，成像系统安装在一架飞行器上，飞行器以恒定的高度和速度飞过被成像的地区。可响应各种电磁波谱波段的一维传感器带按垂直于飞行方向来安装。成像传感器带一次给出一幅图像的一行，传感器带的运动完成二维图像的另一个维度。透镜和其他聚焦方法用于把扫描区域投影到传感器上。

以圆环形方式安装的传感器带用于医学和工业成像，以得到三维物体的剖面(“切片”)图像，如图 2.14(b) 所示。一个旋转的 X 射线源提供照射，射线源对面的传感器则收集穿过物体的 X 射线能量(很明显，这些传感器必须对 X 射线敏感)。这就是 1.2 节和 1.3.2 节中指出的医学和工业计算机轴断层(CAT)成像的基础。注意，传感器的输出必须由重建算法处理，重建算法的目的是把感知数据



转换为有意义的剖面图像(见5.11节)。换句话说,图像不可能单靠传感器的运动直接得到;它们需要进一步的处理。由图像堆叠组成的三维数字物体是由物体与传感器环在相垂直方向的运动产生的。基于CAT原理的其他成像模式包括核磁共振成像(MRI)和正电子发射断层(PET)成像。照射源、传感器和图像的类型是不同的,但概念上它们与图2.14(b)中所示的基本成像方法非常相似。

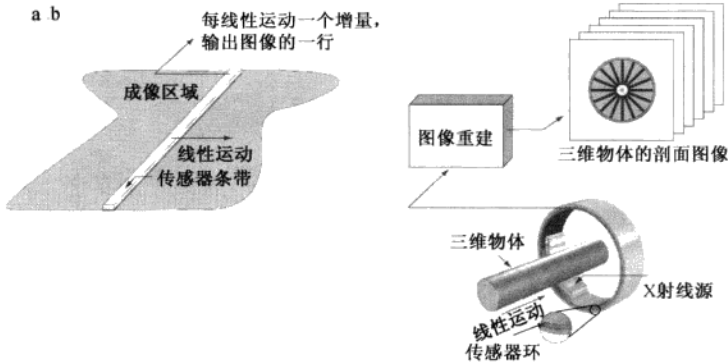


图 2.14 (a) 使用一个线性传感器带获取图像; (b) 使用一个环形传感器带获取图像

### 2.3.3 使用传感器阵列获取图像

图 2.12(c) 显示了以二维阵列形式排列的单独的传感器。大量的电磁波和一些超声波传感装置常以阵列形式排列。这也是我们在数字摄像机中所看到的主要排列方式。这些摄像机的典型传感器是 CCD 阵列, 这种阵列可制造为具有很宽范围的传感特性, 并能封装为具有  $4000 \times 4000$  个单元或更多单元的阵列。CCD 传感器广泛用于数字摄像机和其他光敏设备中。每个传感器的响应正比于投射到传感器表面的光能总量, 这一特性被用于天文学和其他要求低噪声图像的应用中。通过让传感器累积输入光信号超过几分钟甚至几小时的方法, 可达到减少噪声的目的。因为图 2.12(c) 所示的传感器阵列是二维的, 故其主要优点是通过将能量聚焦到阵列表面, 以得到一幅完整的图像。很明显, 像前面两节中讨论的传感器排列的运动是不需要的。

这样的传感器阵列所用的主要方法示于图 2.15。该图显示了来自照射源的能量是场景元素的反射(如本节开始提到的那样, 该能量也可以通过场景元素透射)。图 2.15(c) 所示成像系统执行的第一个功能是收集入射能量, 并将它聚焦到一个图像平面上。如果照射的是光, 则成像系统的前端是一个光学透镜, 该透镜把观察的场景投影到透镜的聚焦平面上, 如图 2.15(d) 所示。与焦平面重合的传感器阵列产生与每个传感器接收到的光的总量成正比的输出。数字或模拟电路扫描这些输出, 并把它们转换成模拟信号, 然后由成像系统的其他部分数字化, 输出是一幅数字图像, 如图 2.15(e) 的图解所示。把一幅图像转换为数字形式是 2.4 节的主题。

在某些情况下, 我们直接对源成像, 就像得到太阳的图像一样。

### 2.3.4 简单的图像形成模型

正如在 1.1 节中介绍的那样, 我们用形如  $f(x, y)$  的二维函数来表示图像。在空间坐标  $(x, y)$  处,  $f$  的值或幅度是一个正的标量, 其物理意义由图像源决定。当一幅图像由物理过程产生时, 其亮度值正比于物理源(如电磁波)所辐射的能量。因此,  $f(x, y)$  一定是非零的和有限的, 即

$$0 < f(x, y) < \infty \quad (2.3-1)$$

在处理过程中或是作为解译的结果, 图像灰度可以为负值。例如, 在雷达成像中, 移向雷达系统的物体通常解释为速度为负, 而远离雷达系统的物体总是解释为速度为正。这样, 一幅速度图像就可编码为既有正值又有负值。在存储和显示图像时, 我们通常会标定灰度, 以便最小的负值变为 0(灰度标定将在 2.6.3 节介绍)。

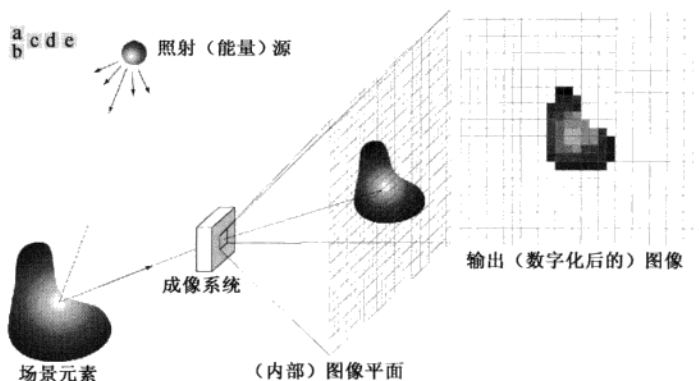


图 2.15 数字图像获取过程的一个例子: (a) 能量 (“照射”) 源; (b) 场景元素; (c) 成像系统; (d) 场景到图像平面的投影; (e) 数字化后的图像

函数  $f(x, y)$  可由两个分量来表征: (1) 入射到被观察场景的光源照射总量; (2) 场景中物体所反射的光照总量。这两个分量分别称为入射分量和反射分量, 且分别表示为  $i(x, y)$  和  $r(x, y)$ 。两个函数作为一个乘积合并形成  $f(x, y)$ , 即

$$f(x, y) = i(x, y)r(x, y) \quad (2.3-2)$$

其中,

$$0 < i(x, y) < \infty \quad (2.3-3)$$

和

$$0 < r(x, y) < 1 \quad (2.3-4)$$

式 (2.3-4) 指出反射分量限制在 0 (全吸收) 和 1 (全反射) 之间。  $i(x, y)$  的性质取决于照射源, 而  $r(x, y)$  的性质则取决于成像物体的特性。注意, 这种表示方式还可用于照射光通过一个媒体形成图像的情况, 如胸透 X 射线照片。在这种情况下, 我们应该用透射系数代替反射函数, 但其限制应该与式 (2.3-4) 相同, 且形成的图像函数会建模为式 (2.3-2) 中的乘积。

### 例 2.1 照射和反射的一些典型值。

式 (2.3-3) 和式 (2.3-4) 中给出的值是理论界限。对于可见光, 下面的平均数值用数字说明了  $i(x, y)$  的一些典型范围。在晴朗的白天, 太阳在地面上可能会产生超过  $90\,000 \text{ lm/m}^2$  的照度。在有云的白天, 这个数值下降到  $10\,000 \text{ lm/m}^2$ 。在晴朗的夜晚, 满月情况下的照度大约为  $0.1 \text{ lm/m}^2$ 。商用办公室的典型照度约为  $1000 \text{ lm/m}^2$ 。类似地, 下面是  $r(x, y)$  的一些典型值: 黑天鹅绒为 0.01, 不锈钢为 0.65, 白色墙为 0.80, 镀银金属为 0.90, 雪为 0.93。

令单色图像在任何坐标  $(x_0, y_0)$  处的强度 (灰度) 表示为

$$\ell = f(x_0, y_0) \quad (2.3-5)$$

由式 (2.3-2) 到式 (2.3-4) 可知  $\ell$  的取值范围为

$$L_{\min} \leq \ell \leq L_{\max} \quad (2.3-6)$$

理论上, 仅要求  $L_{\min}$  为正, 而要求  $L_{\max}$  为有限值。实际上,  $L_{\min} = i_{\min} r_{\min}$  和  $L_{\max} = i_{\max} r_{\max}$ 。参照前面提到的办公室的平均照度和反射值范围, 我们希望  $L_{\min} \approx 10$  和  $L_{\max} \approx 1000$  作为没有附加照明的室内值的典型限制。

区间  $[L_{\min}, L_{\max}]$  称为灰度级 (或强度级)。实际情况常常令该区间为  $[0, L-1]$ , 其中  $\ell = 0$  为黑色,  $\ell = L-1$  在灰度级中为白色。所有中间值是从黑色到白色之间变化的灰度色调。

## 2.4 图像取样和量化

由前一节的讨论, 我们看到有多种获取图像的方法, 但在所有这些方法中我们的目的是相同的, 就是从感知的数据生成数字图像。多数传感器的输出是连续的电压波形, 这些波形的幅度和空间特性都与感知的物理现象有关。为了产生一幅数字图像, 我们需要把连续的感知数据转换为数字形式。这种转换包括两种处理: 取样和量化。

本节中关于取样的讨论较为直观。第4章将深入探讨该主题。

### 2.4.1 取样和量化的基本概念

图2.16说明了取样和量化的基本概念。图2.16(a)显示了一幅连续图像 $f$ , 我们想把它转换为数字形式。一幅图像的 $x$ 和 $y$ 坐标及幅度可能都是连续的。为将它转换为数字形式, 必须在坐标上和幅度上都进行取样操作。对坐标值进行数字化称为取样, 对幅值数字化称为量化。

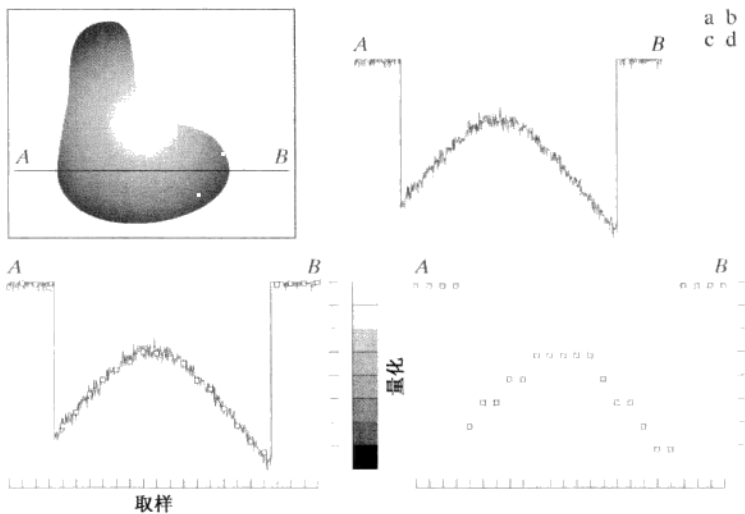


图 2.16 生成一幅数字图像: (a) 连续图像; (b) 连续图像中从 A 到 B 的一条扫描线, 用于说明取样和量化的概念; (c) 取样和量化; (d) 数字扫描线

图2.16(b)中的一维函数是图2.16(a)中沿线段 $AB$ 的连续图像幅度值(灰度级)的曲线。随机变化是由图像噪声引起的。为了对该函数取样, 我们沿线段 $AB$ 等间隔地对该函数取样, 如图2.16(c)所示。每个样本的空间位置由图形底部的垂直刻度指出。样本用放在函数曲线上的白色小方块表示。这样的一组离散位置就给出了取样函数。然而, 样本值仍(垂直)跨越了灰度值的连续范围。为了形成数字函数, 灰度值也必须转换(量化)为离散量。图2.16(c)的右侧显示了已分为8个离散区间的灰度标尺, 范围从黑到白。垂直刻度标记指出了赋予8个灰度的每一个特定值。通过对每一样本赋予8个离散灰度级中的一个来量化连续灰度级。赋值取决于该样本与一个垂直刻度标记的垂直接近程度。取样和量化操作生成的数字样本如图2.16(d)所示。从该图像的顶部开始逐行执行这一过程, 则会产生一幅二维数字图像。图2.16意味着除了所用的离散级数外, 量化所达到的精度强烈地依赖于取样信号的噪声。

按刚刚讨论的方式取样, 假设我们有一幅在两个坐标方向和幅度上都连续的图像。实践中, 取样方法由用于生成该图像的传感器配置决定。如图2.13所示, 当一幅图像由单个传感单元与机械运

动相结合生成时,传感器的输出可用前面讨论的方式量化。然而,空间取样是通过选择各个机械增量的数量并在取样处激活传感器来收集数据完成的。机械运动可以非常精确地实现,因此从原理上讲,使用这种方法对一幅图像取样可以到达的精细程度并没有限制。实践中,取样精度的限制由其他因素决定,譬如系统的光学元件的质量。

当我们使用带状传感器来获取图像时,带中传感器的数量决定了图像中一个方向上的限制。另一个方向上的机械运动可控制得更精确一些,但试图在一个方向上实现超过另一个方向上被传感器数量限制的取样密度是没有意义的。对传感器输出的量化就完成了产生数字图像的过程。

当传感阵列用于图像获取时,没有运动且阵列中传感器的数量决定了两个方向上的取样限制。传感器输出的量化与前述相同。图2.17说明了这个概念。图2.17(a)显示了投影到一个阵列传感器平面上的连续图像。图2.17(b)显示了取样和量化后的图像。很明显,数字图像的质量在很大程度上取决于取样和量化中所用的样本数和灰度级。然而,正如2.4.3节中表明的那样,在选择这些参数时,图像内容是一个重要的考虑因素。

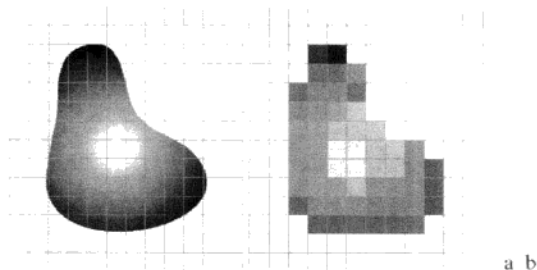


图2.17 (a)已投影到一个传感器阵列上的连续图像; (b)图像取样和量化后的结果

## 2.4.2 数字图像表示

令  $f(s, t)$  表示一幅具有两个连续变量  $s$  和  $t$  的连续图像函数。如前一节解释的那样,通过取样和量化,我们可把该函数转换为数字图像。假如我们把该连续图像取样为一个二维阵列  $f(x, y)$ , 该阵列包含有  $M$  行和  $N$  列, 其中  $(x, y)$  是离散坐标。为表达清楚和方便起见,我们对这些离散坐标使用整数:  $x=0, 1, 2, \dots, M-1$  和  $y=0, 1, 2, \dots, N-1$ 。这样,数字图像在原点的值就是  $f(0,0)$ , 第一行中下一个坐标处的值是  $f(0,1)$ 。这里,符号  $(0,1)$  表示第一行的第二个样本,它并不意味着是对图像取样时的物理坐标值。通常,图像在任何坐标  $(x, y)$  处的值记为  $f(x, y)$ , 其中  $x$  和  $y$  都是整数。由一幅图像的坐标张成的实平面部分称为空间域,  $x$  和  $y$  称为空间变量或空间坐标。

如图2.18所示,有三种基本方法表示  $f(x, y)$ 。图2.18(a)是一幅函数图,用两个坐标轴决定空间位置,第三个坐标是以两个空间变量  $x$  和  $y$  为函数的  $f$ (灰度)值。虽然我们可以在这个例子中用该图来推断图像的结构,但是,通常复杂的图像细节太多,以至于很难由这样的图去解译。在处理的元素是以  $(x, y, z)$  三坐标的形式表达的灰度集时,这种表示是很有用的,其中  $x$  和  $y$  是空间坐标,  $z$  是  $f$  在坐标  $(x, y)$  处的值。在2.6.4节中,我们将探讨这种表示。

图2.18(b)是更一般的表示。它显示了  $f(x, y)$  出现在监视器或照片上的情况。这里,每个点的灰度与该点处的  $f$  值成正比。该图中仅有三个等间隔的灰度值。如果灰度被归一化到区间  $[0, 1]$  内,那么图像中每个点的灰度都有  $0, 0.5$  或  $1$  这样的值。监视器或打印机简单地把这三个值分别变换为黑色、灰色或白色,如图2.18(b)所示。第三种表示是将  $f(x, y)$  的数值简单地显示为一个阵列(矩阵)。在这个例子中,  $f$  的大小为  $600 \times 600$  个元素,或  $360\,000$  个数字。很清楚,打印整个矩阵是很麻烦的,且传达的信息也不多。然而,在开发算法时,当图像的一部分被打印并作为数值进行分析时,这种表示相当有用。图2.18(c)以图形方式传达了这一概念。

