

基于内容图像检索的关键技术研究

作者：牛一帆、成刚虎

【内容提要】总的来说，随着多媒体和网络技术的发展，图像信息管理问题越来越成为一个具有重大现实意义和挑战性问题。而基于内容的图像检索技术是解决目前图像信息爆炸的有效方法之一。

基于文本的图像检索技术 20 世纪 70 年代末期承印材料，基于文本的图像检索技术（text-based image retrieval）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有词典支持。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。首先媒体，目前的计算机视觉技术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的，还不可避免地带有主观偏差。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。

基于内容的图像检索技术

20 世纪 90 年代初期包装总论，随着大规模数字图像库的出现，上述问题越来越严重。为了克服这些问题，基于内容的图像检索技术(content-based image retrieval, 简称 CBIR)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索引西北人，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。

基于内容的图像检索系统具有与传统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图例来构造查询防伪印刷，然后由系统查找与图例在视觉内容上比较相似的图像，按相似度大小排列返还给用户。这就是所谓的通过图例的检索(query

by image xample)。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交互，以使用户能够方便地构造查询、评估检索结果和改进检索结果。

图 1 是基于内容图像检索系统的体系结构。系统的核心是图像数据库。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印刷配件，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 1 CBIR 系统的体系结构软包装

图像检索技术的组成

1. 图像特征的提取

图像特征的提取是基于内容的图像检索技术。从广义上讲糊盒，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。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地研究，本文主要讨论图像视觉特征的提取。

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，与具体的应用紧密相关区域报道，例如人的面部或指纹特征等。

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,对某个特征并不存在一个所谓最佳的表达方式。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。

2. 图像相似度的度量

(1)视觉特征的相似度模型

基于文本的检索方法采用文本的精确匹配,而基于内容的图像检索则通过计算查询和候选图像之间在视觉特征上的相似度匹配进行。因此机构/组织,定义一个合适的视觉特征相似度衡量方法对检索的效果有很大影响。由于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。常用的度量方式有绝对值距离度量、欧拉距离度量、二次式距离度量、马式距离度量等。

(2)图像检索的性能评价

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其他,系统返回查询结果中相关图像的数目占有所有返回图像数目的比例;查全率则指系统返回查询结果中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。

3.相关反馈技术

系统根据用户提交的查询条件,在图像数据库中找到一些最相似的图像返回给用户。提交的查询将首先转换成由一些特征结合起来的向量,采用线性组合的方法结合起来再分别与数据库中图像的向量计算得到最后的相似度。相似度高的图像成为检索的结果。

虽然特征和特征相结合起来进行检索包装容器,在一定程度上提高了图像检索的效果,不过检索系统的性能还是不太令人满意,其主要原因有以下几个方面:

(1)由于一张图像中存在很多信息,而且不同的用户对同一张图像的认识差异也很大;这就使得即使是相同的查询,不同用户希望得到的结果差异也很大。

(2)由于低层特征并不反映图像真正的语义信息故障分析与排除,因此当用户提交一张图像作为查询时,系统很难找到用户真正想要寻找的图像。

这些问题导致自动的图像检索效果不能令人满意。因此,许多系统都引入了人的交互,通过用户的相关反馈来逐渐提高检索的结果。借鉴文本信息检索的方法,我们在系统中也引入了相关反馈来修改用户提交的查询防伪印刷,使得修改后的查询逐步接近用户真正的需求,来提高系统的检索性能。

通过相关反馈对用户提交的查询的修改,检索的性能比原先有了一定提高。不过,大多数相关反馈并不具有记忆能力,每次反馈后的结果只能提高本次查询结果。因此我们引入了语义网络包装总论,把每次反馈的结果记录到语义网络中,使得系统的效果随着使用次数的增加而逐步提高。

4.索引技术

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,图像特征向量维数的数量级是 102。

(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他的相似性度量方法。

要解决上述这些问题图像处理,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）。常用的两种缩减维数方法是 Karhunen-Loeve 变换法(KLT)和聚类法。尽管经过了维数缩减，图像特征向量的维度仍然较高，因此我们还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，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

树和 R*树等。除了以上几种方法印刷商巡礼，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。

总的来说，随着多媒体和网络技术的发展，图像信息管理问题越来越成为一个具有重大现实意义和挑战性问题。而基于内容的图像检索技术是解决目前图像信息爆炸的有效方法之一，从而上述几项关键技术就显得尤为重要。