

## 第二部分 基于内容的图像检索技术

近年来，随着多媒体技术和计算机网络的飞速发展，全世界的数字图像的容量正以惊人的速度增长。无论是军用还是民用设备，每天都会产生容量相当于数千兆字节的图像。这些数字图像中包含了大量有用的信息。然而，由于这些图像是无序地分布在世界各地，图像中包含的信息无法被有效地访问和利用。这就要求有一种能够快速而且准确地查找访问图像的技术，也就是所谓的图像检索技术。自从 20 世纪 70 年代以来，在数据库系统和计算机视觉两大研究领域的共同推动下，图像检索技术已逐渐成为一个非常活跃的研究领域。数据库和计算机视觉两大领域是从不同的角度来研究图像检索技术的，前者基于文本的，而后者是基于视觉的。

基于文本的图像检索技术（text-based image retrieval）的历史可以追溯到 20 世纪 70 年代末期。当时流行的图像检索系统是将图像作为数据库中存储的一个对象，用关键字或自由文本对其进行描述。查询操作是基于该图像的文本描述进行精确匹配或概率匹配，有些系统的检索模型还是有词典支持的。另外，图像数据模型、多维索引、查询评价等技术都在这样一个框架之下发展起来。然而，完全基于文本的图像检索技术存在着严重的问题。首先，目前的计算机视觉和人工智能技术都无法自动对图像进行标注，而必须依赖于人工对图像做出标注。这项工作不但费时费力，而且手工的标注往往是不准确或不完整的，还不可避免地带有主观偏差。也就是说，不同的人对同一幅图像有不同的理解方法，这种主观理解的差异将导致图像检索中的失配错误。此外，图像中所包含的丰富的视觉特征（颜色或纹理等）往往无法用文本进行客观地描述的。

90 年代初期，随着大规模数字图像库的出现，上述的问题变得越来越尖锐。为克服这些问题，基于内容的图像检索技术（content-based image retrieval）应运而生。区别于原有系统中对图像进行人工标注的做法，基于内容的检索技术自动提取每幅图像的视觉内容特征作为其索引，如色彩、纹理、形状等。此后几年中，这个研究领域中的许多技术发展起来，一大批研究性的或商用的图像检索系统被建立起来。这个领域的发展主要来归功于计算机视觉技术的进步，在文献[ ]中有对这一领域的详细介绍。

应该认识到，基于内容的图像检索系统具有与传统基于文本的检索系统完全不同的构架。首先，由于图像依赖其视觉特征而非文本描述进行索引，查询将根据图像视觉特征的相似度进行。用户通过选择具有代表性的一幅或多幅例子图像来构造查询，然后由系统查找与例子图像在视觉内容上比较相似的图像，按相似度大小排列返回给用户。这就是所谓的通过例子图像的检索（query by image example）。另外，基于内容的检索系统一般通过可视化界面和用户进行频繁的交互，以便于用户能够方便地构造查询、评估检索结果和改进检索结果。

下图表示了基于内容的图像检索系统的体系结构。系统的核心是图像特征数据库。图像特征既可以从图像本身提取得到，又可以通过用户交互获得，并用于计算图像之间的相似度。用户和系统之间的关系是双向的：用户可以向系统提出查询要求，系统根据查询要求返回查询结果，用户还通过对查询结果的相关反馈来改进查询结果。图中还标出了基于内容的图像检索中的一些关键环节：

- 1) 选择、提取和索引能够充分表达图像的视觉特征。

- 2) 处理基于相似度的图像检索。
- 3) 处理用户对检索结果的相关反馈，改善检索结果。



在这一部分中，我们将主要讨论有关基于内容的图像检索方面的一些相关问题和方法。第 2 章中给出了一系列图像视觉特征的提取、表达和索引方法。第 3 章中讨论了图像相似度衡量方法和其它检索相关技术。在第 4 章中，我们介绍了图像检索中相关反馈的机制和途径。最后，第 5 章中总结了现有的一些图像检索系统，并对这一领域的未来方向作出展望。

## 第2章 图像特征的提取与表达

图像特征的提取与表达是基于内容的图像检索技术的基础。从广义上讲，图像的特征包括基于文本的特征（如关键字、注释等）和视觉特征（如色彩、纹理、形状、对象表面等）两类。由于基于文本的图像特征提取在数据库系统和信息检索等领域中已有深入的研究，本章中我们主要介绍图像视觉特征的提取和表达。

视觉特征又可分为通用的视觉特征和领域相关的视觉特征。前者用于描述所有图像共有的特征，与图像的具体类型或内容无关，主要包括色彩、纹理和形状；后者则建立在对所描述图像内容的某些先验知识（或假设）的基础上，与具体的应用紧密有关，例如人的面部特征或指纹特征等。由于领域相关的图像特征主要属于模式识别的研究范围，并涉及许多专业的领域知识，在此我们就不再详述，而只考虑通用的视觉特征。

对于某个特定的图像特征，通常又有多种不同的表达方法。由于人们主观认识上的千差万别，对于某个特征并不存在一个所谓的最佳的表达方式。事实上，图像特征的不同表达方式从各个不同的角度刻画了该特征的某些性质。在本章中，我们主要介绍那些由实践证明对图像检索比较有效的特征和相应的表达方法。本章的第1、2、3节中我们将分别介绍图像的颜色、纹理和形状特征，第4节中介绍包含有空间信息的图像特征，最后一节简述了多维索引技术和降低维度技术。

### 2.1 颜色特征的提取

颜色特征是在图像检索中应用最为广泛的视觉特征，主要原因在于颜色往往和图像中所包含的物体或场景十分相关。此外，与其他的视觉特征相比，颜色特征对图像本身的尺寸、方向、视角的依赖性较小，从而具有较高的鲁棒性。

面向图像检索的颜色特征的表达涉及到若干问题。首先，我们需要选择合适的颜色空间来描述颜色特征；其次，我们要采用一定的量化方法将颜色特征表达为向量的形式；最后，还要定义一种相似度（距离）标准用来衡量图像之间在颜色上的相似性。在本节中，我们将主要讨论前两个问题，并介绍颜色直方图、颜色矩、颜色集、颜色聚合向量以及颜色相关图等颜色特征的表达方法。

#### 2.1.1 颜色直方图

颜色直方图是在许多图像检索系统中被广泛采用的颜色特征。它所描述的是不同色彩在整幅图像中所占的比例，而并不关心每种色彩所处的空间位置，即无法描述图像中的对象或物体。颜色直方图特别适于描述那些难以进行自动分割的图像。

当然，颜色直方图可以是基于不同的颜色空间和坐标系。最常用的颜色空间是 RGB 颜色空间，原因在于大部分的数字图像都是用这种颜色空间表达的。然而，RGB 空间结构并不符合人们对颜色相似性的主观判断。因此，有人提出了基于 HSV 空间、Luv 空间和 Lab 空间的颜色直方图，因为它们更接近于人们对颜色的主观认识。其中 HSV 空间是直方图最常用的颜色空间。它的三个分量分别代表色彩（Hue）、饱和度（Saturation）和值（Value）。从 RGB 空间到 HSV 空间的转化公式如下所示：

$$\left. \begin{aligned}
 v &= \max(r, g, b) \\
 s &= [v - \min(r, g, b)]/v \\
 h &= \begin{cases} 5 + b' & \text{if } r = \max(r, g, b) \text{ and } g = \min(r, g, b) \\ 1 - g' & \text{if } r = \max(r, g, b) \text{ and } g \neq \min(r, g, b) \\ 1 + r' & \text{if } g = \max(r, g, b) \text{ and } b = \min(r, g, b) \\ 3 - b' & \text{if } g = \max(r, g, b) \text{ and } b \neq \min(r, g, b) \\ 3 + g' & \text{if } b = \max(r, g, b) \text{ and } r = \min(r, g, b) \\ 5 - r' & \text{otherwise} \end{cases} \\
 r' &= [v - r]/[v - \min(r, g, b)] \\
 g' &= [v - g]/[v - \min(r, g, b)] \\
 b' &= [v - b]/[v - \min(r, g, b)]
 \end{aligned} \right\} \quad (1)$$

其中  $r, g, b \in [0 \dots 1]$ ,  $h \in [0 \dots 6]$ , and  $s, v \in [0 \dots 1]$ 。从 RGB 空间到 Luv 空间和到 Lab 空间的转化可以在文献[1]中找到。

计算 HSV 空间中两种颜色的距离由多种不同的方法。例如在[2]中提出了如下的颜色距离计算公式：

$$a(i, j) = 1 - 1/\sqrt{5} [(v_i - v_j)^2 + (s_i \cos(h_i) - s_j \cos(h_j))^2 + (s_i \sin(h_i) - s_j \sin(h_j))^2]^{1/2} \quad (2)$$