由前面几节的讨论,我们可得出一个结论,即图2.18(b)和图2.18(c)中的表示是最有用的。图像显示允许我们快速地观察结果。数值阵列用于处理和算法开发。以公式形式,我们可将一个  $M \times N$  的数值阵列表示为

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, N-1) \\ f(1, 0) & f(1, 1) & \dots & f(1, N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1, 0) & f(M-1, 1) & \dots & f(M-1, N-1) \end{bmatrix} \quad (2.4-1)$$

该式的两边以等效的方式定量地表达了一幅数字图像。右边是一个实数矩阵，该矩阵中的每个元素称为图像单元、图像元素或像素。图像和像素这两个术语将在全书中表示数字图像及其元素。

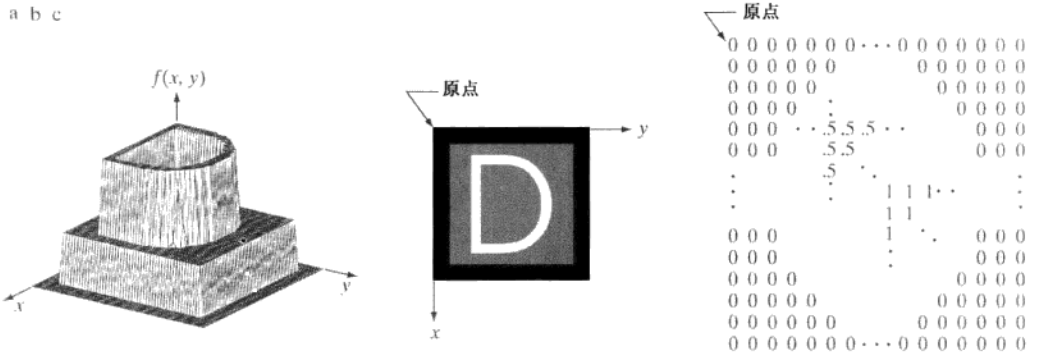


图 2.18 (a) 画为表面图形的图像; (b) 显示为可视灰度阵列的图像; (c) 显示为一个二维数值阵列的图像(0, 0.5 和 1 分别表示黑色、灰色和白色)

在某些讨论中，使用传统的矩阵表示法来表示数字图像及其像素更为方便：

$$A = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,N-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1,N-1} \\ \vdots & \vdots & & \vdots \\ a_{M-1,0} & a_{M-1,1} & \dots & a_{M-1,N-1} \end{bmatrix} \quad (2.4-2)$$

显然， $a_{ij} = f(x=i, y=j) = f(i, j)$ ，因此式(2.4-1)和式(2.4-2)是相同的矩阵。我们甚至可以将一幅图像表示一个向量  $\mathbf{v}$ 。例如，尺寸为  $MN \times 1$  的列向量由  $\mathbf{v}$  的第一  $M$  个元素作为  $A$  的第一列来构成，下一  $M$  个元素作为第二列，等等。我们也可以使用  $A$  的行代替列来形成这样的一个向量。只要一致，哪种表示都是有效的。

简要地回顾图2.18，注意到，数字图像的原点位于左上角，其中正  $x$  轴向下延伸，正  $y$  轴向右延伸。这种方便的表示基于这样的事实：许多图像显示(譬如电视显示器)扫描都是从左上角开始的，然后一次向下移动一行。更重要的事实是矩阵的第一个元素按惯例应在阵列的左上角，因此，将  $f(x, y)$  的原点选择在左上角于数学上是讲得通的。记住，这种表示是您所熟悉的标准的右手笛卡儿坐标系<sup>①</sup>。我们仅说明了指向下方和指向右方的坐标轴来代替向右和向上的坐标轴。

有时，以更正式的数学术语表达取样和量化可能会很有用。令  $Z$  和  $R$  分别表示整数集和实数集。取样处理可看成是把  $xy$  平面分为一个网格的过程，网格中每个单元的中心坐标是笛卡儿积  $Z^2$  中的一对元素， $Z^2$  是所有有序元素对  $(z_i, z_j)$  的集合， $z_i$  和  $z_j$  是  $Z$  中的整数。因此，如果  $(x, y)$  是  $Z^2$  中的整数，且  $f$  是把灰度值(即实数集  $R$  中的一个实数)赋给每个特定坐标对  $(x, y)$  的一个函数，则  $f(x, y)$

<sup>①</sup> 回忆可知，右手坐标系是这样一种系统：当正  $x$  轴方向向右时，在大拇指竖直向上时，正  $y$  轴的方向为垂直于  $x$  轴的中指方向。如图 2.18(a) 所示，事实上这是我们的图像坐标系情形。

就是一幅数字图像。显然,这种赋值过程就是前面描述的量化处理。如果灰度级也是整数(此时及后续章节中通常为这种情况),则用  $Z$  代替  $R$ ,然后,数字图像变成一个二维函数,且它的坐标和幅值都是整数。

数字化过程要求针对  $M$  值、 $N$  值和离散灰度级数  $L$  做出判定。对于  $M$  和  $N$ ,除了必须取正整数外没有其他限制。然而,出于存储和量化硬件的考虑,灰度级数典型地取为 2 的整数次幂,即

$$L = 2^k \quad (2.4-3)$$

我们假设离散灰度级是等间隔的,且它们是区间  $[0, L-1]$  内的整数。有时,由灰度跨越的值域非正式地称为动态范围。这一术语在不同的场合有不同的用法。这里,我们将图像系统的动态范围定义为系统中最大可度量灰度与最小可检测灰度之比。作为一条规则,上限取决于饱和度,下限取决于噪声(见图 2.19)。基本上,动态范围由系统能表示的最低和最高灰度级来确定,因此,也是图像具有的动态范围。与这一概念紧密联系的是图像对比度,我们定义一幅图像中最高和最低灰度级间的灰度差为对比度。当一幅图像中像素可感知的数值有高的动态范围时,那么我们认为该图像具有高的对比度。相反,具有低动态范围的图像就很呆滞,看上去像冲淡了灰度。第 3 章将更加详细地讨论这些概念。

存储数字图像所需的比特数  $b$  为

$$b = M \times N \times k \quad (2.4-4)$$

当  $M = N$  时,该式变为

$$b = N^2 k \quad (2.4-5)$$

表 2.1 中显示了  $N$  和  $k$  取不同值时需要用来存储方形图像的比特数。括号中的数字表示对应于每个  $k$  值的灰度级数。当一幅图像有  $2^k$  个灰度级时,实际上通常称该图像为一幅“ $k$  比特图像”。例如,有 256 个可能的离散灰度值的图像称为 8 比特图像。注意,大小为  $1024 \times 1024$  或更高的 8 比特图像的存储需求是很重要的。

表 2.1  $N$  和  $k$  取不同值时存储所需的比特数。 $L$  是灰度级数

$N/k$	1 ( $L = 2$ )	2 ( $L = 4$ )	3 ( $L = 8$ )	4 ( $L = 16$ )	5 ( $L = 32$ )	6 ( $L = 64$ )	7 ( $L = 128$ )	8 ( $L = 256$ )
32	1 024	2 048	3 072	4 096	5 120	6 144	7 168	8 192
64	4 096	8 192	12 288	16 384	20 480	24 576	28 672	32 768
128	16 384	32 768	49 152	65 536	81 920	98 304	114 688	131 072
256	65 536	131 072	196 608	262 144	327 680	393 216	458 752	524 288
512	262 144	524 288	786 432	1 048 576	1 310 720	1 572 864	1 835 008	2 097 152
1024	1 048 576	2 097 152	3 145 728	4 194 304	5 242 880	6 291 456	7 340 032	8 388 608
2048	4 194 304	8 388 608	12 582 912	16 777 216	20 971 520	25 165 824	29 369 128	33 554 432
4096	16 777 216	33 554 432	50,331 648	67,108,864	83 886 080	100 663 296	117 440 512	134 217 728
8192	67 108 864	134 217 728	201 326 592	268,435,456	335 544 320	402 653 184	469 762 048	536 870 912

通常,将  $L$  个灰度值缩放至区间  $[0, 1]$  对于计算或算法开发目的是有用的,这时,它们不再是整数。但多数情形下,这些值都会被缩放到用于图像存储和显示的整数区间  $[0, L-1]$ 。

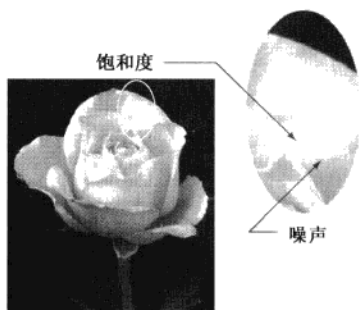


图 2.19 显示了饱和度和噪声的一幅图像。饱和度是指超过这个值的灰度级将被剪切掉这样一个最高值(注意整个饱和区域具有恒定的高灰度级)。这种情况下的噪声表现为粒状纹理模式。噪声,特别是较暗图像区域(譬如玫瑰的茎)中的噪声,掩盖了可检测的最低真实灰度级

### 2.4.3 空间和灰度分辨率

直观上看,空间分辨率是图像中可辨别的最小细节的度量。在数量上,空间分辨率可以有多种方法来说明,其中每单位距离线对数和每单位距离点数(像素数)是最通用的度量。假设我们用交替的黑色和白色垂直线来构造一幅图形,其中线宽为  $W$  个单位( $W$  可以小于 1)。线对的宽度就是  $2W$ , 每单位距离有  $1/2W$  个线对。例如,如果一条线的宽度是 0.1 mm, 每单位距离(mm)就有 5 个线对。广泛使用的图像分辨率的定义是每单位距离可分辨的最大线对数量(譬如每毫米 100 个线对)。每单位距离点数是印刷和出版业中常用的图像分辨率的度量。在美国,这一度量通常使用每英寸点数(dpi)来表示,例如,为给您一个质量概念,报纸用 75 dpi 的分辨率来印刷,杂志是 133 dpi, 光鲜的小册子是 175 dpi, 您正在看的书页是以 2044 dpi 印刷的。

前一段中的关键点是空间分辨率的度量必须针对空间单位来规定才有意义。图像大小本身并不会告诉我们全部内容。如果没有规定图像包含的空间维数,那么我们说一幅图像的分辨率为  $1024 \times 1024$  像素是没有意义的。尺寸本身只是在图像容量间做比较时才有帮助。例如,带有 20 兆像素成像芯片的 CCD 数字摄像机与 8 兆像素的摄像机相比有较高的分辨细节的能力,假定两部摄像机都配备了可比较的镜头,并在相同的距离拍摄可比较的图像。

类似地,灰度分辨率是指在灰度级中可分辨的最小变化。我们有值得考虑的关于产生数字图像的样本数的判断,但是对于灰度级来说这并不真实。基于硬件考虑,正如前一节中提到的那样,灰度级数通常是 2 的整数次幂。最通用的数是 8 比特,在某些特殊的图像增强应用中,用 16 比特也是必要的。灰度量级用 32 比特是很罕见的。有时,我们会发现使用 10 比特或 12 比特来数字化图像灰度级的系统,但这些系统都是特例而不是常规系统。不像空间分辨率必须以每单位距离为基础才有意义,而灰度分辨率指的则是用于量化灰度的比特数。例如,通常说一幅被量化为 256 级的图像有 8 比特的灰度分辨率。因为灰度中可分辨的真实变化不仅受噪声和饱和度值的影响,也受人类感知能力的影响(见 2.1 节),所以说一幅图像有 8 比特的灰度分辨率并不比以灰度幅值为  $1/256$  的固定增量量化为 8 比特的系统能力的规定有更多的内容。

下面的两个例子分别说明了图像尺寸和灰度分辨率对可分辨细节比较的效果。在本节稍后,我们将讨论决定感知图像质量中的两个参数是如何相互影响的。

#### 例 2.2 降低图像空间分辨率的效果说明。

图 2.20 显示了降低一幅图像的空间分辨率的效果。图 2.20(a)到(d)显示了分辨率分别为 1250 dpi, 300 dpi, 150 dpi 和 72 dpi 的图像。很自然,低分辨率的图像与原图像相比要小。例如,原图像的大小为  $3692 \times 2812$  像素,但 72 dpi 图像是一个大小为  $213 \times 162$  的阵列。为便于比较,所有的小图像都放大到了原图像的大小(采用 2.4.4 节中讨论的放大方法)。这对较小图像的归靠来说稍微有些对等之意,因此,我们可就可见细节做可比性的说明。

图 2.20(a)和图 2.20(b)之间有一些较小的视觉差别,最明显的差别是大的黑针稍微有点失真。然而,图 2.20(b)的大部分还是可以接受的。事实上,300 dpi 是书籍印刷所用的最小空间分辨率。因此,这里我们不能期望看到很大的不同。图 2.20(c)显示了可见的退化(例如,计时器的圆形边缘和右侧指向 60 处的小针)。图 2.20(d)显示了图像中多数可见特征的退化。如 4.5.4 节中讨论的那样,当以这样低的分辨率印刷时,印刷和出版业会使用一些“技巧”(例如,像素尺寸的局部变化)来产生比图 2.20(d)更好的结果。此外,如 2.4.4 节所示,选择所用的内插方法也可以改进图 2.20 的结果。

#### 例 2.3 改变数字图像中灰度级数的典型效果。

在这个例子中,我们保持样本数恒定,而将灰度级数以 2 的整数次幂从 256 减少到 2。图 2.21(a)是

一幅大小为  $452 \times 374$  的 CT 投影图像，它以  $k = 8$  (256 个灰度级) 显示。这样的图像是通过将 X 射线源固定在一个位置，然后在任何期望的方向生成二维图像来得到的。投影图像用于指导设置 CT 扫描仪的参数，包括倾斜角、切片数和范围。

图 2.21 (b) 到图 2.21 (h) 是在保持图像大小为  $452 \times 374$  像素不变的情况下，比特数从  $k = 7$  减小到  $k = 1$  时得到的图像。256 级、128 级和 64 级灰度图像对于所有实用目的在视觉上的效果是相同的。然而，图 2.21 (d) 所示的 32 灰度级图像中，在恒定或接近恒度灰度区域 (尤其在头盖骨处) 内有一组不易察觉的细小山脊状结构。这种效果是由数字图像的平滑区域中的灰度级数不足引起的，通常称为伪轮廓，之所以这样称呼，是因为这些山脊状结构类似于地图中的地形轮廓。伪轮廓通常在以 16 或更少级数的均匀设置的灰度级显示的图像中十分明显，如图 2.21 (e) 到图 2.21 (h) 所示的图像。

为方便起见，作为一个粗略的经验规则，假设数值为 2 的整数次幂，一般来说，尺寸为  $256 \times 256$  像素、64 灰度级、印刷在大小为  $5 \text{ cm} \times 5 \text{ cm}$  的版式上的图像是具有最低空间和灰度分辨率的图像，一般认为没有令人讨厌的取样棋盘格和伪轮廓。

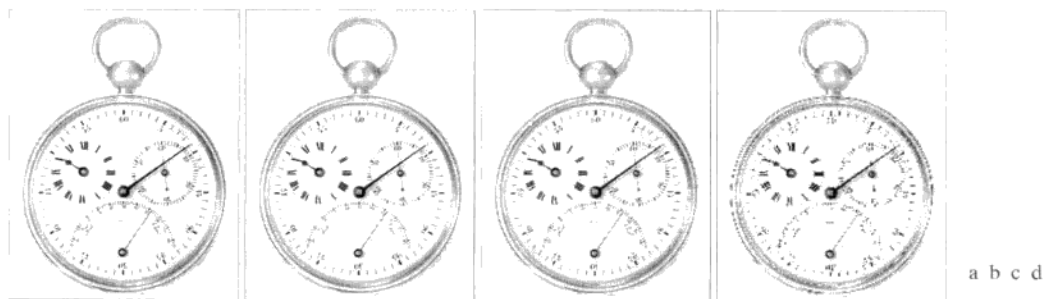


图 2.20 降低空间分辨率的典型效果。图像分别以 (a) 1250 dpi, (b) 300 dpi, (c) 150 dpi, (d) 72 dpi 来显示。为清楚起见，图像添加了黑色的细边框，它们不是数据的一部分

