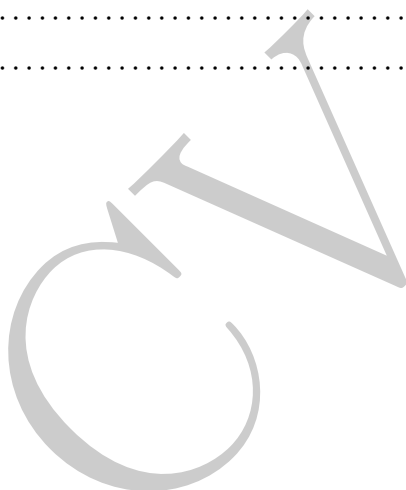


---

# 目录

第一章	图像及其表达与性质 .....	2
1.1	图像表达的若干概念 .....	2
1.2	图像数字化 .....	3
1.3	数字图像性质 .....	4
1.4	彩色图像 .....	11
1.5	摄像机论述 .....	12



# 第 1 章 图像及其表达与性质

## 1.1 图像表达的若干概念

图像可以表述为两个变量的一个连续函数  $f(x, y)$ ，其中  $(x, y)$  是平面的坐标；或者可能是三个变量的连续函数  $f(x, y, t)$ ，其中  $t$  是时间。

当 3D 物体经透视投影映射到摄像机平面后，由于这样的变换不是一一对应的，因而大量的信息丢失了，通过一幅图像来识别和重构 3D 图像是个病态的问题，可以预料这不是一个简单的人物，涉及试图建立图像中点的深度这个中间表达层次。目标是恢复完整的 3D 表达，比如计算机图形学中的表达，及独立于视点的表达，表示在物体坐标系中而不是在观察者坐标系中。如果这样的表达可以恢复，则物体的任何视角的亮度图像都可以用标准的计算机图形学技术合成出来。

计算机化的图像处理使用的数字图像通常表示成矩阵的形式，因此其坐标是整数。图像函数的定义域是一个区域  $R$

$$R = \{(x, y), 1 \leq x \leq x_m, 1 \leq y \leq y_n\} \quad (1.1)$$

其中  $x_m$  和  $y_n$  表示最大的图像坐标。图像函数具有有限的作用域——由于假定图像函数在域  $R$  外的值为零，可以使用无限求和或者积分的形式。

图像函数的值域也是有限的，按照惯例，在单色图像中最低对应于黑，而最高对应于白。它们之间的亮度值是灰阶。

数字图像的品质随着空间、频谱、辐射计量、时间分辨率增长而提高。空间分辨率是由图像平面上图像采样点间的接近程度确定的，频谱分辨率是由传感器获得的光线频率带宽决定的，辐射计量分辨率对应于可区分的灰阶数量，时间分辨率取决于图像获得的时间采样间隔。

图像变换假定图像函数  $f(x, y)$  是“良态的”，意思是指：该函数是可积的、具有可逆的傅里叶变换等。

## 1.2 图像数字化

图像数字化是指将  $f(x, y)$  采样为一个  $M$  行  $N$  列的矩阵。图像量化给每个连续的样本数值一个整数数字，图像函数  $f(x, y)$  的连续范围被划分为  $K$  个区间。采样及量化越精细（即  $M$ 、 $N$ 、 $K$  越大），对连续函数  $f(x, y)$  的近似就越好。

图像函数采样有两个问题，其一是确定采样的间隔，即相邻两个采样图像点的距离；其二是设置采样点的几何排列（采样栅格）。

### 1.2.1 采样

一个连续图像在采样点处被数字化。这些采样点是在平面上排列的，称它们的几何关系为栅格。因此数字图像是一个数据结构，通常是矩阵。在实践中，栅格一般是方的或者是正六边形。

