

图像检索技术综述

石 军, 常义林

(西安电子科技大学 综合业务网国家重点实验室, 陕西 西安 710071)

摘要: 图像检索是很多研究中的关键技术, 图像检索是一种近似检索, 与检索文本相比, 检索图像要困难得多。根据不同的图像索引对现有的各种图像检索技术进行了分析和比较, 包括基于注释的图像检索、基于特征的图像检索以及基于知识的图像检索, 提出了图像检索技术的发展趋势和研究方向。

关键词: 图像检索; 图像索引; 注释; 特征; 知识

中图分类号: TP311.113 文献标识码: A 文章编号: 1001-2240(2003)04-0486-06

Overview of image retrieval

SHI Jun, CHANG Yilin

(National Key Lab. of Integrated Service Networks, Xidian Univ., Xi'an 710071, China)

Abstract: Implementation of image retrieval is based on an underlying index. Conventional techniques for retrieving textual and numerical data based on simple comparisons are no longer adequate for images, since the digitized representation of an image does not convey its semantic content. This paper presents the current state of the art in image retrieval. We first examine the indexing techniques for visual content and information content for image retrieval. Next, we review recent studies of image retrieval from the point of view of image indexes. The current challenges and feature trends for image retrieval are also given.

Key Words: image retrieval; image index; annotation; feature; knowledge

图像检索就是根据对图像内容的描述, 在目标图像集合中找到具有指定特征或包含指定内容的图像^[1]。图像的内容可以分为两类: 视觉内容和信息内容。视觉内容对应图像的物理表示, 如颜色、形状、纹理等。信息内容对应图像的语义, 如主题、人物、场景等。为了检索图像, 首先要描述图像的内容, 对图像的内容进行形式化表示, 即建立图像索引。建立图像索引要用到图像处理、计算机视觉、人工智能、数据库等技术。由于图像内容的复杂性和人类的认知主观性, 建立高效、通用的图像索引是一项很困难的工作。目前, 视觉内容的索引一般可以通过特征提取得到, 语义信息的索引则往往要通过人机交互的方式才能得到。

1 图像索引

索引是一种元数据, 也就是用于说明数据的数据。索引是构造数据解释机制, 实现数据统一理解的关键。建立图像索引的过程就是对图像内容进行形式化描述的过程。需要解决以下几个问题: 首先, 图像数据是非结构化的, 为了建立颜色、形状、纹理等视觉内容的索引, 需要构造相应的数据模型。随着图像处理和计算机视觉技术的发展, 这个问题已经基本得到了解决。针对图像中的各种视觉内容, 已经建立了很多清晰实用的数据模型, 如基于颜色的直方图、色矩、色集; 基于形状的傅里叶描述符、代数不变矩; 基于纹理的共生矩阵、小波变换等。其次, 图像的内容往往受到主观因素的影响, 与人的经验、知识、心理都有关系, 因此很难准确全面地描述图像的内容。目前的各个图像检索系统中缺乏统一的描述方案, 限制了数据的共享和交换, 在互联网环境下, 这个问

收稿日期: 2002-08-210

基金项目: ISN 国家重点实验室资助; 北京大学视觉和听觉信息处理国家实验室资助

作者简介: 石 军 (1975), 男, 西安电子科技大学博士研究生。

题更加突出。为了解决这个问题, MPEG专家组正在制定一个多媒体内容描述方案——MPEG-7 标准。MPEG-7 将对多媒体的内容进行标准化的描述, 并将该描述与所描述的内容相联系, 以实现快速有效的检索。再则, 尽管图像中包含了丰富的信息内容, 但在物理层次上, 图像数据只是二维像素阵列。为了自动建立信息内容的索引, 需要实现从物理内容到信息内容的映射, 完成这一过程要综合图像处理、图像理解、人工智能等技术, 目前还处于起步阶段。建立信息内容的索引一般要通过人机交互的方式。