其中  $(h_i, s_i, v_i)$  和  $(h_j, s_j, v_j)$  分别代表两种 HSV 空间中的颜色。这种相似度量方法相当于一个圆柱形颜色空间中的欧拉距离，该空间中的颜色值表示为  $(s \cosh, s \sinh, v)$ 。在[3]中这样的圆柱空间被进一步变形称为圆锥性空间，其中的颜色表示为  $(s \cosh, s \sinh, v)$ 。这些改变使  $v$  值较小的时候，降低了直方图对  $h$  和  $s$  分量的分辨能力。

计算颜色直方图需要将颜色空间划分成若干个小颜色区间，每个小区间成为直方图的一个 bin。这个过程称为颜色量化 (color quantization)。然后，通过计算颜色落在每个小区间内的像素数量可以得到颜色直方图。颜色量化有许多方法，例如向量量化、聚类方法或者神经网络方法。最为常用的做法是将颜色空间的各个分量 (维度) 均匀地进行划分。相比之下，聚类算法则会考虑到图像颜色特征在整个空间中的分布情况，从而避免出现某些 bin 中的像素数量非常稀疏的情况，使量化更为有效。另外，如果图像是 RGB 格式而直方图是 HSV 空间中的，我们可以预先建立从量化的 RGB 空间到量化的 HSV 空间之间的查找表 (look-up table)，从而加快直方图的计算过程。

上述的颜色量化方法会产生一定的问题。设想两幅图像的颜色直方图几乎相同，只是互相错开了一个 bin，这时如果我们采用  $L_1$  距离或者欧拉距离 (见3.1.1节) 计算两者的相似度，会得到很小的相似度值。为了克服这个缺陷，需要考虑到相似但不相同的颜色之间的相似度。一种方法是采用二次式距离[4] (见3.1.3节)。另一种方法是对颜色直方图事先进行平滑过滤，即每个 bin 中的像素对于相邻的几个 bin 也有贡献。这样，相似但不相同颜色之间的相似度对直方图的相似度也有所贡献。

选择合适的颜色小区间 (即直方图的 bin) 数目和颜色量化方法与具体应用的性能和效率要求有关。一般来说，颜色小区间的数目越多，直方图对颜色的分辨能力就越强。然而，bin 的数目很大的颜色直方图不但会增加计算负担，也不利于在大型图像库中建立索引。而且对于某些应用来说，使用非常精细的颜色空间划分方法不一定能够提高检索效果，特别是对于不能容忍对相关图像错漏的那些应用。另一种有效减少直方图 bin 的数目的办法是只选用那些数值最大 (即像素数目最多) 的

bin 来构造图像特征，因为这些表示主要颜色的 bin 能够表达图像中大部分像素的颜色。实验证明这种方法并不会降低颜色直方图的检索效果。事实上，由于忽略了那些数值较小的 bin，颜色直方图对噪声的敏感程度降低了，有时会使检索效果更好。两种采用主要颜色构造直方图的方法可以在文献 [5,6] 中找到。

### 2.1.2 颜色矩

另一种非常简单而有效的颜色特征是由 Stricker 和 Orengo 所提出的颜色矩(color moments) [7]。这种方法的数学基础在于图像中任何的颜色分布均可以用它的矩来表示。此外，由于颜色分布信息主要集中在低阶矩中，因此仅采用颜色的一阶矩 (mean)、二阶矩 (variance) 和三阶矩 (skewness) 就足以表达图像的颜色分布。与颜色直方图相比，该方法的另一个好处在于无需对特征进行向量化。颜色的三个低次矩在数学上表达为：

$$\mu_i = \frac{1}{N} \sum_{j=1}^N p_{ij} \quad (3.1)$$

$$\sigma_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^2 \right)^{\frac{1}{2}} \quad (3.2)$$

$$s_i = \left( \frac{1}{N} \sum_{j=1}^N (p_{ij} - \mu_i)^3 \right)^{\frac{1}{3}} \quad (3.3)$$

其中  $p_{ij}$  是图像中第  $j$  个像素的第  $i$  个颜色分量。因此，图像的颜色矩一共只需要 9 个分量 (3 个颜色分量，每个分量上 3 个低阶矩)，与其他的颜色特征相比是非常简洁的。在实际应用中为避免低次矩较弱的分辨能力，颜色矩常和其它特征结合使用，而且一般在使用其它特征前起到过滤缩小范围 (narrow down) 的作用。

### 2.1.3 颜色集

为支持大规模图像库中的快速查找，Smith 和 Chang 提出了用颜色集 (color sets) 作为对颜色直方图的一种近似 [8]。他们首先将 RGB 颜色空间转化成视觉均衡的颜色空间 (如 HSV 空间)，并将颜色空间量化成若干个 bin。然后，他们用色彩自动分割技术将图像分为若干区域，每个区域用量化颜色空间的某个颜色分量来索引，从而将图像表达一个二进制的颜色索引集。在图像匹配中，比较不同图像颜色集之间的距离和色彩区域的空间关系 (包括区域的分离、包含、交等，每种对应于不同得分)。因为颜色集表达为二进制的特征向量，可以构造二分查找树来加快检索速度，这对于大规模的图像集合十分有利。

### 2.1.4 颜色聚合向量

针对颜色直方图和颜色矩无法表达图像色彩的空间位置的缺点，Pass [9] 提出了图像的颜色聚合向量 (color coherence vector)。它是颜色直方图的一种演变，其核心思想是将属于直方图每一个 bin 的像素进行分为两部分：如果该 bin 内的某些像素所占据的连续区域的面积大于给定的阈值，则该区域内的像素作为聚合像素，否则作为非聚合像素。假设  $\alpha_i$  与  $\beta_i$  分别代表直方图的第  $i$  个 bin 中聚合像素和非聚合像素的数量，图像的颜色聚合向量可以表达为  $\langle (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_N, \beta_N) \rangle$ 。而  $\langle \alpha_i$

$+\beta_1, \alpha_2 + \beta_2, \dots, \alpha_N + \beta_N >$  就是该图像的颜色直方图。由于包含了颜色分布的空间信息, 颜色聚合向量相比颜色直方图可以达到更好的检索效果。

### 2.1.5 颜色相关图

颜色相关图 (color correlogram) 是图像颜色分布的另一种表达方式[16]。这种特征不但刻画了某一种颜色的像素数量占整个图像的比例, 还反映了不同颜色对之间的空间相关性。实验表明, 颜色相关图比颜色直方图和颜色聚合向量具有更高的检索效率, 特别是查询空间关系一致的图像。

假设  $I$  表示整张图像的全部像素,  $I_{c(i)}$  则表示颜色为  $c(i)$  的所有像素。颜色相关图可以表达为:

$$\gamma_{i,j}^{(k)} = \Pr_{p_1 \in I_{c(i)}, p_2 \in I} [p_2 \in I_{c(j)} \mid |p_1 - p_2| = k] \quad (4)$$

其中  $i, j \in \{1, 2, \dots, N\}$ ,  $k \in \{1, 2, \dots, d\}$ ,  $|p_1 - p_2|$  表示像素  $p_1$  和  $p_2$  之间的距离。颜色相关图可以看作是一张用颜色对  $\langle i, j \rangle$  索引的表, 其中  $\langle i, j \rangle$  的第  $k$  个分量表示颜色为  $c(i)$  的像素和颜色为  $c(j)$  的像素之间的距离小于  $k$  的概率。如果考虑到任何颜色之间的相关性, 颜色相关图会变得非常复杂和庞大 (空间复杂度为  $O(N^2d)$ )。一种简化的变种是颜色自动相关图 (color auto-correlogram), 它仅仅考察具有相同颜色的像素间的空间关系, 因此空间复杂度降到  $O(Nd)$ 。

## 2.2 纹理特征的提取

纹理特征是一种不依赖于颜色或亮度的反映图像中同质现象的视觉特征[12]。它是所有物体表面共有的内在特性, 例如云彩、树木、砖、织物等都有各自的纹理特征。纹理特征包含了物体表面结构组织排列的重要信息以及它们与周围环境的联系[13]。正因为如此, 纹理特征在基于内容的图像检索中得到了广泛的应用, 用户可以通过提交包含有某种纹理的图像来查找含有相似纹理的其他图像。

由于纹理特征对模式识别和计算机视觉等领域的重要意义, 对纹理的分析研究在过去的三十年中取得了重大的成果。在本节中, 我们将着重介绍那些在基于内容的图像检索中所常用的那些纹理特征, 主要有 Tamura 纹理特征、自回归纹理模型、方向性特征、小波变换和共生矩阵等形式。