把栅格和光栅区别开是重要的。光栅是指在点之间定义了相邻关系的栅格<sup>1</sup>。

栅格中一个无限小的采样点对应于数字化图像中的一个相元，也称作像素。从图像分析的角度看，像素是不能再分割的一个单元。

### 1.2.2 量化

在图像处理中，采样的图像数值  $f_s(j\Delta x, k\Delta y)$  用域一个数字表示。将图像函数的连续数值（亮度）转化为其数字等价量的过程称为量化。为了使人们能够觉察图像的细微变化，量化的级别要足够的高。大部分数字图像处理仪器都采用  $k$  个等间隔的量化方式。如果用  $b$  位来表示像素亮度的数值，那么亮度阶就是  $k = 2^b$ 。通常采用每个像素每个通道（红、绿、蓝各一个）8 位的表示方式，尽管也有一些系统使用其他数字（比如 16）。数字图像中亮度值一种有效的计算机表示，需要 8 位、4 位或 1 位，也就是说计算机存储的每个字节分别相应地可以存下 1 个、2 个或 8 个像素的亮度。

在量化级别不够时，图像的主要问题是出现伪轮廓（false contours）。

<sup>1</sup>即如果在方形的栅格上定义了 4-邻域，则得到方形的光栅

## 1.3 数字图像性质

数字图像具有一些度量和拓扑性质，与我们在基础微积分中所熟悉的连续二维函数的性质有所不同。另一个不同点在于人们对图像的感知，因为对图像质量的判断也是重要的。

### 1.3.1 数字图像的度量和拓扑性质

连续图像所具有的一些明显的直觉特性在数字图像领域中没有直接的类似推广。**距离**是一个重要的例子。满足下面三个条件的任何函数是一种“距离”（或度量）：

$$\begin{aligned} D(\mathbf{p}, \mathbf{q}) &\geq 0, \text{ 当且仅当 } \mathbf{p} = \mathbf{q} \text{ 时 } D(\mathbf{p}, \mathbf{q}) = 0 && \text{同一性} \\ D(\mathbf{p}, \mathbf{q}) &= D(\mathbf{q}, \mathbf{p}) && \text{对称性} \\ D(\mathbf{p}, \mathbf{r}) &\leq D(\mathbf{p}, \mathbf{q}) + D(\mathbf{q}, \mathbf{r}) && \text{三角不等式} \end{aligned}$$

坐标为  $(i, j)$  和  $(h, k)$  的两点间的距离可以定义为几种形式。经典的几何学和日常经验中的**欧氏距离** $D_E$  定义为：

$$D_E[(i, j), (h, k)] = \sqrt{(i - h)^2 + (j - k)^2} \quad (1.2)$$

两点之间的距离也可以表示为数字栅格上从起点移动到终点所需要的最少的基本步数。如果只允许横向和纵向的移动，即是  $D_4$  距离。 $D_4$  也称为“**城市街区距离**”，

$$D_4[(i, j), (h, k)] = |i - h| + |j - k| \quad (1.3)$$

在数字栅格中如果允许对角线方向的移动，我们就得到了距离  $D_8$ ，常常称为“**棋盘距离**”。距离  $D_8$  等于

$$D_8[(i, j), (h, k)] = \max\{|i - h|, |j - k|\} \quad (1.4)$$

像素**邻接性**是数字图像的另一个重要概念。任意两个像素如果它们之间的距离  $D_4(\mathbf{p}, \mathbf{q}) = 1$ ，则称彼此**4-邻接**的。类似地，**8-邻接**是指两个像素之间的距离  $D_8(\mathbf{p}, \mathbf{q}) = 1$ 。

由一些彼此邻接的像素组成的重要几何，我们称为“**区域**”，是一个重要的概念。更具体描述性的说法是，如果定义从像素  $P$  到像素  $Q$  的**路径**为一个点序列  $A_1, A_2, \dots, A_n$ ，其中  $A_1 = P, A_n = Q$ ，且  $A_{i+1}$  是  $A_i$  的临界点，

$i = 1, \dots, n - 1$ ; 那么**区域**是指这样的集合, 其中任意两个像素之间都存在完全属于该集合的路径。

如果两个像素之间存在一条路径, 那么这些像素就是**连通的**。因此, 可以说区域是彼此连通的像素的集合。“连通”关系是自反的, 对称的、且传递的, 因此它定义了集合的一个分解, 即等价类。