根据建立索引的方式和索引的内容不同, 可以把图像索引分为 3 类: 基于注释的索引、基于特征的索引和基于知识的索引^[2]。基于注释的索引也称为高级索引, 它是对图像的信息内容建立的定性索引。基于特征的索引又称为低级索引, 它是对图像的视觉内容建立的量化索引。基于注释的索引一般是手工方式生成的, 基于特征的索引则可以在图像模型的引导下通过特征提取的方式自动生成。基于知识的索引是针对某个应用领域建立的索引, 首先建立该领域知识的逻辑模型, 然后对图像进行特征提取和分析, 将结果与模型匹配。一旦确定了匹配关系, 就可以把模型所包含的语义信息赋给相应的图像。这种索引方式实质上是以领域知识为中介, 实现视觉内容到信息内容的映射, 提高系统处理信息内容的能力。

2 基于注释的图像检索

基于注释图像检索是在注释索引的基础上实现的。图像注释就是描述图像内容的文本信息, 分为语句注释和关键字注释。语句注释可以详细描述图像的内容, 但随意性较大, 相同的图像内容在不同情况下可能会得到相差很多的注释结果。关键字注释就是用一组预先定义的关键字描述图像的内容, 比较规范, 但不够灵活, 难以充分表达图像的内容。由于自然语言理解技术的限制, 在图像检索系统中多使用关键字注释。

基于注释的图像检索早在 20 世纪 70 年代就出现了, 最初是在关系数据库中加入描述图像内容的字段, 并在图像的存储路径和这些字段之间建立联系, 然后利用数据库的查询功能实现图像检索^[3]。

早期的检索系统都使用自行定义的关键字和注释结构, 缺少描述图像的统一方案, 数据的共享程度低。随着互联网和多媒体的发展, 检索图像的环境发生了很大的变化, 数据的共享和交换成为一个迫切的要求。为了在互联网环境下用一致的方式检索不同的数据源, 需要创建一种简单通用的元数据模型。1995 年 3 月在都柏林召开的第一届元数据研讨会上产生了一个简单的元数据集——都柏林核心元数据集 (Dublin Core Element Set), 简称都柏林核心 (Dublin Core)。Dublin Core 具有简洁、易于理解、可扩展的特性。它的内核是一个很小的应用集合, 并规定了可供选择的数据内容和数据格式。经过几年来的发展, Dublin Core 从内容到形式都大大丰富了。不仅能描述网络资源, 而且可以很好地应用在大多数的电子资源描述中。此外, 万维网协会 W3C 还制定了资源描述框架 RDF。RDF 采用可扩展标识语言 XML 作为处理和交换元数据的通用语法结构体系, 是一个能够对结构化的元数据进行编码、交换和再利用的体系框架, 使不同的用户能够在这一框架下定义它们自己的元数据。这些元数据模型为在互联网环境下对图像进行一致的描述提供了可能^[4]。

基于注释的图像检索可以充分利用图像中丰富的信息内容, 与人类认知图像的方式相一致, 因而检索的准确率比较高。但是, 由于目前计算机技术的局限, 完全自动的图像注释无法实现, 因此目前的图像注释多采用手工完成, 存在着以下不足: 注释图像的工作量太大; 由于人的认知主观性, 造成了注释结果的模糊性。

到了 20 世纪 90 年代, 由于图像数据的迅速增加, 这两个问题越发突出。为此, 出现了基于特征的图像检索。由于系统能够根据图像的视觉内容自动建立特征索引, 因此基于特征的图像检索具有较高的速度, 适用于海量的图像集合。

3 基于特征的图像检索

基于特征的图像检索是在特征索引的基础上实现的。图像中包含信息内容和视觉内容, 虽然目前图像处理技术不能自动识别信息内容, 但是可以自动识别视觉内容。根据图像的视觉内容建立特征索引, 基于这些特征索引即可实现基于特征的图像检索。与基于注释的检索不同, 基于特征的检索可以自动进行, 减少了手工操作的工作量, 当图像数量很大时, 这一点尤为重要。可用于特征索引的视觉内容包括颜色、形状、纹理等^[4]。