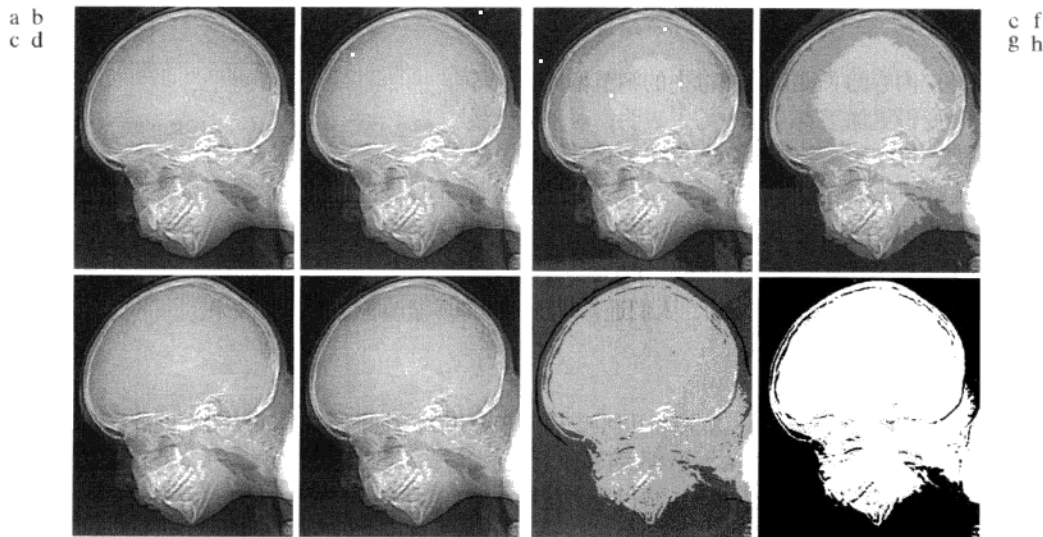


图 2.21 (a) 大小为  $452 \times 374$  的 256 灰度级图像；(b)~(d) 保持图像大小不变的同时以灰度级 128, 64 和 32 显示的图像；(e)~(h) 以 16, 8, 4 和 2 灰度级显示的图像 (原图像由 Vanderbilt 大学医学中心辐射学与放射学系的 David R. Pickens 博士提供)



例 2.2 和例 2.3 中的结果说明了分别改变  $N$  和  $k$  时对图像质量产生的影响。然而, 这些结果也仅仅部分地回答了改变  $N$  和  $k$  如何影响图像的问题, 因为我们还未考虑这两个参数之间可能存在的任何关系。Huang[1965]早期的研究试图通过实验来量化同时改变  $N$  和  $k$  时对图像质量所产生的影响。实验由一组主观测试组成, 采用了类似于图 2.22 中所示的图像。妇女脸庞图像是包含较少细节的一幅有代表性的图像, 摄像师图像包含了中等程度的细节, 人群图像相比之下包含大量的细节。



图 2.22 (a)具有少量细节的图像; (b)具有中等程度细节的图像  
(图像由麻省理工学院提供); (c)具有大量细节的图像

通过改变  $N$  和  $k$ , 生成了一组这三种类型的图像, 然后要求观察者主观地按图像的质量对图像排序。最终结果以  $Nk$  平面中的等偏爱曲线形式汇总(图 2.23 中显示了对应于图 2.22 中图像的平均等偏爱曲线)。 $Nk$  平面中的每一点表示一幅图像, 该图像的  $N$  值和  $k$  值等于该点的坐标。位于等偏爱曲线上的点对应于有相等的主观判定质量的图像。实验发现等偏爱曲线倾向于向右上方移动, 但这三类图像的等偏爱曲线形状与图 2.23 中所示的那些曲线类似。这并不意外, 因为曲线向右上方移动仅意味着更大的  $N$  值和  $k$  值, 而这又意味着更好的图像质量。

当前讨论内容的关键是当图像中的细节增加时, 等偏爱曲线会变得更加垂直。这一结果表明, 对于有大量细节的图像, 可能只需要较少的灰度级。例如, 图 2.23

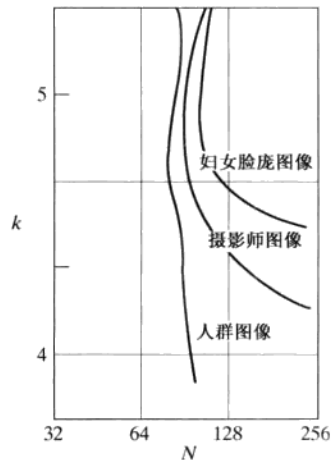


图 2.23 图 2.22 中三类图像的典型等偏爱曲线

中对应于人群图像的等偏爱曲线接近垂直。这指出当  $N$  值固定时, 这类图像的感觉质量与所用的灰度级数(对于图 2.23 所示的灰度级范围)是近似独立的。更有趣的是, 还要注意, 另两类图像在样本数增加的某些间隔内感觉质量保持相同, 但灰度级数实际上降低了。对于这一结果, 最可能的原因是  $k$  的减小倾向于对比度的明显增加, 人们通常感受到图像质量改善了的视觉效果。

#### 2.4.4 图像内插

内插是在诸如放大、收缩、旋转和几何校正等任务中广泛应用的基本工具。这一节我们的主要目标是介绍内插并用它调整图像的大小(收缩和放大), 这是基本的图像重取样方法。在诸如旋转和几何校正的应用中, 内插的用法将在 2.6.5 节讨论。在第 4 章, 我们还将回到这一主题, 以详细地讨论图像重取样。

从根本上看, 内插是用已知数据来估计未知位置的数值的处理。我们用一个简单的例子开始讨论该话题。假设一幅大小为  $500 \times 500$  像素的图像要放大 1.5 倍到  $750 \times 750$  像素。一种简单的放大方法是创建一个假想的  $750 \times 750$  网格, 它与原始图像有相同的间隔, 然后将其收缩, 使它准确地与原图

像匹配。显然，收缩后的  $750 \times 750$  网格的像素间隔要小于原图像的像素间隔。为了对覆盖的每一个点赋以灰度值，我们在原图像中寻找最接近的像素，并把该像素的灰度赋给  $750 \times 750$  网格中的新像素。当我们完成对网格中覆盖的所有点的灰度赋值后，就把图像扩展到原来规定的大小，得到放大后的图像。

刚刚讨论的方法称为最近邻内插法，因为这种方法把原图像中最近邻的灰度赋给了每个新位置（2.5 节将正式讨论邻近像素）。这种方法简单，但是，本节稍后将显示该方法有产生不希望的人为缺陷的倾向，如某些直边缘的严重失真。由于这一原因，实际上该方法并不常用。更实用的方法是双线性内插，在该方法中，我们用 4 个最近邻去估计给定位置的灰度。令  $(x, y)$  为我们想要赋以灰度值的位置（把它想象为前面描述的网格点）的坐标，并令  $v(x, y)$  表示灰度值。对于双线性内插来说，赋值是由下面的公式得到的：

$$v(x, y) = ax + by + cxy + d \quad (2.4-6)$$

与该方法名称相反的是，双线性内插不是一种线性内插方法，因为它其中包含有  $xy$  项。

其中，4 个系数可由 4 个用  $(x, y)$  点最近邻点写出的未知方程确定。

正如您马上将看到的那样，双线性内插给出了比最近邻内插好得多的结果，但随之而来的是计算量的增加。

下一种复杂度较高的方法是双三次内插，它包括 16 个最近邻点。赋予点  $(x, y)$  的灰度值是使用下式得到的：

$$v(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j \quad (2.4-7)$$

其中，16 个系数可由 16 个用  $(x, y)$  点最近邻点写出的未知方程确定。观察式 (2.4-7)，如果求和的上下限分别为 0 和 1，则其简化为式 (2.4-6) 的形式。通常，双三次内插在保持细节方面比双线性内插相对要好。双三次内插是商业图像编辑程序的标准内插方法，例如 Adobe Photoshop 和 Corel Photopaint。

#### 例 2.4 用于图像收缩和放大的内插方法的比较。

图 2.24 (a) 是与图 2.20 (d) 相同的图像，它是由图 2.20 (a) 中分辨率为 1250 dpi 的图像先将分辨率降低到 72 dpi (大小由原来的  $3692 \times 2812$  收缩到  $213 \times 162$  像素)，然后将该缩小的图像再放大到原来的尺寸得到的。图 2.20 (d) 是采用最近邻内插法进行收缩与放大产生的。正如前面讨论的那样，图 2.24 (a) 的结果比较差。图 2.24 (b) 和图 2.24 (c) 是采用相同的过程，但分别采用双线性内插和双三次内插收缩和放大得到的。使用双线性内插得到的结果较最近邻内插有重大改进。双三次内插的结果较双线性内插的结果稍微清晰一些。图 2.24 (d) 与图 2.20 (c) 相同，它们都是由最近邻内插方法通过收缩与放大得到的。我们前面讨论的结论是分辨率降低到 150 dpi 时，图像开始显现退化。图 2.24 (e) 与图 2.24 (f) 分别显示了使用双线性内插和双三次内插收缩和放大图像的结果。尽管分辨率从 1250 dpi 降低到了 150 dpi，但与原图像相比，这两种内插方法再次显示了其强大的性能。与以前一样，双三次内插得到了稍微清晰一些的结果。

在内插中，有可能采用更多的邻点和更复杂的技术，如采用样条和小波，在某种情况下，可以得到比刚刚讨论的方法更好的结果。对于三维图形 (Watt[1993], Shirley[2002]) 和医学图像处理 (Lehmann et al.[1999]) 保留精细细节是一个特别重要的考虑，对于一般目的的数字图像处理，较少考虑额外的计算负担是合理的，所以双线性内插和双三次内插是人们选择的典型方法。

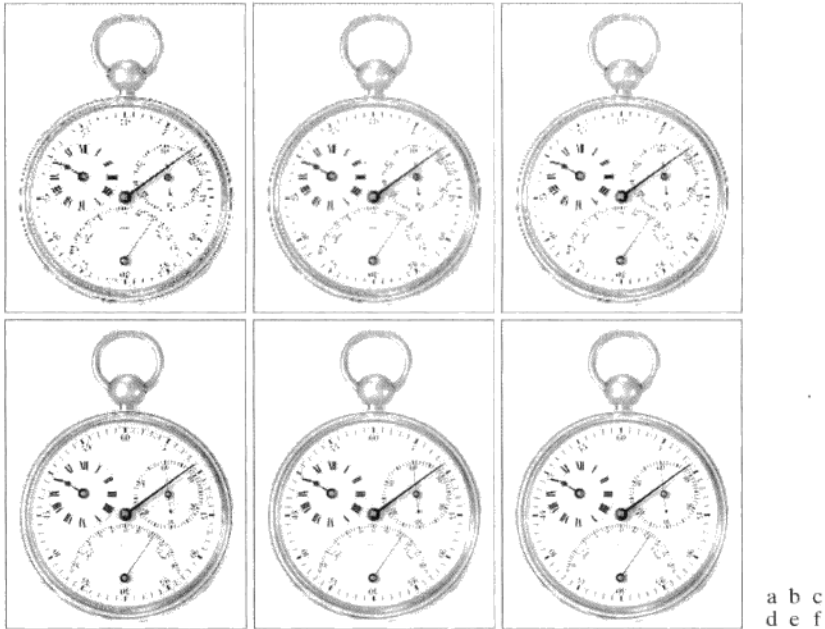


图 2.24 (a)降低到 72 dpi, 并使用双线性内插方法将其放大到原大小(3692x2812 像素)的图像。该图像与图 2.20(d)相同; (b)使用双线性内插收缩和放大的图像; (c)与图像(b)相同但使用双三次内插的图像; (d)~(f)相同的图像序列, 但它们是收缩到 150 dpi 而不是收缩到 72 dpi 的图像 [图 2.24(d)与图 2.20(c)相同]。将图 2.24(e)和(f)与原图 2.20(a)比较, 特别是后者

## 2.5 像素间的一些基本关系

在这一节, 我们考虑数字图像中像素间的几个重要关系。正如前面提及的那样, 图像由  $f(x, y)$  表示。当在本节中引用某个特殊的像素时, 我们通常会使用小写字母, 如  $p$  和  $q$ 。

### 2.5.1 相邻像素

位于坐标  $(x, y)$  处的像素  $p$  有 4 个水平和垂直的相邻像素, 其坐标由下式给出:

$$(x+1, y), (x-1, y), (x, y+1), (x, y-1)$$

这组像素称为  $p$  的 4 邻域, 用  $N_4(p)$  表示。每个像素距  $(x, y)$  一个单位距离, 如果  $(x, y)$  位于图像的边界上, 则  $p$  的某些相邻像素位于数字图像的外部。我们将在第 3 章处理这个问题。

$p$  的 4 个对角相邻像素的坐标如下:

$$(x+1, y+1), (x+1, y-1), (x-1, y+1), (x-1, y-1)$$

并用  $N_D(p)$  表示。这些点与 4 个邻点一起称为  $p$  的 8 邻域, 用  $N_8(p)$  表示。与前面一样, 如果  $(x, y)$  位于图像的边界上, 则  $N_D(p)$  和  $N_8(p)$  中的某些邻点会落入图像的外边。

### 2.5.2 邻接性、连通性、区域和边界

令  $V$  是用于定义邻接性的灰度值集合。在二值图像中, 如果把具有 1 值的像素归诸于邻接像素, 则  $V = \{1\}$ 。在灰度图像中, 概念是一样的, 但是集合  $V$  一般包含有更多的元素。例如, 具有可能的灰度值范围为 0 到 255 的邻接像素中, 集合  $V$  可能是这 256 个值的任何一个子集。考虑三种类型的邻接:

- (a) 4邻接。如果  $q$  在集合  $N_4(p)$  中, 则具有  $V$  中数值的两个像素  $p$  和  $q$  是 4 邻接的。
- (b) 8邻接。如果  $q$  在集合  $N_8(p)$  中, 则具有  $V$  中数值的两个像素  $p$  和  $q$  是 8 邻接的。
- (c)  $m$  邻接(混合邻接)。如果 (i)  $q$  在  $N_4(p)$  中, 或 (ii)  $q$  在  $N_D(p)$  中, 且集合  $N_4(p) \cap N_4(q)$  中没有来自  $V$  中数值的像素, 则具有  $V$  中数值的两个像素  $p$  和  $q$  是  $m$  邻接的。

我们分别使用符号  $\cap$  和  $\cup$  来表示集合的交与并。假设有集合  $A$  和  $B$ , 回忆可知它们的交集是既在集合  $A$  中又在集合  $B$  中的成员的元素集。这两个集合的并集是集合  $A$  的成员、集合  $B$  的成员或两者的成员的元素集。详细讨论见 2.6.4 节。

混合邻接是 8 邻接的改进。混合邻接的引入是为了消除采用 8 邻接时产生的二义性。例如, 考虑图 2.25(a) 中对于  $V = \{1\}$  的像素排列。位于图 2.25(b) 上部的 3 个像素显示了多重(二义性)8 邻接, 如虚线所示。这种二义性可以通过  $m$  邻接消除, 如图 2.25(c) 所示。

从具有坐标  $(x, y)$  的像素  $p$  到具有坐标  $(s, t)$  的像素  $q$  的通路(或曲线)是特定的像素序列, 其坐标为

$$(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$$

其中  $(x_0, y_0) = (x, y), (x_n, y_n) = (s, t)$ , 且像素  $(x_i, y_i)$  和  $(x_{i-1}, y_{i-1})$  对于  $1 \leq i \leq n$  是邻接的。在这种情况下,  $n$  是通路的长度。如果  $(x_0, y_0) = (x_n, y_n)$ , 则通路是闭合通路。可以依据特定的邻接类型定义 4 邻接、8 邻接或  $m$  邻接。例如, 如图 2.25(b) 所示, 右上点和右下点之间的通路是 8 通路, 而图 2.25(c) 中的通路是  $m$  通路。

令  $S$  是图像中的一个像素子集。如果  $S$  的全部像素之间存在一个通路, 则可以说两个像素  $p$  和  $q$  在  $S$  中是连通的。对于  $S$  中的任何像素  $p$ ,  $S$  中连通到该像素的像素集称为  $S$  的连通分量。如果  $S$  仅有一个连通分量, 则集合  $S$  称为连通集。

令  $R$  是图像中的一个像素子集。如果  $R$  是连通集, 则称  $R$  为一个区域。两个区域, 如果它们联合形成一个连通集, 则区域  $R_i$  和  $R_j$  称为邻接区域。不邻接的区域称为不连接区域。在谈到区域时, 我们考虑的是 4 邻接和 8 邻接。为使我们的定义有意义, 必须指定邻接的类型。例如, 如果仅使用 8 邻接时, 则图 2.25(d) 中的两个区域(由 1 组成的)是邻接的(根据前一段的定义, 两个区域之间不存在 4 通路, 它们的并集不是连通集)。