假设  $R_i$  是“连通”关系产生的不相交的区域, 进一步假设 (为了避免特殊的情况) 这些区域与图像边界不接触。设区域  $R$  是所有这些区域  $R_i$  的并集,  $R^C$  是区域  $R$  相对于图像的补集合。我们称包含图像边界的  $R^C$  的连通集合为**背景**, 而称补集合  $R^C$  的其它部分称为**孔**。如果区域中没有孔, 则称为**简单连通**区域。等价地, 简单连通区域的补集合是连通的。有孔的区域称为**复连通**。

请注意, 区域概念只使用了“连通”性。我们可以给区域赋予第二属性, 这些源于对图像数据的理解。我们常称图像的一些区域为**物体**, 决定图像中哪些区域对应于世界中的物体的过程是图像**分解**。

像素的亮度是一种非常简单的性质, 在有些图像中可以用于寻找物体, 例如, 如果一个像素比给定的值 (阈值) 黑就属于物体。所有这样的点的连通集构成一个物体。一个孔由非物体的点组成且被物体所围, 所有其他的点就构成了背景。

**距离变换**, 也叫做**距离函数**或**斜切算法**或**简单地斜切**, 它是距离概念的一个简单应用。距离变化提供了像素与某个图像子集 (可能表示物体或某些特征) 的距离。所产生的图像在该子集元素位置处的像素值为 0, 邻接的像素具有较小的值, 而离它远的数值就大, 该技术的命名源于这个阵列的外观。理解的例子参考桑卡书籍第 15 页。

对于距离度量  $D_4$  和  $D_8$ , Rosenfeld 和 Pfaltz 提出了一个计算距离变换的两遍算法。其想法是用一个小的局部掩膜遍历图像。第一遍从左上角开始, 水平从左到右直至图像边界, 然后返回到下一行开始处继续。第二遍从右下角开始, 使用一个不同的局部掩膜, 从右到左, 从下到上。该算法的有效性源于以“波浪状”的方式传播前一步的勘测的数值。

#### Computer Vision (算法 1). 距离变换

1. 按照一种距离度量  $D$ ,  $D$  是  $D_4$  或  $D_8$ , 对大小为  $M \times N$  的图像的一个子集  $S$  计算距离变换, 建立一个  $M \times N$  的数组  $F$  并做初始化, 子集  $S$  中的元素置为 0, 其他置为无穷。

2. 按行遍历图像，从上到下，从左到右。对于上方和左面的邻接像素，设

$$F(p) = \min_{q \in AL} [F(p), D(p, q) + F(q)]$$

3. 按行遍历图像，从下到上，从右到左。对于下方和右面的邻接像素，设

$$F(p) = \min_{q \in BR} [F(p), D(p, q) + F(q)]$$

4. 数组  $F$  中得到的是子集  $S$  的斜切。

**边缘**是另一个重要的概念。它是一个像素和其直接邻域的局部性质，它是一个有大小和方向的矢量。边缘告诉我们在一个像素的小邻域内图像亮度变化有多块。边缘计算的对象是具有很多亮度级别的图像，计算边缘的方式是计算图像函数的梯度。

区域的**边界**是图像分析中的另一个重要概念。区域  $R$  的边界是它自身的一个像素集合，其中的每个点具有一个或更多个  $R$  外的邻接点。该边界的定义与我们的直观理解相同，即边界是区域边界点的集合。有时我们称这样的边界为**内部边界**，同样地，我们可以定义类似的**外部边界**，**外部边界**是指区域的背景（即区域的补集）的边界。

“边界”和“边缘”虽然相关，但是它们不是同一个概念。边界是与区域有关的全局概念，而边缘表示图像函数的局部性质。一种可能的寻找边界的方法是链接显著的边缘。

一个区域是**凸**的是指如果区域内的任意两点连成一条线段，那么这条线段完整地处于区域内部。凸性将所有的区域划分成两个等价类：凸和非凸。

一个区域的**凸包**是指包含输入区域的一个最小凸区域。参看桑卡书籍的第 17 页。

**拓扑性质**不是基于距离的概念。相反，它们对于同胚的 (homeomorphic) 变换具有不变性，对图像而言，homeomorphic 变化可以解释为**橡皮面变换**。想像一下一个其表面上绘制了物体的小橡皮球，物体的拓扑性质是指在橡皮表面任意伸展时具有不变性的那部分性质。伸展不会改变物体各部分的连通性，也不会改变区域中孔的数目。

