1 图像检索概述

1.1 引 言

1990 年，万维网(World Wide Web)出现，并在随后的几年中获得了空前的发展，而且互联网上的信息量以指数形式飞速增长。现在，Internet 已成为一个浩瀚的海量信息源，人们之间的信息交流也因此达到了空前的广泛，由于人们生存在数字化的环境中，地理和民族的界限正在缩小，“数字化地球”正从梦想变为现实。由于 Web 上包含的信息异常丰富，这使得 Web 成为人们查找信息以及信息交互的一个重要媒介。为有效提高在 Web 上信息查找以及信息利用的效率，人们迫切需要能够从互联网上快速、有效地发现资源的工具。一般来讲，Web 信息资源具有以下的特点。

1.1.1 数量巨大，增长迅速

随着网络覆盖范围的不断扩大以及网络技术的发展，存在于网络上的信息资源以飞快的速度传播并迅速增长。据统计，截至 2007 年全球在互联网上的网站总数达到 1.5 亿个。

1.1.2 内容丰富，形式多样

数量巨大的网络信息资源来源于各行各业，包括不同学科、不同领域、不同地区、不同语言的各种信息。此外，互联网上的信息类型也变得更加丰富，即由单一的文本方式逐步变为以图

形、图像、动画、视频3D模型等多媒体信息为主的表现方式。

1.1.3 不稳定性，强交互性

网络信息资源的地址、链接及内容本身处于经常变动之中，使得信息资源的更迭、消亡无法预测，因此整个网上信息资源目前的状态都是不稳定的。同时，与传统的媒介相比，交互性是网络信息传播的一大特点，具体表现在主动性、参与性和可操作性等方面。

一方面，Web信息资源的这些特点对有效地信息挖掘构成了巨大的挑战。另一方面，信息本质上是多模式的，人类自身对信息的处理也是多模式的。随着数字存储、数字检索和数字传输等技术的巨大突破，人们可以用更有效的方法来数字化、存储、检索和传输令人感兴趣的视听内容，进一步加强人类对多媒体系统的需求。俗话说，“百闻不如一见”，与文本信息相比，图像具有更为直观逼真、形象生动、易于理解等优点，它越来越受到人们的青睐，成为互联网上重要的多媒体表现形式。作为结果，目前各种针对Internet的图像搜索引擎纷纷问世，极大地方便了用户对Internet图像进行检索。由于Internet上的图像大多数都是嵌入在Web网页中的图像，本书的研究也将是针对于Web网页中的图像，相应的系统即为Web图像检索系统。

Web图像的检索不同于传统的基于文本的图像检索(Text Based Image Retrieval，TBIR)和基于内容的图像检索(Content Based Image Retrieval，CBIR)，网络环境中的图像一般是嵌入在Web网页中发布的，其所在网页的上下文信息为更好地分析和提取图像特征提供了丰富的外部信息。本书通过互联网中图像的多类型关联数据(多模特性)之间的相互作用来弥合“语义鸿沟”，建立以用户为中心的检索体系，根据用户的兴趣模型组织检索结果，实现面向用户的个性化图像检索。

1.2 国内外研究现状

随着 Internet 的普及发展，人们已经进入了一个信息化社会，信息对日常的生活起到了越来越重要的作用，人们可以方便地接触到大量的信息，信息资源不足的问题已不存在。但是人们也感觉到，目前最大的问题不是信息的缺乏或不足，而是信息量的严重膨胀，人们突然发现自己所面对的信息远远超出其处理能力。现代社会已经进入了一个“数据爆炸”和“信息丰富，但有用信息获取困难”的社会[1]。

目前，多媒体信息正逐渐成为网上最重要的信息载体，并且正在以惊人的速度快速膨胀，每天都有数以千兆字节的多媒体信息产生。这些多媒体信息主要由数字图像、音频、视频和 3D 模型组成，其中图像占据网络多媒体信息的主导地位。Web 图像中包含了大量有用的视觉信息，它们构成了人类认知世界的重要功能手段。

用户最终希望的是能够基于概念和语义模式来访问多媒体信息。这就要求有一种能够从 Web 海量数据中快速准确地查找和访问图像的技术，即图像检索技术。迄今为止，Web 图像检索技术和模型层出不穷，但根据被应用于索引和检索图像的内容的不同，可以大致将其分为四种类型，即基于文本信息的图像检索、基于内容的图像检索、基于语义的图像检索和基于文本和视觉信息融合的 Web 图像检索[2]。下面分别对其进行介绍。

