

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学号: 19920111152739

UDC \_\_\_\_\_

廈門大學

碩 士 學 位 論 文

基于文本语义和视觉内容的图像检索技术  
研究

Research on Image Retrieval Based on Textual Semantic and  
Visual Content

顾昕

指导教师姓名: 方正 副教授

专 业 名 称: 机械工程

论文提交日期: 2014 年 月

论文答辩时间: 2014 年 月

学位授予日期: 2014 年 月

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2014 年 月

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为( )课题(组)的研究成果，获得( )课题(组)经费或实验室的资助，在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人(签名):

年 月 日

# 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，  
于 年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

## 摘要

随着数字成像、数据存储等技术的飞速发展以及互联网的普及,各种各样的图像正以惊人的速度增长,日益丰富的图像资源使用户难以在浩如烟海的数据库中找到其真正需要的信息,因而有效的图像检索技术成为近年来研究界关注的热点。

现有的图像检索技术主要可分为两种:基于文本信息的图像检索与基于视觉信息的图像检索。前者主要依赖于图像的文本标注信息进行检索,但面对数以万计的图像,手工标注的代价太过昂贵,使得此种检索方案渐已不能满足现实的应用需要;后者主要利用视觉特征提取和高维索引技术进行检索,但因为语义鸿沟的存在,视觉特征相似的图像很可能在语义上是不相关的,使得很多情况下此种检索方案难以满足用户的信息需求。为了充分发挥两种检索方案各自的优势同时降低它们的缺陷,学者们研究了多种方法来结合这两种技术以进行检索,这些工作都显著地提高了图像检索的性能。本文在此基础上就如何融合视觉内容与文本语义进行图像检索做了一些研究工作,主要分为以下几点:

- 1 研究了图像视觉内容的描述方法。针对使用传统的 SIFT 特征来描述图像视觉内容速度较慢、效率较低等缺陷,提出了使用改进的 DSIFT 特征结合词袋模型以构造视觉单词的方式来描述图像的视觉内容。

- 2 研究了一种如何根据图像视觉内容自动获取其语义信息的方法。针对依赖人工标注来获取图像语义信息需耗费较大代价的问题,提出了一种基于概率隐语义分析模型的图像语义自动获取方法,它将图像视为一系列文本信息与视觉信息的潜在主题混合,先通过拟合两个 PLSA 模型来分别学习文本信息和视觉信息,再使用一种自适应不对称学习方式来建立两种信息间的联系,如此即可以根据图像的视觉内容来自动获取其语义信息。

- 3 研究了一种将文本信息与视觉信息结合起来进行图像检索以提高检索精度的方法。提出了分层检索的思想,先依据自动获取的语义信息初步检索筛选出语义相关的图像,再根据提取的视觉特征进行第二层细化检索对图像按相关度由大到小排序输出。

在自主设计的检索原型系统上,针对图像数据集 Corel1K 进行的一系列实验

表明，本文所做的研究工作确实提高了图像检索的性能。

**关键词：**图像检索；视觉语义；信息融合

厦门大学博硕士论文摘要库

## Abstract

With the rapid development of digital imaging, data storage, network and other technologies, all kinds of images are increasing at an alarming rate. The increasingly richness of images resource make it difficult for users to find the information they really need from the vast database, so that efficient image retrieval technology becomes a hot spot of research community in recent years.

Nowadays, image retrieval technologies mainly include text-based image retrieval and content-based image retrieval. The former largely relies on the annotated textual information to retrieval, however, it would cost heavy burden of manual annotation when facing billions of images, so that the retrieval method gradually cannot meet the requirement for reality application; the latter largely relies on the technology of feature extraction and high-dimensional indexing, however, the images with similar feature may have unrelated semantic because of the semantic gap, so that the retrieval method cannot meet the information needs of users in many cases. In order to give full pay to the two technologies' strength and reduce their defects, many researchers devised various methods to combine the two technologies to retrieval which dramatically improve the performance of image retrieval. On these basics, the thesis makes some research works on how to fuse the visual content and textual semantic to retrieval images; it mainly divides into following three parts:

- 1 The thesis studies the method of describing image's visual content. In order to overcome the shortcomings of slow extract speed and so on in traditional SIFT algorithm, the method combines an improved DSIFT feature extraction algorithm with BOW model to describe the image's visual content by the way of constructing visual words.

