

基于内容的图像检索技术研究

黄祥林, 沈兰荪

(北京工业大学信号与信息处理研究室, 北京 100022)

摘 要: 在对海量的图像数据进行检索时, 传统的基于数值/ 字符的信息检索技术并不能满足要求. 因此, 基于内容的图像检索技术 (CBIR: Content-Based Image Retrieval) 的研究应运而生, 并引起了广泛关注. 本文主要讨论 CBIR 研究中的一些关键问题: 图像的内容特征及其提取、特征之间的相似度计算、查询条件的表达、检索性能的评价、压缩域的图像检索技术等等, 并指出了一些可值得深入研究的方向.

关键词: 基于内容的图像检索; 示例查询; 压缩域图像检索; 多媒体技术

中图分类号: TN919 **文献标识码:** A **文章编号:** 0372-2112 (2002) 07-1065-07

Research on Content-Based Image Retrieval Techniques

HUANG Xiang-lin, SHEN Lan-sun

(Signal and Information processing lab, Beijing Polytechnic University, Beijing 100022, China)

Abstract: With the rapid increasing of image data, an effective and efficient image retrieval system is needed for finding useful images in large databases or on Internet. Because of the complexity of images, the traditional information retrieval techniques, which are based on value and/or text comparison, are not sufficient. Then the research is shifted to the techniques of content-based image retrieval (CBIR). In this paper, some key issues on CBIR are discussed: image features extraction, similarity measures, the expression of query conditions, retrieval techniques in compressed domain, the performance evaluation, and so on. Some future research directions are also given in the conclusion.

Key words: content-based image retrieval; query by example (QBE); image retrieval in compressed domain; multimedia techniques

1 引言

目前, 急剧增长的图像数据主要来源于卫星系统、各类监测系统、科学试验和生物医学等领域, 如 NASA 的地球观测系统每天产生约 1T 的图像数据^[1], 而这些图像数据又通过网络传送到世界各地. 因此, 如何从海量的图像数据中高效、快速地检索出所需信息是当前许多应用领域的重要问题^[29].

传统的信息检索是基于数值/ 字符的, 并不能客观反映图像内容的多样性, 其数学模型、系统结构、查询方式及用户接口等也不具备有效管理和检索图像数据的能力^[2]:

传统的信息检索技术不能有效管理具有时空关系的多媒体数据. 时空关系是多媒体数据的一个重要特征, 只有进行有效的管理才能确保各对象间的时空同步, 它直接影响到数据库的数据模型和索引方式.

传统的信息检索技术没有图像特征提取的工具. 在基于内容的图像检索中, 必须要对图像进行基于内容的识别和解释, 特征提取在图像检索中占有重要地位.

传统的信息检索技术缺乏能表达图像特征多样性的查询方式. 图像的许多特征并不能用简单的数值和字符来精确表达, 因此需要具有能表达图像内容特征的查询方式, 这将影响传统检索系统的用户接口.

可见, 基于内容的图像检索技术 (CBIR) 需要进行包括数据模型、图像内容的获取、人机接口、以及完成检索框架等方面的研究, 一个典型的图像检索系统的基本组成如图 1.

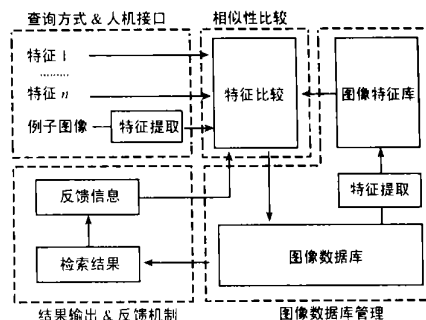


图1 图像检索系统的基本组成

一些研究机构对 CBIR 进行了深入的研究, 并开发了相应的检索系统, 如 Photobook^[57]、Chabot^[59]、WebSEEK^[60]、QBIC^[56]、Virage^[12, 58]等等. 本文准备对 CBIR 的一些关键技术即: 图像的内容特征及其提取、特征之间的相似度计算、查询方式、检索性能的评价策略等进行讨论, 并探讨硬件发展对 CBIR 的影响和压缩域的图像检索技术研究, 最后指出了一些可值得深入研究的方向.

2 图像的内容特征及提取

图像内容可以理解为一个简化了的层次模型, 如图 2 所示: 第一层为原始数据层, 即图像的原始像素点; 第二层为物理特征层, 反映了图像内容的低层物理特征, 如颜色、纹理、形状、轮廓、图像内容的空间关系和时间关系 (对视频来说) 等; 第三层为语义特征层, 是人们对图像内容概念级的反映, 一般

收稿日期: 2001-03-19; 修回日期: 2001-09-29

基金项目: 国家自然科学基金 (No. 60172045); 国家 863 计划 (No. 2001AA121061); 北京市基金 (No. 4002002)

是对图像内容的文字性描述。

图像特征的表示方法有三种:数值表示、关系表示和语义表示。譬如,图像的颜色可用 R、G、B 三种数值表示,图像中对象之间的位置关系就要用到关系表示,而语义表示方法需要对物体进行识别和解释,往往要借助人类的知识推理。采用的表示方式不同,查询时进行相