1.2.1 基于文本信息的图像检索

图像检索的历史可以追溯到 20 世纪 70 年代，由于数据库技术的进步而建立和发展了基于文本的图像检索技术，并取得了一定成果，例如数据建模、多维数据索引、查询优化和查询评估

等。图像数据研究者们在对图像进行文本标注的基础上，对图像进行基于关键字的检索。其基本步骤是先对图像文件建立相应的关键字或描述字段，并将图像的存储路径与该关键字对应起来，然后用基于文本的数据库管理系统来进行图像检索。该方法实质是把图像检索转换为与该图像对应的文本检索，文献[3]对该技术进行了较为全面的综述。

早期的图像检索系统采用手工方式对图像进行关键字注释，建立图像索引。这种人工标注的方法仅适用于有限范围的图像库管理系统，它存在着两个主要问题[4,5]：一是对多媒体数据进行人工标注费时费力，尤其是面对海量的多媒体数据库时，人工标注工作量巨大；二是图像内容非常丰富，人工注释所采用的少量文字很难充分表达图像的内涵。更重要的是由于人们对图像理解存在着主观性，这种主观性导致注释的模糊性，直接影响到检索结果的准确性。

但在互联网环境中，Web 图像数据是海量的且在动态更新，采用人工方式对图像进行广泛的关键信息标注是无法实现的。现有相关研究主要集中在如何采用合适的算法，从 Web 中相关的文本信息中提取图像的主题，实现图像的自动标注。显然，对图像自动标注的准确性依赖于 Web 中图像关键信息的提取算法。通常可以从图像的以下几个外部信息中提取这些关键信息。

1.2.1.1 图像的文件名及网址

大多数作者直接通过文件名来表示图像的内容，如 car. png 和 tiger. jpg 等，这样图像在上传后，其内容体现在文件名之中。同时，图像的网址信息有时也提供了一些相关的语义信息，如“http：//www. mingcha1. com/images/animal/tiger/ tiger1. jpg”就提供了图像所属的类别信息及其语义信息。

1.2.1.2 图像的替代文字

替代文字 Alt 在网页中通常用来表示图像的语义信息，它常

常用于由于网络或者其他故障导致图片不能在客户端浏览器正常显示时的替代显示文本。

1.2.1.3 图像周围的文字

通常 Web 图像周围的文字是最可能表达图像所包含的内容的，因此常被选择成为图像的语义特征之一。然而图像周围的文字同样包含很多无用信息，对图像的语义构成噪声。

1.2.1.4 图像所在页面的标题

有些图像被用来加强作者的意图，此时它们的内容同页面的标题内容直接相关，因此页面的标题也可作为图像的语义特征之一。

1.2.1.5 图像的超链接

图像的超链接信息很多时候与图像的内容相关，因此一些语义特征可以通过对超链接的分析计算得到。

1.2.1.6 图像所在网页彼此间的链接

通过对网页与网页间的链接分析，网页内所包含的图像彼此间语义上的相似性可以在一定程度上计算得到。

目前一些搜索网站提供了图像搜索功能。这些网站采用搜索引擎从 Web 网页的相关文本中提取图像的关键信息，建立图像索引数据库，并让用户利用关键词进行检索。下面介绍几个典型的图像搜索网站。

(1) GoogleT 图像检索(http：//images. google. com/)

在 Web 空间，Google 是最全面、最好用的图像搜索工具之一。其图像搜索的工作原理是利用网络蜘蛛技术，通过分析页面上图像附近的文字、图像标题以及许多其他元素来确定图像内容，并使用复杂的算法删除重复信息，并确保在搜索结果中首先显示质量最好的图像。用户在关键词框内输入描述图像内容的关键词，便可得到最贴切的相关内容。

(2)Yahoo 图像检索(http://images.search.yahoo.com/)

Yahoo 以分类目录的形式将标引内容分为艺术、商业与经济、计算机和 Internet、教育等 14 大类，用户可以利用它的搜索引擎以关键词的方式查询它的目录。Yahoo 目录的最大特点在于信息的分类工作由专家手工进行。与其他由计算机自动分类的搜索引擎相比，Yahoo 目录更具科学性。

(3)AltaVista 图像检索(http://www.altavista.com/image/)