- 2 The thesis researches a method to automatically acquire semantic information according to the image's visual contents. In order to overcome the shortcoming that semantic acquired by manually annotate would cost expensively, the method proposes an automatically acquire image's semantic algorithm based on PLSA model. The

algorithm regards an image as the fusion of some textual information topics and visual information topics, and it first studies the textual information and visual information by fitting the two PLSA models, then it uses an adaptive asymmetric learning approach to make the association between the two types of information, so that it can automatically acquire semantic information according to the visual content through the association.

3 The paper researches an image retrieval method which combined the textual information with visual information to improve the precision of retrieval. The method proposes the idea of layered retrieval, it first preliminarily searches and selects the semantically related images from database according to the semantic information automatically acquired, then it particularly retrievals at the second level to sort the filtered images according to the similarity of the extracted visual features.

Experimental results in Corel1K database on the self-developed retrieval prototype system demonstrate that the research works of the thesis really improve the performance of image retrieval.

**Key words:** Image Retrieval; Visual Semantic; Information Fusion.

摘要.....	I
Abstract.....	III
<b>第一章 绪论</b> .....	1
1.1 研究背景.....	1
1.2 图像检索技术综述.....	1
1.2.1 基于文本信息的图像检索.....	2
1.2.2 基于视觉信息的图像检索.....	4
1.2.3 两种检索方式的结合.....	7
1.3 论文的研究内容.....	8
1.4 论文的组织结构.....	9
<b>第二章 图像的视觉内容描述</b> .....	11
2.1 图像的视觉内容特征研究.....	11
2.1.1 全局特征研究.....	11
2.1.2 局部特征研究.....	15
2.2 图像的内容特征组织方法.....	26
2.3 本章小结.....	28
<b>第三章 语义分析</b> .....	29
3.1 问题提出背景.....	29
3.2 隐语义分析.....	30
3.3 概率隐语义分析.....	32
3.4 概率隐语义分析的应用.....	34
3.5 本章小结.....	35
<b>第四章 基于文本语义和视觉内容的图像检索算法</b> .....	36
4.1 概述.....	36
4.2 数据建模与学习.....	37
4.3 检索算法描述.....	40
4.4 实验设计.....	41
4.4.1 实验数据集.....	41
4.4.2 实验过程.....	42
4.4.3 实验结果展示.....	43
4.4.4 性能评价方法.....	45
4.4.5 实验结果分析.....	45
4.5 本章小结.....	46

<b>第五章 基于文本语义和视觉内容的图像检索原型系统</b> .....	48
5.1 概述 .....	48
5.2 原型系统的框架结构 .....	48
5.3 原型系统实现 .....	51
5.4 本章小结 .....	53
<b>第六章 总结和展望</b> .....	54
6.1 工作总结 .....	54
6.2 工作展望 .....	55
<b>参 考 文 献</b> .....	57
<b>攻读硕士学位期间的科研成果</b> .....	62
<b>致 谢</b> .....	63

厦门大学博硕士论文摘要库

## Contents

<b>Chapter 1 Introduction</b> .....	1
<b>1.1 Research Background</b> .....	1
<b>1.2 Image Retrieval Technology Review</b> .....	1
1.2.1 Image Retrieval Based on Textual Information .....	2
1.2.2 Image Retrieval Based on Visual Information.....	4
1.2.3 Combination of Two Kinds of Retrieval Technology .....	7
<b>1.3 Research Contents of The Thesis</b> .....	8
<b>1.4 Organization Structure of The Thesis</b> .....	9
<b>Chapter 2 Description of Image’s Visual Content</b> .....	11
<b>2.1 Research on the Visual Content Feature of Image</b> .....	11
2.1.1 Research on the Global Feature .....	11
2.1.2 Research on the Local Feature .....	15
<b>2.2 Organization Analysis of Image’s Content Feature</b> .....	26
<b>2.3 Summary of the Chapter</b> .....	28
<b>Chapter 3 Semantic Analysis</b> .....	29
<b>3.1 Background of Question</b> .....	29
<b>3.2 Latent Semantic Analysis</b> .....	30
<b>3.3 Probability Latent Semantic Analysis</b> .....	32
<b>3.4 Application of Probability Latent Semantic Analysis</b> .....	34
<b>3.5 Summary of the Chapter</b> .....	35
<b>Chapter 4 Image Retrieval Algorithm Based on Textual Semantic and Visual Content</b> .....	36
<b>4.1 Overview</b> .....	36
<b>4.2 Data Modeling And Learning</b> .....	37
<b>4.3 Description of Retrieval Algorithm</b> .....	40
<b>4.4 Experimental Design</b> .....	41
4.4.1 Experimental Data Sets.....	41
4.4.2 Experimental Process.....	42
4.4.3 Show of Experimental Results.....	43
4.4.4 Method of Performance Assessment.....	45
4.4.5 Analysis of Experimental Results .....	45
<b>4.5 Summary of the Chapter</b> .....	46
<b>Chapter 5 Image Retrieval Prototype System Based On Textual</b>	