311 基于颜色特征的检索

在图像检索中最常用的物理特征是颜色,这是因为颜色比较直观,并且对图像的位置和大小的变化具有较好的鲁棒性。

最常见的颜色表达式是直方图^[13],直方图反映了各种颜色在图像中的分布情况。图像的直方图是一个一维的离散函数: $H(k) = n_k/n, k = 0, 1, \dots, L-1$, 其中 L 是颜色空间的量化数目, n_k 是具有第 k 种颜色的像素个数, n 是像素总数。为了度量直方图的相似性,可以采用直方图相交法^[5],令 $H_q(k)$ 和 $H_d(k)$ 分别为查询图像 Q 和范例图像 D 的直方图,则它们之间的相似度为

$$P(Q, D) = \frac{\sum_{k=0}^{L-1} \min[H_q(k), H_d(k)]}{\sum_{k=0}^{L-1} H_q(k)}$$

当某个颜色分量值没有在图像中出现时,直方图中会产生零值,用欧氏距离度量直方图的相似度时,对直方图中的每个颜色分量平等对待,没有考虑颜色间的相似性,不符合人类的视觉感觉,这种情况会对直方图交叉法的精确度带来影响^[7]。为了解决这个问题,文献[8]提出了累计直方图。累计直方图也是一个一维的

离散函数: $I(k) = \sum_{i=0}^k n_i/n, k = 0, 1, \dots, L-1$, 式中参数的含义与直方图表达式中相同。累计直方图减少了直方图中的零值,使颜色分量之间的距离与它们之间的相似程度成正比。

除了直方图之外,还有其他形式的颜色表达式。文献[6]在分析已有的基于颜色的图像检索方法的基础上,提出了一种新的基于颜色的图像检索算法,该方法对每个图像单独进行颜色量化,计算其颜色直方图并排序,根据颜色在图像中出现的频率和图像中各种颜色的对比强度确定图像之间的相似度。

312 基于纹理特征的检索

纹理是指图像中所具有的局部不规则而宏观有规律的特性。以往纹理多用于模式识别和计算机视觉,近年来在图像检索中也得到了广泛的应用。纹理对图像灰度变化的特征进行量化,与对象的位置、走向、大小、形状有关,与平均灰度级无关。纹理分为随机纹理和模式纹理。随机纹理用统计性质表征,如灰度级的标准偏差或自相关宽度;而模式纹理可通过抽取某些度量进行进一步表征。

纹理特征的一个有效表达方法是共生矩阵^[7],共生矩阵可以表示出两个特定灰度的像素在相距 (x, y) 时同时出现的联合概率分布。若图像的灰度级为 N ,则共生矩阵为 N 阶方阵,可以表示为 $M(x, y)(h, k)$,其中位于 (h, k) 的元素 $M(h, k)$ 表示一个灰度为 h 的像素和另一个灰度为 k 的像素相距为 (x, y) 时出现的次数。共生矩阵的几个统计量可以作为纹理特征的度量,包括纹理二阶矩 W_M 、熵 W_E 、对比度 W_C 和均匀性 W_H :

$$W_M = \sum_h \sum_k M^2(h, k); \quad W_E = - \sum_h \sum_k M^2(h, k) \log M(h, k);$$

$$W_C = \sum_h \sum_k |h - k| M(h, k); \quad W_H = \sum_h \sum_k M(h, k) / (\sum_h \sum_k |h - k|);$$

其中 W_M 对应图像的均匀性或平滑性,当所有的 $M(h, k)$ 相等时, W_M 最小; W_E 是图像随机性的度量,当 $M(h, k)$ 都相等, W_E 最大; W_C 是共生矩阵各元素灰度值差的一阶矩,当矩阵中大的元素接近矩阵对角线时, W_C 比较大; W_H 可看作是 W_C 的倒数,的作用是避免分母为零。通过这些统计量或它们的线性组合就可以度量不同图像间的纹理相似度^[7]。