Alta Vista 属于非专职类图像搜索引擎，它提供基于文件名和扩展名的图像搜索功能。它可以查出 HTML 的 IMG SRC 字段，该字段包括图像路径和文件名。如果它找到一个 GIF 或 JPG 扩展名，就会试图将该文件名或目录名与用户搜索的词汇进行比较。在 Alta Vista 的多媒体信息检索页，用户可对图像、MP3/Audio、视频等进行选择检索，并可对检索结果、来源进行限定。

(4)Baidu 图像检索(http://image.baidu.com/)

百度搜索引擎是世界上最大的中文搜索引擎，图像搜索是百度搜索引擎的新增功能，它从 1.5 亿个中文网页中提取各类图片建成中文图片库，库存图片 26 万件。它的最大特色是“图片搜索分类目录”、“明星、人物图片搜索指南”和“风景图片搜索指南”功能。

(5)卡内基梅隆大学的 SLIF 检索系统

SLIF(Subcellular Location Image Finder)系统自动对生物医学文献进行处理，分别提取生物医学文献的文本特征和文献内嵌图片的视觉特征，将文献的文本特征和文献内嵌图片的视觉特征分别储存到特征数据库。系统通过文献内嵌图片的视觉特征对文献进行索引，用户可以通过图像来检索相应的文献[6]。SLIF 系统的相关资料可以从 http://murphylab.web.cmu.edu /services/SLIF/ 下载。

(6)东北林业大学的明察 1 号(mingcha1)图像检索系统

明察1号(http://mingcha1.com/soutu/soutu.jsp)支持基于内容的图像检索，用户可以通过提交查询示例图像来检索图像。系统可以对用户的兴趣进行学习，提供面向用户的个性化图像检索服务。

由于基于文本信息的图像检索系统仅仅考虑网页的文本信息而忽略了图像本身的内容，因此从大量的检索结果中只能依靠用户浏览再次选择满意的图像，工作量很大。而Web图像检索的特点是用户通常只对排列在返回结果队列前面的图像有兴趣，因此提高检索结果队列的靠前结果的精确率是一个迫切需要解决的问题。

1.2.2 基于内容的图像检索

为了克服基于文本的图像检索在检索图像内容上存在的不足，20世纪90年代出现了基于内容的图像检索(Content-Based Image Retrieval，称CBIR)技术[7]。CBIR系统主要采用图像低层视觉特征来描述图像内容，通过特征匹配算法检索图像。在许多情况下，也引入相关反馈等交互式学习方法来提高图像检索的性能。CBIR仅采用客观视觉特征来寻找视觉上相似的图片，图像相似性无需人的理解，从而无需或者仅需少量人工干预，故大量应用于自动化场合。

CBIR系统的通用框架大致可以用图1-1描述，主要分成四大部分：图像存储、特征提取、匹配机制以及用户系统。原始图像经过预处理后存储在图像数据库中，然后对图像库中的数据进行特征提取，得到特征库，经过一定的高维索引机制处理之后，采用某种匹配算法将图像库中的图像与用户查询要求进行匹配，并将结果通过用户接口返回给用户。

1.2.2.1 图像低层视觉特征

特征(或内容)的自动提取和表示是基于内容图像检索的基

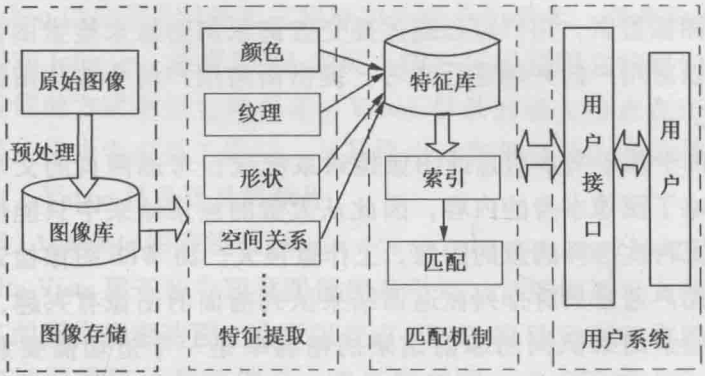

图 1-1 CBIR 系统框架图

础。图像的特征主要包括低层特征和语义特征两个方面。低层特征主要包括图像的颜色、纹理、形状和空间关系等定量的特征；图像的语义特征则是一种定性特征，是对图像内容与人类主观概念吻合的抽象描述。由于计算机视觉和图像理解技术在自动提取图像的语义特征方面尚未完全成熟，因此目前的 CBIR 系统主要使用图像低层特征来索引和检索图像。