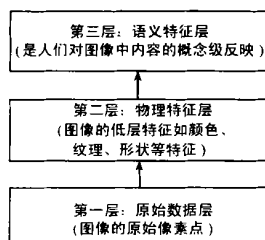


图2 图像内容的层次

似性比较的算法也不一样,如数值式的特征比较可采用多维空间中点的距离来计算,而对基于文字性的语义特征的查询,可直接采用文本信息检索技术。

图像检索所用到的基本特征大多属于第二层和第三层特征,即颜色、纹理、形状、轮廓、空间关系、时间关系、图像的语义等等。

2.1 颜色

颜色是图像的一个重要特征,通常采用 RGB 三色表示,但是这种表示法并不符合人眼视觉特性。颜色比较时一般要把 RGB 空间转换到其他颜色空间中处理,如 HSI 表示法就更能反映人们对颜色的感知。

颜色是图像检索中最先被采用的特征,主要方法有^[22]:

颜色直方图,颜色一致性矢量(CCV:color coherence vectors),颜色相关图,颜色矩等等。颜色直方图的比较是最基本方法^[14],但缺乏图像的空间信息。而 CCV 方法^[33]不仅统计了整幅图像中各颜色的像素值,还统计了图像中各颜色最大区域的像素值,效果较好,但 CCV 并没有强调各颜色最大区域的形状以及与背景的关系,于是 Zachary 等人又考虑了图像的边缘信息,提出了 CCV-TEV(Threshold Edge Vector)算法^[58]。颜色相关图法^[30]强调同一颜色在图像中的空间距离相关性,其检索效果比上述几个方法都要好,但计算量比较大。颜色矩算法主要是采用图像中各颜色的均值和方差作比较,处理简单,可用它作为图像检索的初检,为下一步的细检缩小搜索范围。

颜色特征在图像检索中占有重要的位置,但其固有的光照敏感性也限制了其使用范围。

2.2 纹理

纹理是图像的又一重要特征,虽然尚无精确的定义,但却受到广泛研究。一般认为纹理就是纹理元素有规律的排列组合,而把具有重复性、形状简单、强度一致的区域看作纹理元素。视觉纹理特性主要有:粗糙度、对比度、方向度、线象度、规整度、粗略度等^[4]。图像检索用到的纹理特征表示方法主要有^[22]:Tamura 法(反映了粗糙度、对比度、方向度等)、MRSAR (multi-resolution simultaneous auto-regressive model)、Canny 角直方图法、Gabor 变换、塔式小波变换(PWT:pyramid wavelet transform)、树式小波变换(TWT:tree wavelet transform)等。Tamura 法^[27]是对整幅图像进行处理,效果较差,改进方法是采用直方图^[22]。和其他方法(纹理的主分量分析、小波变换特征等)的比较^[53]可知,MRSAR^[28]区分不同纹理模式的能力较强,但

计算开销也较大。文献^[22]详细比较了各种方法,MRSAR 与 Gabor 方法虽然性能较好,但其复杂度较高、处理时间也较长,在图像检索中使用基于 TWT、PWT 的方法比较合适^[22]。

2.3 形状

形状是物体的一个重要特征,但由于物体形状的自动获取很困难,基于形状的检索一般仅限于非常容易识别的物体^[5,6]。形状可用面积、离心率、圆弧度、形状矩、曲率、分形维等全局和局部特征来表示。其主要的分析方法有不变矩、Fourier 描述符、自回归模型、向心链码、CSS(Curvature Scale Space)、VSW(Variable Scale Wavelet)等,它们都各自具有优缺点^[3,36,39,63]。而基于向心链码的方法^[63]既具有形状的编码能力又同时支持检索,它首先采用向心链码对形状进行编码,再在编码流中直接提取形状的“相对凸数”及“凸度”,以此作为形状检索的依据。由于形状的向心链码具有旋转、平移、尺度的不变性,因此这种检索算法具有一定的抗“相似性形变”能力。这些方法并不涉及到形状的提取算法,均假设形状已被完整提取。

2.4 轮廓

轮廓图就是图像中的各物体外形所形成的图,一般要经过边缘检测、细化等处理后才能得到。QBIC^[56]系统中采用分块相关法进行图像轮廓之间的比较,这种方法不具有尺度和方向的不变性,并且边缘检测、细化等预处理直接决定了图像轮廓的质量,对检索结果具有很大的影响。

2.5 空间关系

空间关系是图像中各对象的基本关系之一,直角坐标系是其最直接的表示法。在 GIS(geographical information system)的应用中,人们对空间关系的表征和检索技术进行了研究,用一种 2D 串^[66]的索引技术来表征空间关系。Liu 和 Sun^[67]还使用直角坐标的方法对物体之间的关系进行了分级性表示,既有粗略的表示法又有精细的表示法。Sistla 等^[68]定义了一套基本关系集来表达空间关系,主要包括:左、右、前、后、上、下、内、外、重叠等,这些定义简化了物体间的空间关系,但是忽略了距离概念,使其应用场合受到限制。

2.6 时间关系、时空关系

时间关系、时空关系是视频序列的一个基本特征。在 VOD 和视频数据库等多媒体应用中,提出了基于点的和基于段的表达时间关系的方法。基于点的方法采用时间轴上的点来代表对象的位置,而基于段的方法则用对象出现的时间间隔来表达对象间的关系,这两种表达方法可以相互转换,它们之间的详细比较见参考文献^[69]。一般将空间关系和时间关系结合起来处理时空关系^[70],目前对于时空模型的研究并不深入,有待进一步探索。

2.7 语义特征

语义特征一般用文字描述,可以用传统的数据库方法和基于知识的模型来管理和检索。目前一般用半自动或手动的方法提取语义特征,在大型数据库或动态数据库中这种提取方法并不太现实,因此只适合特定的应用。另外,图像/视频中的一些文字具有一定的语义信息,可对其自动定位^[62]、识别后进行检索。提取语义特征的自动化程度越高,通常使得适用

