

基于内容的图像检索研究进展 *

Research on Development of Content - based Image Retrieval

肖 明 王永红 石 勇

(北京师范大学管理学院信息技术与信息管理学系 北京 100875)

摘 要 概述了基于内容的图像检索(CBIR)的含义和研究进展,阐述了 CBIR 相关技术,介绍了当前流行的几种 CBIR 系统,CBIR 的当前应用以及存在的主要问题。

关键词 基于内容的检索 图像检索 CBIR

1 CBIR 概述

图像检索的传统方法是基于文本的。使用关键字注释是最常用的方法,对图像的检索变成了对关键字的查找。这种方法简单易行,能够从用户角度表达图像内容的高层语义。但是基于文本检索存在着两大困难,尤其是当图像的数量非常大的时候,采用基于文本描述的检索方法具有如下缺点^[1]: a. 文本描述难以充分表达图像的丰富内容。俗话说“一图胜千言”,而文本描述是一种定性的描述,描述能力有限,图像中则往往含有大量需要定量描述的信息。而且许多图像的特征难以用文本描述表达,如图像中的不规则形状、散布的纹理等就很难用文本来描述。文本描述具有一定的主观性。由于图像内容的丰富性以及不同人理解和感兴趣方面的不同,导致内容描述的建立具有很大的主观性(如不同的人对颜色的感知就不尽相同),这样采用这种检索方法就会带来一定的歧义性。b. 文本描述难以实现基于图像视觉特征的相似性检索。采用文本描述的检索方法,本质就在于计算检索请求与媒体文本描述之间的相似度,这就涉及到目前尚未解决的自然语言理解问题,尽管目前实现的系统中主要通过采用同义词词典(Thesaurus)来使问题得到简化,但同时也使检索的表达力受到了较大限制。手工输入文本描述效率低,难以满足大容量媒体库的要求。由于数据规模不断膨胀,人工标注的开销就显得太大。文本描述一般需要手工输入,效率较低,所以不能满足海量环境下多媒体信息库的检索要求。

为了克服基于文本方法的局限性,20 世纪 90 年代出现了基于内容的图像检索(Content - based Image Retrieval,简称 CBIR)。CBIR 主要是指将图像的可视特征(如颜色、纹理、形状、位置关系等)作为图像的内容进行匹配、查找。利用已有的算法,进行特征提取和匹配,这些完全可以由机器自动完成,克服了手工注释的低效和二义性。总之,基于内容的图像检索技术融合了图像理解技术,从而可以提供更加有效的检索途径。

基于内容的图像检索的发展过程可以用表 1 来表示^[2]。20 世纪 70 年代主要研究图像文件的管理;20 世纪 70 年代到 80 年代是采用关键词等描述方法为主,同时出现了面向特定应用领域的小规模的图像数据库系统;20 世纪 90 年代以后,人们转向研究面向网络环境的支持基于内容检索的大规模图像数据库系统;

2000 年以后,则进入到标准化发展阶段。

表 1 CBIR 的研究阶段

时间范围	研究方向	主要特点
20 世纪 70 年代	基于文件系统的图像管理	直接对图像文件进行访问,操作简单;对于大型的图像库,需逐一打开图像文件进行浏览,效率较低
	基于文本方式的图像检索	通过对图像标注关键词来建立图像的索引;一般用传统数据库管理图像的索引,图像数据采用外部文件进行存储,按间接方式访问图像;主要以数据库专家的研究为主;其主要缺陷是基于特征的相似检索会导致检索结果的不确定性,整个检索过程是一个逐步逼近和相关反馈的过程
20 世纪 70 年代到 20 世纪 80 年代	基于知识和视觉特征的图像检索	通过对图像或图像对象加注知识或特定领域的视觉特征来建立图像的索引,这个时期的图像数据库主要面向特定的应用领域,并取得了许多成果,如指纹、人脸识别系统;图像数据用外部文件存储,以间接方式对图像进行访问;主要以人工智能和模式识别的研究为主,以图像检索的精确特征匹配为主
	Web 环境下基于内容的图像检索	主要以面向 Web 环境下的图像信息检索,通过对图像所包含的内容进行自动或者半自动的特征提取来建立图像索引;对图像内容的统一特征存在着多种表达方法;图像检索主要面向更加通用的领域,图像查找是一种相似性查询,检索过程是一个逐步逼近和相关反馈的过程;对图像文件的存储既可以采用外部文件方式,也可以采用内部二进制对象方式;主要以模式识别领域为主,并得到其他领域研究人员的广泛关注
20 世纪 90 年代	Web 环境下基于内容的图像检索	图像内容描述将遵循统一标准;基于内容的图像检索技术将朝商业化方向迈进
2000 年以后	标准化阶段	