(1)颜色

颜色是图像的一个重要特征，是图像检索中具有最为广泛应用的特征之一。颜色特征的提取也受到极大重视并得到深入研究。相对几何特征而言，颜色特征稳定性好，对大小、方向等均不太敏感，而且颜色特征也是人类认知物体的一个主要特征[8]。

典型的颜色特征有：

①颜色直方图。

颜色直方图是表示图像中颜色分布的一种统计值。颜色直方图检索简单，具有平移、尺度以及旋转不变性，但它丢失了颜色的空间分布信息，在一定程度上容易造成误检现象[9]，因此较多

改进的算法被提出。文献[10]采用彩色图像的主要颜色来构造颜色直方图，由于忽略了那些数值较小的颜色区间，改进后的颜色直方图对图像噪声的敏感程度降低了，从而使检索效果更好。文献[11]提出的采用累加直方图方法提高了采用直方图特征进行图像检索的效率。

②颜色矩。

Stricker等提出的颜色矩[12]算法，依据颜色分布信息主要集中在低阶矩中，因此仅采用颜色直方图特征的一阶中心矩、二阶中心矩和三阶中心矩就足以表达图像的颜色特征。颜色矩计算简单，可以有效表示图像中颜色分布。黄朝兵等[13]进一步提出了多统计矩直方图特征，一阶中心矩表征了邻域颜色的“平均值”，而二阶中心矩则表征了邻域颜色与“平均值”的偏差程度。

③颜色信息熵。

为克服直方图维数较高的问题，Zachary[14]提出采用图像颜色的信息熵表示图像的颜色特征，从而将图像的颜色直方图由多维降低到一维。Sun等[15]将颜色信息熵和颜色直方图融合在一起，提高了算法的检索率，同时使得特征具有了空间分布特性。

④颜色相关图。

颜色相关图是利用图像中像素间的颜色关系来描述图像颜色空间分布的另一种表达方式[16]。颜色相关图不但刻画了某一颜色的像素占整个图像的比例，还反映了不同颜色对之间的空间相关性。

(2)纹理

纹理是描述图像中同质现象的视觉特征，目前还没有正式统一的定义，通常表现为局部不规则但宏观有规律的特征，例如云彩、树木、砖等。纹理特征刻画了邻域像素灰度的分布规律，包含了关于物体表面组织结构排列的重要信息以及它们与周围环境的联系，在基于内容的图像检索中得到了广泛应用。常用的纹理

描述方法有统计法、频谱法、结构法和模型法四种[17]。

①统计法。

统计法分析纹理的主要思想是通过图像中灰度级分布的随机属性来描述纹理特征。Haralick 等提出关于纹理特征的共生矩阵表示[18]，该方法研究的是灰度级纹理的空间依赖关系，首先根据图像像素之间的方向和距离构造一个共生矩阵，然后从该矩阵中提取出有意义的统计信息作为纹理表达。Tamura 等在人类对纹理视觉感知的心理学研究基础上，提出了 Tamura 纹理特征[19]。Tamura 纹理特征的六个分量对应于心理学上纹理特征的六个属性，分别是粗糙度(Coarseness)、对比度(Contrast)、方向度(Directionality)、线像度(Linelikeness)、规整度(Regularity)以及粗略度(Roughness)。Tamura 纹理特征在视觉上是有意义的，故被应用到许多图像检索系统之中，如 QBIC 系统[20]。

②频谱法。

频谱法主要借助于频率特性来描述纹理特征。常用的频谱法主要包括傅立叶功率谱法，Gabor 变换[21]、塔式小波变换(Pyramid Wavelet Transform)[22]、树式小波变换(Tree Wavelet Transform)[23]等。

③结构法。

结构法分析纹理的基本思想是假定纹理模式由纹理基元以一定的、有规律的形式重复排列组合而成，特征提取就变为确定这些基元并定量地分析它们的排列规则。由于纹理基元描述了局部纹理特征，因此对整幅图像中不同纹理基元的分布统计可获得图像的全面纹理信息。

④模型法。

模型法主要利用一些成熟的图像模型来描述纹理特征，如基于随机场统计学的马尔可夫随机场[24]、分形模型[25]、自回归模型等。这些模型的共同特点是通过少量的参数来表征纹理。