的领域越小,如何从图像的物理特征中自动提取语义特征,是一个难题。

国际标准 MPEG7 也规定这些特征的描述(如颜色、纹理、形状等)^[13],为图像检索提供标准化接口。下面讨论如何进行图像特征之间的相似度计算。

3 相似度计算

相似度计算是图像检索的一个重要方面。特征空间中点与点之间的距离度量是一个典型的计算方法,此外还有相关计算、关联系数计算等^[8]。(下列公式中 X 、 Y 为图像, x_i 、 y_i 分别为其特征。)

3.1 距离度量

距离度量就是求特征空间中点与点之间的距离, Minkowski 距离度量为:

$$d(X, Y) = \left(\sum_{i=1}^n |x_i - y_i|^r \right)^{1/r} \quad (1)$$

其中, $r=1$ 时称为 Manhattan 距离,也叫 city-block 距离。 $r=2$ 时为 Euclidean 距离,这是人们通常所说的距离概念。 r 时为 Chebyshev 距离,此时的表达是为:

$$d(X, Y) = \max_{1 \leq i \leq n} |x_i - y_i| \quad (2)$$

另外在聚类分析中经常用到的广义距离为 (W 是协方差矩阵):

$$d(X, Y) = (X - Y)^T W^{-1} (X - Y) \quad (3)$$

在图像检索中,采用哪种距离要看具体情况,但 Manhattan 距离被广泛采用,因为其计算简单,效果也比较好^[9]。

3.2 相关计算

相关计算是计算两个特征矢量之间的相关性,相关值越大,说明越相似。下面是一些常用的相关方法,有余弦相关、Pearson 积矩相关等。

$$\text{余弦相关: } \cos = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \quad (4)$$

若 $\cos = 1$,表示两矢量互为标量倍数,精确匹配,若 $\cos = 0$,则相互垂直,不匹配。

Pearson 积矩相关 (product-moment correlation):

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

\bar{x}_i 、 \bar{y}_i 是数据库中所有第 i 个特征的均值。这种相关是居于均值的矩 (均值 & 方差),常用于聚类分析和多媒体检索^[10]。另外,进行颜色直方图比较时,采用归一化直方图相交的方法^[71]:

$$r_{XY} = \frac{\sum_{i=1}^N \min(x_i, y_i)}{\sum_{i=1}^N x_i} \quad (6)$$

实际这也是一种相关计算,对图像的大小、直方图维数、部分遮挡等都不太敏感^[22],比较颜色直方图时效果较好。

3.3 关联系数计算

图像中的有些特征是二值型的(如状态型特征等),其相似性比较一般采用关联系数的计算方法, Gower 方法是经常使用的方法之一(这里 x_i 、 y_i 分别为二值特征):

$$S_G = \frac{\sum_{i=1}^n s_i}{\sum_{i=1}^n 1} \quad (7)$$

其中, $s_i = x_i \odot y_i$,即如果 x_i 、 y_i 相匹配则 $s_i = 1$,否则 $s_i = 0$ 。 s_i 是一个权重因子,如果是有效特征则置为 1,否则为 0(当然不能全为零)。

图像的特征是一个复杂体,可以有形状、颜色、纹理及其语义特征,还可能具有二值特征等等。假设有共 I 个非二值特征(第 i 个特征的相似度为 s_i), J 个二值特征(第 j 个特征的相似度为 s_j ,匹配时为 1,不匹配时为 0),则整个相似度为:

$$S = \sum_{j=1}^J s_j(X, Y) + \sum_{i=1}^I s_i(X, Y) \quad (8)$$

其中 $s_i > 0$,表明第 i 个特征特征的重要程度,一般由试验决定。在公式(8)中,二值特征对整个结果的影响很大,如果二值特征中有一个不匹配,则认为整个匹配失败。这种计算方法已用于人脸图像的检索中^[11]。实际上图像间的相似度计算是一个很复杂的问题,上述计算方法与人眼视觉系统所具有的测量性能存在很大差距,对图像相似度计算方法的研究仍是一个难点。

4 查询方式

由于图像特征本身的复杂性,查询方式(对查询条件的表达)也应具有多样性,特征不同,查询方式也不尽一样。目前查询方式可归纳为以下几种^[1,12]:

4.1 低层物理特征查询

查询时可以用图像的单个物理特征,也可以同时用多个物理特征,并可以根据特征的不同,赋予不同的权值。如用单一特征“检索 25 %红色、70 %蓝色、5 %绿色的图像”,用多个特征“检索包含蓝色和云彩纹理的图像,其中蓝色的权重为 40 %,纹理为 60 %”。在检索时还可利用图像中各对象的位置关系,如查询“上面为蓝色天空,下面为绿色草地”的图像。

4.2 感兴趣区查询

不是查询整幅图像的内容,而是图像中用户感兴趣的内容,如查询“含有红色汽车”的图像。

4.3 语义特征查询

与低层特征查询不同,语义特征查询是基于文字型的查询方式,包含了自然语言处理和传统信息检索技术。由于图像语义特征的提取目前只限于半自动或手动方式,并存在主观成分,因此目前的应用范围受到很大限制。

4.4 自定义特征查询

自定义特征主要是对图像的一些客观属性的描述,如图像拍摄的日期、地点等,这种查询可采用传统的结构化查询语言(SQL)。如查询“拍摄时间为 2001 年 2 月 22 日,拍摄地为北京”的图像。

上述查询方式都是由用户直接给出特征。而用户要提供这些特征有时是比较困难的(除了语义特征查询和自定义特征查询)。譬如对于基于轮廓的查询,并不是用户都具有勾画出目标图像轮廓的能力。因此,在实际中检索系统一般都希望采用下面的查询方式:

4.5 示例查询

示例查询就是根据给出的例子图像,提取其特征,然后在

图像库中找出和例子图像特征相似的图像. 例子图像的特征可以是单个或多个, 得到这些特征后就可以采用上述几种查询方法. 示例查询为用户提供了一个方便、简单的方式来表达图像的内容, 因此被大多数图像检索系统采用.