2 CBIR 的检索类型

图像检索是对静止图像所进行的检索,由于图像存在着颜色、形状和纹理等方面的特征,所以基于内容的图像检索也可以细分为颜色检索、形状检索和纹理检索等多种类型。

2.1 颜色检索 颜色(color)具有一定的稳定性。在许多情况下,颜色是描述一幅图像最简便而有效的特征。例如,在检索海滨景物图像时,指定图像中的主要颜色(如对应天空的蓝色和白云对应的白色)的大致比例后,即可以依据检索与该颜色分布类似的图像。

对颜色进行检索主要从图像颜色分布、相互关系、组成出发,主要是利用颜色空间直方图进行匹配。常见的颜色坐标空间有

基金项目:国家社会科学基金资助项目(编号:02CTQ002)。

作者简介:肖 明,男,1969 年生,副教授,研究方向为信息资源管理;王永红,女,1976 年生,工程师,研究方向为计算机网络;石 勇,男,1983 年生,硕士研究生,研究方向为情报学、信息资源管理。

(R、G、B)(红、绿、蓝)、(H、S、V)(色调 Hue、饱和度 Saturation、亮度 Value)等,它反映出图像中的颜色和出现这种颜色的概率之间的关系。根据颜色数据进行查询时,数据库中的图像和被查询图像之间的距离可用加权欧几里德距离表示,采用基于颜色分布的匹配将获得视觉效果上更接近被检索实体的查询结果。常用的方法有互补颜色空间直方图、直方图交叉法、直方图距离比较法、二次距离算法等。

2.2 形状检索 形状(Shape)是刻画物体的本质特征之一。基于形状的检索从图像的轮廓组成、形状、大小出发,它既包括传统意义的基于二维形状的检索,也包括在三维图像中的基于三维形状的检索,可以针对面积(可用像素点的个数计算)、环形性(即周长 \times 周长/面积,周长也用像素点的个数表示)、主轴方向、偏心率、圆形率、连通性、正切角等形状特征进行匹配。形状检索方法主要有针对图像边缘轮廓线进行的检索和针对矢量特征进行的检索两种方法。

2.3 纹理检索 同颜色、形状一样,纹理(Texture)也是刻画物体的重要特征。虽然图像的纹理特征在局部区域内可能没有规则,但在整体上却往往呈现出一定的规律性,这也正是基于内容检索的一条主要线索。纹理特征主要由纹理的均匀度、对比度和方向的特征量表示。均匀度反映纹理的尺寸,对比度反映纹理的清晰度,方向反映实体是否有规则的方向性。

2.4 草图检索 草图(Sketch)特征是由原始图像实体的边界映射而成。根据草图进行查询时,通常用手画出图像的大致边界,然后利用相似性原则查询出数据库中具有相似边界的图像。

2.5 对象检索 对象检索主要从图像中子对象的关系、数量、属性、旋转等特征出发进行检索。

3 CBIR 的索引技术

美国匹兹堡大学的张系国教授在研究图像信息系统时指出,对于图像数据其索引要从三个方面(包括索引的表示、索引的组织)进行研究,并用一个三维坐标来表示(如图 1 所示)^[2]。