假设一幅图像包含有  $K$  个不连接的区域, 即  $R_k, k = 1, 2, \dots, K$ , 且它们都不接触图像的边界<sup>①</sup>。令  $R_u$  代表所有  $K$  个区域的并集, 并且令  $(R_u)^c$  代表其补集(回忆可知, 集合  $S$  的补集是不在  $S$  中的点的集合)。我们称  $R_u$  中的所有点为图像的前景, 而称  $(R_u)^c$  中的所有点为图像的背景。

区域  $R$  的边界(也称为边缘或轮廓)是这样的点的集合, 这些点与  $R$  的补集中的点邻近。换一种方式说, 一个区域的边界是该区域中至少有一个背景邻点的像素集合。这里再强调一下, 我们必须指定用于定义邻接的连通性。例如, 图 2.25(e) 中被圈出的点如果在区域及其背景间使用 4 连通, 就不是 1 值区域边界的成员。基于这一规则, 为处理这种情况, 一个区域及其背景中的点之间的邻接要根据 8 连通来定义。



图 2.25 (a) 像素的排列; (b) 8 邻接像素(邻接性由虚线所示, 注意二义性); (c)  $m$  邻接; (d) 采用 8 邻接时, 两个值为 1 的区域是邻接的; (e) 如果在区域和背景间采用 8 邻接, 则加圆圈的点是仅赋 1 值的像素的边界点的一部分; (f) 1 值区域的内部边界不形成闭合通路, 但其外部边界可以形成闭合通路

<sup>①</sup> 进行这种假设的目的在于避免处理特殊情况。这样做不会丧失一般性, 因为如果一个或多个区域接触到图像的边界, 我们可简单地使用 1 像素宽的背景值边界来填充图像。

前述定义有时称为区域的内边界,以便与其外边界相区分,外边界对应于背景边界。在开发追踪边界的算法时这一区别很重要。这种算法为了保证结果形成一个闭合通路,通常是沿外边界确立的。例如,在图 2.25(f)中,1 值区域的内边界是该区域本身。这一边界并不满足先前给出的闭合通路的定义。另一方面,区域的外边界确实形成了一个围绕该区域的闭合通路。

如果  $R$  恰巧是整幅图像(我们假设这幅图像是像素的方形集合),则边界由图像第一行、第一列和最后一行、最后一列的像素集合来定义。这个附加定义是需要的,因为一幅图像超过边界范围之外没有邻点。正常情况下,当我们提到一个区域时,指的是一幅图像的子集,并且区域边界中任何与图像边缘吻合的像素都作为区域边界的一部分全部包含于其中。

边缘的概念在涉及区域和边界的讨论中常常遇到。然而,这些概念之间有一个关键区别。一个有限区域的边界形成一条闭合通路,并且是“整体”概念。正如第 10 章详细讨论的那样,边缘由具有某些超过预先设定的阈值的导数值的像素形成。这样,边缘的概念就是基于在进行灰度级度量时不连续点的“局部”概念。把边缘点连接成边缘线段是可能的,并且有时以与边界对应的方法连接线段,但并不总是这样。边缘和边界吻合的一个例外是二值图像的情况。根据连通类型和所用的边缘算子(在第 10 章中讨论),从二值区域提取边缘与区域边界是一样的,这很直观。在概念上,我们在第 10 章才会接触到,把边缘考虑为灰度不连续和边界是闭合通路是有帮助的。

### 2.5.3 距离度量

对于坐标分别为  $(x, y)$ ,  $(s, t)$  和  $(v, w)$  的像素  $p, q$  和  $z$ , 如果

- (a)  $D(p, q) \geq 0$  [ $D(p, q) = 0$ , 当且仅当  $p = q$ ]
- (b)  $D(p, q) = D(q, p)$  且
- (c)  $D(p, z) \leq D(p, q) + D(q, z)$ 。

则  $D$  是距离函数或度量。 $p$  和  $q$  间的欧几里得(欧氏)距离定义如下:

$$D_e(p, q) = \left[ (x-s)^2 + (y-t)^2 \right]^{\frac{1}{2}} \quad (2.5-1)$$

对于距离度量,距点  $(x, y)$  的距离小于或等于某个值  $r$  的像素是中心在  $(x, y)$  且半径为  $r$  的圆平面。

$p$  和  $q$  间的距离  $D_4$  (又称为城市街区距离)由下式定义:

$$D_4(p, q) = |x-s| + |y-t| \quad (2.5-2)$$

在这种情况下,距  $(x, y)$  的距离  $D_4$  小于或等于某个值  $r$  的像素形成一个中心在  $(x, y)$  的菱形。例如,距中心点  $(x, y)$  的距离  $D_4$  小于或等于 2 的像素,形成固定距离的下列轮廓:

$$\begin{array}{ccccccc} & & & & 2 & & \\ & & & & 2 & 1 & 2 \\ & & & 2 & 1 & 0 & 1 & 2 \\ & & & 2 & 1 & 2 & & \\ & & & & 2 & & & \end{array}$$

其中  $D_4 = 1$  的像素是  $(x, y)$  的 4 邻域。

$p$  和  $q$  间的  $D_8$  距离(又称为棋盘距离)由下式定义:

$$D_8(p, q) = \max(|x-s|, |y-t|) \quad (2.5-3)$$

在这种情况下,距  $(x, y)$  的  $D_8$  距离小于或等于某个值  $r$  的像素形成中心在  $(x, y)$  的方形。例如,距中心点  $(x, y)$  的  $D_8$  距离小于或等于 2 的像素形成下列固定距离的轮廓:

$$\begin{array}{ccccc}
 2 & 2 & 2 & 2 & 2 \\
 2 & 1 & 1 & 1 & 2 \\
 2 & 1 & 0 & 1 & 2 \\
 2 & 1 & 1 & 1 & 2 \\
 2 & 2 & 2 & 2 & 2
 \end{array}$$

其中  $D_8 = 1$  的像素是  $(x, y)$  的 8 邻域。

注意,  $p$  和  $q$  之间的  $D_4$  距离和  $D_8$  距离与任何通路无关, 通路可能存在于各点之间, 因为这些距离仅与该点的坐标有关。然而, 如果选择考虑  $m$  邻接, 则两点间的  $D_m$  距离用点间的最短通路定义。在这种情况下, 两个像素间的距离将依赖于沿通路的像素值及其邻点值。例如, 考虑如下排列的像素并假设  $p, p_2$  和  $p_4$  的值为 1,  $p_1$  和  $p_3$  的值为 0 或 1:

$$\begin{array}{cc}
 & p_3 & p_4 \\
 p_1 & & p_2 \\
 p & & 
 \end{array}$$

假设我们考虑值为 1 的像素邻接 [即  $V = \{1\}$ ]。如果  $p_1$  和  $p_3$  是 0, 则  $p$  和  $p_4$  间的最短  $m$  通路的长度 ( $D_m$  距离) 是 2。如果  $p_1$  是 1, 则  $p_2$  和  $p$  将不再是  $m$  邻接的 (见  $m$  邻接的定义), 并且最短  $m$  通路的长度变为 3 (通路通过点  $p, p_1, p_2, p_4$ )。类似地评论, 如果  $p_3$  是 1 (且  $p_1$  是 0), 在这种情况下, 此时最短的  $m$  通路距离也是 3。最后, 如果  $p_1$  和  $p_3$  都为 1, 则  $p$  和  $p_4$  间的最短  $m$  通路的长度为 4, 在这种情况下, 通路通过点  $p, p_1, p_2, p_3, p_4$ 。

## 2.6 数字图像处理中所用数学工具的介绍

这一节有两个主要目的: (1) 向您介绍我们在本书中所用的各种数学工具; (2) 通过把它们用于各种基本图像处理任务来帮助您开始建立如何使用这些工具的“感觉”。它们之中的一些在后续讨论中将被使用多次。在后面的章节中, 我们将根据需要来扩展这些工具的范围及应用。

### 2.6.1 阵列与矩阵操作

包含一幅或多幅图像的阵列操作是以逐像素为基础执行的。本章前面我们曾提到, 图像可以等价地被看成是矩阵。事实上, 在很多情况下, 图像间的操作是用矩阵理论执行的 (见 2.6.6 节)。基于这个原因, 阵列与矩阵操作间的区别必须搞清楚。例如, 考虑下面的  $2 \times 2$  图像:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \text{ 和 } \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

这两幅图像的阵列相乘是

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

另一方面, 矩阵相乘由下式给出:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

我们假设阵列操作贯穿全书, 除非另做说明。例如, 当我们谈到一幅图像的求幂时, 意味着每个像素均进行求幂操作; 当我们谈到一幅图像除以另一幅图像时, 意味着在相应的像素对之间进行相除, 等等。

在继续阅读前, 建议读者从本书网站上的 Tutorials 部分下载并学习复习材料。复习材料包括关于矩阵和向量、线性系统、集合论和概率的简介性知识。

## 2.6.2 线性操作与非线性操作

图像处理方法的最重要分类之一是它是线性的还是非线性的。考虑一般的算子  $H$ ，该算子对于给定的输入图像  $f(x, y)$ ，产生一幅输出图像  $g(x, y)$ ：

$$H[f(x, y)] = g(x, y) \quad (2.6-1)$$

如果

$$H[a_i f_i(x, y) + a_j f_j(x, y)] = a_i H[f_i(x, y)] + a_j H[f_j(x, y)] = a_i g_i(x, y) + a_j g_j(x, y) \quad (2.6-2)$$

则称  $H$  是一个线性算子，其中  $a_i, a_j, f_i(x, y)$  和  $f_j(x, y)$  分别是任意常数和图像(大小相同)。式(2.6-2)指出输出是线性操作，因为两个输入的和与分别对输入进行操作然后再求和得到的结果相同。另外，输入乘以常数的线性操作的输出与乘以该常数的原始输入的操作的输出是相同的。第一个特性称为加性，第二个特性称为同质性。

作为一个简单的例子，假设  $H$  是求和算子  $\Sigma$ ，即该算子的功能是对输入简单地求和。为检验其线性，我们从式(2.6-2)的左侧开始，并且试图证明它与右侧相等：

$$\begin{aligned} \sum [a_i f_i(x, y) + a_j f_j(x, y)] &= \sum a_i f_i(x, y) + \sum a_j f_j(x, y) \\ &= a_i \sum f_i(x, y) + a_j \sum f_j(x, y) \\ &= a_i g_i(x, y) + a_j g_j(x, y) \end{aligned}$$

这些是阵列求和，而不是图像的所有元素的求和。因此，单幅图像的求和是该图像本身。

其中第一步遵循求和是分布式的这样一个事实。因此，左边的展开等于式(2.6-2)的右边，从而我们得出该求和算子是线性的这一结论。

另一方面，考虑最大值操作，其功能是在图像中寻找像素的最大值。针对这一目的，证明该操作是非线性的最简方法是寻找一个测试式(2.6-2)时失败的例子。考虑下列两幅图像：

$$f_1 = \begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix} \text{ 和 } f_2 = \begin{bmatrix} 6 & 5 \\ 4 & 7 \end{bmatrix}$$

并假设令  $a_1 = 1$  和  $a_2 = -1$ 。为了对线性进行测试，我们再次从式(2.6-2)的左侧开始：

$$\max \left\{ (1) \begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix} + (-1) \begin{bmatrix} 6 & 5 \\ 4 & 7 \end{bmatrix} \right\} = \max \left\{ \begin{bmatrix} -6 & -3 \\ -2 & -4 \end{bmatrix} \right\} = -2$$

下一步，做右边，我们得到

$$(1) \max \left\{ \begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix} \right\} + (-1) \max \left\{ \begin{bmatrix} 6 & 5 \\ 4 & 7 \end{bmatrix} \right\} = 3 + (-1)7 = -4$$

在这种情况下，式(2.6-2)的左边和右边并不相等，因此我们证明了通常求最大值的操作是非线性的。

正如您在后面三章中将要看到的那样，特别是在第4章和第5章中，线性操作特别重要，因为它们是以大量的可用于图像处理的理论和实践结果为基础的。非线性系统很难被很好地理解，因此其应用范围受到了更多的限制。然而，在后面的几章中，我们会碰到一些性能远优于线性操作的非线性图像处理操作。

## 2.6.3 算术操作

图像间的算术操作是阵列操作，如2.6.1节中讨论的那样，其意思是算术操作在相应的像素对之间执行。4种算术操作表示为

$$\begin{aligned}
 s(x, y) &= f(x, y) + g(x, y) \\
 d(x, y) &= f(x, y) - g(x, y) \\
 p(x, y) &= f(x, y) \times g(x, y) \\
 v(x, y) &= f(x, y) \div g(x, y)
 \end{aligned}
 \tag{2.6-3}$$

它可理解为是在  $f$  和  $g$  中相应的像素对之间执行操作，其中  $x=0, 1, 2, \dots, M-1$ ， $y=0, 1, 2, \dots, N-1$ 。通常， $M$  和  $N$  是图像的行和列。很清楚， $s, d, p$  和  $v$  是大小为  $M \times N$  的图像。注意，按照刚才定义的方式，图像算术操作涉及相同大小的图像。下面的几个例子指出在数字图像处理中，算术操作起着重要的作用。

### 例 2.5 针对降噪的带噪图像相加(平均)。

令  $g(x, y)$  是无噪声图像  $f(x, y)$  被加性噪声  $\eta(x, y)$  污染后的图像，即

$$g(x, y) = f(x, y) + \eta(x, y) \tag{2.6-4}$$

这里假设在每一对坐标  $(x, y)$  处，噪声是不相关的<sup>①</sup>，并且其均值为零。以下步骤的目的是通过一组带噪图像  $\{g_i(x, y)\}$  的相加减少噪声。

如果噪声满足刚才声明的约束，那么可以证明(见习题 2.20)，如果图像  $\bar{g}(x, y)$  是通过将  $K$  幅不同的噪声图像进行平均形成的：

$$\bar{g}(x, y) = \frac{1}{K} \sum_{i=1}^K g_i(x, y) \tag{2.6-5}$$

然后，遵循

$$E\{\bar{g}(x, y)\} = f(x, y) \tag{2.6-6}$$

和

$$\sigma_{\bar{g}(x, y)}^2 = \frac{1}{K} \sigma_{\eta(x, y)}^2 \tag{2.6-7}$$

式中  $E\{\bar{g}(x, y)\}$  是  $\bar{g}$  的期望值， $\sigma_{\bar{g}(x, y)}^2$  和  $\sigma_{\eta(x, y)}^2$  分别是  $\bar{g}$  和  $\eta$  在所有坐标  $(x, y)$  处的方差。在平均图像中的任意一点处的标准差(方差的均方根)是

$$\sigma_{\bar{g}(x, y)} = \frac{1}{\sqrt{K}} \sigma_{\eta(x, y)} \tag{2.6-8}$$

随着  $K$  的增大，式(2.6-7)和式(2.6-8)指出，在每个位置  $(x, y)$  处的像素值的变化(就像方差或标准差度量的那样)将减小。因为  $E\{\bar{g}(x, y)\} = f(x, y)$ ，这意味着在求平均过程中所使用的带噪图像的数量增加时， $\bar{g}(x, y)$  将逼近  $f(x, y)$ 。实际上，为了避免输出图像带来模糊和其他人为缺陷，图像  $g_i(x, y)$  必须已配准(对齐)。

图像平均的一种重要应用是在天文学领域，在该领域，由于在非常低的照度下成像常常会导致传感器噪声，以至于单幅图像无法分析。图 2.26(a) 显示了一幅 8 比特图像，其中加入了均值为 0、标准差为 64 个灰度级的高斯噪声模拟退化。这些典型的低照度条件下获取的图像，对所有实用目的来说都是不可用的。图 2.26(b)~(f) 显示了分别对 5、10、20、50 和 100 幅图

该例中显示的图像来自称为 NGC 3314 的 Pair 星系，由 NASA 的哈勃空间望远镜拍摄。NGC 3314 离地球约 1.4 亿光年，位于南半球长蛇座方向。最亮的几颗星在由星尘组成的前星系的中心处形成了风车形状。

① 回忆可知，均值为  $m$  的随机变量  $z$  的方差定义为  $E[(z-m)^2]$ ，其中  $E\{\cdot\}$  是该参量的期望值。两个随机变量  $z_i$  和  $z_j$  的协方差定义为  $E[(z_i - m_i)(z_j - m_j)]$ 。如果变量是不相关的，则它们的协方差为零。



像平均的结果。我们看到,图 2.26(e),由  $K=50$  得到的结果已经相当清晰。图 2.26(f)的图像是由 100 幅带噪图像平均得到的结果,与图 2.26(e) 比较仅稍微有些改进。

