基于综合特征的图像检索研究

**摘要：**基于内容的检索是多媒体数据查询和检索的重要方法，是当前研究的热点。本文对基于内容的多媒体信息检索概念、特征、基于内容检索的体系结构、处理过程、检索方法，用户查询接口的分类、查询说明的形式等也进行了一定的探讨，分析了多媒体数据的特征提取、检索索引方法、检索方法等内容。在分析当前图像检索系统的研究方向、基于内容图像检索技术的应用现状和主要技术等基础上，设计实现了一个小型的基于内容图像检索系统原型。该系统运用基于文本与基于内容相结合(关键词与颜色直方图和多级分块相结合)的查询方法，克服了单一的基于内容检索计算复杂、占用时间长的缺陷，提高了检索的速度与查准度。

基于内容的查询和检索是一个逐步求精的过程，存在一个特征调整、重新匹配的循环过程。相关性反馈技术是改善查询的重要方法。根据相关性反馈有关理论，该系统提供了一种简单的、直观的反馈途径，即通过改变检索结果中每幅图像与目标图像的相似程度后再刷新，提高了查询质量。另外，为了在图像特征间比较和度量，本文在图像特征向量和在某特征上的相似距离的归一化方面进行了一定的探讨与研究，给出了一种综合特征归一化方法，并将其作为两图像相似性度量依据。

**关键词:** 基于内容的检索 特征提取 相似性匹配 示例查询

STUDY OF IMAGE RETRIVAL BASED ON INTEGRATED FEATURES

**Abstract:** Contend-based retrieval is one of the most important methods of query and research in Multimedia data, and it is also the research hot spot. At the same time, the concept, feature, system structure, process, search methods of CBIR, and classify of query interface, form of query specification are discussed. I also analyze feature extraction of multimedia data, search index methods, search methods etc. On the base of analyzing research direction, application status and main technology of the current CBIR is system, I design and realized a prototype of small CBIR system. In this system, because the use of combined methods of text---based retrieval, the shortcoming of complex and long-time calculation is got over, the speed and veracity are improved.

CBIR is a gradually precision process. There is a feature adjustment and rematch process in it . Technology is one of the most important methods in improving query. According to relative feedback theory, I put forward a single, straight feedback approach, namely by set similitude degree between every image in query result and result and target image, the query quality is improved. In this paper, normalization of image eigenvector and similarity distance based on a feature are studied and discussed. a method on normalization of Integrated features is also put forward, it is the basis of similarity measure in two images.

**Keywords:** Content-based Retrieval; Feature extraction; Comparability matching; QBE

目录

[1 前言 1](#_Toc4683915)

[1.1 问题的提出 1](#_Toc4683916)

[1.2 文章的组织 1](#_Toc4683917)

[2 设计思想及依据 2](#_Toc4683918)

[2.1 搜索引擎简介 2](#_Toc4683919)

[2.2 开发工具 3](#_Toc4683920)

[3 图像的综合特征归一化方法 4](#_Toc4683921)

[3.1 图像基本特征及基于单特征检索时的缺陷 4](#_Toc4683922)

[3.2 综合特征归一化 4](#_Toc4683923)

[3.3 图像间的相似性度量 4](#_Toc4683924)

[4 基于内容的图像检索系统设计 5](#_Toc4683925)

[4.1 基于内容图像检索概述 5](#_Toc4683926)

[4.2 基于内容图像检索技术现状 5](#_Toc4683927)

[4.3 基子内容图像检索系统设计思想 5](#_Toc4683928)

[4.4 实验分析 5](#_Toc4683929)

[结束语 6](#_Toc4683930)

[参考文献 7](#_Toc4683931)

[附录 8](#_Toc4683932)

[致谢 10](#_Toc4683933)

1 前言

1.1 问题的提出

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

1.2 文章的组织

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

2 设计思想及依据

2.1 搜索引擎简介

2.1.1 搜索引擎发展史

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

2.1.2 搜索引擎的结构

从上面的介绍我们能看出一个完整的搜索引擎基本上分成以下三个部分：

1．Robot

Robot也叫做Spider、Web Crawler或Web Wanderer。Robot从一个事先制定好的URLs列表出发，这个列表中的URLs通常是从以往访问记录中提取出来的，特别是一些热门站点和"What＇s New"网页，从Usenet等地方检索得到的URLs也常被用作起始URLs，此外，很多搜索引擎还接受用户提交的URLs，这些URLs也会被安排在列表中供Robot访问。

2．搜索软件

搜索软件为用户提供一个界面，当用户查询一个关键词时，搜索软件将搜索Index，找出所有与关键词相符合的网页，有时候这些网页可能有成千上万，搜索软件将按照等级值从高到低的顺序把搜索结果送回到用户的浏览器中。图像检索的基本原理如图2.1所示。