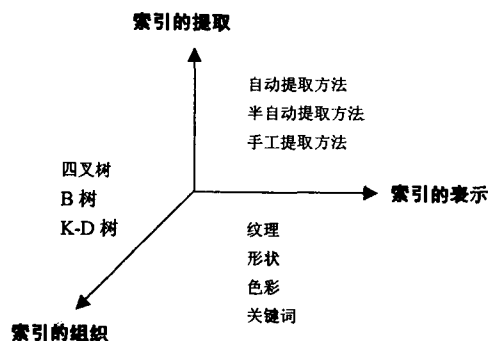


图 1 图像索引的三维表示

图像索引的提取可以分为手工提取、半自动提取和自动提取,这在很大程度上依赖于图像处理技术的发展,对于不同情况应该区别对待。图像索引可以细分为关键词索引、色彩特征索引、形状特征索引和纹理特征索引等。由于图像索引的表示一般维数都比较高,有的高达上百维,传统数据库采用的数据结构(如哈希表、B 树等)不能很好地组织这些高维的特征索引。因此,研究出一种高效的数据结构就显得迫切和必要。

在基于内容的图像检索中,图像索引表示主要由图像的视觉

和形象特征来表示,从色彩、纹理以及对象形状和空间关系这几方面研究。目前的主要技术包括:基于色彩特征的索引技术、基于形状特征的索引技术、基于纹理特征的索引技术和基于空间关系的索引技术。基于内容检索的图像索引技术可以用分类树来进行表示(如图 2 所示)^[2]。

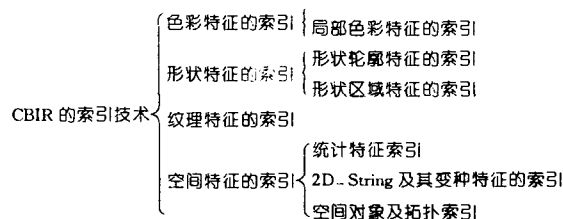


图 2 CBIR 的索引技术

4 CBIR 典型系统

目前,世界各国的相关研究人员沿着基于内容的图像检索这个研究方向成功地开发出许多带有商业性质或者研究性质的 CBIR 系统。

4.1 商业性 CBIR 系统 QBIC、VIR 以及 RetrievalWare 等商用原型系统的出现,意味着基于内容的图像分析与检索技术正开始逐步地从实验室的研究阶段走向实用阶段。

IBM 的 QBIC^[3]是第一个走向商用化的图像检索系统。目前 QBIC 的技术已经应用到 IBM 的商用产品中,如 DB2 数据库"Image Extender"模块。QBIC 的系统框架和技术对后来的图像检索系统产生了深远的影响。QBIC 支持基于图例、视觉特征以及关键词等查询方式,实现了用颜色直方图、颜色布局、纹理和特殊混合色四种特征表示方式。颜色直方图使用了 256 种颜色,它们是通过 MTM 转换得到的。颜色布局则采用对图像进行 12×12 固定为 256 预定义的色之一。纹理特征则采用 Tamura 纹理表示。此外,QBIC 还考虑了高维特征索引的问题。

VIR Image Emage^[4]是由 Virage 公司开发的图像搜索引擎,其特征分为通用特征和领域相关的特征。通用特征包括全局颜色、颜色布局、纹理和结构以及它们的任意组合。VIR 还提供了一个图像管理的开放框架,开发者可以往框架中添加与领域相关的特征处理模块。VIR 还被扩展到视频管理。在 Oracle 和 Informix 的数据库产品中应用了 VIR 的技术,AlataVista 图像搜索引擎也使用了 VIR 的技术。

RetrievalWare^[5]是由 Excalibur 技术公司开发的,它使用了颜色、形状、纹理、亮度、颜色布局以及图像的方式等作为查询特征。它同时还支持将这些查询特征组合起来,并可以由用户来指定各自的权重。RetrievalWare 的技术已经部分应用到 Yahoo 的 Image Surfer 图像搜索引擎中。