20 世纪 90 年代以后,小波变换开始应用于纹理表示中,文献[9]将小波变换和共生矩阵相结合以同时实现基于统计和基于变换的纹理分析。

313 基于形状特征的检索

在基于形状的图像检索中有 3 个问题需要解决。首先,形状通常与图像中的特定目标对象有关,因此形状比颜色和纹理的语义性更强。要获得目标的形状参数,先要进行图像分割,所以形状特性会受到图像分割效果的影响。其次,描述目标的形状是一个非常复杂的问题。人对形状的感觉是视网膜感受和现实世界的知识相结合的结果。事实上,目前还没有找到与人的主观感觉相一致的形状模型。第三,从不同的视角获得的形状可能会有很大差别,为了准确进行形状匹配,需要解决平移、缩放、旋转中的不变性问题^[7]。

形状表示方法有两类 : 基于边界的表示和基于区域的表示 . 这两种表示方法的典型代表分别是傅里叶描述子和不变矩 . 傅里叶描述子的基本思想是用对图像进行傅里叶变换得到的边界作为形状描述 . 其中一个优点就是把二维问题简化为一维问题 . 傅里叶变换的高频分量对应细节而低频分量对应总体形状 , 所以可以只用一些对应低频分量的傅里叶系数来近似描述边界形状 . 文献 [10] 比较了基于区域的方法、基于边界的方法和两种方法的综合使用 , 实验结果表明 , 两种方法综合使用效果更好 .

为了在互联网环境下检索图像 , 文献 [14] 提出了一种自动从互联网上获取图像的方法 , 使用的是关键字和图像的视觉内容 . 文献 [15] 提出了在互联网上查询和浏览图像的要求和方法 .

4 基于知识的图像检索

基于特征的图像检索提高了检索速度 , 但是这些低级图像特征对图像的描述与人类对图像的描述存在较大的差异 , 所以直接利用这些特征作为检索的依据常得不到满意的结果 , 用户也不熟悉特征值的变化对视觉效果带来的影响 . 为了提高图像检索的智能性 , 应当在检索过程中充分利用图像的信息内容 . 图像中包括视觉内容和信息内容 , 视觉内容是信息内容的载体 , 信息内容是视觉内容的内涵 . 为了实现基于内容的检索 , 需要对图像进行解释 , 也就是建立视觉内容和信息内容之间的映射 , 这个过程称为图像解释或图像理解 . 人类在理解世界时用到很多经验和知识 , 因此在图像解释中也可以借用知识提高解释的可靠性和效率 . 对知识的利用是通过建立领域知识的逻辑模型实现的 . 对图像进行特征提取和分析 , 将结果与逻辑模型匹配 , 如果确立了匹配关系 , 就可以把模型所包含的语义信息赋给相应的图像^[7] . 这样 , 既可以实现对信息内容的检索 , 又避免了大量的手工工作 .

4.1.1 知识的种类

在图像检索过程中用到的知识可以分为两类 : 事实性知识和规则性知识 .

事实性知识是描述性的 , 如描述图像语义的关键字 . 基于注释的图像检索事实上也是一个利用事实性知识检索图像的例子 . 人工定义的描述图像内容的文本就是一种应用领域中知识的逻辑模型 , 模型中的内容是无法通过图像处理从图像中得到的 . 在这种情况下 , 图像检索是通过关键字匹配实现的 . 这种方法的优点是容易实现并且精确度比较高 , 但是不适用于大型图像库 . 地理信息系统是一种利用事实性知识检索图像的例子 , 它用类似 E2R 模型的方法表示地图 , 地图中所标注的城市或街道的名字就是与地图相关的事实性知识 .

另一种利用知识检索图像的方法是建立一个规则库或知识库^[21] . 这一方法在文本数据库中很早就出现了 . 与事实性知识相比 , 规则性知识不仅用于内容匹配 , 还可用于语义特征的提取、相似度计算、隐含信息的挖掘等方面 . 规则性知识是推理性的 , 能够表示各种复杂的图像 . 在文献 [12] 中 , 知识的逻辑模型包括从图像中提取语义特征的方法 , 把检索要求转化为与信息内容相关的计算的规则 , 把查询条件转化为与提取的图像特征具有相同表示方法的规则 . 还定义了把查询条件同描述图像信息内容的知识相关联的方法 . 在文献 [16] 中 , 用描述性的关键字声明图像的特征 , 如颜色、位置、区域等 , 用状态迁移模型表示颜色区域和由这些颜色区域所表示的语义内容之间的关系 . 文献 [18] 是使用基于知识的方法检索医学图像的例子 , 利用知识推理得到对象的形状和位置 .