4.6 浏览检索

当用户检索目的比较含糊、或不太熟悉图像库中的信息结构和信息类型时, 就可以用浏览检索. 用户首先浏览数据库中的图像, 发现感兴趣的目标图像后把其作为例子图像进行检索. 这与示例查询的区别在于获得例子图像的途径不同.

一个图像检索应用系统并不需要包含所有的查询类型, 但需要综合多种查询方式, 并且对用户是透明的. 此外, 图像并不是孤立存在的, CBIR 还要在广泛的信息检索上下文中进行考虑. 例如, 医学图像的检索不仅要根据其内容, 还要根据相关的其他信息如医生的诊断记录、治疗计划、最后结果等进行查询, 因为这些诊断记录与图像本身同样重要.

由于图像的物理特征与图像内容的语义信息具有一定差距, 检索结果往往不尽人意, 于是人们就提出用相关反馈技术来优化检索结果^[50]. 相关反馈是一个指导性的学习过程, 由用户对每次的检索结果进行确认, 选择一组符合要求和一组不符合要求的图像分别作为正、负反馈提交给系统, 系统据此进行检索优化并输出结果, 如此反复, 直到满意为止. 相关反馈技术对图像检索起着指导性作用, 具有优化检索结果的能力, 实用性较强^[17, 57].

5 检索性能评价^[48]

目前需要一个公共的试验平台来评价 CBIR 系统的性能, 基本的图像检索试验平台应该具有一个图像集 (含 N 幅图像)、一套基准 (benchmark) 查询 (M 个)、一套评价准则.

图像集应具有代表性强、容量大、多样化、不受使用局限等特点. 目前采用的图像集的内容主要有动物、植物、建筑、艺术、飞机、人脸、纹理图像、卫星图像等. 试验中用得最多的是美国加州水资源部门^[50]的图像集和 Corel 的专业图片集. 这两类图像集是否符合试验平台的要求, 还需要得到更多研究者的认可.

基准查询所涉及的内容既要全面又要简练. 目前主要采用低层特征进行查询, 与人们所期望的在较高语义级上检索图像的目标还存在着相当大的差距. 处理办法就是利用多级别的方法逐步实现人们的愿望: 第一步采用低层特征 (如颜色、纹理、形状、空间位置等) 进行查询, 随着图像分析、索引和分类等技术不断发展, 基准查询将更多地包含语义信息. 这样, 可利用试验平台的同一图像集实现不同级别的基准查询. 下面讨论评价方法^[48]:

用 $V_n \in \{0, 1\}$ 表示第 n 幅图像与查询条件的相关性 (1 表示相关, 0 表示不相关). 假设进行某个基准查询时得出了 k 幅图像, 则可计算出正确检测数 $A_k = \sum_{n=0}^{k-1} V_n$ 、错误检测数 $B_k = \sum_{n=0}^{k-1} (1 - V_n)$ 、漏检数 $C_k = A_N - A_k$ 、正确淘汰数 $D_k = B_N - B_k$. 进一步可计算: 正确检测率 $R_k = A_k / A_N$: 所检出的相关图像数与图像库中所有相关图像数的比值. 检测精度 $P_k = A_k / k$: 所检出的相关图像数与所检出的图像总数的比值. 错

误接受率 $F_k = B_k / B_N$: 所检出的非相关图像数与数据库中所有非相关图像数的比值.

当前有多套评价参数, 如 A_k 与 B_k , A_k 与 F_k , A_k / B_k 与 A_k , 在 $R_k = 0.2, 0.5, 0.8$ 处的 P_k , P_k 与 R_k 等, 但采用第套 P_k 与 R_k 评价参数比较直观, 是值得推荐的方法之一.

建立试验平台的一个较大困难是决定图像之间的相关性. 如果图像库中有 N 幅图像和 M 个基准查询, 则需要 MN 次观察才能得到他们的相关性, 而这种观察往往需要人工完成, 因此投入是巨大的, 这时可采用 TREC (Text Retrieval Conference) 的相应方法处理^[49].

6 硬件对检索性能的影响^[51]

80 年代, 在小型机上建立一个 800MB 的文本索引需要 23 天, 查询时间需要 4 秒^[52], 而现有台式机建立 1GB 的文本索引不到 1 小时, 查询时间不到 1 秒^[51]. 可见, 计算机硬件技术的发展对 CBIR 的性能提高起着重要作用^[51].

计算机的硬件发展主要表现为: CPU 的速度每 18 个月大约提高 2 倍, 并且其性能的提高也非常快; 计算机内存配置增大, 大约两年扩大两倍; 磁盘容量增长迅速. 但是硬件的发展并不平衡, 主要表现为: 磁盘的等待时间和寻道时间发展很慢, 1994 年以前每年以 8% 的速度发展^[54], 自此以后发展更慢; 总线宽度和速度的提高总跟不上磁盘容量和处理器速度的发展; 内存的存取时间的提高, 落后于 CPU 速度的增长.