### 2.2.1 Tamura 纹理特征

基于人类对纹理的视觉感知的心理学的研究, Tamura 等人提出了纹理特征的表达[14]。Tamura 纹理特征的六个分量对应于心理学角度上纹理特征的六种属性, 分别是粗糙度 (coarseness)、对比度 (contrast)、方向度 (directionality)、线像度 (linelikeness)、规整度 (regularity) 和粗略度 (roughness)。其中, 前三个分量对于图像检索尤其重要[e.g,15]。接下来我们就着重讨论粗糙度、对比度和方向度这三种特征的定义和数学表达。

- **粗糙度**。粗糙度的计算可以分为以下几个步骤进行。首先, 计算图像中大小为  $2^k \times 2^k$  个像素的活动窗口中像素的平均强度值, 即有

$$A_k(x, y) = \frac{1}{2^{2k}} \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i, j)$$

其中  $k = 0, 1, \dots, 5$  而  $g(i, j)$  是位于  $(i, j)$  的像素强度值。然后, 对于每个像素, 分别计算它在水平

和垂直方向上互不重叠的窗口之间的平均强度差。

$$\left. \begin{aligned} E_{k,h}(x, y) &= |A_k(x+2^{k-1}, y) - A_k(x-2^{k-1}, y)| \\ E_{k,v}(x, y) &= |A_k(x, y+2^{k-1}) - A_k(x, y-2^{k-1})| \end{aligned} \right\}$$

其中对于每个像素，能使  $E$  值达到最大（无论方向）的  $k$  值用来设置最佳尺寸  $S_{best}(x, y) = 2^k$ 。最后，粗糙度可以通过计算整幅图像中  $S_{best}$  的平均值来得到，表达为

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i, j)$$

粗糙度特征的另一种改进形式是采用直方图来描述  $S_{best}$  的分布，而不是像上述方法一样简单地计算  $S_{best}$  的平均值。这种改进后的粗糙度特征能够表达具有多种不同纹理特征的图像或区域，因此对图像检索更为有利。

- **对比度**。对比度是通过对像素强度分布情况的统计得到的。确切地说，它是通过  $\alpha_4 = \mu_4 / \sigma^4$  来定义的，其中  $\mu_4$  是四次矩而  $\sigma^2$  是方差。对比度是通过如下公式衡量的：

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}}$$

该值给出了整个图像或区域中对比度的全局度量。

- **方向度**。方向度的计算需要首先计算每个像素处的梯度向量。该向量的模和方向分别定义为

$$\begin{aligned} |\Delta G| &= (|\Delta_H| + |\Delta_V|) / 2 \\ \theta &= \tan^{-1}(\Delta_V / \Delta_H) + \pi / 2 \end{aligned}$$

其中  $\Delta_H$  和  $\Delta_V$  分别是通过图像卷积下列两个 3x3 操作符所得的水平和垂直方向上的变化量。

$$\begin{array}{ccc|ccc} -1 & 0 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & -1 & -1 \end{array}$$

当所有像素的梯度向量都被计算出来后，一个直方图  $H_D$  被构造用来表达  $\theta$  值。该直方图首先对  $\theta$  的值域范围进行离散化，然后统计了每个 bin 中相应的  $|\Delta G|$  大于给定阈值的像素数量。这个直方图对于具有明显方向性的图像会表现出峰值，对于无明显方向的图像则表现得比较平坦。最后，图像总体的方向性可以通过计算直方图中峰值的尖锐程度获得，表示如下：

$$F_{dir} = \sum_p \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi) \quad (11)$$

上式中的  $p$  代表直方图中的峰值， $n_p$  为直方图中所有的峰值。对于某个峰值  $p$ ， $w_p$  代表该峰值所包含的所有的 bin，而  $\phi_p$  是具有最高值的 bin。

## 2.2.2 自回归纹理模型

最近二十年中大量的研究集中在应用随机场模型表达纹理特征,这方面 Markov 随机场(MRF)模型取得了很大的成功。自回归纹理模型(simultaneous auto-regressive, 或 SAR)就是 MRF 模型的一种应用实例。

在 SAR 模型中,每个像素的强度被描述成随机变量,可以通过与其相邻的像素来描述。如果  $s$  代表某个像素,则其强度值  $g(s)$  可以表达为它的相邻像素强度值的线性叠加与噪音项  $\varepsilon(s)$  的和,如下所示:

$$g(s) = \mu + \sum_{r \in D} \theta(r)g(s+r) + \varepsilon(s) \quad (12)$$

其中  $\mu$  是基准偏差,由整幅图像的平均强度值所决定,  $D$  表示了  $s$  的相邻像素集。 $\theta(r)$  是一系列模型参数,用来表示不同相邻位置上的像素的权值。 $\varepsilon(s)$  是均值为 0 而方差为  $\sigma^2$  的高斯随机变量。通过上式可以用回归法计算参数  $\theta$  和标准方差  $\sigma$  的值,它们反映了图像的各种纹理特征。例如较高的  $\sigma$  表示图像具有很高的精细度,或较低的粗糙度。又比如,如果  $s$  正上方和正下方的  $\theta$  很高,表明图像具有垂直的方向性。最小误差法(least square error)和极大似然估计(maximum likelihood estimation)可以用来计算模型中的参数。此外, SAR 的一种变种称为旋转无关的自回归纹理特征(rotation-invariant SAR 或 RISAR),具有与图像的旋转无关的特点。

定义合适的 SAR 模型需要确定相邻像素集合的范围。然而,固定大小的相邻像素集合范围无法很好地表达各种纹理特征。为此,有人提出过多维度的自回归纹理模型(multi-resolution SAR 或 MRSAR) [16],能够在多个不同的相邻像素集合范围下计算纹理特征。文献[17,18]中给出了 MRSAR 纹理特征和其他纹理特征对于图像检索的性能对比。实验结果表明 MRSAR 纹理特征能够较好地识别出图像中各种纹理特征。

### 2.2.3 小波变换

小波变换(wavelet transform)也是一种常用的纹理分析和分类方法[19,20]。小波变换指的是将信号分解为一系列的基本函数  $\psi_{mn}(x)$ 。这些基本函数都是通过对母函数  $\psi(x)$  的变形得到,如下所示:

$$\psi_{mn}(x) = 2^{-m/2} \psi(2^{-m}x - n) \quad (19)$$

其中  $m$  和  $n$  是整数。这样,信号  $f(x)$  可以被表达为:

$$f(x) = \sum_{m,n} c_{mn} \psi_{mn}(x) \quad (20)$$

二维小波变换的计算需要进行递归地过滤和采样。在每个层次上,二维的信号被分解为四个子波段,根据频率特征分别称为 LL、LH、HL 和 HH。有两种类型的小波变换可以用于纹理分析,其中是金字塔结构的小波变换(pyramid-structured wavelet transform 或 PWT)和树桩结构的小波变换(tree-structured wavelet transform 或 TWT)。PWT 递归地分解 LL 波段。但是对于那些主要信息包含在中频段范围内的纹理特征,仅仅分解低频的 LL 波段是不够的。因此,TWT 被提出来克服上述的问题。TWT 区别于 PWT 的主要之处在于它除了递归分解 LL 波段之外,还会分解其它的 LH、HL 和 HH 等波段。

小波变换表示的纹理特征可以用每个波段的每个分解层次上能量分布的均值和标准方差。例如



三层的分解，PWT 表达为  $3 \times 4 \times 2$  的特征向量。TWT 的特征向量取决于每个子波段是分解方式。一般来说，由 PWT 所得的特征是由 TWT 所得特征的一个子集。此外，根据在文献[21]中所作的性能对比，不同的小波变换在对纹理分析方面没有很显著的差别。

### 2.2.4 其它纹理特征

除了上述的 Tamura 特征、SAR 模型和小波变换等纹理特征之外，还有许多其它的纹理特征。早在 70 年代，Haralick 等研究人员就提出了用共生矩阵 (co-occurrence matrix) 表示纹理特征的方法[13]。该方法从对数学角度研究了图像纹理中灰度级的空间依赖关系。它首先建立一个基于像素之间方向性和距离的共生矩阵，然后从矩阵中提取有意义的统计量作为纹理特征。许多其他研究人员沿着这个方向提出了扩展的方案。例如 Gotlieb 和 Kreyszig 研究了[13]中提出的统计特征，在实验中得出能量、相关性、惯量和熵是最有效的特征[22]。

Gabor 过滤法[23]能够最大程度地减少空间和频率的不确定性，同时还能够检测出图像中不同方向和角度上的边缘和线条。[24,25]中提到了很多方法根据过滤输出结果来描述图像特征。

此外，小波变换也常常与其它技术结合以获得更好的效果，例如 Gross 等人用小波变换与 KL 展开式和 Kohonen 多处理机系统来进行纹理分析[26]。Thyagarajan 等[27]用小波变换和共生矩阵来进行纹理分析，结合了统计和变换两者的优点。