非规则形状的物体可以用一组它的拓扑分量来表示。凸包中非物体的部分



称为**凸损**；它可以分解为两个子集：**湖**，完全被物体所包围；**海湾**，与物体凸包的边界连通。

### 1.3.2 直方图

图像的**亮度直方图** $h_f(z)$  给出图像中亮度值  $z$  出现的频率，一幅有  $L$  个灰阶的图像的直方图由具有  $L$  个元素的一维数组表示。

#### Computer Vision (算法 2). 计算亮度直方图

1. 数组  $h_f$  的所有元素赋值为 0。
2. 对于图像  $f$  的所有像素，做  $h_f[f(x, y)] + 1$ 。

直方图在图像和概率之间建立了一个自然的桥梁。我们可能需要考虑找一个一阶概率函数  $p_q(z; x, y)$  来表示像素  $(x, y)$  的值  $z$  概率。在直方图中感兴趣的不是像素的位置，而是密度函数  $p_1(z)$ ，亮度直方图就是它的估计。直方图通常用柱状图来显示。

直方图通长是关于图像的唯一可得到的全局信息。在寻找最佳的照明条件以便抓取图像、进行灰阶变换以及将图像分割为物体和背景这些场合，都要用到直方图。

数字图像的直方图一般都有很多和极大值和极小值，这会使进一步的处理变得复杂。这个问题可以通过对直方图进行局部平滑来避免，比如，可以用相邻直方图元素的局部平均来做，因此新的直方图按下式来计算：

$$h'_f(z) = \frac{1}{2K+1} \sum_{j=-K}^K h_f(z+j) \quad (1.5)$$

其中  $K$  是一个常量，代表平滑所使用的邻域的大小。这个算法需要某种边界调整，也不能保证去除所有的局部极小值。还有一些其他平滑技术，重要的高斯模糊，在直方图的情况下，它是  $2D$  高斯模糊的简化。

### 1.3.3 熵

如果知道概率密度  $p$ ，用**熵**  $H(\text{entropy})$  就可以估计出图像的信息量，而与其解释无关。熵的信息论的形成源于香农，常称作**信息熵**。

信息熵的直觉理解与关联于给定概率分布的事件的不确定性大小有关。熵可作为失调的度量。当失调水平上升时，熵就增加而事件就难于预测。

假设离散随机变量  $X$  的可能结果（也称作状态）为  $x_1, \dots, x_n$ ，设  $p(x_k)$  是出现  $x_k (k = 1, \dots, n)$  的概率，熵定义为

$$H(X) = \sum_{k=1}^n p(x_k) \log_2 \left( \frac{1}{p(x_k)} \right) = - \sum_{k=1}^n p(x_k) \log_2 p(x_k) \quad (1.6)$$

随机变量  $X$  的熵是  $X$  所有可能的出现  $x_k$  的图下乘积的累加和：出现  $x_k$  的概率于  $x_k$  概率的倒数之对数的乘积。 $\log_2(\frac{1}{p(x_k)})$  也称作出现  $x_k$  的惊异 (surprisal)。随机变量  $X$  的熵是其出现惊异的期望。

### 1.3.4 图像的视觉感知

我们在设计或使用数字图像处理算法或设备时，应该考虑人的图像感知原理。

#### 对比度 (contrast)

对比度是亮度的局部变化，定义为物体亮度的平均值与背景亮度的比值。严格地说，如果我们的目的是要在物理上精确，应该讲的是辐射率而非亮度。人的眼睛对亮度的敏感性称对数关系，意味着对于同样的感知，高亮度需要高的对比度。

表现上的亮度很大程度上取决于局部背景的亮度，这种现象被称为条件对比度。

#### 敏锐度

敏锐度是觉察图像细节的能力。人的眼睛对于图像平面中的亮度的缓慢和快速变化敏感度差一些，而对于其间的中等变换较为敏感。敏锐度也随着离光轴的距离增加而降低。

### 1.3.5 图像品质

在图像的捕获、传输和处理过程中可能使图像退化，图像品质的度量可以用来估计退化的程度。

估计图像品质的方法可分为两类：主观的和客观的。度量图像品质的客观定量方法对我们更重要。理想的情况是，这样的方法同时也提供了主观的测试，且易于使用；这样我们就可以将标准用于参数优化。图像  $f(x, y)$  的品质通常