<b>Semantic And Visual Content</b> .....	48
<b>5.1 Overview</b> .....	48
<b>5.2 Frame Structure of Prototype System</b> .....	48
<b>5.3 Implementation of Prototype System</b> .....	51
<b>5.4 Summary of The Chapter</b> .....	53
<b>Chapter 6 Summary And Prospects</b> .....	54
<b>6.1 Work Summary</b> .....	54
<b>6.2 Work Prospects</b> .....	55
<b>References</b> .....	57
<b>Publications</b> .....	62
<b>Acknowledgements</b> .....	63

厦门大学博硕士论文摘要库

## 第一章 绪论

### 1.1 研究背景

“A picture is worth a thousand words”，图像作为一种蕴含丰富语义的信息载体，传递着比其它任何媒体更加生动、丰富、直观和形象的信息，因此图像在报刊杂志、广告媒体、医学、建筑、工程设计等众多领域中扮演着重要角色。近年来，随着数字成像技术的飞速发展，数码相机、带摄像功能的手机等移动设备正日趋流行，X射线计算机断层扫描系统、超声成像等专业领域的成像装置也得到了推广，图像的采集方法变得更加方便；另一方面，互联网技术的兴起以及大容量存储技术的成熟，使得图像传播的速度大大加快。全球的数字图像数据正呈现爆炸式增长，巨大的数据量便利了人们的生活，但也随之出现了一个问题：如何让用户从众多的图像中搜索到他们真正想要的信息。从而，研究高效的图像检索技术以快速准确地查询到所需的图像已经成为学术界和产业界的一个具有突出意义的课题<sup>[1]</sup>。

### 1.2 图像检索技术综述

图像的内容信息主要包括两类：文本语义信息和视觉内容信息，按照检索对象的不同，可将现有的图像检索技术分为两种：一种是基于文本信息的图像检索<sup>[2]</sup> (Text-Based Image Retrieval, TBIR)，另一种是基于图像视觉信息的图像检索<sup>[3]</sup> (Content-Based Image Retrieval, CBIR)。在基于文本信息的图像检索中系统利用图像文件名及周围的文本构造索引，从而将对图像的检索转化为对文本信息的检索，根据用户输入的关键字和图像库索引进行匹配，即可返回所需图像，这种检索方式被形象地描述成“以字搜图”。在基于视觉信息的图像检索中，系统通过提取其本身的底层视觉特征（颜色分布直方图、纹理描述特征、边缘形状特征等）来为图像建立索引<sup>[4]</sup>，由于这些特征往往表示为高维向量，检索时用户难以直接输入，一般是按照提供的样图、手绘图或图像特征的描述等来搜索，通过比较特征间的相似性，按相似性大小来排序给出最终的检索结果，这种检索方

式被形象地描述成“以图搜图”。两种图像检索方式各有自身的特点和适用的场景，接下来，本文将分别对它们进行详细的介绍。

### 1.2.1 基于文本信息的图像检索

随着计算机技术和数据存储技术的飞速发展，传统的文本检索方式得以高效地在计算机系统进行。基于文本信息的图像检索也就可以直接利用已比较成熟的文本检索技术，简单而快捷地实现；并且图像的语义一般能够由文本信息近似地表达出来，使得最终的结果常常可以达到用户的查询需求。然而，图像文本信息的获取多依赖于人工标注，随着图像数据量的迅猛增长，基于文本信息的图像检索的弊端也就不言而喻，主要有以下三方面：