## 2.3 形状特征的提取

物体和区域的形状是图像表达和图像检索中的另一重要的特征。但不同于颜色或纹理等底层特征，形状特征的表达必须以对图像中物体或区域的划分为基础。由于当前的技术无法做到准确而鲁棒的自动图像分割，图像检索中的形状特征只能用于某些特殊应用，在这些应用中图像包含的物体或区域可以直接获得。另一方面，由于人们对物体形状的变化、旋转和缩放主观上不太敏感，合适的形状特征必须满足对变换、旋转和缩放无关，这对形状相似度的计算也带来了难度。

通常来说，形状特征有两种表示方法，一种是轮廓特征的，一种是区域特征的。前者只用到物体的外边界，而后者则关系到整个形状区域。这两类形状特征的最典型方法分别是傅立叶描述符和形状无关矩。我们在下文中将详细介绍这两种方法，同时还简单介绍了其它的形状特征。

### 2.3.1 傅立叶形状描述符

傅立叶形状描述符 (Fourier shape descriptors) 的基本思想是用物体边界的傅立叶变换作为其形状描述。假设一个二维物体的轮廓是由一系列坐标为  $(x_s, y_s)$  的像素组成，其中  $0 \leq s \leq N-1$ ，而  $N$  是轮廓上像素的总数。从这些边界点的坐标中可以推导出三种形状表达，分别是曲率函数 (curvature function)、质心距离 (centroid distance) 和复坐标函数 (complex coordinates function)。

轮廓线上某一点的曲率定义为轮廓切向角度相对于弧长的变化率。曲率函数  $K(s)$  可以表示为：

$$K(s) = \frac{d}{ds} \theta(s) \quad (21)$$

其中  $\theta(s)$  是轮廓线的切向角度，定义为：

$$\left. \begin{aligned} \theta(s) &= \tan^{-1} \left( \frac{y'_s}{x'_s} \right) \\ y'_s &= \frac{dy_s}{ds} \\ x'_s &= \frac{dx_s}{ds} \end{aligned} \right\} \quad (22)$$

质心距离定义为从物体边界点到物体中心  $(x_c, y_c)$  的距离，如下所示：

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2} \quad (23)$$

复坐标函数是用复数所表示的像素坐标：

$$Z(s) = (x_s - x_c) + j(y_s - y_c) \quad (24)$$

对这种复坐标函数的傅立叶变换会产生一系列复数系数。这些系数在频率上表示了物体形状，其中低频分量表示形状的宏观属性，高频分量表达了形状的细节特征。形状描述符可以从这些变换参数中得出。为了保持旋转无关性，我们仅仅保留了参数的大小信息，而省去了相位信息。缩放的无关性是通过将参数的大小除以 DC 分量（或第一个非零参数）的大小来保证的。请注意变换无关性是基于轮廓的形状表示所固有的特点。

对于曲率函数和质心距离函数，我们只考虑正频率的坐标轴，因为这时函数的傅立叶变换是对称的，即有  $|F_{-i}| = |F_i|$ 。基于曲率函数的形状描述符表示为

$$f_K = \left[ |F_1|, |F_2|, \dots, |F_{M/2}| \right] \quad (24)$$

其中  $F_i$  表示傅立叶变换参数的第  $i$  个分量。类似的，由质心距离所导出的形状描述符为

$$f_R = \left[ \frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \dots, \frac{|F_{M/2}|}{|F_0|} \right] \quad (25)$$

对于复坐标函数，正频率分量和负频率分量被同时采用。由于 DC 参数与形状的所处的位置有关而被省区。因此，第一个非零的频率分量被用来对其它变换参数进行标准化。复坐标函数所导出的形状描述符为

$$f_Z = \left[ \frac{|F_{-(M/2-1)}|}{|F_1|}, \dots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \dots, \frac{|F_{M/2}|}{|F_1|} \right] \quad (26)$$

为保证数据库中所有物体的形状特征都有相同的长度，在实施傅立叶变换之前需要将所有边界点的数目统一为  $M$ 。例如， $M$  可以取为  $2^n = 64$ ，这样就可以采用快速傅立叶变换来提高算法效率。

### 2.3.2 形状无关矩

形状无关矩 (Moment Invariants) 是基于区域的物体形状表示方法。假设  $R$  是用二值图像表示的物体, 则  $R$  形状的第  $p+q$  阶中心矩为:

$$\mu_{p,q} = \sum_{(x,y) \in R} (x-x_c)^p (y-y_c)^q \quad (28)$$

其中  $(x_c, y_c)$  是物体的中心。为获得缩放无关的性质, 可以对该中心矩进行标准化操作[1]:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma}, \quad \gamma = \frac{p+q+2}{2} \quad (29)$$

基于这些矩, Hu [28]提出了一系列分别具有变换、旋转和缩放无关性的 7 个矩:

$$\left. \begin{aligned} \phi_1 &= \mu_{2,0} + \mu_{0,2} \\ \phi_2 &= (\mu_{2,0} - \mu_{0,2})^2 + 4\mu_{1,1}^2 \\ \phi_3 &= (\mu_{3,0} - 3\mu_{1,2})^2 + (\mu_{0,3} - 3\mu_{2,1})^2 \\ \phi_4 &= (\mu_{3,0} + \mu_{1,2})^2 + (\mu_{0,3} + \mu_{2,1})^2 \\ \phi_5 &= (\mu_{3,0} - 3\mu_{1,2})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] \\ &\quad + (\mu_{0,3} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2] \\ \phi_6 &= (\mu_{2,0} - \mu_{0,2})[(\mu_{3,0} + \mu_{1,2})^2 - (\mu_{0,3} + \mu_{2,1})^2] + 4\mu_{1,1}(\mu_{3,0} + \mu_{1,2})(\mu_{0,3} + \mu_{2,1}) \\ \phi_7 &= (3\mu_{2,1} - \mu_{0,3})(\mu_{3,0} + \mu_{1,2})[(\mu_{3,0} + \mu_{1,2})^2 - 3(\mu_{0,3} + \mu_{2,1})^2] \\ &\quad + (\mu_{3,0} - 3\mu_{2,1})(\mu_{0,3} + \mu_{2,1})[(\mu_{0,3} + \mu_{2,1})^2 - 3(\mu_{3,0} + \mu_{1,2})^2] \end{aligned} \right\} \quad (30)$$

除了上述的七种无关矩以外, 还有许多计算形状无关矩的方法。在文献[29]中, Yang 和 Albregtsen 在 Green 定理的基础上提出了在二值图像中快速计算矩的方法。由于许多有效的不变量都是从反复的实验中得到的, Kapur 等开发了一系列算法用来系统地寻找特定的几何不变性[30]。Gross 和 Latecki 还开发出了一种方法, 能够在图像数字化的过程中保持物体边缘的定性微分几何[30]。另外, 文献[31]还提到了一种代数曲线和不变量的框架, 用来在混杂的场景中表示复杂物体。它用多项式拟合来表示局部几何信息, 用几何不变量进行对象的匹配和识别。

### 2.3.3 基于内角的形状特征

在文献[32]中提出了一种基于内角的形状特征表达方法。与傅立叶描述符一样, 我们首先将物体近似的表达成多边形的形式。多边形内角的对形状的表达和识别非常重要, 可以表达为

$$\text{Intra-angle} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$$

显然, 基于内角的形状描述与形状所在位置、旋转和大小无关, 因此它非常适于图像检索系统。

以下是一系列从内角导出的形状特征定义:

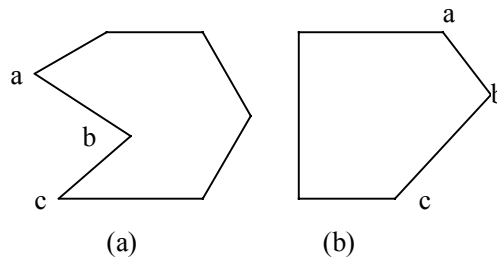
- **顶点数:** 多边形的顶点数目越多, 形状就越复杂。把具有不同顶点数目的两个形状当作很不相似的两个形状是有一定的合理性的。

- **内角平均值:** 多边形所有内角的平均值从一定程度上反映了多边形的形状属性。例如三角形的内角平均值为 60 度, 与矩形的内角平均值 90 度之间有较大差别。
- **内角标准方差:** 多边形内角的标准方差为

$$\delta = \sqrt{\sum_{i=1}^n (a_i - \bar{a})^2}$$

其中  $\bar{a}$  是内角的平均值。该标准方差  $\delta$  是多边形的总体描述。多边形越规则,  $\delta$  值越小。因此, 它可以用来分辨正多边形和不规则多边形。

- **内角直方图 (Intra-angle histogram):** 首先将  $0^\circ$  - $360^\circ$  的角度范围等分成  $k$  个区间, 作为直方图的  $k$  个 bin, 然后统计每个角度区间中的内角数目。内角直方图反映了内角的总体分布。