通过与一个已知的参考图  $g(x, y)$  作比较来估计。为这一目的，常常使用合成的图像作为参考图像。有一类方法使用简单的度量，比如均方差  $\Sigma(g - f)^2$ 。

### 1.3.6 图像中的噪声

实际的图像常受一些随机误差的影响而退化，我们通常称这个退化为**噪声**。噪声一般由其概率特征来刻画。理想的噪声，称作**白噪声**，经常会用到。白噪声具有是常量的功率谱，也就是说噪声在所有的频率上出现且强度相同。例如，白噪声的强度并不随着频率的增加而衰减，这在实际世界的信号中是典型的。白噪声是常用的模型，作为退化的最坏估计。

白噪声的一个特例是**高斯噪声**。服从高斯型的随机变量具有高斯曲线型的概率密度。在一维的情况下，密度函数是

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.7)$$

其中  $\mu$ ,  $\sigma$  分别是随机变量的均值和标准差。在很多实际情况下，噪声可以很好地用高斯噪声来近似。

当图像通过信道传输时，噪声一般与其出现的图像信号无关。这种独立于信号的退化被称为**加性噪声**，可以用下列模型来表示

$$f(x, y) = g(x, y) + v(x, y) \quad (1.8)$$

其中，噪声  $v$  和传输图像  $g$  是相互独立的变量。

#### Computer Vision (算法 3). 产生加性零均值高斯噪声

1. 假定图像的灰阶范围是  $[0, G - 1]$ 。取  $\sigma > 0$ ；它的值小时，相应的噪声也小。
2. 针对每对水平相邻的像素  $(x, y)$ ,  $(x, y + 1)$  产生一对位于  $[0, 1]$  范围的独立的随机数  $r, \phi$ 。
3. 计算

$$\begin{aligned} z_1 &= \sigma \cos(2\pi\phi) \sqrt{-2 \ln r} \\ z_2 &= \sigma \sin(2\pi\phi) \sqrt{-2 \ln r} \end{aligned} \quad (1.9)$$

(这是 Box-Muller 变换，假定  $z_1, z_2$  是独立的具有 0 均值和  $\sigma$  方差的正态分布。)

4. 置  $f'(x, y) = g(x, y) + z_1$  和  $f'(x, y + 1) = g(x, y + 1) + z_2$ , 其中  $g$  是输入图像。

5. 置

$$f(x, y) = \begin{cases} 0 & \text{当 } f'(x, y) < 0 \\ G - 1 & \text{当 } f'(x, y) > G - 1 \\ f'(x, y) & \text{其它} \end{cases} \quad (1.10)$$

$$f(x, y + 1) = \begin{cases} 0 & \text{当 } f'(x, y + 1) < 0 \\ G - 1 & \text{当 } f'(x, y + 1) > G - 1 \\ f'(x, y + 1) & \text{其它} \end{cases} \quad (1.11)$$

6. 跳转到步骤 3, 直到扫描完所有的像素。

式 (1.10) 和式 (1.11) 的截断会减弱噪声的高斯性质, 特别是当  $\sigma$  值与  $G$  比起来大时更加显著。

根据式 (1.9), 可以定义**信噪比 SNR**。计算噪声贡献的所有平方和

$$E = \sum_{(x,y)} v^2(x, y) \quad (1.12)$$

将它与观察到的信号的所有平方和作比较

$$F = \sum_{(x,y)} f^2(x, y) \quad (1.13)$$

信噪比就是

$$SNR = \frac{F}{E} \quad (1.14)$$

(严格地说, 我们将平均观测值相比于平均误差, 计算显然是一样的)。SNR 是图像品质的一个度量, 值大的就“好”。

信噪比常用对数尺度来表示, 单位为分贝:

$$SNR_{db} = 10 \log_{10} SNR \quad (1.15)$$

噪声的幅值在很多情况下与信号本身的幅值有关:

$$f = gv \quad (1.16)$$

这种模型表达的是**乘性噪声**。

**量化噪声**会在量化级别不足时出现，例如，仅有 50 个级别的单色图像。

**冲击噪声**是指一幅图像被个别噪声像素破坏，这些像素的亮度与其邻域显著不同。

**胡椒盐噪声**是指饱和的冲击在噪声，这是图像被一些白的或黑的像素所破坏，胡椒盐噪声会使二值图像退化。

## 1.4 彩色图像

人的色彩感知在电磁辐射的波长这一客观的物理性质上加了一主观层次。因此，色彩可以被认为是一种心里物理现象。

### 1.4.1 色彩物理学

只有狭窄的一段电磁波谱对人是可见的，大致对应于 380nm–740nm 的一段。色彩可以表达为基色的组合，基色即红、绿、蓝，为了标准化它们分别定义为 700nm、546.1nm 和 435.8nm。然而，这并不意味着所有的彩色都可以通过这三个基色的组合合成出来。

为什么我们看到的世界是带色彩的？当照射表面时，有两种占主导地位的物理机制来解释会发生什么。首先，**表面反射**以类似镜子的方式弹回进来的能量。反射光的光谱和入射光的光谱一模一样，与表面无关。其次，能量扩散进入材料内并随机地从其内部的颜料反射。这种机制称为**体反射**。色彩是颜料粒子的性质引起的，它们从入射光谱中吸收了某些波长。

### 1.4.2 人所感知的色彩

人类和一些动物在进化中发展了一种间接色彩感知机制。人建立起了对入射辐照光波长敏感的三种类型的传感器，即**三色觉**。人类视网膜上颜色敏感的感受器是**锥状体**。视网膜上另一种光敏感受器是**杆状体**，专注于在周边光照强度低的情况下的单色感知。锥状体按照感知的波长范围分为三类：S(短)，敏感度最大出现在  $\approx 430\text{nm}$ ；M(中) 在  $\approx 560\text{nm}$ ；L(长) 在  $\approx 610\text{nm}$ 。锥状体 S、M、L 偶尔也分别称作锥状体 B、G、R，但是这有点误导。当锥状体 L 激发时，我们看到的不单独时红。

感光器的反应或摄像机种的传感器的输出可以以数学建模。设  $i$  是某个传

传感器类型,  $i = 1, 2, 3$  (在人的情况下是视网膜锥状体的类型 S、M、L)。设  $R_i(\lambda)$  是传感器的光敏度,  $I(\lambda)$  是光照的谱密度,  $S(\lambda)$  表达表面元如何反射照明光的每个波长。第  $i$  个传感器的光谱响应可以用一定波长范围内的积分来建模:

$$q_i = \int_{\lambda_1}^{\lambda_2} I(\lambda) R_i(\lambda) S(\lambda) d\lambda \quad (1.17)$$

一种现象称为**条件等色**。一般而言, 条件等色是指两件物理上不同的事看起来却相同。红加绿产生黄就是一种条件等色, 因为黄也可以由一个光谱颜色产生。人的视觉系统受愚弄将绿感知成与黄一样。

由于彩色几乎可以用任意的基色集合来定义, 国际社会协商确定广泛应用的基色和色彩匹配函数。引进**色彩模型**作为数学抽象, 使我们可以将色彩表达为数字的元组, 通常是颜色分量的三或四数值的元组。受报刊和彩色电影发展的驱动, 于 1931 年 CIE 提出了一个技术标准, 称作**XYZ 色彩空间**。CIE 标准是绝对标准的一个例子, 它定义了色彩的无歧义的表达, 不依赖其它外部因素。

XYZ 色彩标准满足三个要求:

- 不同于色彩匹配实验中产生色彩匹配函数负波瓣的情况, XYZ 色彩空间的色彩匹配函数必须是非负的;
- $Y(\lambda)$  的数值应该于亮度 (照度) 相等;
- 实施规范化以确保对应于三种色彩匹配函数的功率像等。

实际的色彩是如下的混合

$$c_X X + c_Y Y + c_Z Z \quad (1.18)$$

其中  $0 \leq c_X, c_Y, c_Z \leq 1$  是混合权重。

## 1.5 摄像机论述