相加是连续积分的离散形式。在天文观测中,一种与刚刚描述的方法等同的处理是使用 CCD(见 2.3.3 节)或类似传感器的累积能力,通过长时间地观察同一场景来达到降噪的目的。冷却也常常用于降低传感器噪声。然而,最终结果类似于对一组噪声数字图像进行平均操作。

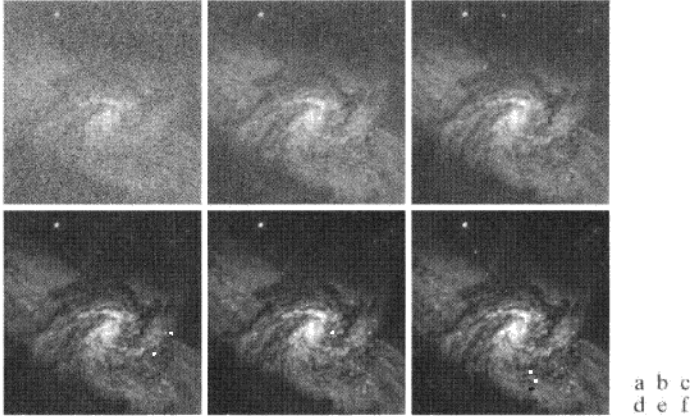


图 2.26 (a) 被加性高斯噪声污染的 Galaxy Pair NGC 3314 图像; (b)~(f) 分别对 5, 10, 20, 50, 100 幅噪声图像平均的结果(原图像由 NASA 提供)

#### 例 2.6 增强差别的图像相减。

图像相减经常用于增强图像之间的差。例如,图 2.27(b)中的图像是通过把图 2.27(a)中的每一个像素的最低有效比特置 0 得到的。在视觉上,这些图像都不可辨别。然而,如图 2.27(c)所示,从一幅图像减去另一幅图像,则清楚地显示了它们的差别。在差值图像中,黑(0)值指出了在该位置图 2.27(a)和图 2.27(b)之间没有差别。

作为另一个例子,我们简单地讨论称之为模板模式放射成像的医学成像领域,这是图像相减在商业上成功的和高受益的应用。考虑如下形式的图像差:

$$g(x, y) = f(x, y) - h(x, y) \quad (2.6-9)$$

在这种情况下,模板  $h(x, y)$  是病人身体一个区域的 X 射线图像,该图像由放在 X 射线源对面的、强化了电视摄像机(代替传统的 X 射线胶片)获取。过程由以下几步组成:将一种 X 射线造影剂注入病人的血流中,摄取一系列与  $h(x, y)$  解剖区域相同的所谓的活体图像 [它们的样品表示为  $f(x, y)$ ], 从一系列注射过造影剂的活体图像中减去模板  $h(x, y)$ 。从每一幅活体图像中减去模板的最终结果是这样的区域,即出现在输出图像  $g(x, y)$  中用以区分  $f(x, y)$  和  $h(x, y)$  之间的差别的区域,目的是增强细节。因为图像可以按电视速率获取,这一过程基本上给出了一段显示造影剂怎样在被观察区域通过动脉传播的影片。

通过图像相减来更改检测也用于图像分割,图像分割是第 10 章的主题。

图 2.28(a)显示了在碘介质注射到血流中前,病人头部上方的一幅模板的 X 射线图像,图 2.28(b)是在注射碘介质后得到的一幅活体图像。图 2.28(c)是图 2.8(a)与图 2.8(b)的差。在这幅图像中,一些细的血管结构清晰可见。图 2.28(d)的差别很清楚,它是由(c)增强了对比度后得到的(下一章中将讨论对比度增强)。图 2.28(d)是介质在大脑内怎样通过血管传播的一幅清晰的“图”。

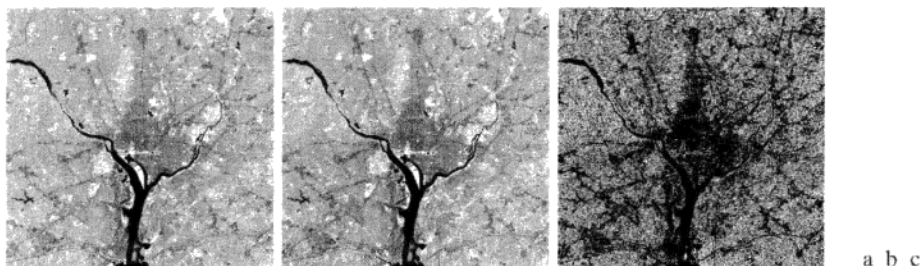


图 2.27 (a)华盛顿特区的红外图像；(b)将图像(a)中的每一个最低阶比特置 0 得到的图像；(c)两幅图像的差，为清楚起见，图像已标定到区间[0, 255]

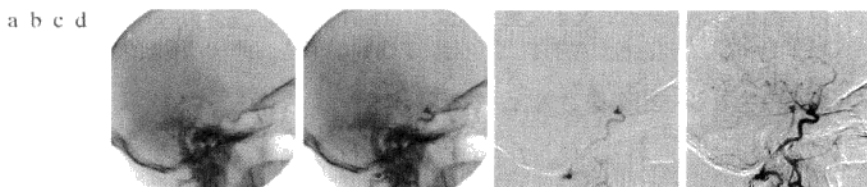


图 2.28 数字减影血管造影：(a)模板图像；(b)活体图像；(c)图(a)和图(b)的差值图像；(d)增强后的差值图像 [图(a)和图(b)由荷兰 Utrecht 大学图像科学研究所医学中心提供]

### 例 2.7 使用图像相乘和相除来校正阴影。

图像相乘(或相除)的一种重要应用是阴影校正。假设一个成像传感器产生可由  $f(x, y)$  表示的完美图像与阴影函数  $h(x, y)$  的乘积来建模的图像，即  $g(x, y) = f(x, y)h(x, y)$ 。如果  $h(x, y)$  已知，那么我们可以使用  $h(x, y)$  的反函数 [即  $g(x, y)/h(x, y)$ ] 乘以感知图像的方法得到  $f(x, y)$ 。如果  $h(x, y)$  未知，但图像系统可以访问，我们可以通过对具有恒定灰度的目标成像得到一个近似的阴影函数。当传感器不可用时，我们常常如在 9.6 节讨论的那样直接从图像估计阴影模式。图 2.29 显示了阴影校正的一个例子。

图像相乘的另一种普通应用是模板操作，也称为感兴趣区域 (ROI) 操作。图 2.30 中说明的处理仅将模板图像与一幅给定的图像相乘，模板图像的 ROI 区域为 1，其他区域为 0。模板图像中的 ROI 可能不止一个，虽然由于实现容易而常常使用矩形形状，但 ROI 的形状可以是任意的。

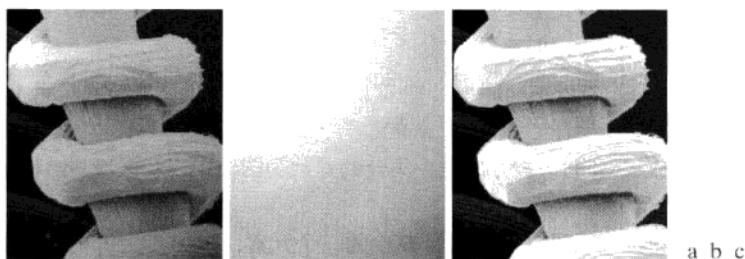


图 2.29 阴影校正：(a)钨丝及其支撑物的阴影 SEM 图像，该图像放大了近 130 倍；(b)阴影模式；(c)图(b)的倒数与图(a)相乘的结果(原图像由俄勒冈 Eugene 大学地质科学系的 Michael Shaffer 提供)

在结束这一节的讨论之前，对图像算术操作的实现做些评述是有益的。实际上，大多数图像使用 8 比特显示(即使 24 比特的彩色图像也是由分立的三个 8 比特通道组成的)。这样，我们认为图像的灰度值范围为 0~255。当图像以标准格式存储时，如 TIFF 或 JPEG，图像的灰度值将自动转换到该范围。然而，转换所用的方法取决于所用的系统。例如，两幅 8 比特图像以不同的取值范围可以是最小的-255 到最大的+255，图像之和的值可能在 0 到 510 的范围。许多软件包在把图像转换为 8 比特

图像时, 仅简单地把所有的负值转换为 0, 而把超过 255 的值转换为 255。给定一幅图像  $f$ , 保证图像间算术操作的整个值域落入某个固定比特数的方法如下。首先, 我们执行操作

$$f_m = f - \min(f) \quad (2.6-10)$$



图 2.30 (a) 牙齿的 X 射线数字图像; (b) 使用填充物隔离牙齿的 ROI 模板 (白色对应于 1, 黑色对应于 0); (c) 图(a)和图(b)相乘的结果

该操作生成其最小值为 0 的一幅图像。然后, 执行操作

$$f_s = K [f_m / \max(f_m)] \quad (2.6-11)$$

该操作生成一幅标定的图像  $f_s$ , 其值在  $[0, K]$  范围内。在处理 8 比特图像时, 置  $K = 255$ , 我们得到一幅灰度范围从 0 到 255 的全部 8 比特的满标度图像。类似的方法可用于 16 比特或更高比特的图像。这种方法可用于所有的算术操作。在执行除法操作时, 我们还有额外的要求, 即需要将一个较小的数加到除数图像的像素上, 以避免用 0 去除。

## 2.6.4 集合和逻辑操作

在这一节, 我们简单介绍一些重要的集合和逻辑操作。我们还将介绍模糊集合的概念。

### 基本集合操作

令  $A$  为一个实数序对组成的集合。如果  $a = (a_1, a_2)$  是  $A$  的一个元素, 则将其写成

$$a \in A \quad (2.6-12)$$

同样, 如果  $a$  不是  $A$  的一个元素, 则写成

$$a \notin A \quad (2.6-13)$$

不包含任何元素的集合称为空集, 用符号  $\emptyset$  表示。

集合由两个大括号中的内容表示, 即  $\{\cdot\}$ 。例如, 当我们将一个表达式写成  $C = \{w | w = -d, d \in D\}$  的形式时, 所表达的意思是: 集合  $C$  是元素  $w$  的集合, 而  $w$  是通过用  $-1$  与集合  $D$  中的所有元素相乘得到的。该集合用于图像处理的一种方法是令集合的元素为图像中表示区域(物体)的像素的坐标(整数序对)。

如果集合  $A$  中的每个元素又是另一个集合  $B$  中的一个元素, 则称  $A$  为  $B$  的子集, 表示为

$$A \subseteq B \quad (2.6-14)$$

两个集合  $A$  和  $B$  的并集表示为

$$C = A \cup B \quad (2.6-15)$$

这个集合包含集合  $A$  和  $B$  中的所有元素。类似地, 两个集合  $A$  和  $B$  的交集表示为

$$D = A \cap B \quad (2.6-16)$$

这个集合包含的元素同时属于集合  $A$  和  $B$ 。如果  $A$  和  $B$  两个集合没有共同的元素, 则称这两个集合是不相容的或互斥的。此时,

$$A \cap B = \emptyset \quad (2.6-17)$$

全集  $U$  是给定应用中的所有元素的集合。根据这一定义，给定应用的所有集合元素是对于该应用所定义的全部成员。例如，如果处理实数集合，则集合的全集是实数域，它包含所有的实数。在图像处理中，我们一般将全集定义为包含图像中所有像素的正方形。

集合  $A$  的补集是不包含于集合  $A$  的元素所组成的集合，表示为

$$A^c = \{w | w \notin A\} \quad (2.6-18)$$

集合  $A$  和  $B$  的差表示为  $A - B$ ，定义为

$$A - B = \{w | w \in A, w \notin B\} = A \cap B^c \quad (2.6-19)$$

我们可以看出，这个集合中的元素属于  $A$ ，而不属于  $B$ 。例如，我们可以根据  $U$  并做集合的差操作来定义  $A^c$ ，即  $A^c = U - A$ 。

图 2.31 说明了前述概念，全集是包含在所示正方形中的坐标的集合，并且集合  $A$  和  $B$  是包含在所示边界内的坐标的集合。在每一幅图中，集合操作的结果用灰色显示<sup>①</sup>。

在前面的讨论中，集合成员是以位置(坐标)为基础的。当处理图像时，一个隐含的假设是集合内所有像素的灰度是相同的，因为我们还没有定义关于灰度值的集合操作(即我们还未指定两个集合的交集的灰度是什么)。图 2.31 所示操作的唯一方法可以这样确定，如果包含这两个集合的图像是二值的，在这种情况下，我们可以说集合成员是基于坐标的，并假设这两个集合的所有成员具有相同的灰度。我们将在下一节做更详细的讨论。

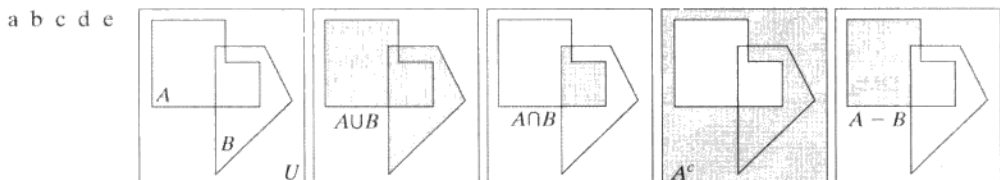


图 2.31 (a) 二维空间中的两个坐标集合  $A$  和  $B$ ; (b)  $A$  和  $B$  的并集; (c)  $A$  和  $B$  的交集; (d)  $A$  的补集; (e)  $A$  和  $B$  的差。在 (b) 到 (e) 中，阴影区域表示指定集合操作的成员

在处理灰度图像时，前述的概念就不适用了，因为我们必须指定来自集合操作结果的所有像素的灰度。事实上，正如您将在 3.8 节和 9.6 节中看到的那样，灰度值的并集操作和交集操作通常分别定义为相应像素对的最大和最小，而补集操作定义为常数与图像中每个像素的灰度间的两两之差。我们处理相应像素对的事实告诉我们，灰度级集合操作是如 2.6.1 节定义的阵列操作。下面的例子是涉及灰度级图像的集合操作的简单说明。我们将进一步讨论上面两节提到的概念。

### 例 2.8 图像灰度的集合操作。

令灰度级图像的元素用集合  $A$  来表示，如 2.4.2 节提到的那样，这些元素是三元组形式  $(x, y, z)$ ，其中  $x$  和  $y$  是空间坐标， $z$  是灰度。我们可以将  $A$  的补集定义为  $A^c = \{(x, y, K - z) | (x, y, z) \in A\}$ ，它简单地表明像素集合  $A$  中的灰度已从常数  $K$  中减去。该常数等于  $2^k - 1$ ，其中  $k$  是用于表示  $z$  的灰度的比特数。令  $A$  表示图 2.32 (a) 中的 8 比特灰度图像，并且假设我们想要用集合操作形成一个负  $A$ 。简单地形成集合  $A_n = A^c = \{(x, y, 255 - z) | (x, y, z) \in A\}$ 。注意，坐标依旧，因此  $A_n$  是与  $A$  大小相同的图像。图 2.32 (b) 示出了这一结果。

<sup>①</sup> 式 (2.6-12) 至式 (2.6-19) 中的操作是代数集的基础，它从  $A \cup B = B \cup A$  和  $A \cap B = B \cap A$  这样的交换律开始，形成了集合操作的广泛理论基础。尽管代数集的处理超出了当前讨论的范围，但读者应了解这一点。

两个灰度集合  $A$  和  $B$  的并集可定义为集合

$$A \cup B = \{\max(a, b) \mid a \in A, b \in B\}$$

也就是说, 两个灰度集合(图像)的并集是一个由空间相应元素对间的最大灰度形成的阵列。再一次注意到坐标同样不变, 因此  $A$  和  $B$  的并集是大小与这两幅图像相同的图像。正如说明的那样, 假设  $A$  仍表示图2.32(a)中的图像, 并且令  $B$  表示与  $A$  大小相同的方形阵列, 但其中所有的  $z$  值等于  $A$  中元素的平均灰度  $m$  的3倍。图2.32(c)显示了执行这种并集操作的结果, 其中所有超过  $3m$  的值呈现为  $A$  的值, 而所有其他像素的值为  $3m$ , 即中间灰度值。