1. 人工标注耗时耗力。目前，Google、Yahoo、Baidu等搜索引擎都提供了图像搜索的功能，其索引的图像数目已达到了数十亿张，通过人工手段对全部的图像进行标注所耗费的人力、物力等资源是不能承受的；另外人工标注的速度慢也难以适应每天大量新资源的出现。

2. 人工标注具有主观差异性和不确定性。因为不同人的知识构造和理解能力有异，所以不同的人对相同的图像会产生差异的认知和领会<sup>[5]</sup>；此外限定的文本信息难以全面的表达出内容繁杂的图像具有的含义。

3. 由于资源的国际化和网络化趋向，难以通过同一种语言对不同民族或国家的图像数据进行标注，所以统一的检索方法无法形成。

虽然存在上述问题，但是基于文本信息的检索系统便于理解、速度快，用户只需输入关键词便能查询获得相关图像。这种便捷性使得基于文本信息的图像检索仍旧作为现时应用最为广泛的检索方式<sup>[6]</sup>。下面，本文按照文本信息的来源方式分别介绍若干基于文本信息的图像检索系统。

首先，非常多的图像检索系统，特别是一些专业的图像管理人员或者商业的图像搜索引擎，多采用对图像进行人工标注的方法来进行分类检索。在初期的图像检索系统中，图像多数是依据如类型、时代、作者等与其有关的文本信息有序分类储存的，对这些图像进行标注分类的人员一般是相关领域的专业人士。随着图像数量的急剧增长，采用完全依赖于专业人士手工标注的方式显然是不可行的，在现有的一些大型商业图像网站如Instagram、SmugMug、Photobucket等对图像进行标注多数是通过互联网用户协同工作完成的。网站许可用户在上传图像

时可附带少许可以表达图像含义的文本信息，系统就会通过分析这些文本词来分类图像。

其次，网络中的图像往往不是孤立的个体，它周围往往伴随着许多与图像相关的文本信息，所以，现有的网络图像库在标注图像时往往使用半自动的标注方法，这种方法的有效性很大程度上是依赖于对图像周围关键文本信息进行提取的算法，图像的关键文本信息的来源主要有：

#### 1. 图像的文件名及网址

很多图像提供者是通过直接命名的方式来表示图像的内容，如football.jpg、animal.bmp等，可直接通过图像文件名知晓图像的含义。另外，图像的网址信息也包含了一些与图像有关的关键信息，比如<http://www.baidu.com/images/landscape/mountain.jpg>，就指示了该图像所属类型和具体的对象。

#### 2. 图像的替代文字

替代文字在网页中通常用来表示图像的语义信息，而且也是最为准确的一个特征。它常常是图像由于网络或者其它故障导致不能正确显示时用以替代图像的文本信息。

#### 3. 图像周围的文字

网页中的图像，其周围的文字很可能有描述了图像的内容含义，虽然其中有些文字也许与图像本身的相关性不大，但是综合分析这些文字信息还是有助于获取图像所表达的语义信息的，因此被选择成为语义特征之一。

#### 4. 图像所在页面的标题

图像有时是用来加强作者表现意图的一种形式，所以部分图像的内容含义很可能与图像所在页面的标题内容有关联，此时页面的标题也可被选择作为一种语义特征。

#### 5. 图像的超链接

图像的含义所指某种程度上关联于超链接信息，对其进行分析计算可得到图像的语义特征。

#### 6. 图像所在网页彼此间的链接

网页中所镶嵌的图像，其间在语义上的关联性在一定程度上是与网页彼此间

的链接有关的,对网页间的链接进行分析计算,可得到一些能够改善图像检索结果的有用信息。

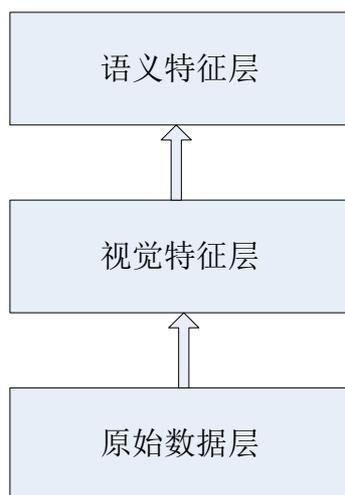
以上所述的有用信息,都可以利用页面分析器自动地从网页中获取得到,对它们赋予不同大小的权重,再利用成熟的文本信息检索的相关技术,构建图像的语义特征向量用以检索。