4.2 研究性 CBIR 系统 Netra^[6]是 USCB 的亚历山大数字图书馆工程研制的一个原型系统。目前,Netra 正在 UCSB Alexandria Digital Library(ADL)项目中得到进一步完善,它完全基于 JAVA 语言环境,利用分割后的图像区域中的颜色、纹理、形状和空间关系从数据库中搜索和检索相似的区域。Netra 系统的主要研究特征基于 Gabor 过滤器的纹理分析、基于神经网络的图像词典构造和基于边缘流的区域分割。

Photobook^[7]是由 MIT 媒体实验室研究开发的 CBIR 试验系统,它提供了一系列浏览和搜索图像的工具。最初的 Photobook 包含了 FourEyes,它是一个计算机辅助的人工图像标注和分组工

具。Minka 和 Picard 等提出了“模型族”(society of model)方法,将用户的反馈融入到交互式学习方法中。试验表明,这种方法在交互式图像标注中是非常有效的。

MARS^[7]多媒体分析和检索系统是由 UIUC 的 Y. Rui 等人研究开发的。MARS 的一个最大特色是提供了功能强大的联机相关反馈机制,并提出了一个相关反馈框架结构。MARS 所关注的并不是寻找一个“最佳”的特征表示方式,而是考虑如何将各种视觉特征组织起来并使之动态地适应不同的应用和不同的用户。

VisualSEEK^[7]是哥伦比亚大学研究开发的一个图像检索系统。在 VisualSEEK 中,通过颜色集的背投(back-projection)过程可以自动将图像中显著的颜色区域抽取出来,从而实现了基于颜色区域的视觉特征和空间布局的查询方式。VisualSEEK 的研究人员还研究开发出一个支持动作查询(motion query)的视觉搜索引擎 VideoQ,以及基于 Web 的图像元搜索引擎 MataSEEK。

Blobworld^[8]是加州大学 Berkley 分校数字图书馆项目中的一部分。Blobworld 的不同之处在于支持“同构区域”(即 BlobWorld 中的 blob)的分割、表示和查询。Blobworld 在多特征抽取的基础上应用 EM 和 MDL 来获得 blob。实际查询时,用户可以指定只针对图像中的某个 blob 发出查询。Blobworld 所收集的图像库中图像的数目超过了 1 万多幅。

4.3 其他系统 除了上述几种著名的 CBIR 系统以外,国外较著名的 CBIR 系统还包括:日本电子技术综合研究所开发的 ART-MUSEUM、普林斯顿大学研制的 CAETIML、Massachusetts 大学 CIIR 的图像检索系统、荷兰 Leiden 大学的肖像数据库及英国科特大学研制的多媒体检索系统(由虚拟数据库和查询主机构成)等。

国内享有一定知名度的 CBIR 系统主要有中国科学院计算技术研究所研制的 MIREs 以及清华大学计算机系研制的 ImgRetr 和 VideoBase。

基于特征的多媒体信息检索系统 MIREs 由中国科学院计算技术研究所和国家图书馆联合研制,该系统可以按照图像的颜色、纹理、形状等特征进行检索,对中文信息能够进行全文检索,具有布尔检索、截词模糊匹配检索和完全字符串等多种检索技术。

ImgRetr 是一个基于 Java 的图像检索系统,它可对静态图像的颜色(直方图、分布)、形状、纹理等特征进行直接检索,其用户界面采用 100% 的 Java,可以直接通过 Web 进行访问,应用 Internet 之上。

VideoBase 是正在研究的一个智能化的视频信息管理系统,它可以对 MPEG-1 和 MPEG-2 视频档案进行镜头分割、关键帧提取和形成视频图标,为用户浏览查找视频提供方便。利用它,甚至可以将一段视频资料进行结构化组织,形成 HSTG 图(分层场景转换图),从而能够让用户直观看到各镜头之间的组织关系等。