4.1.2 知识的表示方法

为了有效地利用知识 , 对知识的表达有一定的要求 . 知识的表达是数据结构和解释过程的有机结合 . 一个用于表达某一领域中结构复杂知识的良好方法应当具有表达充分、有效推理和有效地获取知识的特点 . 对图像知识进行形式化表示的常用方法包括语义网络、谓词逻辑、框架等^[17] .

语义网络是编了号的有向图 . 在人工智能中应用广泛 . 语义网络给出了图像中各元素间联系的直观表示 , 可以在视觉上有效地表示知识 . 语义网络的一个重要特性是可以用一小组基元构造大量复杂的关系 . 文献 [18] 用 E2R 模型表示和识别图像中对象间的相互关系 . 在文献 [19] 中 , 用一种称为类型抽象层的语义网络表示图像的语义信息 . 类型抽象层能够把专家知识和对象及其语义结合起来 , 从而提高了图像检索的效率 .

谓词逻辑是一种对知识进行形式化表示的有力工具 , 能够表达命题以及借助事实的规则库推出新的事实 . 在图像检索中 , 谓词逻辑多用于表示图像对象间的空间关系 . 在文献 [20] 中 , 将对象投影到直角坐标系

上,并用二维字符串表示.这些二维字符串构成了二维的偏序关系,从而表示出了对象间的空间关系.图像间的相似度计算可以通过二维序列的匹配实现.在文献[21]中,用地理信息系统GIS中的一些空间关系谓词表示图像中对象的空间关系,如远、近、上、下、内、外等.这些不精确的谓词可以用模糊函数处理.

框架由一个框架名和若干个槽和侧面组成,适合于表示结构性的知识,能够将图像中对象间的相互关系表示出来.文献[22]用一种称为媒介抽象的框架结构表示图像知识.

5 结 论

随着多媒体技术和互联网的迅速发展,信息的发布方式由单一的文本方式逐步变为以图形、图像、动画、视频等多媒体信息为主的表现方式.整个互联网就像个大型的分布数据库,而这个数据库是无人管理的.如果没有有效的图像检索技术,大量的信息将淹没在信息的海洋中.在这样的驱动下,图像检索的用户、目的、使用方式、交互方式、应用领域、实现方法等都将不断发展变化.根据不同的索引技术对图像检索技术进行分析和比较,它们的性能如表1所示.

表 1 不同图像检索方法的性能比较

	基于注释的图像检索	基于特征的图像检索	基于知识的图像检索
通用性	高	高	低
效 率	低	高	高
智能性	高	低	高
可实现性	好	较好	一般

图像检索是多个领域的研究成果相结合的产物,包括计算机视觉、数据库、人工智能等.当前的图像检索受到图像处理和人工智能技术的限制.为了实现自动化、智能化、通用的图像检索,有3个问题需要研究:图像知识的表示方法、图像检索算法、图像的高层语义模型.就目前而言,将几种方法结合起来使用会得到较好的检索效果.现有的图像检索技术大都是数据源已知时的静态图像检索.在互联网上,数据源是未知的,这种情况下如何实现高效、可靠的图像检索也是一个需要研究的问题.此外,图像检索也是视频检索的基础,实现高效的图像检索对实现视频检索有着重要的意义.

参考文献 :