处理器速度及性能 (如 MMX (MultiMedia eXtension) 技术)、磁盘容量、内存容量等的快速提高为 CBIR 系统的应用普及起到了决定性的作用. 值得注意的是, 由于数据的存取在 CBIR 中占有重要的比例, 而磁盘的操作时间决定了数据的存取效率, 因此磁盘性能也是影响 CBIR 的重要因素, 并已成为影响查询响应时间的关键瓶颈.

计算机硬件的发展, 在一定程度上也影响了检索算法的改善, 例如由于多处理器的使用, 可采用并行的检索算法等. 而新技术的发展与应用对 CBIR 的影响有时是难以预测的, 譬如, 目前全息存储技术研究表明^[55], 全息存储的存储密度极限可达到 $1\text{T}/\text{cm}^3$, 其读出速度已达 $1\text{G}/\text{s}$, 这种技术实用化后, 将极大提高检索系统的性能.

7 压缩域图像检索技术

由于大量的图像数据都是以压缩格式存贮和传输的, 因此在提取压缩图像的特征时一般要先进行解压操作, 显然增加了系统的处理时间和复杂度. 压缩域的 CBIR 是直接压缩域提取特征, 提高了处理速度并减少了对计算资源的需求. 其系统组成如图 3 所示, 它可以不要特征库的支持, 非常适合 Internet 网、动态数据库的图像查询.

压缩域图像检索技术的主要研究思路是根据压缩域的特点提取图像特征, 可分为三类: 变换压缩域的检索技术、空域压缩域的检索技术、变换域与空域相融合的检索技术.

7.1 变换压缩域的图像检索技术

许多图像编码方法的基本技术都是以某种数学变换为基础的, 如 DFT 变换、KLT 变换、DCT 变换、子带/小波变换等. 在

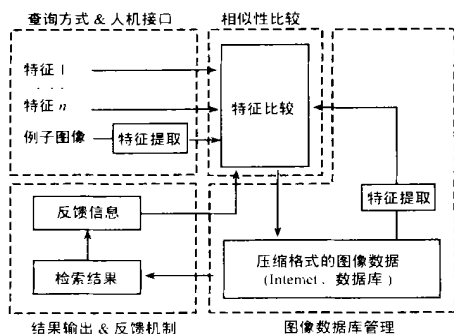


图3 基于压缩域的图像检索系统的基本组成

DFT压缩域中,傅里叶系数的最大幅度、平均幅度、幅度能量、系数的方差、极坐标下的傅里叶系数分布等都可作为图像检索、分类的特征^[16]。KLT变换的实现比较复杂,但在人脸的识别上具有较好的效果^[17]。

在JPEG、MPEG1/2、H.261/263等标准中都采用了DCT变换,研究DCT压缩域的图像检索技术具有重要意义。可用DCT域的各个滤波通道的方差与均值作为图像的匹配特征^[18],也可用DCT前几个AC系数的方差作为检索特征^[19]。文献[20]中将整个图像分成非重叠的K个窗口对(共2K个窗口),根据每个窗口对中DCT系数的相似性组成特征向量,来进行图像检索。另外根据DCT域的区域能量方向性可进行纹理图像的分类^[64,65]和图像字符的提取^[62],这都可利用到图像的检索技术中。

子带/小波变换在图像编码中得到了广泛应用^[7,23](JPEG2000等标准已采用小波编码算法^[47]),其所具有的图像匹配、分类等特点也比较突出,可利用各分辨率的能量及灰度直方图进行图像匹配与分类^[24,26],利用小波和隐马尔科夫模型还可做到图像匹配的灰度、旋转不变性^[25]。各种算法基本上是利用小波各方向子带多分辨率的特点,提取各子带的期望、方差、组织各种直方图等进行图像匹配^[29,31,32]。文献[35]和[18]还提出了直接支持压缩域检索的小波图像编码算法,这也是新一代图像编码的趋势。

7.2 空域压缩域的检索技术

矢量量化和分形是两个重要的图像编码方法。矢量量化本身就是一种索引机制,有利于图像检索^[34],可利用码书的使用图(usage map)^[38]和索引值的直方图^[37]进行图像匹配。而分形编码算法是一种自我索引技术,可直接利用分形码进行图像匹配^[15,40],并优于小波方法^[41](但小波具有更好的纹理表征能力)。文献[42]还提出了一种基于分形码的图像分割方法,可进一步实现图像检索。

7.3 融合检索技术

许多图像编码技术都采用了几种基本的编码算法,如基于矢量量化的小波编码^[23,44]、基于分形的小波编码^[21,43]等,这些编码技术提供了更好的图像检索能力。文献[44]提出了一种在小波域利用矢量量化的图像索引技术,根据小波各子带的特点,对其系数进行矢量量化,以便进行图像检索。Swanson等^[45]设计了一种基于分割的图像编码算法,对所分割的

物体对象利用小波、矢量量化进行编码以提供可直接面向对象的图像索引能力,而对残差图像进行JPEG编码。Podilchuk和Zhang^[46]对DCT域的频率系数进行矢量量化,并利用矢量量化的索引技术进行人脸识别,其结果令人满意。

8 展望

信息时代的一个突出特点就是信息量的爆炸式增长和信息传播的网络化,对海量信息查询,必须有一个有效、快速的搜索引擎。目前的文本搜索引擎对文本信息的查找起着举足轻重的作用,而对于图像信息来说,就必须研究基于内容的图像检索技术,本文对目前的研究现状进行了阐述。虽然研究已取得很大进展,并且还开发了许多CBIR系统,但离真正的实用阶段还有一定的距离,许多方面还值得深入研究:

如何提高查询速度?由于在实际应用中图像数据库很大,速度是优先考虑的问题之一,必须研究优化的检索算法。除计算机硬件的限制外,特征矢量的高维数和现有多维索引方法的低效性是影响查询速度的主要原因,所以要深入研究如何降低特征矢量维和建立有效高速的多维索引方法。另外,研究压缩域的检索方法也是提高查询速度的有效途径。

如何建立简单实用的用户查询接口?目前认为QBE是一个较好的查询方式,但对一般用户来说例子图像的获取(或绘制)仍然比较困难。调查发现^[48],多数用户希望用语义的方式进行查询,因此,建立一个有效的高层语义与低层特征相结合的查询方式是值得思考的方向之一。

目前使用的相似度计算基本上是基于数学比较,与人的视觉感知特性还具有一定的差距,因此,如何建立符合人眼特性的相似度计算模型,是一个不可回避的研究课题。

需要深入研究图像数据库的数据模型。因为目前基于关系和对象的数据库模型不能有效地管理图像(视频)的空间关系、时间关系、空间-时间关系等之间的同步,给图像数据的检索和管理带来了一定的困难。

如何实现从低层物理特征到高层语义特征的转换?虽然这是计算机视觉所研究的传统问题,但并没有突破性进展。在CBIR中如何实现这种转换仍然是一个关键难题。

在CBIR中研究由高层语义导出低层特征更具有实际意义,此时CBIR系统组成如图4。首先,用户的查询方式可以是语义级的,将用户提供的语义内容转换成物理特征后,就可进行查询过程。其次,不用改变现有图像特征库、匹配方式等。因此,这种从语义特征导出图像物理特征的方法应该是在CBIR研究中要引起足够重视的研究方向。

计算机网络已成为我们获取图像信息的重要来源之一,如何利用网络特点来查询图像数据是十分必要的,并已有了一些研究成果^[61~63]。随着网络技术的飞速发展,面向网络的图像查询技术也应作深入的研究。

图像检索技术的研究是一个涉及到多个领域的综合性课题,是对计算机视觉技术、数据库检索和管理技术、用户界面技术、数学建模技术、图像分析处理技术、计算机网络技术以及人类心理学等各个方面成果的融合,只有这样才能建立一个高效、快速、实用的图像检索系统。

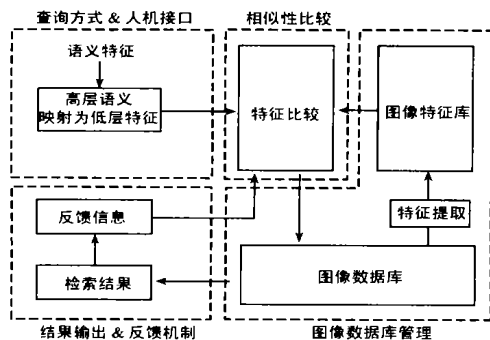


图4 高层语义导出低层特征的图像检索基本结构

参考文献:

- [1] V N Gudivada, V V Raghavan. Content based image retrieval systems [J]. IEEE Computer, 1995, 28(9): 18 - 22.
- [2] A Yoshitaka, T Ichikawa. A survey on content-based retrieval for multimedia databases [J]. IEEE Transactions on Knowledge and Data Engineering, 1999, 11(1): 81 - 93.
- [3] Lifang Wu, Lansun Shen. Variable Scale Wavelet Shape Descriptor [Z]. Beijing China: ISO/IEC JTC1/SC29/WG11 MPEG9/ m6139, 2000.
- [4] 庄越挺, 潘云鹤等. 基于内容的图像检索综述 [J]. 模式识别与人工智能, 1999, 12(2): 170 - 177.
- [5] C Faloutsos, R Barber, M Flickner, et al. efficient and effective querying by image content [J]. J intelligent information systems, 1994, 3(1): 231 - 262.
- [6] A Pentland, R Picard, S Sclaroff. Photobook: Tools for content-based manipulation of image databases [A]. SPIE Storage and Retrieval of Image and Video Databases II [C]. San Jose CA: 1994: 34 - 47.
- [7] 黄祥林, 魏海, 沈兰荪. 基于小波分解的视频分级编码方法 [J]. 高技术通讯, 2000, 10(4): 41 - 44.
- [8] M S Kankanhalli. Introduction to multimedia information retrieval [DB/OL]. 2000, <http://comp.nus.edu.sg>.
- [9] R Brunelli, O Mich. On the use of histograms for image retrieval [A]. IEEE International Conference on Multimedia Computing and Systems [C], Florence, 1999. 2: 143 - 147.
- [10] J K Wu, A D Narasimhalu. Fuzzy retrieval of image databases [A]. Proceedings of the First Asian Fuzzy Systems Symposium [C]. Singapore, 1993. 38 - 40.
- [11] K Wu, A D Narasimhalu. Identifying faces using multiple retrievals [A]. IEEE Multimedia, Proceedings of the First Asian Fuzzy Systems Symposium [C]. Singapore, 1993. 27 - 38.
- [12] Y A Aslandogan, C T Yu. Techniques and systems for image and video retrieval [J]. IEEE Transactions on Knowledge and Data Engineering, 1999, 11(1): 56 - 63.
- [13] R Koenen. MPEG7: Context and objective version 5.0 [Z]. Fribourg witzerland: ISO/IEC JTC1 SC29/WG11 N1920, 1997.
- [14] M J Swain, D H Ballard. Color indexing [J]. Int Journal of Computer Vision, 1991, 7(1): 11 - 32.
- [15] Wei Hai, Shen Lansun. Fractal-based image storage and indexing [A]. Proceedings of SPIE: Storage and Retrieval for Media Databases [C]. San Jose CA, USA: 2000, 3972: 421 - 429.
- [16] A Celentano, V D Lecce. A FFT based technique for image signature generation [A]. Proc SPIE: Storage and Retrieval for Image and Video Databases V 3022 [C]. Newport Beach CA USA: 1997. 457 - 466.
- [17] R Brunelli, O Mich. Image retrieval by examples [J]. IEEE Transactions on Multimedia, 2000, 2(3): 164 - 171.
- [18] 魏海, 沈兰荪. 小波变换域内基于方向梯度相角直方图的图像检索算法 [J]. 电路与系统学报, 2001, 6(2): 20 - 24.
- [19] R Reeves, K Kubik, W Osberger. Texture characterization of compressed aerial images using DCT coefficients [A]. Proc SPIE: Storage and Retrieval for Image and Video Databases V3022 [C]. Newport Beach CA USA: 1997. 398 - 407.
- [20] M Shneier, M S Mottaleb. Exploiting the JPEG compression scheme for image retrieval [J]. IEEE Trans, Pattern Analysis and Machine intelligence, 1996, 18(8): 849 - 853.
- [21] Wei Hai, Shen Lansun. Fractal coding of wavelet image based on human's contrast masking effect [A]. Proceedings of SPIE: Video/ Image Communication and Processing [C]. San Jose CA USA: 2000. 3974: 584 - 594.
- [22] Ma Weiyang, Zhang Hongjiang. Benchmarking of image features for content-based retrieval [A]. The Thirty-Second Asilomar Conference on Signals, Systems & Computers [C]. California, USA: 1998. 1: 253 - 257.
- [23] M Antonini, M Barlaud, P Mathieu, et al. Image coding using wavelet transform [J]. IEEE Trans Image Processing, 1992, 1: 205 - 220.
- [24] T Chang, C C J Kuo. Texture analysis and classification with tree-structured wavelet transform [J]. IEEE Trans Image Processing, 1993, 2(4): 429 - 441.
- [25] J L Chen, A Kundu. Rotation and gray scale invariant texture identification using wavelet decomposition and hidden Markov model [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 1994, 16(2): 208 - 214.
- [26] M K Mandal, T Aboulnasr, S Panchanathan. Image indexing using moments and wavelets [J]. IEEE Trans Consumer Electronics, 1996, 42(3): 557 - 565.
- [27] H Tamura, S Mori, T Yamawaki. Texture features corresponding to visual perception [J]. IEEE Trans on Systems, Man, and Vypernetics, 1978, 8(6): 460 - 473.
- [28] J Mao, A K Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models [J]. Pattern Recognition, 1992, 25(2): 173 - 188.
- [29] Yong Rui, T S Huang, S F Chang. Image retrieval: current techniques, promising directions, and open issues [J]. Journal of Visual Communication and Image Representation, 1999, 10: 39 - 62.
- [30] J Huang. Image indexing using color correlogram [A]. IEEE int Conf on Computer Vision and Pattern Recognition [C]. Puerto Rico: 1997. 762 - 768.
- [31] C E Jacobs, A Finkelstein, D H Salesin. Fast multiresolution image querying [A]. Proc of ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques [C]. Los Angeles: 1995. 277 - 286.
- [32] J Z Wang, G Wiederhold, O Firschein, et al. Wavelet-based image indexing techniques with sketch retrieval capability [A]. Proc of the Forum on Research and technology advances in digital libraries [C].