我们以上图中内角  $\theta = \angle abc$  的计算为例介绍内角的计算方法。设  $a$ 、 $b$ 、 $c$  三点的中心为  $p$ , 则有

$$o\vec{p} = \frac{o\vec{a} + o\vec{b} + o\vec{c}}{3}$$

其中  $o$  为原点。如  $p$  在多边形内部, 则  $\theta$  小于  $180^\circ$ , 否则  $\theta$  大于  $180^\circ$ 。当  $\theta \leq 180^\circ$  时,

$$\theta = \arccos\left(\frac{|ab|^2 + |bc|^2 - |ac|^2}{2|ab||bc|}\right)$$

当  $\theta > 180^\circ$  时,

$$\theta = 360 - \arccos\left(\frac{|ab|^2 + |bc|^2 - |ac|^2}{2|ab||bc|}\right)$$

### 2.3.4 其它形状特征

近年来在形状表示和匹配方面的工作, 包括有限元法 (Finite Element Method 或 FEM)、旋转函数 (Turning Function)、和小波描述符 (Wavelet Descriptor) 等方法。FEM [33] 定义了一个稳定性矩阵来描述物体上的每一个点与其它点之间的联系。这个稳定性矩阵的特征向量被称作特征空间的模合基。所有的形状都首先映射到这个特征空间, 再在特征值的基础上计算形状相似性。类似于傅立叶描述符的思路, Arkin 等提出了旋转函数用来比较凹面或凸面多边形的相似性[34]。在文献[35]中, Chuang 和 Kuo 用小波变换来描述物体形状。它几乎包含了符合我们要求的所有性质, 如不变性、单一性、稳定性和空间位置等。在众多的形状匹配算法中, Chamfer 匹配方法吸引了不少研究者的兴趣。Barrow 等首先提出了 Chamfer 比较法[36], 该方法能够以线性的时间复杂度比较两个的形状块集合。为加快匹配的速度, Borgerfos 提出了分层 Chamfer 匹配算法[37], 这种匹配算法可以在不同

的精确层次上进行,从粗糙到精确。另外,IBM所开发的QBIC图像检索系统[38]采用曲率、离心率和主轴方向等参数作为形状特征。

除了二维形状表示法外,还有许多用于三维形状表达的方法。在[39]中,Wallance和Wintz提出了傅立叶描述符的标准化方法,它包含了所有形状信息,而且计算效率很高。他们还利用了傅立叶描述符的良好插补能力,有效地表示了三维空间中的形状。在文献[40]中,Wallance和Mitchell提出了兼顾结构和统计方法的局部形状分析法来表达三维形状特征。Taubin提出了用一套代数无关矩来同时表示二维空间的形状特征和三维空间的形状特征[41],大大减少了形状匹配的计算量。

尽管计算上述的形状特征并不复杂,但发明一种符合人们主观判断的形状相似性度量算法还是一个有待解决的难题。同时,要在图像检索中充分使用形状特征,还必须有鲁棒的图像自动分割的通用算法。最近的研究工作中,有人提出了适于处理大规模图片库的图像自动分割算法[42]。

## 2.4 图像的空间关系特征

上述的颜色、纹理和形状等多种特征反映的都是图像的整体特征,而无法体现图像中所包含的对象或物体。事实上,图像中对象所在的位置和对象之间的空间关系同样是图像检索中非常重要的特征。打个比方,蓝色的天空和蔚蓝的海洋的颜色直方图上是接近而难以辨别。但如果我们指明是“处于图像上半部分的蓝色区域”,则一般来说就可以区分天空和海洋。由此可见,包含空间关系的图像特征对检索有很大帮助。

空间关系特征可以分为两类:一类方法首先对图像进行自动分割,划分出其中所含的对象或颜色区域,然后根据这些区域对图像索引;另一类方法则简单地将图像均匀划分若干规则子块,对每个图像子块提取特征建立索引。下面我们分别介绍这两类方法。

### 2.4.1 基于图像分割的方法

这类方法中的图像空间关系特征主要包括二维符号串,空间四叉树和符号图像。匹兹堡大学的Chang提出了用二维符号串(2D-String)[43]的方法,其基本思想是将图像沿x轴和y轴方向进行投影,然后按2D子串匹配进行图像空间关系的检索。该方法比较简单,但缺点在于仅利用对象质心不足以表达对象的空间位置关系,而且描述的空间关系太简单,实际的空间关系要复杂得多。符号图像[44](symbolic image)方法是基于图像中全部有意义的对象已经被预先分割出来的假设,将每个对象用质心坐标和一个符号名字代表,从而构成整幅图像的索引。这些方法都假设所有对象都可以通过一定的特征被精确地识别出来,因而只需要关注如何匹配对象的空间关系即可。然而,对象并非总是由某些确定的特征构成的。此外,除了少数特殊应用外,图像自动分割对大多数应用来说是相当困难的。下面我们介绍一些常用的图像分割算法。

在[45]中,Lybanon等用基于形态学动作进行自动图像分割。他们用各种类型的图像来测试算法效果,包括光学天文图、红外线的海洋图和磁力图等。这种模拟方法在处理以上科技图像有良好的效果,但处理一般图像的效果还有待进一步证实。Li[46]等提出了基于的模糊熵的分割算法。这种方法是这样的:以这样的事实为前提,即熵的局部最大值对应于图像上各个区域之间的不确定性。它对于那些直方图上没有明显起伏的图像是非常有效的。

所有以上提到的算法都是自动的，其主要优点是可以从大量的图像中提取边界而不占用用户的时间和精力。然而，如果通用领域内没有经过预处理的图像，这种自动的分割技术效果就不太好。通常，算法所划分的仅仅是区域而不是对象。如果想在图像检索中获得高层语义上的对象（实体），就需要人工的辅助。Samadani 和 Han[47]提出计算机辅助下的边界提取法，将用户手工输入和计算机图像边界生成算法结合起来。Daneels 等[48]提出了一种有关有效轮廓的更完善的方法。该方法首先在用户出入的基础上，用贪婪法获得快速初始收敛，然后再动态地改进边框轮廓。Rui 等[49]提出了基于色彩、纹理空间中的聚类算法。首先由用户指出图像上感兴趣的区域，再用这个算法将该区域聚合成有意义的对象。

### 2.4.2 基于图像子块的方法

为了避免准确的图像自动分割的困难，同时又要提供一些有关图像区域空间关系的基本信息，一种折衷的方法是将图像预先等分成若干子块（可能是重叠的），然后分别提取每个子块的各种特征。在检索中，我们首先根据特征计算图像的相应子块之间的相似度，然后通过加权计算总的相似度。类似的方法还有四叉树方法，即将整个图看成四叉树的结构，每一个分支都拥有的直方图来描述颜色特征。此方法可以支持对象空间关系的检索方法，例如“找出草地上是蓝色天空的图片”的查询可以通过把图像分割成上下两个子块，在每个子块中匹配相应的特征来实现。

虽然这类方法从概念上来说非常简单，但这种普通规则的分块并不能精确的给出局部色彩的信息，而且计算和存储的代价都比较昂贵。因此，在这类方法在实际中应用较少，从而给基于对象空间关系的图像检索带来了困难。

## 2.5 多维图像特征的索引

为了使基于内容的图像检索技术能够扩展到应用于大规模的图像库，我们必须采用有效的多维索引技术。存在的难题有两个方面：

- 1) 高维数：通常情况下，图像特征向量的维数的数量级是 $10^2$ 。
- 2) 非欧拉的相似度度量：由于欧拉度量方法可能无法有效地模仿人类对视觉内容的所有感知，我们经常需要采用其它的相似性度量方法，例如直方图的交、余弦、相关性等非欧拉的相似度衡量方法。

要解决上述这些问题，可行的途径是首先采用维数缩减技术降低特征向量的维数，然后使用适当的多维索引技术（通常能够支持非欧拉的相似度衡量方法）。

### 2.5.1 维数缩减技术

虽然图像特征向量的维数非常高，但内在必需的维数并不高[50,51]。因此我们在使用任何索引技术以前，最好首先进行维数缩减。常用的两种缩减方法是 Karhunen-Loeve 变换（KLT）和按列聚类。

KL 变换和它的变种在面部识别、特征图像、信息分析和主成分分析（Principal Component Analysis）等领域内的应用已经得到了深入的研究。Ng 和 Sedighian [52]曾采用特征图像的方法来实