(3)形状特征

图像的形状特征比图像的颜色、纹理具有更高的语义，在图像检索中也具有更重要的作用[26]。但是，由于自然图像中的对象和区域的分割原本就比较困难，因此一般情况下图像的形状特征是难以自动提取的。此外，如何寻找一种符合人们主观判断的形状相似性度量算法仍是一个有待解决的难题。通常图像形状特征的表示方法主要有两类：基于边界的表示方法和基于区域的表示方法[27]。

基于边界的形状特征是在边缘检测的基础上，用形状数、周长、角点、链码、兴趣点等来描述物体的形状。傅立叶描述子即是一种最典型的基于边界的形状描述方法，它将经过傅立叶变换后的边界作为形状特征，这样可以用较少的参数表示很复杂的边界[28]。Rui 等[29]提出的改进傅立叶描述子算法，不仅对噪声具有很好的鲁棒性，而且具有几何变换不变性。Mokhtarian 等提出的 CSS(Curvature Scale Space)方法[30]是一种对平面曲线曲率零交叉点的多尺度描述，并将图像转化为尺度空间图像的匹配，这种方法的复杂度较高。

基于区域的表示方法将区域形状当做一个整体看待，描述了区域像素的统计分布特征，受噪声和形状变化的影响相对较小。基于区域的方法中常用面积、重心和偏心率等来对区域形状做最基本的描述，使用最普遍的描述方法是矩方法。Hu[31]证明了利用二阶和三阶中心几何矩组成的矩组，在物体平移、缩放和旋转时保持不变。Khotanzad 等[32]将 Zernike 矩用于图像的识别取得了不错的效果。The 等[33]比较了各种正交矩和非正交矩，发现 Zernike 矩性性能更为出众。

Manjunath 等[34]比较了基于边界的表示(链码，傅立叶描述子，UNL 傅立叶描述子)、基于区域的表示(矩不变量，Zernike 矩，pseudo - Zernike 矩)和联合表示(矩不变量和傅立叶描述子，

矩不变量和 UNL 傅立叶描述子)的性能。他们的实验表明，边界和区域联合表示的性能优于简单的边界或区域表示法。

1.2.2.2 图像相似性匹配

图像检索的匹配策略大致可以分为两种，一种是完全匹配，另外一种是相似性匹配。当两个图像的特征完全相同时图像匹配成功，称之为完全匹配。当两个图像特征间的距离小于某一个阈值时，图像匹配成功，称之为相似匹配。在基于内容的图像检索中占主导地位的是建立在图像低层视觉特征对比基础上的相似性检索匹配。在颜色、纹理和形状等图像低层视觉特征被提取出来以后，可采用相应的相似性度量策略来进行特征匹配，即通过确定检索图像同数据库目标图像特征向量间的距离来判定二者间的相似性。

显然，一个合适的相似性度量方法对图像检索结果影响很大。相似性度量方法的好坏不仅会影响到图像检索的性能，也会影响到图像检索的用户响应时间。理想的相似性度量方法不仅应该满足人的视觉特性，还应具有较低的计算复杂度。

目前，大多数的图像检索系统的相似性度量是基于欧氏距离函数，它也是最常见的距离度量函数。它的缺点是事先假定了图像特征的各分量之间是正交无关的，而且各维数的重要程度相同。Mahalanobis 距离在基于欧氏距离的基础上加入了协方差矩阵的权重影响，它适用于特征向量的各个分量间具有相关性或者具有不同权重的情形。

另一种常用的距离度量是直方图相交法(histogram intersection)，是由 Swain 等于 1991 年首次提出的[35]。直方图相交法计算简单快速并且能较好地抑制背景的影响，但计算量偏大。二次式距离[36]由于考虑了不同颜色之间存在的相似度及颜色之间的相关性，因此检索结果更加符合人的视觉感觉，但其相关性对称矩阵的计算量较大。其他的相似性度量方法还有余弦距

离、相关系数、Kullback - Leibler 散度、Jefrrey 散度[37]和 x^2 距离[38]等。

在上述的度量方法中没有任何一种方法可以适用于所有特征向量间的相似性度量，其主要原因是上述度量方法具有特征依赖的特点[39]。不同的特征应该应用不同的度量方法，例如，直方图相交法不适合于非直方图的特征，虽然二次式距离可以有效的度量颜色直方图的距离，但在对其他特征向量的距离度量效果却没有采用欧氏距离度量的效果好[40]。