5 CBIR 总结与展望

基于内容的图像检索已经深入到社会和生活的许多方面。目前,CBIR 的主要应用领域包括^[7]:电子会议、远程教学、艺术收藏、博物馆管理和电子图书馆、建筑工程设计、遥感和地球资源管理、地球信息系统、科学数据库管理、天气预报、商业零售、织物时装设计、商标和版权数据库管理、公共安全和犯罪调查、图像档案和通讯系统、远程医疗、智能群体决策、防汛指挥。

针对数据库系统研究与开发中的三个范畴的不确定性,即数

据模型的不精确、检索方法(用户目标、交互式工具等)以及数据库的内容的含糊/不确定性的根本问题,还需要重点解决以下几个问题^[9]。

5.1 从自动分析到人机交互 基于计算机视觉的模式识别系统和图像检索系统的根本区别在于后者中的人的参与是不可缺少的。早期人们强调全自动意义上的单一最佳特征选择,现在更强调交互式系统。QBIC 用交互式进行区域分割;MIT 在从 Photobook 向 FourEyes 发展中也引入了交互性;WebSEEK 允许在用户反馈下动态计算特征向量;UCSB 在纹理分析中结合了有监督的分析方法;MARS 在图像检索中提出了适当反馈结构,利用人机交互提高检索性能;VideoQ 及 Informedia 在很多处理上引入了人的辅助。

5.2 从低级视觉特征到高级概念 人们在实际应用中倾向于用高级的概念,而不是低级的视觉特征来检索图像,如新闻、摄影、绘画的检索,但目前的计算机视觉特征仅支持这些低级特征。从低级特征到高级概念并没有直接的联系,所以需要更加智能化的处理,必要时可以借助人机交互方式来完成。

5.3 高效存储及检索技术 网络扩展的一个副产品是多媒体,特别是图像和视频数据的海量膨胀。为了保证检索速度和效率,需要不断地研究出各种高性能的数据存储和检索方法,包括高维索引技术、面向一般应用的存储和检索模型等。

5.4 从单机、局域网到因特网 随着计算机和网络技术的发展,信息存储和检索已经从单机、局域网向因特网发展。因特网的成功是令人震惊的,其发展速度及其对于人类生活、工作观念和方式的影响是革命性的。每天都会有大量的文件、图像和视频等加入到网络中,所以基于网络的多媒体信息检索已经变得十分迫切。常用的 Infoseek 等搜索器都是基于文字等高级属性去查询,对于图像等特殊信息的查询则需要利用更为专业的搜索器才行。

5.5 性能评价及测试标准 任何技术都是由其相应领域的性能评价的准则而向前推动的。由于图像和视频内容不但丰富,而且具有很大程度上的主观性,所以其检索性能的评价标准制定起来具有相当大的难度。另一方面,对于图像检索而言,其标准测试集的研究和制定也非常重要。通过精心选择有代表性的测试集,在一定意义上可以解决性能评价中的主观性。

总之,基于内容的图像分析与检索技术还有很漫长的探索道路,还存在许多值得深入研究的课题。正如 John Eakins 博士所言:基于内容检索的图像分析与检索技术的研究不会是昙花一现,它不像某些过份热情洋溢的学者所期望的那么有效,它也不像某些批评者所想的那么糟糕,它会继续得到发展和改进。

参考文献

- 1 白雪生. 基于内容检索及其相关技术的研究. 清华大学博士论文, 1998
- 2 李向阳. 基于内容的图像数据库检索技术及其模型的研究. 浙江大学博士论文, 2000
- 3 许天兵. 基于内容的图像检索方法研究. 山东工业大学硕士论文, 2000
- 4 何清法. 基于内容的图像分析与检索关键技术的研究. 中科院博士论文, 2001
- 5 <http://vrw.convera.com:8015/cst>
- 6 <http://vision.ece.ucsb.edu/netra/Netra.html>
- 7 梁永全. 基于内容的多媒体信息检索的研究. 中科院博士论文, 1999
- 8 <http://elib.cs.berkeley.edu/photos/blobworld>
- 9 任金昌. 基于内容的图像存储检索技术研究. 西北工大博士论文, 2000

(责编: 柯钧阳)