现维数缩减, Faloutsos 和 Lin 在[53]中提出了用于维数缩减的 KLT 快速逼近算法。研究表明, 大多数的实数集合(视觉特征向量)可以大量的缩减维数, 并且对检索效果不会产生明显的影响。Chandrasekaran [54]等开发了低秩奇异值分解(Singular Value Decomposition)更新算法, 它能够被高效而稳定地应用 KL 变换。由于图像检索系统是一个的动态系统, 不断会有新的图像加入到系统中, 索引结构也需要相应地进行动态更新。奇异值分解算法就提供了这样一种动态更新索引结构的工具。

除了 KL 变换, 聚类是另一种实现维数缩减的有力工具。聚类技术被广泛地应用于模式识别、语音分析、信息检索等领域。通常的聚类方法是将相似的对象(如模式、信号和文档等)聚合在一起, 以实现识别或分组等功能, 即所谓的按行聚类。同样, 聚类也可以按列进行, 从而缩减特征空间的维数[55]。实验表明这种方法非常简单有效。

值得一提的是, 盲目的缩减维数是非常危险的, 因为如果维数被缩减到必要的维数以下, 图像特征的信息就有可能丢失。为了避免盲目的维数缩减, 事后的验证阶段是十分必要的。在众多的验证方法中, Fisher 判别式[11]可以提供有效的监督。

## 2.5.2 多维索引技术

尽管经过了维数缩减, 图像特征向量的维度仍然较高, 因此我们需要选择一个合适的多维索引算法来为特征向量建立索引。有三个研究领域对多维索引技术做出过贡献, 它们分别是计算几何、数据库管理系统和模式识别。现在较流行的多维索引技术包括 Bucketing 成组算法、k-d 树、优先级 k-d 树、四叉树、K-D-B 树、HB 树、R 树以及它的变种 R<sup>+</sup> 树和 R\* 树等等。除了以上几种方法, 在模式识别领域有广泛应用的聚类和神经网络技术也是可能的索引技术。

多维索引技术的历史可以追溯到 20 世纪 70 年代中期。就在那个时候, 诸如 Cell 算法、四叉树和 k-d 树等各种索引技术纷纷问世, 但它们的效果都不尽人意。在 GIS 和 CAD 系统对空间索引技术的需求推动下, Guttman 于 1984 年提出了 R 树索引结构[56]。在他的工作的基础上, 许多 R 树的变种被开发出来, 例如 Sellis 等提出了 R<sup>+</sup>树[58], Greene 也提出了他的 R 树的变种[57]。在 1990 年, Beckman 和 Kriegel 提出了最佳动态 R 树的变种—R\*树[59]。然而, 即便是 R\*树也无法处理维数高于 20 的情况。

文献[51,52]中给出了对各种索引算法的综述和对比。其中 White 和 Jain [51]的研究目标是提供通用的或者领域相关的索引算法。受到 k-d 树和 R 树的启发, 他们提出了 VAM k-d 树和 VAMSplit R 树。实验表明 VAMSplit R 树具有最优的算法效率, 但失去了 R 树所具有的动态特点。在[52]中, Ng 和 Sedighian 提出了一种用于图像检索的三阶段检索技术, 即维数缩减、现有索引算法的评估和选择、对所选索引算法的优化。

考虑到几乎所有的树状结构索引技术都是为传统的数据库查询(点查询和范围查询)所设计, 而并非为了图像检索所设计, 因此有必要探讨能够符合图像检索要求和特点的新的索引结构。Tagare 在[60]中研究了这样的技术, 他提出了一种树结构的调整方法, 能够通过删减妨碍相似度查询效率的树节点来优化树结构。

上述的方法仅仅讨论了如何对图像检索中的高维数据进行索引, 而没有涉及到非欧拉的相似度度量问题。图像检索中的许多相似度计算方法都不是基于欧拉距离的, 比如颜色直方图的求交。在

这个方面上有两种很有希望的技术, 分别是聚类技术和神经网络。Charikar 等提出了适用于动态信息检索增加的聚类技术[61]。这种技术具有动态的结构, 能够处理高维数据, 同时支持非欧拉的相似度量。在[62]中, Rui 和 Chakrabarti 等在速度、精确性和对非欧拉相似度量度的支持等方面进一步扩展了这种技术。

在[63]中, Zhang 和 Zhong 提出了用自组织 (Self-Organization Map, 或 SOM) 神经网络构造树状索引结构的方法。SOM 的好处在于具有无人监督的记忆能力、动态聚类功能和支持任意的相似度量。通过在纹理集上进行的实验表明, SOM 是一种很有希望的索引技术。

## 参考文献

1. A. K. Jain, *Fundamentals of Digital Image Processing*, Englewood Cliffs, Prentice Hall, 1989.
2. J. R. Smith and S. F. Chang, "VisualSEEK: a fully automated content-based image query system," *ACM Multimedia 96*, Boston, MA, Nov. 1996.
3. S. Belongie, et al., "Recognition of images in large databases using a learning framework," *Technical Report 97-939*, U.C. Berkeley CS Division, 1997
4. M. Ioka, "A method of defining the similarity of image on the basis of color information," *Technical Report RT-0030*, IBM Research, November 1989.
5. M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, Vol. 7, No. 1, pp.11-32, 1991.
6. Y. Gong, H. J. Zhang and T. C. Chua, "An image database system with content capturing and fast image indexing abilities", *Proc. IEEE International Conference on Multimedia Computing and Systems*, Boston, 14-19 May 1994, pp.121-130.
7. M. Stricker and M. Orengo, "Similarity of color images," *SPIE Storage and Retrieval for Image and Video Databases III*, vol. 2185, pp.381-392, Feb. 1995.
8. John R. Smith and Shih-Fu Chang. Tools and techniques for color image retrieval. In *Proc. of SPIE: Storage and Retrieval for Image and Video Database*.vol 2670, 1995.
9. G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," *IEEE Workshop on Applications of Computer Vision*, pp. 96-102, 1996.
10. J. Huang, et al., "Image indexing using color correlogram," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 762-768, Puerto Rico, June 1997.
11. John. R. Smith and Shih-Fu Chang. Transform features for texture classification and discrimination in large image databases. In *Proc.IEEE Int. Conf. on Image Proc.* 1995.
12. John R. Smith and Shih-Fu Chang. Automated binary texture feature sets for image retrieval. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.*, May 1996.
13. Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein. Texture features for image classification. *IEEE Trans. On Sys, Man, and Cyb*, SMC-3(6):610-621, 1973.
14. H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. On Systems, Man, and Cybernetics*, vol. Smc-8, no. 6, June 1978.
15. Niblack, et al., "The QBIC project: querying images by content using color, texture, and shape," *Proc. of SPIE, Storage and Retrieval for Image and Video Databases*, Vol. 1908, February 1993, San Jose, pp. 173-187.
16. J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognition*, Vol. 25, No. 2, pp. 173-188, 1992.
17. R. W. Picard, T. Kabir, and F. Liu, "Real-time recognition with the entire Brodatz texture database," *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 638-639, New York, June 1993.
18. B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, No. 8, pp. 837-842, Aug. 1996.
19. T. Chang and C. C. Jay Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. On Image Processing*, vol. 2, no. 4, pp. 429-441, October 1993.
20. A. Laine and J. Fan, "Texture classification by wavelet packet signatures," *IEEE Trans. Pattern Analysis and Machine*