1.2.2.3 经典的 CBIR 系统

鉴于基于内容的图像数据库检索系统的重要性，各大公司和科研机构陆续推出了一些商用或研究用的图像检索系统，下面简单加以介绍。

(1) QBIC

QBIC[41,42](Query by image content)检索系统是由 IBM 公司 Almaden 研究中心开发的。该系统是第一个真正的功能较为齐全的 CBIR 系统，它通过友好的图形界面，为使用者提供了颜色、纹理、形状等多种检索方法。QBIC 的索引子系统中，首先用 KLT 变换来完成降维，然后采用 R^+ 树来构造多维索引结构。QBIC 的最新版系统采用了基于文本的关键字查找方式和基于内容的图像检索相结合的方式。QBIC 系统建立较早，技术成熟，功能全面，为基于内容的图像检索技术的验证和推广作出了很大贡献。QBIC 的演示见 http：//www. qbic. almaden. ibm. com。

(2) Visualseek 和 Webseek

Visualseek[43]和 Webseek[44]系统是美国哥伦比亚大学开发的姊妹系统，主要利用图像区域空间关系进行查询和从压缩域提取视觉特征来进行检索。系统中使用颜色特征和基于小波变换的纹理特征，利用基于 Quad - Tree 和 R - Tree 的索引结构以提高检索速度。Webseek 主要是面向 Web 的搜索引擎，支持关键词检索，

并使用户相关反馈技术来改善检索结果。

(3) Virage

Virage[45]是由 Virage 公司开发研制的基于内容的图像搜索引擎。其特点是提供完善的用户开发功能，如提供用于开发用户界面的工具包；提出 Primitive 概念，用来支持用户定义新的图像视觉特征；支持 5 种抽象数据结构便于图像特征的描述，用户还可以根据需要来调整一些基本图像特征的权重；提供用户相关反馈检索机制。因此，该系统比较适合用来进行特定应用领域图像数据库的二次开发。Virage 已经和多种商业数据库进行了集成。

(4) Netra

Netra[46]是 UCSB 数字图书馆项目中图像检索原型系统。它在分割后的图像区域内使用颜色、纹理、形状等特征，以及图像分割后的子区域的位置信息来描述图像，它的特点是使用 Gabor 滤波器进行纹理分析，使用人工神经网络用于分类。

(5) Photobook

Photobook[47,48]是 MIT 多媒体实验室开发的图像检索系统，Photobook 有三个子模块，分别用于提取形状、纹理和面部特征，用户可以在每一个子模块中使用相应的特征进行查询。在 Photobook 的最新版本 FourEyes 中，图像标注和检索过程中加入了人机交互以提高检索精度。

(6) MARS

MARS[49]系统是 UIUC 大学开发的支持图像低层视觉特征复合检索的图像检索系统。其特点是使用比较全面的图像低层特征，提供基于树结构的多特征的组合检索。采用的图像特征有：使用 HSV 空间中 HS 上的颜色直方图描述图像的颜色；抽取图像纹理的粗糙度和方向性以及对比度等特征描述纹理；采用图像的规则分割方法对图像特征的空间进行描述；根据纹理对图像进行分割来实现图像中的对象描述，并对分割后的图像按照敏感性进

行分组；使用 Fourier 描述子对图像中对象的形状进行描述。MARS 最早将相关反馈思想引入了图像检索系统，在不同层次上使用了相关反馈技术，包括查询矢量优化、自动匹配工具选择、特征自适应等，使得检索精确率得到很大改善。

1.2.3 基于语义的图像检索

从目前基于内容的图像检索演示系统的检索结果看，检索效果并不理想，其根本原因是低层的视觉特征与高层的图像语义之间存在的“语义鸿沟”[50]。低层的视觉特征不能代表图像丰富的内涵，用户搜索图像更关心的是概念层次上图像的内容和图像表现的寓意，也就是图像的高层语义。因此，图像检索的理想方式是根据图像的语义进行检索下面对语义层次模型、语义提取方法以及语义表示方法进行概述。