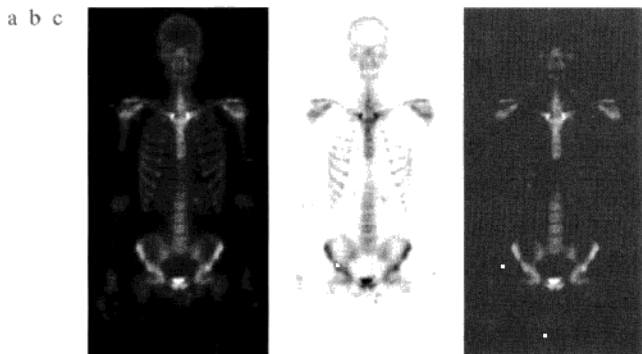


图 2.32 灰度级图像的集合操作: (a)原图像; (b)使用补集操作得到的负像; (c)图(a)与常数图像的并集(原图像由 G. E. Medical Systems 公司提供)

### 逻辑操作

在处理二值图像时, 我们可以把图像想象为像素集合的前景(1 值)与背景(0 值)。然后, 如果我们将区域(目标)定义为由前景像素组成, 则图 2.31 中说明的集合操作就变成了二值图像中目标坐标间的操作。处理二值图像时, OR、AND 和 NOT 逻辑操作就是指普通的并、交和求补操作, 其中“逻辑”一词来自逻辑理论, 在逻辑理论中, 1 代表真, 0 代表假。

考虑由前景像素组成的区域(集合) $A$  和  $B$ 。这两个集合的 OR(或)操作结果不是属于  $A$ , 就是属于  $B$ , 或者属于两者。AND 操作是共同属于  $A$  和  $B$  的元素的集合。集合  $A$  的 NOT 操作是不在  $A$  中的元素的集合。因为我们要处理图像, 如果  $A$  是给定的前景像素的集合, 那么 NOT( $A$ )是图像中不在  $A$  中的所有像素的集合, 这些像素是背景像素, 并有可能是其他前景像素。我们可以将该操作想象为: 把  $A$  中的所有像素转换为 0(黑色), 并把所有不在  $A$  中的元素转换为 1(白色)。

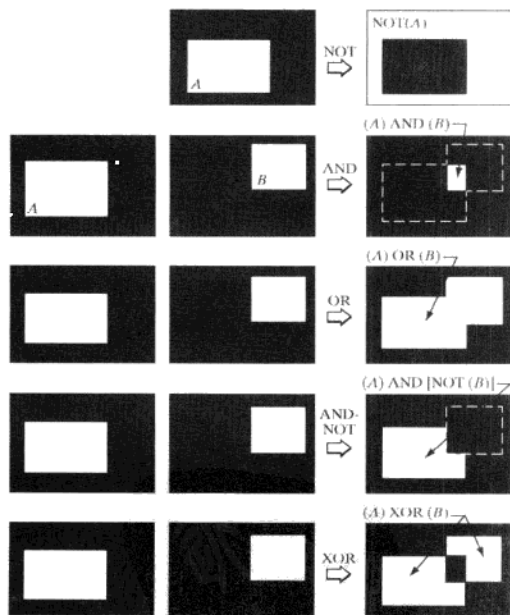


图 2.33 涉及前景(白色)像素的逻辑操作的说明。黑色代表二进制数 0, 白色代表二进制数 1。虚线不是结果的一部分, 仅作为参考显示

图2.33 说明了这些操作。注意，显示该操作结果的第4行是前景像素集合，它属于A但不属于B，这就是式(2.6-19)中集合差的定义。图中的最后一行是XOR(异或)操作的结果，该结果是属于A或B的前景像素的集合，但不是两者的前景像素的集合。可以说，前述操作是区域间的操作，很清楚，区域可以是不规则的和大小不同的。这与早些时候讨论的灰度级操作对立了，前面的灰度操作是阵列操作并要求空间维数相同。也就是说，灰度集合操作涉及整个图像，而不是图像的区域。

我们仅在实现AND, OR和NOT逻辑算子时需要涉及理论知识，因为这三个算子是功能完备的。换句话说，任何其他逻辑操作都可以仅用这三个基本函数执行，如图2.33中的第4行所示，在这里，我们用AND和NOT实现了差集操作。逻辑操作广泛用于图像形态学处理，这是第9章的主题。

### 模糊集合

前述的集合和逻辑结果是“干脆”的概念，在这种意义下，元素要么是或不是集合中的成员。这在某些应用中受到严重限制。考虑一个简单的例子。假如我们希望把世界上所有人分为年轻人和非年轻人。使用“干脆”的集合，令 $U$ 代表所有的人，令 $A$ 是 $U$ 的子集，我们把它称为年轻人集合。为了形成集合 $A$ ，我们需要一个隶属度函数，该函数可以对 $U$ 中的每个元素(人)赋1值或0值。如果 $U$ 中的元素赋值为1，则该元素就是 $A$ 的一个成员，否则就不是 $A$ 的一个成员。因为我们在处理二值逻辑，隶属度函数简单地定义为一个阈值，低于该阈值的人考虑为年轻人，高于该阈值的人考虑为非年轻人。假如我们定义任何20岁或更年轻的人是年轻人。我们面临一个直接的困难：年龄为20岁零1秒的人将不是年轻人集合的成员。这一限制出现在我们用于分类年轻人年龄的阈值上。什么是我们意味的“年轻”需要更多的灵活性，即需要从年轻到非年轻逐步过渡。模糊集合理论使用隶属度函数来实现这种概念，该函数在数值1(定义为年轻)和0(定义为非年轻)之间逐步过渡。使用模糊集合，我们可以声明一个人的年轻度为50%(年轻和非年轻过渡的中间)。换句话说，年龄是一个不精确的概念，而模糊逻辑提供处理这种概念的工具。我们将在3.8节中详细探讨模糊集合。

## 2.6.5 空间操作

空间操作直接在给定图像的像素上执行。我们把空间操作分为三大类：(1)单像素操作；(2)邻域操作；(3)几何空间变换。

### 单像素操作

我们在数字图像中执行的最简单的操作就是以灰度为基础改变单个像素的值。这类处理可以用一个形如下式的变换函数 $T$ 来描述：

$$s = T(z) \quad (2.6-20)$$

其中， $z$ 是原图像中像素的灰度， $s$ 是处理后的图像中相应像素的(映射)灰度。例如，图2.34示出了得到一幅8比特负图像的变换，图2.32(b)中的图像是利用集合操作得到的。我们将在第3章中讨论指定灰度变换功能的一些技术。

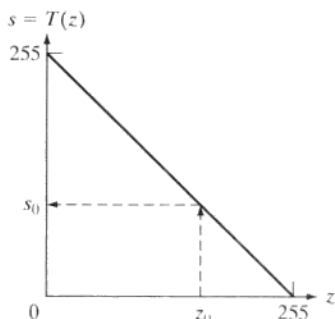


图2.34 用于得到8比特图像的负图像的灰度变换函数。虚线箭头显示了任意输入灰度值 $z_0$ 到对应输出值 $s_0$ 的变换

### 邻域操作

令 $S_{xy}$ 代表图像 $f$ 中以任意一点 $(x, y)$ 为中心的一个邻域的坐标集。邻域处理在输出图像 $g$ 中的相同坐标处生成一个相应的像素，该像素的值由输入图像中坐标在 $S_{xy}$ 内的像素经指定操作决定。例如，

假设指定的操作是计算在大小为  $m \times n$ 、中心在  $(x, y)$  的矩形邻域中的像素的平均值。这个区域中像素的位置组成集合  $S_{xy}$ 。图 2.35 (a) 和 (b) 说明了这一过程。我们可以用公式的形式将这一操作描述为

$$g(x, y) = \frac{1}{mn} \sum_{(r,c) \in S_{xy}} f(r, c) \quad (2.6-21)$$

其中,  $r$  和  $c$  是像素的行和列坐标, 这些坐标是  $S_{xy}$  中的成员。图像  $g$  是这样得到的: 改变坐标  $(x, y)$ , 以便邻域的中心在图像  $f$  中从一个像素到另一个像素移动, 并在每一个新位置重复邻域操作。例如, 图 2.35 (d) 中的图像是使用大小为  $41 \times 41$  的邻域按这种方式创建的。最终结果是在原图像中执行局部模糊。例如, 这种类型的处理可消除小的细节, 并在图像中相应的大区域实施“斑点”补偿。我们将在第 3 章、第 5 章和本书的其他几处地方讨论邻域处理。

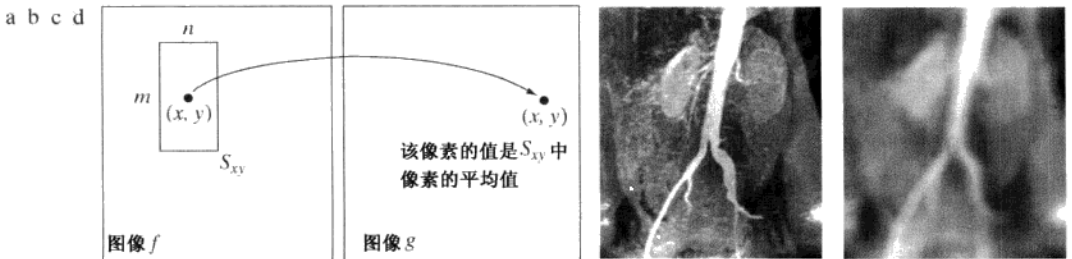


图 2.35 使用邻域处理的局部平均: (a)~(b) 说明了矩形邻域的过程; (c) 1.3.2 节讨论过的大动脉血管造影; (d) 将  $m=n=41$  代入式(2.6-21)后的结果。图像大小为  $790 \times 686$  像素

### 几何空间变换和图像配准

几何变换改进图像中像素间的空间关系。这些变换通常称为橡皮膜变换, 因为它们可看成是在一块橡皮膜上印刷一幅图像, 然后根据预定的一组规则拉伸该薄膜。在数字图像处理中, 几何变换由两个基本操作组成: (1) 坐标的空间变换; (2) 灰度内插, 即对空间变换后的像素赋灰度值。

坐标变换可由下式表示:

$$(x, y) = T\{(v, w)\} \quad (2.6-22)$$

其中,  $(v, w)$  是原图像中像素的坐标,  $(x, y)$  是变换后图像中像素的坐标。例如, 变换  $(x, y) = T\{(v, w)\} = (v/2, w/2)$  在两个方向上把原图像缩小一半。最常用的空间坐标变换之一是仿射变换 (Wolberg[1990]), 其一般形式如下:

$$[x \ y \ 1] = [v \ w \ 1] \mathbf{T} = [v \ w \ 1] \begin{bmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{bmatrix} \quad (2.6-23)$$

这个变换可根据矩阵  $\mathbf{T}$  中元素所选择的值, 对一组坐标点做尺度、旋转、平移或偏移。表 2.2 说明了用于实现这些变换的矩阵值。式(2.6-23)中矩阵表示的真实作用提供了一个把一系列操作连接在一起的框架。例如, 如果我们想要调整一幅图像的大小, 对其旋转, 把处理结果移动到某一位置, 我们可以简单地构成一个  $3 \times 3$  矩阵, 去乘来自表 2.2 的尺度、旋转和平移矩阵。

前面的变换把一幅图像上的像素重新定位到一个新位置。为了完成该处理, 我们还必须对这些新位置赋灰度值。该任务可用灰度内插的方法完成。灰度内插这一主题已在 2.4.4 节讨论过。在那一节开始, 我们用一个放大图像的例子讨论了给新像素位置赋灰度值的问题。正像在表 2.2 中第二行详细

表述的那样，放大是简单的尺度变换，并且其分析类似于我们对图像放大那样，可用于从表 2.2 的其他变换得到的对重定位像素赋灰度值问题。如 2.44 节那样，在做这些变换时，我们可考虑最近邻、双线性和双三次内插技术。

表 2.2 基于式(2.6-23)的仿射变换

变换名称	仿射矩阵 $T$	坐标公式	例子
恒等变换	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x &= v \\ y &= w \end{aligned}$	
尺度变换	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x &= c_x v \\ y &= c_y w \end{aligned}$	
旋转变换	$\begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x &= v \cos \theta - w \sin \theta \\ y &= v \sin \theta + w \cos \theta \end{aligned}$	
平移变换	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ t_x & t_y & 1 \end{bmatrix}$	$\begin{aligned} x &= v + t_x \\ y &= w + t_y \end{aligned}$	
(垂直) 偏移变换	$\begin{bmatrix} 1 & 0 & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x &= v s_v + w \\ y &= w \end{aligned}$	
(水平) 偏移变换	$\begin{bmatrix} 1 & s_h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{aligned} x &= v \\ y &= s_h v + w \end{aligned}$	

实际上，我们可以按两种基本方法来使用式(2.6-23)。第一种方法称为前向映射，它由扫描输入图像的像素，并在每个位置 $(v, w)$ 用式(2.6-23)直接计算输出图像中相应像素的空间位置 $(x, y)$ 组成。前向映射算法的一个问题是输入图像中的两个或更多个像素可被变换到输出图像中的同一位置，这就产生了如何把多个输出值合并到一个输出像素的问题。另外，一种可能是某些输出位置可能完全没有要赋值的像素。第二种方法，称为反向映射，扫描输出像素的位置，并在每一个位置 $(x, y)$ 使用 $(v, w) = T^{-1}(x, y)$ 计算输入图像中的相应位置。然后，内插(使用 2.44 节讨论的技术之一)是用最近的输入像素之一决定输出像素的灰度值的。对于实现来说，反向映射比前向映射更有效，因而被许多空间变换的商业实现所采用(如 MATLAB 就采用该方法)。

### 例 2.9 图像旋转与灰度内插。

本例的目的是用仿射变换说明图像旋转。图 2.36(a)显示了一幅 300 dpi 的图像，图 2.36(b)~(d)是将原图像旋转  $21^\circ$  并分别使用最近邻、双线性和双三次内插后的结果。旋转是在保持直线特性方面最苛求的几何变换之一。正如我们在图中所看到的那样，最近邻内插产生了最大的锯齿边缘，如 2.4.4 节那样，双线性内插得到了明显改进的结果。如以前那样，使用双三次内插产生了稍微清晰一些的结果。事实上，



如果您比较图 2.36(c) 和图 2.36(d) 中放大的细节, 您将在子图像的中间注意到由亮到暗变换所引入的大量垂直灰度块, 而图 2.36(c) 中的垂直灰度块数量要比图 2.36(d) 中的相应垂直灰度块数量多, 说明字母的边缘更清晰。类似的结果可用表 2.2 中的其他空间变换按要求的内插得到(恒等变换不会得到类似的结果, 增量为整数个像素的平移变换也不会得到类似的结果)。这个例子是使用前一段讨论的反向映射方法实现的。



图 2.36 (a) 字母 T 的 300 dpi 图像; (b) 旋转 21° 并用最近邻内插给经空间变换后的像素赋灰度值后的图像; (c) 旋转 21° 并用双线性内插赋灰度值后的图像; (d) 旋转 21° 并用双三次内插赋灰度值后的图像。放大部分显示了三种内插方法的边缘细节

图像配准是数字图像处理的一种重要应用, 用于对齐两幅或多幅相同场景的图像。在前述讨论中, 已经知道了为达到希望的几何变换的变换函数形式。在图像配准中, 我们有可用的输入图像和输出图像, 但是, 从输入图像产生输出图像的特定变换通常是不知道的。这样, 问题就是估计变换函数, 然后用它配准两幅图像。为搞清楚该术语, 输入图像是我们希望变换的图像, 而我们所谓的参考图像是我们想要配准输入图像的图像。

例如, 感兴趣的是对齐(配准)在近似相同时间但使用不同成像系统获取的两幅或多幅图像, 如一幅是用 MRI(核磁共振成像)扫描器获取的图像, 另一幅是用 PET(正电子断层)扫描器获取的图像; 或在不同的时间使用相同的设备获取的图像, 如给定位置间隔几天、几个月甚至几年的卫星图像。无论哪种情况, 合并这些图像或执行定量分析, 以及做比较, 都要求对由观察视角、距离、方向、传感器分辨率、目标位置移动及其他因素导致的几何畸变进行修正。

解决刚才讨论的这个问题的主要方法之一是使用约束点(也称为控制点), 这些点是在输入图像和参考图像中其位置恰好已知的相应点。有很多选择约束点的方法, 可以交互式地选择这些点, 也可以使用试图自动检测这些点的算法。在某些应用中, 成像系统在成像传感器中会嵌入有形的人为物质(如小的金属物)。这会在系统获取的图像上直接产生一系列已知点(称为网状标记), 这些点都用做建立约束点的指导。

估计变换函数问题是建模问题之一。例如, 假设在输入图像和参考图像中有一组 4 个约束点。基于双线性近似的简单模型由如下两式给出:

$$x = c_1v + c_2w + c_3vw + c_4 \quad (2.6-24)$$

和

$$y = c_5v + c_6w + c_7vw + c_8 \quad (2.6-25)$$

其中, 在估计阶段,  $(v, w)$  和  $(x, y)$  分别是输入图像和参考图像中约束点的坐标。如果在两幅图像中我们有 4 对相应的约束点, 则可用式(2.6-24)和式(2.6-25)写出 8 个方程, 然后用它们解出 8 个未知参数  $c_1, c_2, \dots, c_8$ 。这些参数就组成了一个模型, 该模型把一幅图像的像素变换为另一幅需要配准的图像的像素位置。