- Intelligence*, vol. 15, no. 11, pp. 1186-1191, Nov. 1993.
21. W. Y. Ma and B. S. Manjunath, "A comparison of wavelet features for texture annotation," *Proc. of IEEE Int. Conf. on Image Processing*, vol. II, pp. 256-259, Washington D.C., Oct. 1995.
  22. Calvin C. Gotlieb and Herbert E. Kreyzig. Texture descriptors based on co-occurrence matrices. *Comput. Vis., Graphics, and Image Proc.*, 51:70-86, 1990.
  23. J. G. Daugman, "Complete discrete 2D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. ASSP*, vol. 36, pp. 1169-1179, July 1998.
  24. A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 55-73, January 1990.
  25. A. K. Jain and F. Farroknia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, 24(12), pp. 1167-1186, 1991.
  26. M. H. Gross, R. Koch, Li. Lippert, and A. Dreger. Multiscale image texture analysis in wavelet spaces. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.
  27. K. S. Thyagarajan, Tom Nguyen, and Charles Persons. A maximum likelihood approach to texture classification using wavelet transform. In *Proc. IEEE Int. Conf. on Image Proc.*, 1994.
  28. M. K. Hu, "Visual pattern recognition by moment invariants," in J. K. Aggarwal, R. O. Duda, and A. Rosenfeld, *Computer Methods in Image analysis*, IEEE computer Society, Los Angeles, CA, 1977.
  29. L. Yang and F. Algrejtsen, "Fast computation of invariant geometric moments: A new method giving correct results," *Proc. IEEE Int. Conf. On Image Processing*, 1994.
  30. Deepak Kapur, Y. N. Lakshman, and Tushar Saxena. Computing invariants using elimination methods. In *Proc. IEEE Int. Conf. on Image roc.*, 1995.
  31. David Copper and Zhibin Lei. On representation and invariant recognition of complex objects based on patches and parts. In *Springer Lecture Notes in Computer Science series, 3D Object Representation for Computer Vision*. M. Hebert, J. Ponce, T. Boulton, A. Gross, Eds., New York,: Springer, 1995, pp. 139-153.
  32. Yueting Zhuang, "Intelligent multimedia information analysis and retrieval with applications to visual design", Ph.D thesis, Zhejiang University, 1998. (庄越挺, "智能多媒体信息分析与检索的研究", 博士论文, 浙江大学, 1998.)
  33. A. Pentland and R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *Int. J. Comput. Vis.*, 18(3):233-254, 1996.
  34. Esther M. Arkin, L. Chew, D. Huttenlocher, K. Kedem, and J. Mitchell. An efficiently computable metric for comparing polygonal shapes. *IEEE Trans. Patt. Recog. And Mach. Intell.*, 13(3), March 1991.
  35. Gene C. H. Chuang and C. C. Jay Kuo. Wavelet descriptor of planar curves: Theory and applications. *IEEE Trans. Image Proc.*, 5(1): 56-70, Jan 1996.
  36. H. G. Barrow. Parametric correspondence and chamfer matching: Two new techniques for image matching. In *Proc. 5<sup>th</sup> Int. Joint Conf. Artificial Intelligence*, 1997.
  37. Gunilla Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Trans. Patt. Recog. And Mach. Intell.*, 10(6): 849-865, 1988.
  38. ACM. Proc. int. conf. on multimedia. 1993-1997.
  39. I. Wallace and O. Mitchell. Three-dimensional aircraft recognition algorithm using normalized Fourier descriptors. *Comput. Vis., Graphics, and Image Proc.*, 13:99-126, 1980.
  40. I. Wallace and P. Wintz. An efficient three-dimensional shape analysis using local shape descriptors. *IEEE Trans. Patt. Recog. And Mach. Intell.*, PAMI-3(3):310-323, MAY 1981.
  41. G. Taubin. Recognition and positioning of rigid objects using algebraic moment invariants. In *Proc. SPIE Vol. 1570 Geometric Methods in Computer Vision*, 1991.
  42. W. Y. Ma and B. S. Manjunath, "Edge flow: a framework of boundary detection and image segmentation," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 744-749, Puerto Rico, June 1997.
  43. S.-K. Chang, Q. Y. Shi, and C. Y. Yan, "Iconic indexing by 2-D strings," *IEEE Trans. Pattern Anal. Machine Intell.*, 9(3), pp. 413-428, May 1987.
  44. V. N. Gudivada and V. V. Raghavan, "Design and evaluation of algorithms for image retrieval by spatial similarity," *ACM Trans. on Information Systems*, vol. 13, no. 2, pp. 115-144, April 1995.
  45. M. Lybanon, S. Lea, and S. Himes. Segmentation of diverse image types using opening and closing. In *Proc. IEEE Int. Conf. on Image Proc.* 1994.
  46. X. Q. Li, Z. W. Zhao, H. D. Cheng, C. M. Huang, and R. W. Harris. A fuzzy logic approach to image segmentation. In *Proc. IEEE Int. Conf. on Image Proc.* 1994.
  47. Ramin Samadani and Cecilia Han. Computer-assisted extraction of boundaries from images. In *Proc. SPIE Storage and*

- Retrieval for Image and Video Databases*, 1993.
48. Dirk Daneels, D. Campenhout, Wayne Niblack, Will Equitz, Ron Barber, Erwin Bellon, and Freddy Fierens. Interactive outlining: An improved approach using active contours. In *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 1993.
  49. Yong Rui, Alfred C. She, and Thomas S. Huang. Automated shape segmentation using attraction-based grouping in spatial-color-texture space. In *Proc. IEEE Int. Conf. on Image Proc.*, 1996.
  50. D. White and R. Jain. Algorithm and strategies for similarity retrieval. In *TR VCL-96-101*. University of California, San Diego, 1996.
  51. D. White and R. Jain. Similarity indexing: Algorithms and performance. In *Proc. SPIE Storage and Retrieval for Image and Video Database*, 1996.
  52. R. Ng and A. Sedighian. Evaluating multi-dimensional indexing structures for images transformed by principal component analysis. In *Proc. SPIE Storage and Retrieval for Image and Video Database*, 1996.
  53. C. Faloutsos and King-Ip (David) Lin. Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In *Proc. Of SIGMOD*, pages 163-179, 1995.
  54. S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image analysis. *Comput. Vis., Graphics, and Image Proc.* 1997.
  55. G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, New York, 1982.
  56. A. Guttman. R-tree: a dynamic index structure for spatial searching. In *Proc. ACM SIGMOD*, 1984.
  57. D. Greene. An implementation and performance analysis of spatial data access. In *Proc. ACM SIGMOD*, 1989.
  58. T. Sellis, N. Roussopoulos, and C. Faloutsos. The R+ tree: A dynamic index for multi-dimensional objects. In *Proc. 12<sup>th</sup> VLDB*, 1987.
  59. N. Beckmann, H. P. Kriegel, R. Schenier, and B. Seeger. The R\*-tree: an efficient and robust access method for points and rectangles. In *Proc. ACM SIGMOD*, 1990.
  60. Hemant Tagare. Increasing retrieval efficiency by index tree adaption. In *Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with IEEE CVPR '97*, 1997.
  61. Moses Charikar, Chandra Chekur, Tomas Feder, and Rajeev Motwani. Incremental clustering and dynamic information retrieval. In *Proc. of the 29<sup>th</sup> Annual ACM Symposium on Theory of Computing*, pages 626-635, 1997.
  62. Yong Rui, Kaushik Chakrabarti, Sharad Mehrotra, Yunxin Zhao, and Thomas S. Huang. Dynamic clustering for optimal retrieval in high dimensional multimedia databases. In *TR-MARS-10-97*, 1997.
  63. HongJiang Zhang and Di Zhong. A scheme for visual feature based image retrieval. In *Proc. SPIE Storage and Retrieval for Image and Video Database*, 1995.

## 第3章 像相似度方法

### 3.1 视觉特征的相似度模型

在基于文本的检索方法采用的是文本的精确匹配，而基于内容的图像检索则通过计算查询（例子图像）和候选图像之间在视觉特征上的相似度匹配进行。因此，定义一个合适的视觉特征相似度度量方法对检索的效果有很大的影响。由于上一章中的视觉特征大都可以表示成向量的形式，常用的相似度方法都是向量空间模型（vector space model），即将视觉特征看作是向量空间中的点，通过计算两个点之间的接近程度来衡量图像特征间的相似度。

#### 3.1.1 $L_1$ 距离和 $L_2$ 距离

如果图像特征的分量之间是正交无关的，而且各维度的重要程度相同，两个特征向量  $A$  和  $B$  之间距离可以用  $L_1$  距离或者  $L_2$  距离（也称为欧拉距离）来度量。其中  $L_1$  距离可以表示为：

$$D_1 = \sum_{i=1}^N |A_i - B_i| \quad (32)$$

其中  $N$  是特征向量的维数。类似地， $L_2$  距离可以表示为：

$$D_2 = \sum_{i=1}^N (A_i - B_i)^2 \quad (33)$$

#### 3.1.2 直方图相交

上述两种距离度量方法常用来计算颜色直方图之间的距离。度量直方图距离的另一种方法是直方图相交（histogram intersection）。假设  $I$  和  $Q$  是两个含有  $N$  个 bin 的颜色直方图，则它们之间的相交距离表示为：

$$\sum_{j=1}^N \min(I_j, Q_j) \quad (34)$$

直方图的相交是指两个直方图在每个 bin 中共有的像素数量。有时，该值还可以通过除以其中一个直方图中所有的像素数量来实现标准化，从而使其值属于  $[0,1]$  的值域范围。

$$S(I, Q) = \sum_{j=1}^N \min(I_j, Q_j) / \sum_{j=1}^N Q_j \quad (35)$$

#### 3.1.3 二次式距离

对于基于颜色直方图的图像检索来说，二次式（quadratic form）距离已被证明比使用欧拉距离或是直方图相交距离更为有效。原因在于这种距离考虑到了不同颜色之间存在的相似度。两个颜色直方图  $I$  和  $Q$  之间的二次式距离可以表示为：

$$D = (Q - I)^t A (Q - I) \quad (36)$$

这种方法通过引入颜色相似性矩阵  $A$ ，使其能够考虑到相似但不相同的颜色间的相似性因素。

其中  $A = [a_{ij}]$ ,  $a_{ij}$  表示直方图中下标为  $i$  和  $j$  的两个颜色 bin 之间的相似度。颜色相似性矩阵  $A$  可以通过对色彩心理学的研究中获得[64]。与此等价的另一种做法是先对颜色直方图进行求闭包操作,使每个颜色 bin 的值都受到来自它相邻颜色 bin 的影响。这样,颜色直方图本身就包含了不同颜色之间的相似性因素,因此可以直接地使用欧拉距离或直方图相交距离。这种对直方图预处理的方法的好处在于在检索过程中计算相似度的代价较小。