- [1] 高文,刘峰,黄铁军,等.数字图书馆——原理与技术实现[M].北京:清华大学出版社,2000.
- [2] 周洞汝,胡宗斌.视频数据库管理系统导论[M].北京:科学出版社,2000.
- [3] Chang S K, Yan C W, Dimitroff D C, et al. An Intelligent Image Database System[J]. IEEE Trans on Software Eng, 1988, 14(5): 412-421.
- [4] Rui Y, Huang J S. Image Retrieval: Current Techniques, Promising Directions, and Open Issues[J]. Visual Communication and Image Representation, 1999, 19(1): 39-62.
- [5] Swain M, Ballard D. Color Indexing[J]. International of Computer Vision, 1991, 7(1): 11-32.
- [6] 王宇生,陈纯.一种新的基于色彩的图像检索算法[J].计算机研究与发展,2002,39(1): 105-109.
- [7] 章毓晋.图像理解与计算机视觉[M].北京:清华大学出版社,2000.
- [8] Stricker M, Orengo M. Similarity of Color Images[J]. Storage and Retrieval for Image and Video Database, 1995, 2(4): 381-392.
- [9] Kundu A, Chen J L. Texture Classification Using QMF Bank-based Subband Decomposition[J]. CVGIP: Graphical Models and Image Processing, 1992, 54(5): 369-384.
- [10] Mehre B M, Anhalii M K, Lee W F. Shape Measures for Content Based Image Retrieval: a Comparison[J]. Information Processing and Management, 1997, 33(3): 319-337.
- [11] Yoshitaka A, Ichikawa J. A Survey on Content-based Retrieval for Multimedia Database[J]. IEEE Trans on Knowledge and Data Engineering, 1999, 11(1): 81-93.
- [12] Yoshitaka A, Kishida S, Hirakawa M, et al. Knowledge Assisted Content-based Retrieval for Multimedia Database[J]. IEEE Multimedia, 1994, 1(4): 1-21.
- [13] 田玉敏,林高全.基于颜色特征的图像检索算法[J].西安电子科技大学学报,2002,21(1): 43-46.

- [14] Yanai K, Shindo M, Noshita K. A Fast Imagegathering System on the WorldWide Web Using a PC Cluster, Web Intelligence: Research and Development First Asia2Pacific Conference[A]. WI 2001[C]. Heidelberg: Springer, 2001. 3242334.
- [15] Celentano A, Gaggi O. Querying and Browsing Multimedia Presentations, Multimedia Databases and Image Communication Second International Workshop[A]. MDIC 2001[C]. Heidelberg: Springer, 2001. 1052116.
- [16] Ono A, Amano M, Hakaridnai M, et al. A Flexible Content2based Image Retrieval System with Combined SceneDescription Keyword [A]. Proc Intl Conf Multimedia Computing and Systems[C]. Heidelberg: Springer, 1996. 20122081.
- [17] Khatib W A, Day Y F. SemanticModeling and Knowledge Representation in Multimedia Database[J]. IEEE Trans on Knowledge and Data Engineering, 1999, 11(1) : 64280.
- [18] Hsu C C, Chu W W, Jaiara R K. A Knowledge2based Approach for Retrieving Images by Content[J]. IEEE Trans on Knowledge and Data Engineering, 1996, 8(4) : 5222532.
- [19] Chu W W, Chsu C, Cardnas A F, et al. Knowledge2based Image Retrieval with Spatial and Temporal Constructs[J]. IEEE Trans on Knowledge and Data Engineering, 1998, 10(6) : 8122888.
- [20] Chang S K, Shi Q T, Tan C W. Inonic Indexing by 2D Strings[J]. IEEE Trans on Pattern Analysis and Machine Intlligence, 1987, 9(3) : 4132427.
- [21] Abella A, Kender J R. Qualitatively Describing Objects Using Spatial Prepositions[A]. Proc 11th Natl Conf Artificial Intelligence[C]. Heidelberg: Springer, 1993. 5362540.
- [22] Brink A, Marcus S, Manian V S. HeterogeneousMultimedia Reasoning[J]. Computer, 1995, 28(9) : 33239.

(编辑 : 李维东)

法国 Thomson 公司访问我校

2003年2月24日下午,法国 Thomson 公司专家一行 3人前来我校参观考察。

Thomson 公司专家参观了“综合业务网国家重点实验室”和“雷达信号处理国家重点实验室”的展室和部分课题组,听取了两个重点实验室的情况,并同这两个实验室的部分专家进行了深入交流。

Thomson 公司拟在西安市设立该公司的研发中心,此行的目的就是考察了解西安的科技发展现状,以便有针对性地制定其工作规划。通过此次参观考察,Thomson 公司的专家对我校留下了深刻、良好的印象,并且表示了初步与我校合作的意向。

摘自《西电科大报》200313115