最后,很多科研工作者们把目光放到了如何采用完全自动的方式来获取图像的文本信息的研究上去,该类方式通过分析图像的内容从而自动产生相应的语义特征,具有客观、高效、省时等优势,是目前最具有研究意义的方向。由于仍然存在很多技术难题,该领域目前已取得的科研成果距离到真正的实际应用还比较远,特别是在面对大规模的图像数据库时现有的多数算法性能都会快速降低。近几年来,研究人员试着从各种不同的方向去探索,进行了大量的实验工作,是目前受到广泛关注的科研热点。在后续的第三章中,本文还会就如何自动地获取图像的文本语义信息问题进行详细地讨论和分析。

### 1.2.2 基于视觉信息的图像检索

为了克服基于文本信息的图像检索的一些问题,基于视觉信息的图像检索技术应时而生,在文<sup>[7]</sup>中有关于CBIR技术的研究综述。基于视觉信息的图像检索具体是指依据图像视觉特征来检索所需图像的过程,它是一门涉及多个领域的交叉性学科,包括:信号处理、图像处理、机器视觉、数据库、信息检索和模式识别等相关技术。它的基本思路是将提取到的图像视觉特征转化为矢量来作为图像的索引,计算目标图像与库里图像的特征向量间的距离,按相似度匹配进行检索。

图像的视觉特征是在图像原始像素数据基础上提取的,其提取主要采用计算机视觉和数字图像处理与分析技术,它可用于基于图像直观形象的检索。图像的语义是人们对图像深层次的理解,是在图像的视觉特征基础上结合人的先验知识推理得到的,并且随着人的认识的不同表现出不同的含义,能用来基于图像概念级的检索。于是,被用来进行图像检索的内容能够剖析成一个简单的三层模型<sup>[8]</sup>,可由图1-1表示。目前,主流的CBIR技术主要是使用图像第二层次的内容索引和检索图像。

图 1-1 图像内容的三层模型<sup>[8]</sup>

基于视觉信息的图像检索系统的结构能够分成图像特征提取和图像查询两个子系统<sup>[9]</sup>，如图 1-2 所示。

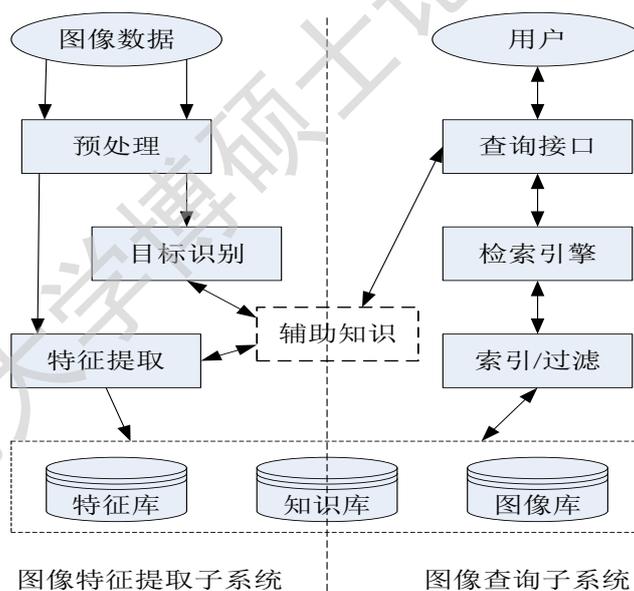


图 1-2 基于视觉信息的图像检索系统体系结构

- (1) 预处理操作主要是指对图像进行去噪声、区域增强、规一化等处理。
- (2) 目标标识即将用户感兴趣的区域在图中标记出来，从而减少无谓的计算，仅针对目标区域进行特征提取并准确查询。
- (3) 特征提取是指抽取图像的各种底层视觉特征作为图像的索引。
- (4) 数据库包括有特征库、知识库以及图像库。特征库包括有自动提取的

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库