### 3.1.4 马氏距离

如果特征向量的各个分量间具有相关性或者具有不同的权重,可以采用马氏距离 (Mahalanobis distance) 来计算特征之间的相似度。马氏距离的数学表达为:

$$D_{\text{mahal}} = (A - B)' C^{-1} (A - B) \quad (37)$$

其中  $C$  是特征向量的协方差矩阵。该距离标准常用来计算 SAR 特征的相似度。

当特征向量的各分量间没有相关性,马氏距离还可以进一步简化,因为这时只需要计算每个分量的方差  $c_i$ 。简化后的马氏距离如下所示:

$$D_s = \sum_{i=1}^N \frac{(A_i - B_i)^2}{c_i} \quad (38)$$

对某个图像特征选择一种合适的相似度衡量方法是获取满意的检索效率的重要保证。然而,更为重要和困难的是确定不同特征之间或是同一特征的不同分量之间的权重。

### 3.1.5 非几何的相似度方法

上述的各种方法都是基于向量空间模型的,采用几何距离作为相似度度量。这样的距离函数通常要满足距离公理的自相似性、最小性、对称性和三角不等性等条件。然而,早在 1950 年,Attneave[65]用几何距离对一组四边形的感知相似性进行了实验,发现距离度量方法和人对相似性的感知判断之间存在一定差距。Tversky 值出相似性的最小性原理在一些识别中并不一定成立。同时,对于相似对称性原则,在一定情况下存在着方向性[66]。对相似三角不等性也同样存在着一些争议。

1977 年, Tversky 提出了著名的特征对比模型 (contrast model) [67]。与几何距离不同,该模型不把每个实体看作特征空间中的一个点,而将每个实体用一个特征集来表示。设两个实体  $a$  和  $b$ , 它们对应的特征集分别为  $A$  和  $B$ , 则两个特征间应当满足匹配性、单调性和独立性假设。基于这样的假设, Tversky 提出了对比模型定理: 对于满足上述的 3 个假设的度量函数  $s$ , 一定存在一个相似度度量函数  $S$  和一个非负函数  $f$ , 以及两个常量  $\alpha, \beta > 0$ , 对于实体  $a, b, c, d$  和它们的特征集  $A, B, C, D$ , 有如下公式:

$$\begin{aligned} S(a, b) &> S(c, d) \Leftrightarrow s(a, b) > s(c, d) \\ S(a, b) &= f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \end{aligned}$$

$f$  是一个反映特征显著性的函数,衡量指定特征对相似度的贡献。当  $\alpha \neq \beta$  时,相似函数是不对称的。Tversky 的理论扬弃了几何模型下相似度度量的优缺点,提出了一个广泛的理论衡量方法。这个方法缺点是还不够实用,只适合那些特征明显的对象,而且对于  $f$  函数的表示形式并不是唯一的,在

具体的应用环境中还需要进一步的明确。

## 3.2 图像特征的性能评价

通过上面的章节可知，基于内容的图像检索中采用多种不同的图像特征和相似度算法。对于某个特定的图像库，我们需要选择一种或多种最有效的图像特征和相似度算法。这需要对不同条件下的检索效果进行全面地评价，比较不同方法的优劣，找出最好的方法。

对检索效果的评价在于检索结果的正确与否，主要使用的是查准率 (precision) 和查全率 (recall) 两个指标。查准率的含义是在一次查询过程中，系统返回的查询结果中的相关图像的数目占所有返回图像数目的比例；查全率则指系统返回的查询结果中的相关图像的数目占图像库中所有相关图像数目（包括返回的和没有返回的）的比例。用户在评价查询结果时，可以预先确定某些图像作为查询的相关图像，然后根据系统返回的结果来计算查准率和查全率。这两个指标的值越高说明检索方法的效果越好。

在[]中 Ma 和 Zhang 对一些常用的颜色特征和纹理特征的检索效率进行了全面的对比。他们的实验是基于 Corel 照片库和 Brodatz 图片集[68]中的 20,000 幅彩色图像进行的。这些图像在内容上有很大的差异，因此可以保证实验的客观公正。为简化起见，所有的图像特征都是从整幅图像中提取的，而没有将图像划分为区域来分别提取特征。检索性能是在一定返回结果数量的前提下通过查全率来衡量的。下面是他们详细的实验结果。

### 3.2.1 颜色特征的比较

他们的实验中对比了四种基本的颜色特征，它们分别是颜色直方图、颜色矩、颜色聚合向量和颜色相关图。

**颜色直方图：**实验对比了不同颜色空间中的颜色直方图对检索效果的影响。这些颜色空间包括 Luv 空间、Lab 空间和 HSV 空间和它的两个变形 (HSV 圆柱形和 HSV 圆锥形空间)。没有进行 RGB 颜色空间的实验是因为它不适于度量颜色之间的主观相似度。实验结果表明 HSV 空间和 HSV 圆柱形空间能够提供最优的检索效果。

**颜色矩：**和颜色直方图的实验结果相反，对于颜色矩来说采用 Luv 和 Lab 颜色空间比用 HSV 颜色空间的效果更好。同时，使用三个低次矩的平均效果好于只使用两个低次矩（均值和标准方差）的情况。但是，有时候使用第三阶矩可能使特征对图像中的场景变化过于敏感，因此有时反而会降低检索效率。另外，实验表明颜色矩的维数尽管远远低于颜色直方图，而检索效果却很接近。

**颜色聚合向量：**由于颜色聚合向量是从颜色直方图演变而来，因此不难理解它在 HSV 颜色空间中的应用效果最优。因为颜色聚合向量包含了空间相关信息，从理论上来说它应该比颜色直方图的检索效果更好。然而，实验结果表明颜色聚合向量只对那些颜色或纹理比较统一的图像有更好的效果，对一般的图像而言没有优势。同时，对于颜色聚合向量和颜色直方图来说，用  $L_1$  距离比使用  $L_2$  距离的效果更好。

**颜色相关图：**实验主要观察了颜色直方图和其他颜色特征相对检索效果的影响。结果表明，颜色相关图能够提供最佳的检索效果，但它的计算也最为复杂。

### 3.2.2 纹理特征的比较

这一节中将给出2.2节中描述的各种纹理特征的性能对比。基于 Corel 图片库的实验表明, 这些纹理特征的性能从最好到最差分别是多维自回归纹理特征 (MRSAR)、Gabor 特征、TWT 小波变换、PWT 小波变换、改进的 Tamura 特征、粗糙度直方图、方向性直方图和传统 Tamura 特征。注意, 所谓改进的 Tamura 特征是指用直方图代替单个数值来表示传统 Tamura 特征中的粗糙度指标。实验结果证明, 这个改进大幅度提高了 Tamura 纹理特征的效果。另外, 所谓粗糙度直方图和方向性直方图都是改进后的 Tamura 特征的分量。

除了使用 Corel 图片外, 实验中也采用了 Brodatz 纹理库来测试纹理特征的性能。这个纹理库常常被用以纹理分析和分类算法的测试, 因此是用来评价纹理特征很好的一个测试集。目前这个库中有 116 幅图片, 每一幅被等分成 16 张互不重叠的小图, 则总共有 1856 张小图。来自于同一幅原图的小图被认为是互相相关的, 则每张用于检索的小图都有 15 张相关图像。基于 Brodatz 的实验结果和前面基本一致, 但 Gabor 特征的性能提高到 MRSAR 特征的水平。另外, 传统 Tamura 特征的性能也优于粗糙度和方向性直方图的水平, 这说明它比较适于描述像 Brodatz 这类图片纹理比较均匀一致的图像集。

## 3.3 其它图像检索技术

### 参考文献

64. H. J. Zhang and D. Zhong, "A Scheme for visual feature-based image indexing," Proc. of SPIE conf. on Storage and Retrieval for Image and Video Databases III, pp. 36-46, San Jose, Feb. 1995.
65. Santini S, Jain R. Similarity measures. <http://www-cse.ucsd.edu/users/ssantini>.
66. D. R. Xu. Research on the imagery generation in Design. Ph.D dissertation, Zhejiang University, Hangzhou, 1995.
67. Tversky A. Feature of similarity. Psychological Review, 1977, 84 (4): 327~352.
68. P. Brodatz, "Textures: A photographic album for artists & designers," Dover, NY, 1966.