图像检索引擎

图像获取与分析

索引图像库

网上图像资源

用户计算机用户

用户检索界面

检索需求

检索结果

用户反馈

检索视图

命中记录

图2.1 图像检索的基本原理

2.2 开发工具

2.2.1 开发工具简介

在我的程序开发过程中，我主要使用了C＃和SQL Server两个工具。其中C＃是主要的语言工具，SQL Server是辅助工具。下面我重点介绍一下C＃。

表2-1 C#的变量类型

|  |  |
| --- | --- |
| **变量类型** | **可能的内容** |
| object | null引用、任何引用类型的对象引用或任何值类型装箱值的引用 |
| 类类型 | null引用、该类类型实例的引用或由该类类型派生类的实例的引用 |
| 接口类型 | null引用、实现该接口的类类型实例的引用或实现该接口的值类型装箱值的引用 |
| 数组类型 | null引用、该数组类型实例的引用或兼容数组类型实例的引用 |
| 委托类型 | null引用或该委托类型实例的引用 |

2.2.2 图像间的相似性度量

提取了图像的某一特征并对特征向量元素及某一特征上的相似距离进行归一化后，再按公式（2-1）计算这些综合特征值，作为最终两图像间的相似性距离。

 （2-1）

其中N为图像特征个数，wj为某一特征上的距离值在综合特征中所占权重。

3 图像的综合特征归一化方法

3.1 图像基本特征及基于单特征检索时的缺陷

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

3.2 综合特征归一化

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

3.3 图像间的相似性度量

4 基于内容的图像检索系统设计

4.1 基于内容图像检索概述

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

4.2 基于内容图像检索技术现状

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

4.3 基子内容图像检索系统设计思想

在互联网发展初期，网站相对较少，信息查找比较容易。然而伴随互联网爆炸性的发展，普通网络用户想找到所需的资料简直如同大海捞针，这时为满足大众信息检索需求的专业搜索网站便应运而生了。

4.4 实验分析

结束语

主要包含本研究结果说明了什么问题，得出了什么规律性的东西，或解决了什么实际问题；本研究的不足之处、尚待解决的问题或提出研究设想和改进建议。

参考文献

[1]钟玉琢,沈洪,吕小星等.多媒体技术及其应用[M].北京:机械工业出版社.2013

[2]周向东,施伯乐等.基于长期学习的多媒体数据库相似性检索[J].软件学报.2014,15（1）：32-39

[3]沈琦,张虹.多媒体数据库模型与存储模型研究[J].计算机工程.2007,23（6）：23-30

[4]刘玉照,黄蕾.多媒体数据库模型及实现途径之比较研究[J].情报科学.2011,19（8）：853-857

[5]郑庆华,李人厚,胡俊.多媒体数据库的主要问题、研究内容与实现方法[J].计算机工程与应用.2015,32（1）：23-30

[6]黄蕾.多媒体数据基于内容检索综述[J].情报探索.2000,4（76）：8-13

[7]黄丽娟.基于内容的多媒体信息检索[J],情报杂志.2010,19（5）：47-51

[8]李国辉,曹莉华,柳伟.基于内容的多媒体数据查询和检索[J].小型微型计算机系统.1998,19（4）：23-30

[9]郝小花.多媒体信息中基于内容的检索[J].情报理论与实践.1999,22（3）：23-30

[10]李国辉,胡小峰.基于内容的检索[J]．计算机世界．1998，D1-D3：23-30

[11]徐建华.一种新型的多媒体检索技术----基于内容的检索[J].情报学报.2000,19（4）：53-58

[12]王涛等.融合型多媒体数据库模型的实现技术[J].焦作工学院学报.2003,22（6）：13-20

[13]李国辉,曹莉华等.多媒体信息查询和检索系统MIRC[J].小型微型计算机系统．2009，20（9）：27-40

[14]何清法,李国杰.综合分块主色和相关反馈技术的图象检索方法[D].北京:中国中科学院.2000

[15]王玉波.多媒体信息检索技术概论[J].情报科学.2008,17（2）：56-60

附录

主要分析程序

using System;

using System.IO;

using System.Text.RegularExpressions;

using System.Windows.Forms;

namespace crawler

{

public class mycountword

{

public int htmlcountword;//超链接单词数

public int allcountword;//总单词数

public int htmlnumble;//超链接数

public int titlewordnumble;//title词数

public void analysefile(string filepath)

{//==============================将文件形成字符串流

htmlcountword=0;

allcountword=0;

htmlnumble=0;

titlewordnumble=0;//初始化变量

FileStream htmloldfilestream;//建立未处理文件流

htmloldfilestream=new FileStream(filepath,FileMode.Open,FileAccess.Read);//打开文件

int temp;

string mystring="";

char tempchar;

for(int i=0;i<htmloldfilestream.Length;i++)

{

temp=htmloldfilestream.ReadByte();

tempchar=(char)temp;//类型转换

//MessageBox.Show(tempchar.ToString());

mystring+=tempchar;

}

//＝＝＝＝＝＝＝＝＝＝＝＝＝＝提取超链接的arch

//寻找<a href=..标志;

//建立匹配集合类

//MessageBox.Show(mystring.Length.ToString());//长度正确

MatchCollection mc,mc2;

致谢

毕业设计（论文）不是作者一个人单独完成的，需要老师的指导以及其他人的帮助。因此应增加本部分，以对论文工作直接提供过资金、设备、人力，以及文献资料等支持和帮助的团体和个人表示感谢。