- Washington, 1995.
- [33] G Pass, R Zabih, J Miller. Comparing images using color coherence vectors [A]. 4th ACM Conf. on Multimedia [C]. Boston: ACM, 1996.
- [34] 魏海, 沈兰荪. 基于分类矢量量化的图像压缩和检索算法 [J]. 电子学报, 2001, 29(7): 933 - 936.
- [35] M K Mandal, F Idris, S Panchanathan. A critical evaluation of image and video indexing techniques in the compressed domain [J]. Image and Vision Computing, 1999, 17: 513 - 529.
- [36] S Loncaric. A survey of shape analysis techniques [J]. Pattern Recognition, 98, 31(8): 983 - 1001.
- [37] F Idris, S Panchanathan. Image indexing using vector quantization [A]. SPIE proceedings: Storage and Retrieval for Image and Video Databases III 2420 [C]. San Jose CA, 1995. 373 - 380.
- [38] F Idris, S Panchanathan. Storage and retrieval of compressed images [J]. IEEE Trans Consumer Electronics, 1995, 41: 937 - 941.
- [39] F Mkhitarian. Silhouette-based isolated object recognition through curvature scale space [J]. IEEE trans Pattern Anal Mach Intell, 1995, 17(5): 539 - 544.
- [40] A Zhang, B Cheng, R S Acharya. Approach to query-by-texture in image database systems [A]. Digital Image Storage and Archiving Systems 2606 [C]. Philadelphia Pa, 1995. 338 - 349.
- [41] A Zhang, B Cheng, R S Acharya, et al. Comparison of wavelet transform and fractal coding in texture-based image retrieval [A]. Visual Data Exploration and Analysis III 2656 [C]. San Jose, 1996.
- [42] T Ida, Y Sambonsugi. Image segmentation using fractal coding [J]. IEEE Trans Circuits and Systems for Video Technology, 1995, 5: 567 - 570.
- [43] J Li. Hybrid wavelet-fractal image compression based on a rate-distortion criterion [A]. Visual Communications and Image Processing [C]. Newport Beach, CA USA: SPIE Proceedings, 1997.
- [44] F Idris, S Panchanathan. Image indexing using wavelet vector quantization [A]. Digital Image Storage Archiving Systems [C]. Philadelphia Pa USA: SPIE Proceedings, 1995.
- [45] M D Swanson, S Hsiao, A H Tewfik. Image coding for content-based retrieval [A]. VCIP 2727 [C]. Orlando FL: Proc SPIE, 1996.
- [46] C Podilchuk, X Zhang. Face recognition using DCT-based feature vectors [A]. Speech and Signal Processing 4 [C]. Atlanta Georgia USA: Proc IEEE Intl Conf on Acoustics, 1996. 2144 - 2147.
- [47] ISO/IEC JTC 1/SC 29/WG1 N1646R. JPEG2000 Image Coding System [S].
- [48] J R Smith. Image retrieval evaluation [A]. IEEE Workshop on Content-Based Access of Image and Video Libraries [C]. Santa Barbara California: IEEE, 1998. 112 - 113.
- [49] D Harman. Overview of the first text retrieval conference [A]. Proc Text Retrieval Conference (TREC) [C]. Washington: TREC, 1992.
- [50] R Schettini, G Ciocca, I Gagliardi. Content - based color image retrieval with relevance feedback [A]. Image Processing, ICIP 99 Proceedings [C]. Kobe Japan: ICIP, 1999. 75 - 79.
- [51] J Zobel, H E Williams, S Kimberley. Trends in retrieval system performance [A]. Computer Science Conference, 23rd ACSC [C]. Australian: ACSC, 1999. 241 - 249.
- [52] D Harman, G Candela. Retrieving records from a gigabyte of text on a minicomputer using statistical ranking [J]. Journal of the American Society for Information Science, 1990, 41(8): 581 - 589.
- [53] B S Manjunath, W Y Ma. Texture features for browsing and retrieval of image data [J]. IEEE T-PAMI, 1996, 18(8): 837 - 842.
- [54] P Chen, E Lee, G Gibson, et al. RAID: high performance, reliable secondary storage [J]. Computing Surveys, 1994, 26(2): 145 - 185.
- [55] J Ashley, C M Jefferson, et al. Holographic data storage [J]. IBM Journal of Research and Development, 2000, 44(3): 341 - 368.
- [56] M Flickner, H Sawhney, W Niblack, et al. Query by image and video content: the QBIC system [J]. IEEE Computer, 1995, 28(9): 23 - 32.
- [57] C Youngsik, K Daewon, R Krishnapuram. Relevance feedback for content-based image retrieval using the choquet integral [A]. Multimedia and Expo, ICME 2000 [C]. New York: ICME, 2000.
- [58] J M Zachary, S S Iyengar. Content based image retrieval systems [A]. IEEE Symposium on Application-Specific Systems and Software Engineering and Technology [C]. Richardson Texas: IEEE, 1999.
- [59] V E Ogle, M Stonebraker. Chabot: retrieval from a relational database of images [J]. Computer, 1995, 28(9): 40 - 48.
- [60] J R Smith, S F Chang. Visually searching the web for content [J]. IEEE multimedia, 1997, 4(3): 12 - 20.
- [61] K D Smith, R B Paranjape. Mobile agents for web-based medical image retrieval [A]. IEEE Canadian Conference on Electrical and Computer Engineering [C]. Ottawa, Canada: IEEE, 1999. 966 - 970.
- [62] 黄祥林, 沈兰荪. 基于 DCT 压缩域的图像字符定位 [J]. 中国图像图形学报, 2000, 7(1): 22 - 26.
- [63] 黄祥林, 宋磊, 沈兰荪. 一种基于向心链码的形状检索方法 [J]. 信号采集与处理, 2001, 16(5): 148 - 152.
- [64] 黄祥林, 沈兰荪. 基于 DCT 压缩域的纹理图像分类 [J]. 电子与信息学报, 2002, 24(12): 216 - 221.
- [65] 黄祥林, 沈兰荪. 一种具有旋转不变性的压缩域纹理图像分类方法 [J]. 电子与信息学报, 2002, 24(11): 1190 - 1196.
- [66] S K Chang. Iconic indexing by 2D string [J]. IEEE trans pattern analysis and machine intelligence, 1984, 6(4): 413 - 428.
- [67] Z Q Liu, J P Sun. Structured image retrieval [J]. J Visual languages and computing, 1997, 8(3): 333 - 357.
- [68] A P Sistla, C Yu, C Liu, et al. Similarity based retrieval of pictures using indices on spatial relationships [A]. Proc intl conf very large databases [C]. Zurich Switzerland, 1995. 619 - 629.
- [69] D Toman. Point vs. Interval-based query languages for temporal databases [A]. Proc Fifth ACM SIGACT/ MOD/ ART symp principles of database systems [C]. Montreal, Canada: ACM, 1996. 58 - 67.
- [70] Y Theodoridis, M Vazirgiannis, T Sellis. Spatio-temporal indexing for large multimedia applications [A]. Proc intl conf multimedia computing and systems [C]. Hiroshima Japan, 1996. 441 - 448.
- [71] F Idris, S Panchanathan. Review of image and video indexing techniques [J]. Journal of visual communication and image representation, 1997, 8(2): 146 - 166.

作者简介:

黄祥林 男, 1967 年 5 月出生于湖南省石门县, 北京工业大学工
作, 博士, 主要研究方向为基于内容的图像检索、压缩域图像/视频处
理等, 曾获国家科技进步三等奖、地矿部科技成果二等奖各一次, 已发
表论文 20 余篇。