一旦我们有了系数, 式(2.6-24)和式(2.6-25)就成为我们变换输入图像中的所有像素并生成期望的新图像的工具, 如果约束点选择得正确, 新图像就能与参考图像配准。在 4 个约束点不足以得到令

人满意的配准的情况下,常常使用的一种方法是选择大量的约束点,如同一个子图像那样,用一组4个约束点形成四边形来处理。子图像用上述方法处理,四边形内部的所有像素使用由这些约束点求出的系数进行变换。然后,移动到另一组4约束点,并重复上述过程,直到所有的四边形区域均被处理为止。当然,也可使用比四边形更复杂的区域并采用更复杂的模型,例如使用最小均方算法的拟合多项式。通常,控制点的数量和解决问题所要求的模型的复杂程度取决于几何畸变的严重程度。最后,应牢记式(2.6-24)和式(2.6-25)所定义的变换,或任何针对该问题的其他模型,以及输入图像中像素空间坐标的简单映射。我们还需要使用前面讨论的方法执行灰度内插,以便给这些像素灰度值赋。

### 例 2.10 图像配准。

图2.37(a)显示了一幅参考图像,图2.37(b)显示了相同的图像,但因垂直和水平切变产生了几何畸变。我们的目标是使用参考图像得到约束点,然后用约束点配准这两幅图像。我们(手工)选择的约束点靠近图像的角点,由白色小方形点表示(因为畸变是在两个方向上的线性切变,因此仅需要4个约束点)。图2.37(c)显示了使用前面讨论过的这些约束点来实现配准的结果。我们注意到,配准并不完美,因为图2.37(c)中有明显的黑边。图2.37(d)更清楚地显示了参考图像和校正后的图像间配准的细小不足。导致这种差别的原因是手工选择约束点时的误差。当畸变很严重时,使用约束点很难实现完美的匹配。



图 2.37 图像配准: (a) 参考图像; (b) 输入图像(几何畸变图像), 相应的约束点以靠近角点的白色小方块显示; (c) 配准后的图像(注意边界上的误差); (d) 图(a)和图(c)的差, 显示了更多的配准误差

## 2.6.6 向量与矩阵操作

多光谱图像处理是使用向量和矩阵操作的典型领域。例如,在第6章中,您将学习在RGB彩色空间中如何使用红色、绿色、蓝色分量图像来形成如图2.38说明的彩色图像。这里,我们看到RGB图像的每一个像素都有三个分量,这些分量可组织成一个列向量的形式:

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \quad (2.6-26)$$

其中,  $z_1$  是红色图像中像素的亮度,其他两个元素是相应的绿色图像和蓝色图像中像素的亮度。这样,大小为  $M \times N$  的RGB彩色图像就可用这一大小的三个分量图像来表示,或者总共用  $MN$  个三维向量来表示。普通的多光谱情况包括  $n$  幅分量图像(见图1.10),将形成  $n$  维向量。在第6章、第10章、第11章和第12章,我们都使用这种形式的向量表示。

一旦像素被表示为向量,我们就可以使用向量矩阵理论这一工具。例如,一个像素向量  $\mathbf{z}$  和一个任意点  $\mathbf{a}$  在  $n$  维空间的欧氏距离  $D$  可以用一个向量积来定义:

关于向量和矩阵的简单介绍,请读者参阅本网站上的Tutorial部分。

$$D(\mathbf{z}, \mathbf{a}) = \left[ (\mathbf{z} - \mathbf{a})^T (\mathbf{z} - \mathbf{a}) \right]^{\frac{1}{2}} = \left[ (z_1 - a_1)^2 + (z_2 - a_2)^2 + \cdots + (z_n - a_n)^2 \right]^{\frac{1}{2}} \quad (2.6-27)$$

我们看到,这是式(2.5-1)定义的二维欧氏距离的通式。式(2.6-27)有时称为向量范数,表示为 $\|\mathbf{z} - \mathbf{a}\|$ 。在后面的章节中,我们将多次使用距离计算。

像素向量的另一个重要优点是由下式表示的线性变换:

$$\mathbf{w} = \mathbf{A}(\mathbf{z} - \mathbf{a}) \quad (2.6-28)$$

其中,  $\mathbf{A}$  是大小为  $m \times n$  的矩阵,  $\mathbf{z}$  和  $\mathbf{a}$  是  $n \times 1$  大小的列向量。正如您将要学到的那样,这种类型的变换在图像处理中有很多有用的用途。

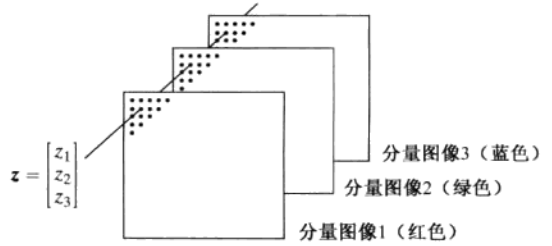


图 2.38 三幅 RGB 分量图像中的相应像素值形成一个向量

如式(2.4-2)表示的那样,整个图像可当做矩阵(或向量)来处理,这是在解决大量数字图像处理问题时有重要意义的事实。例如,我们可以把一幅大小为  $M \times N$  的图像描述为一个  $MN \times 1$  维向量,方法是用图像的第一行作为该向量的前  $N$  个元素,第二行作为下  $N$  个元素,等等。用这种方法形成的图像,我们可以表达应用更广范围的、下列表示的图像线性处理:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \mathbf{n} \quad (2.6-29)$$

其中,  $\mathbf{f}$  表示输入图像的  $MN \times 1$  向量,  $\mathbf{n}$  表示一个  $M \times N$  噪声模式的  $MN \times 1$  向量,  $\mathbf{g}$  表示处理后图像的  $MN \times 1$  向量,  $\mathbf{H}$  是用于对输入图像进行线性处理的  $MN \times MN$  矩阵(见 2.6.2 节关于线性处理的介绍)。例如,像在 5.9 节将要讨论的那样,使用式(2.6-29)为图像复原开发一整套通用技术是可能的。在下一节中,我们会再次接触矩阵这一主题,并在第 5 章、第 8 章、第 11 章和第 12 章显示矩阵在图像处理中的其他用途。

## 2.6.7 图像变换

到目前为止,讨论的所有图像处理方法都是直接在图像像素上进行操作,即直接工作在空间域。在有些情况下,通过变换输入图像来表达图像处理任务,在变换域执行指定的任务,之后再反变换返回到空间域会更好。随着本书学习的深入,您会遇到许多不同的变换。表示为  $T(u, v)$  的二维线性变换是一类特别重要的变换,其通用形式可表达为

$$T(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)r(x, y, u, v) \quad (2.6-30)$$

其中,  $f(x, y)$  是输入图像,  $r(x, y, u, v)$  称为正变换核,式(2.6-30)对  $u = 0, 1, 2, \dots, M-1$  和  $v = 0, 1, 2, \dots, N-1$  进行计算。与以前一样,  $x$  和  $y$  是空间变量,  $M$  和  $N$  是  $f$  的行和列,  $u$  和  $v$  称为变换变量。 $T(u, v)$  称为  $f(x, y)$  的正变换。给定  $T(u, v)$  后,我们可以用  $T(u, v)$  的反变换还原  $f(x, y)$ :

$$f(x, y) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u, v)s(x, y, u, v) \quad (2.6-31)$$

其中,  $x = 0, 1, 2, \dots, M-1, y = 0, 1, 2, \dots, N-1, s(x, y, u, v)$  称为反变换核。式(2.6-30)和式(2.6-31)一起称为变换对。

图 2.39 显示了在线性变换域执行图像处理的基本步骤。首先,变换输入图像,然后用预定义的操作修改该变换,最后,输出图像由计算修改后的变换的反变换得到。这样,我们可以看出,该过程是从空间域到变换域,然后返回到空间域。



图 2.39 线性变换域中操作的一般方法

**例 2.11 变换域图像处理。**

图 2.40 显示了图 2.39 中步骤的一个例子。在这种情况下，所用的变换是傅里叶变换，在本节稍后我们将简单提一下它，第 4 章将对其进行详细介绍。图 2.40(a) 是一幅被正弦干扰污染了的图像，图 2.40(b) 是该图像傅里叶变换的幅度，它是图 2.39 中第一阶段的输出。正如您将在第 4 章学习的那样，空间域中的正弦干扰在变换域以较亮的脉冲方式出现。在这种情况下，脉冲是图 2.40(b) 中可见的圆形模式。图 2.40(c) 显示了一个模板图像(称为滤波器)，白和黑分别代表 1 和 0。对于该例子，图 2.39 第二个方框内的操作是变换的结果与模板相乘，以可靠地消除干扰脉冲。图 2.40(d) 显示了最终结果，该结果是通过计算修改后的变换的反变换得到的。事实上，您恰好可以看到基准标志(淡淡的十字)，它用于图像对准。

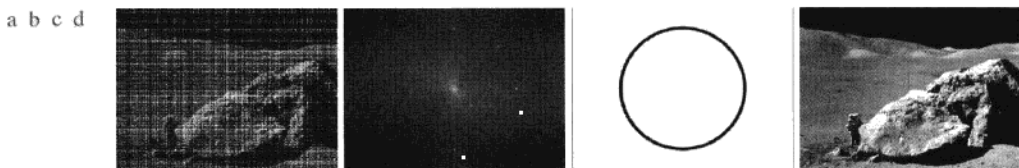


图 2.40 (a) 被正弦干扰污染了的图像；(b) 傅里叶变换的幅度，显示了干扰能量脉冲；(c) 用于消除能量脉冲的模板；(d) 计算修改后的傅里叶变换的反变换的结果(原图像由 NASA 提供)

如果

$$r(x, y, u, v) = r_1(x, u)r_2(y, v) \quad (2.6-32)$$

那么所谓的正向变换核是可分的。另外，如果  $r_1(x, y)$  等于  $r_2(x, y)$ ，则称变换核是对称的，从而有

$$r(x, y, u, v) = r_1(x, u)r_1(y, v) \quad (2.6-33)$$

在前面的公式中，若用  $s$  代替  $r$ ，则同样的说明适用于反变换核。

例 2.11 中讨论的二维傅里叶变换有如下正、反变换核：

$$r(x, y, u, v) = e^{-j2\pi(ux/M+vy/N)} \quad (2.6-34)$$

和

$$s(x, y, u, v) = \frac{1}{MN} e^{j2\pi(ux/M+vy/N)} \quad (2.6-35)$$

其中， $j = \sqrt{-1}$ ，因此这些核是复数。将这些核代入式(2.6-30)和式(2.6-31)给出的通用变换公式中，可得出离散傅里叶变换对：

$$T(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(ux/M+vy/N)} \quad (2.6-36)$$

和

$$f(x, y) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u, v) e^{j2\pi(ux/M+vy/N)} \quad (2.6-37)$$

从基础意义上说，这些公式在数字图像处理中是很重要的，在第 4 章中，我们将投入更多的篇幅从基本原理入手推导它们，然后在较宽的应用范围使用它们。

证明傅里叶核的可分性和对称性并不困难(见习题 2.25), 并且可分的和对称的核允许用一维傅里叶变换计算二维傅里叶变换(见习题 2.26)。当变换对的正、反变换核满足这两个条件, 且  $f(x, y)$  是大小为  $M \times M$  的方形图像时, 式(2.6-30)和式(2.6-31)可表示为矩阵形式:

$$T = AFA \quad (2.6-38)$$

其中,  $F$  是包含元素  $f(x, y)$  的  $M \times M$  矩阵 [ 见式(2.4-2) ];  $A$  是有元素  $a_{ij} = \tau_1(i, j)$  的  $M \times M$  矩阵,  $T$  是  $M \times M$  变换的结果, 其值为  $T(u, v)$ ,  $u, v = 0, 1, 2, \dots, M-1$ 。

为得到反变换, 我们用反变换矩阵  $B$  前乘和后乘式(2.6-38):

$$BTB = BAFAB \quad (2.6-39)$$

如果  $B = A^{-1}$ , 则

$$F = BTB \quad (2.6-40)$$

该式指出  $F$  [ 其所有元素均等于  $f(x, y)$  ] 可完全由其正向变换复原。如果  $B \neq A^{-1}$ , 则用式(2.6-40)可得到一个近似:

$$\hat{F} = BAFAB \quad (2.6-41)$$

除了傅里叶变换之外, 一些重要的变换, 如沃尔什变换、哈达玛变换、离散余弦变换、哈尔变换和斜变换等, 可用式(2.6-30)和式(2.6-31)的形式描述, 或等价地用式(2.6-38)和式(2.6-40)的形式描述。我们将在后续章节讨论这些变换和其他类型的图像变换。

## 2.6.8 概率方法

概率以很多方式用于图像处理工作中。最简单的方式是当我们以随机量处理灰度值时。例如, 令  $z_i, i = 0, 1, 2, \dots, L-1$  表示一幅  $M \times N$  大小数字图像中所有可能的灰度值, 则在给定图像中灰度级  $z_k$  出现的概率  $p(z_k)$  可估计为

$$p(z_k) = \frac{n_k}{MN} \quad (2.6-42)$$

其中,  $n_k$  是灰度  $z_k$  在图像中出现的次数,  $MN$  是像素总数。显然,

$$\sum_{k=0}^{L-1} p(z_k) = 1 \quad (2.6-43)$$

一旦我们知道了  $p(z_k)$ , 就可以得出许多重要的图像特性。例如, 平均灰度由下式给出:

$$m = \sum_{k=0}^{L-1} z_k p(z_k) \quad (2.6-44)$$

类似地, 灰度的方差是

$$\sigma^2 = \sum_{k=0}^{L-1} (z_k - m)^2 p(z_k) \quad (2.6-45)$$

方差是  $z$  值关于均值的展开度的度量, 因此它是图像对比度的有用度量。通常, 随机变量  $z$  关于均值的第  $n$  阶矩定义为

$$\mu_n(z) = \sum_{k=0}^{L-1} (z_k - m)^n p(z_k) \quad (2.6-46)$$

我们看到,  $\mu_0(z) = 1, \mu_1(z) = 0$  且  $\mu_2(z) = \sigma^2$ 。反之, 均值和方差对于图像的视觉特性有明显的直接

关于概率论的简介, 请读者  
查阅本书网站上的 Tutorials 部分。

关系，高阶矩更敏感。例如，一个正三阶矩指出其灰度倾向于比均值高，负三阶矩则指出相反的条件，并且零三阶矩告诉我们灰度近似相等地分布在均值的两侧。这些特性对于计算目的很有用，但它们一般不能告诉我们更多图像外观的内容。

### 例 2.12 图像灰度对比度度量时标准差值的比较。

图 2.41 显示了分别展示低、中、高对比度的三幅 8 比特图像。在这三幅图像中，像素灰度的标准差分别是 14.3、31.6 和 49.2 个灰度级。相应的方差值分别是 204.3、997.8、2424.9。两组数值告诉我们相同的内容，但这三幅图像可能的灰度值范围是  $[0, 255]$ ，与该范围相关的标准差值与方差相比要直观得多。

方差的单位是灰度值的平方。在比较对比度值时，我们通常替代使用标准差  $\sigma$  (方差的均方根)，因为其维数直接取决于灰度值。

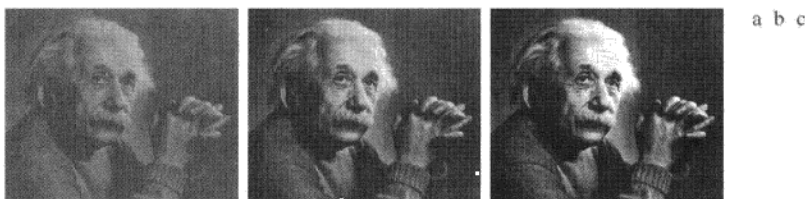


图 2.41 展示了 (a) 低对比度、(b) 中对比度和 (c) 高对比度的图像

在本书的进展中，您将看到，在开发图像处理算法中，概率的概念起着核心角色的作用。例如，在第 3 章，我们利用式 (2.6-42) 的概率度量推导灰度变换算法。在第 5 章中，我们使用概率和矩阵公式开发图像复原算法。在第 10 章中，概率被用于图像分割，而在第 11 章中，我们用概率描述纹理。在第 12 章中，我们以概率公式为基础，导出最佳目标识别技术。