1.2.3.1 语义层次模型

图像的语义是层次化的，也可以说图像的语义是有粒度的，不同层次的语义粒度不同，可以采用多层结构进行分析。

Hong 等[51]将图像内容定义为三层结构，即特征层(basic visual content)、对象层(object content)和场景层(scene content)。第一层为特征层，由图像的视觉特征集合组成，如颜色、纹理、形状等特征。该层的语义主要对应于视觉特征语义。第二层为对象层，是通过对图像中的对象的视觉特征分析理解得到的对对象的语义描述。这一层需要先获取图像中的对象，如“帆船”、“树”、“水”等，然后从对象的视觉特征、空间关系、位置等信息中推导出对象语义。该层主要对应于对象语义和空间关系语义。第三层是对多个对象和场景的语义描述，称为场景层，例如“城市”、“乡村”等。该层是对一组对象语义进行分析得到整个场景的语义，对应于场景语义。从实质上而言，对象层和场景层真正利用了图像的语义，是图像语义研究关注的重点。

王惠锋等[4]则将语义层次模型进一步划分为六层，自下而上依次为特征语义、对象语义、空间关系语义、场景语义、行为语义以及情感语义。其中前四层的语义分类与以前的方法类似，后两层则包含了更高级更抽象的语义。其中行为语义指图像所代表的行为或活动，如一场足球赛中的各种行为；情感语义是图像带给人的主观感受，如让人喜悦等。

Jaime 和 Chang[52]把图像内容概括成五层，包括区域层(region)、感知区域层(perceptual - area)、对象部件层(object - part)、对象层(object)以及场景层(scene)。其中对象层和场景层的含义与 Hong 等的类似，区域层是指图像中分割出来的连通的区域；感知区域层是相邻且感知相似的区域的集合；对象部件层由多个感知区域组成。该模型的前四个层次大致对应于对象语义和空间关系语义，而场景层则对应于场景语义。

1.2.3.2 语义提取方法

限于目前计算机对图像内容理解的技术水平，直接从图像的像素数据或低层视觉特征推理得出图像的高层语义很困难。目前将低层图像特征映射到高层语义的图像语义提取方法主要分为两种，分别是基于知识的语义提取和人工交互语义提取。

(1)基于知识的语义抽取

基于知识的语义提取通过有监督和无监督的学习将图像归并到某种语义类，从而在一定程度上获得图像的语义信息。Luo 等[53]利用 Bayes 网络将低层特征和语义特征结合起来进行图像语义理解。Aksoy 等[54]在图像区域分割和分类的过程中采用 Bayes 分类器对用户给定的正例和反例进行训练，得到语义特征。Han 和 Qi[55]利用 MIL - based SVM 和 global - feature - based SVM 来标注图像语义。Kun - seok 等[56]用多策略小波提取图像的颜色和纹理特征作为图像的特征向量，在此基础上用自组织特征映射算法将图像进行聚类。一些识别算法在对象识别过程中结合了物体的

空间位置关系，如对象间空间关系的表示及它们的相似性匹配可以用2D String[57]、空间方向图(Spatial Orientation Graph)[58]、R-String[59]等方法。但这些空间关系还只限在空间拓扑的层次上，即上下、左右等，更高层的空间语义关系需要应用进一步的领域和外部的知识。

根据识别的对象、获得的对象间的空间关系以及图像的背景，结合场景语义的知识来进行场景分类是一种直观的想法。Fung等[60]将图像划分成固定大小的子块，然后对这些图像子块分别确定其各自对应的语义类别，最后根据子块语义的组合关系来确定整幅图像的语义。

(2)人工交互语义提取

目前通用的完全自动的图像语义处理，还存在一些难以逾越的障碍。要在机器视觉、人工智能现有发展水平上进行语义处理，必须充分考虑到人的作用。人工交互的语义提取主要体现在反馈学习方面。相关反馈方法的基本思想是在检索过程中，允许用户对检索结果进行评价和标记，指出结果中哪些是用户希望得到的查询图像，哪些是不相关的，然后将用户标记的相关信息作为训练样本反馈给系统进行学习，指导下一轮检索，从而使检索结果更符合用户的需要。用户的参与使系统能更好地揣测用户的意图，也使得在低层可视特征和高层语义概念之间建立某种联系成为可能。

1997年，Yong等[61]首次将相关反馈的思想引入基于内容的图像检索，并提出了一种修改查询向量和相似性度量公式的相关反馈方法，基本思想是使相关图像靠近正例中心，远离反例中心。Selim Aksoy等[62]进一步改进了加权相关反馈算法，认为权重是相应图像特征值标准差的比值，并在实验中取得了较好的效果。Meilhac等[63]根据用户的反馈，学习相关图像的先验概率和类条件概率，并用贝叶斯判别式计算每个图像的后验概率，从而