到目前为止，我们已经讨论了将概率应用到单幅二维图像上的单个随机变量(灰度)的问题。如果考虑序列图像，我们可能解释成为时间的第三个变量。处理这一附加复杂性的工具是随机图像处理技术(“stochastic”一词来自希腊语，其大意是在处理结果中随意地瞄准目标)。我们可以进一步将整个图像考虑为一个空间随机事件(相对于一个点)。基于这一概念所需要的处理工具是来自随机域的技术。在 5.8 节中，我们将给出一个例子，说明如何像随机事件那样处理整个图像，但关于随机过程和随机域的讨论超出了本书的范围。本节末尾提供的参考文献是阅读这些主题的起点。

## 小结

本章内容为后续讨论提供了主要的背景知识。对人类视觉系统的论述虽然简单，但提供了人眼感知图像信息能力的一个基本概念。关于光和电磁波谱的讨论是理解本书所用的许多图像来源的基础。类似地，2.3.4 节给出的图像模型也是第 4 章中称为同态滤波图像增强技术的基础。

2.4 节介绍的取样和内插的概念是您在实践中可能遇到的许多数字现象的基础。当您掌握了傅里叶变换和频率域的概念后，在第 4 章中我们还将回到取样及其分支问题的讨论。

2.5 节介绍的概念是基于像素邻域处理技术的基本组成部分。例如，如下一章和第 5 章所示的那样，邻域处理方法是许多图像增强和复原过程的核心。在第 9 章中，我们将采用邻域操作对图像做形态学处理，在第 10 章中，我们用邻域处理进行图像分割，在第 11 章中，我们用它们进行图像描述。在应用方面，邻域处理由于处理速度快，且硬件和/或固件实现简单，故在图像处理商业应用方面得到了人们的偏爱。

2.6 节的内容在您阅读全书时始终很好地为您提供帮助。虽然讨论的水平确实是介绍性的，但对

于您建立数字图像处理的概念是合适的。正如在该节中我们提到的那样,介绍的那些工具在后续章节中是必需的。不是一个章节或附录这样一个地方就能对数学概念充分地论述,您将发现,在后续章节中对来自 2.6 节的数学工具进行必要的扩展,及如何运用它们解决图像处理的问题来说更有意义。

## 参考文献

2.1 节中有关人眼结构的阅读资料可在 Atchison 和 Smith[2000]及 Oyster[1999]中找到。阅读资料见 Regan[2000]和 Gordon[1997]。Hubel[1988]的书和 Cornsweet[1970]的经典书籍也很重要。Born 和 Wolf[1999]是讨论光和电磁波理论的基本参考资料。电磁能传播包含在 Felsen 和 Marcuvitz[1994]的详细论述中。

图像感知领域非常宽泛且发展非常快。关于光传感器和其他成像传感器的一个优秀信息源是国际光学工程学会(SPIE)。下面是 SPIE 在这一领域具有代表性的出版物:Blouke 等人[2001]、Hoover 和 Doty[1996]及 Freeman[1987]的著述。

2.3.4 节给出的图像模型来自 Oppenheim, Schafer 和 Stockham[1968]的论著。该节中所用的照度值和反射值的参考文献是(IESNA 光学手册)IESNA *Lighting Handbook*[2000]提供的。关于图像取样及其一些效应(譬如混淆效应)的其他读物见 Bracewell[1995]。我们将在第 4 章中更详细地讨论这个主题。2.4.3 节提到的作为取样和量化的函数的感知图像质量方面的早期实验是由 Huang[1965]报告的。在降低图像取样数和灰度级的同时,尽量减少随之而来的退化仍是当前人们感兴趣的课题,如 Papamarkos 和 Atsalakis[2000]的例证。关于图像收缩和放大方面的进一步读物见 Sid-Ahmed[1995], Unser 等人[1995], Umbaugh[2005]和 Lehmann 等人[1999]的文献。2.5 节涵盖主题的进一步读物,见 Rosenfeld 和 Kak[1982], Marchand-Maillet 和 Sharaiha [2000]及 Ritter 和 Wilson[2001]提供的材料。


图像处理中关于线性系统(见 2.6.2 节)的附加读物可在 Castleman[1996]中找到。使用图像平均来消除噪声的方法(见 2.6.3 节)由 Kohler 和 Howell[1963]首次提出。关于随机变量求和的均值和方差的期望值的材料见 Peebles[1993]。图像相减(见 2.6.3 节)是一种广泛用于检测变化的通用图像处理工具。为使得图像相减有意义,配准被减图像是必要的,或者必须识别出由于移动导致的人为缺陷。由 Meijering 等人[1999, 2001]撰写的两篇论文给出了用于实现这些目的的技术。

2.6.4 节的基本参考文献是 Cameron[2005]。这一主题的更高级的读物见 Tourlakis[2003]的著作。关于模糊集合的介绍见 3.8 节和第 3 章中的相关参考文献。单点和邻域处理(见 2.6.5 节)的进一步细节见 3.2 节到 3.4 节及第 3 章中关于这些主题的参考文献。关于几何空间变换见 Wolberg[1990]的材料。

Noble 和 Daniel[1988]的著述是矩阵和向量操作(2.6.6 节)的基本参考材料。关于傅里叶变换(见 2.6.7 节)的详细讨论见第 4 章,数字图像处理中使用的其他类型变换的例子见第 7 章、第 8 章和第 11 章。关于概率和随机变量(见 2.6.8 节)的基本介绍见 Peebles[1993], Papoulis[1991]的著述是该话题的更进一步论述。关于随机和随机域图像处理的基础材料,见 Rosenfeld 和 Kak[1982], Jähne[2002]和 Won 及 Gray[2004]的著述。

本章中给出的许多技术的软件实现细节见 Gonzalez, Woods 和 Eddins[2004]的书籍。

## 习题

 标有星号的习题的详细答案可在本书的网站上找到。该网站上还包含有基于本章内容建议的项目。

- ★2.1 使用 2.1 节提供的背景信息,如果纸上的一个打印点离眼睛 0.3 m 远,请采用纯几何方法,估计眼睛能辨别的最小打印点的直径。为简单起见,假设在中央凹处的像点变得远视视网膜区域的感受器(锥状体)的直径小时,视觉系统已不能检测到该点。进一步假设中央凹可建模为直径为 1.5 mm 的圆形阵列,并且锥状体间的间隔在该阵列上均匀分布。

- 2.2 当我们在白天进入一家黑暗的剧场时,在能看清并找到空座时要用一段时间适应。2.1节描述的视觉过程在这种情况下起什么作用?
- ★2.3 虽然图 2.10 中未显示,但交流电的确是电磁波谱的一部分。美国的商用交流电频率是 77 Hz。问这一波谱分量的波长是多少米?
- 2.4 设计一个研究细胞、细菌、病毒和蛋白质边界形状的图像系统前端。在这种情况下,前端由光源和相应的图像摄像机组成。要求围成各个类别的圆形的直径分别为 25  $\mu\text{m}$ , 0.5  $\mu\text{m}$ , 0.05  $\mu\text{m}$  和 0.005  $\mu\text{m}$ 。
- (a) 能用单个传感器和摄像机解决这一成像问题吗?如果回答是肯定的,试给出照射波长和所需的摄影机类型。这里的“类型”指的是摄像机最敏感的电磁波谱波段(譬如红外波段)。
- (b) 如果(a)中您的回答是否定的,那么您建议采用哪种类型的照明光源和相应的图像传感器?指出(a)中所要求的光源和摄像机。要用最少数量的照明光源和摄像机解决这个问题。“解决这个问题”是指能够分别检测直径为 25  $\mu\text{m}$ , 0.5  $\mu\text{m}$ , 0.05  $\mu\text{m}$  和 0.005  $\mu\text{m}$  的环形细节。
- 2.5 14 mm×14 mm 的 CCD 摄像机芯片有 2048×2048 个元素,将它聚焦到相距 0.5 m 远的一个方形平坦区域。该摄像机每毫米能分辨多少线对?摄像机配备了一个 35 mm 镜头。(提示:成像处理模型如图 2.3 所示,但使用摄像机镜头的焦距替代眼睛的焦距。)
- ★2.6 一家汽车制造商想在一条限量版跑车生产线上将一些零件自动安放到跑车的保险杠上。这一部件就是彩色坐标,为选择合适的保险杠零件,机器人需要知道每辆车的颜色。模型有 4 种颜色:蓝色、绿色、红色和白色。请提出一种基于成像的解决方案。您怎样解决自动确定每辆跑车颜色的问题?记住,在选择零件时价格是最主要的考虑因素。
- 2.7 假设中心在  $(x_0, y_0)$  的平坦区域被一个强度分布为

$$i(x, y) = Ke^{-[(x-x_0)^2 + (y-y_0)^2]}$$

的光源照射。为简单起见,假设该区域的反射是恒定的并且等于 1.0,令  $K = 255$ 。如果图像使用  $k$  比特的强度分辨率进行数字化,且眼睛可检测相邻像素间的 4 种灰度突变,请问  $k$  取何值时会出现可见的伪轮廓?

- 2.8 画出习题 2.7 中  $k = 4$  时的图像。
- ★2.9 数字数据传输通常用波特率度量,其定义为每秒钟传输的比特数。通常,传输是以一个开始比特、一个字节(8 比特)的信息和一个停止比特组成的包完成的。利用这些事实,回答下列问题:
- (a) 使用 33.6K 波特的调制解调器传输一幅大小为 2048×2048 的 256 灰度级的图像,需要几分钟时间?
- (b) 波特率为 3000K 时,这是典型的电话 DSL(数字用户线)的媒体速度,传输要用多长时间?
- 2.10 高清晰度电视(HDTV)使用 1080 条水平电视线隔行扫描来产生图像(每隔一行在显像管表面画一条线,每两场形成一帧,每场用时 1/60 秒)。图像的宽高比是 16:9。在水平行数固定的情况下,求图像的垂直分辨率。一家公司已经设计了一种图像获取系统,该系统由 HDTV 图像生成数字图像。在该系统中,每条(水平)电视行的分辨率与图像的宽高比成正比,彩色图像的每个像素都有 24 比特的灰度分辨率,红色、绿色、蓝色图像各 8 比特。这三幅原色图像形成彩色图像。存储 90 分钟的一部 HDTV 电影需要多少比特?
- ★2.11 考虑两个图像子集  $S_1$  和  $S_2$ , 如下页右图所示。对于  $V = \{1\}$ , 确定这两个子集是(a)4 邻接的、(b)8 邻接的,还是(c) $m$  邻接的?
- ★2.12 提出将一个像素宽度的 8 通路转换为 4 通路的一种算法。



2.13 提出将一个像素宽度的  $m$  通路转换为 4 通路的一种算法。

2.14 参阅 2.5.2 节末尾的讨论, 在该讨论中, 我们将背景定义为  $(R_u)^c$ , 即图像中所有区域的并集的补集。在某些应用中, 定义一个像素  $(R_u)^c$  的子集为背景是有利的, 该子集不是空洞像素区域(非正式地, 认为空洞是由区域像素围绕着的背景像素子集)。如何修改这一定义, 以便从  $(R_u)^c$  中排除空洞像素? 像“背景是非空洞像素的  $(R_u)^c$  像素的子集”这样的答案是不能接受的。(提示: 利用连通性这一概念。)

		$S_1$				$S_2$				
0	0	0	0	0	0	0	1	1	0	
1	0	0	1	0	0	1	0	0	1	
1	0	0	1	0	1	1	0	0	0	
0	0	1	1	1	0	0	0	0	0	
0	0	1	1	1	0	0	1	1	1	

2.15 考虑右图图像分割。

★(a) 令  $V = \{0, 1, 2\}$ , 计算  $p$  和  $q$  间 4、8 和  $m$  通路的最短长度。如果在这两点间不存在一个特殊通路, 试解释原因。

(b) 令  $V = \{2, 3, 4\}$ , 重复问题(a)。

2.16 ★(a) 对于点  $p$  和  $q$  间的  $D_4$  距离等于这两点间最短 4 通路的情况, 给出需要的条件。

(b) 这个通路唯一吗?

	3	4	1	2	0
	0	1	0	4	2( $q$ )
	2	2	3	1	4
( $p$ )	3	0	4	2	1
	1	2	0	3	4

2.17 对于  $D_8$  距离, 重做习题 2.16。

★2.18 在下一章, 将讨论一些算子, 其功能是在一个很小的子图像区域  $S$  中计算像素值的总和。证明这些算子都是线性算子。

2.19 一个数集的中值  $\xi$  定义为这样一个值, 该数集中的一半数值比它小, 另一半数值比它大。例如, 数集  $\{2, 3, 8, 20, 21, 25, 31\}$  的中值是 20。试证明计算子图像区域  $S$  的中值的算子是非线性的。

★2.20 证明式(2.6-6)和式(2.6-7)的正确性。[提示: 从式(2.6-4)开始, 并且利用这样一个事实: 一个和的期望值就是期望值的和。]

2.21 考虑两幅 8 比特图像, 它们的灰度级跨越从 0 到 255 的全部范围。

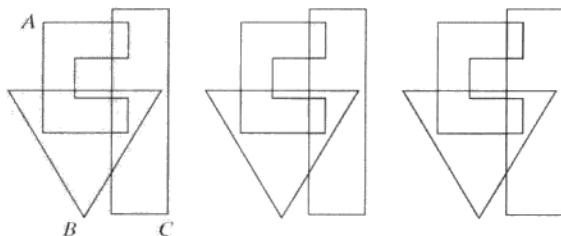
(a) 讨论从图像(1)中重复减去图像(2)的限制效应。假设结果仍然用 8 比特表示。

(b) 如果颠倒图像的顺序, 会产生不同的结果吗?

★2.22 图像相减通常用于在生产线上检测缺失的元件。方法是存储一幅对应于正确组装的“金”图像; 然后, 从相同产品的传入图像中减去该图像。理想情况下, 如果新产品组装正确, 则差值应为零。对于缺失元件的产品, 在不同于金图像的位置, 差值图像将不为零。在实际中, 您认为在什么条件下使用这种方法工作才是合适的?

2.23 ★(a) 参考图 2.31, 画出集合  $(A^c - B) \cup (B - A)$ 。

(b) 给出下图中集合  $A, B, C$  中所示的阴影集合的表达式。每幅图中的阴影区域组成一个集合, 因此应该对这三幅图的每一个阴影区域给出一个表达式。



2.24 使用三角形区域替代四边形区域时, 所得到的类似于式(2.6-24)和式(2.6-25)的等式是什么?

2.25 证明式(2.6-34)和式(2.6-35)中的傅里叶核是可分的和对称的。

- ★2.26 证明具有可分性、对称性的二维变换核可以用(1)沿输入的行(列)的一维变换来计算;接着(2)用第一步结果的列(行)的一维变换来计算。
- 2.27 一家工厂生产半透明的微型聚合物方形产品。要求进行 100%视觉检测来严格控制质量,工厂管理者发现人工检测越来越昂贵,于是检测改为半自动方式。在每个检测位置,机械装置把聚合物方形产品放在一盏灯的上方,灯位于光学系统的下方,该光学系统产生方形产品的放大图像,图像完全落在大小为  $80\text{ mm} \times 80\text{ mm}$  的观察屏上。缺陷以暗的斑点的形式出现,检测者的工作是观察屏幕。如果在屏幕上一个样品中出现了多个直径大于或等于  $0.8\text{ mm}$  的黑色斑点,则丢弃该样品。管理者相信,如果能找到一种全自动的处理方法,利润将提高 50%。她还相信项目的成功会是她升迁为法人的阶梯。经过大量调研之后,管理者确定了解决这一问题的方法,即用 CCD 电视摄像机去观察每个检测屏幕,并把摄像机的输出送到能够检测斑点的图像处理系统中,测量它们的直径,并触发以前由检测员操作的接收和丢弃按钮。她可以找到一个系统,只要最小的缺陷在数字图像中至少为  $2 \times 2$  像素的区域发生,就可以完成该工作。假设管理者聘任您来帮助她确定摄像机和镜头系统的参数,但要求您使用现有的元件。对于镜头,假设该约束条件意味着任何  $25\text{ mm}$ 、 $35\text{ mm}$  直至  $200\text{ mm}$  焦距的整数倍。对于摄像机,这意味着分辨率为  $512 \times 512$  像素、 $1024 \times 1024$  像素或  $2048 \times 2048$  像素。这些摄像机中的各个成像元素是大小为  $8\text{ }\mu\text{m} \times 8\text{ }\mu\text{m}$  的方形,成像元素间的间距  $2\text{ }\mu\text{m}$ 。对于该应用,摄像机要远贵于镜头,因此该问题应以选择镜头为基础,尽可能使用最低分辨率的摄像机来解决。作为一名顾问,请您提供一份书面建议,对您的结论给出合理且详细的分析。使用习题 2.5 中建议的相同成像几何。