分类号 密级

UDC注1



**硕士专业学位论文**

**基于Android平台的服饰图像搜索**

**系统的设计与实现**

**李刚**

**指导教师姓名 赵春霞 教 授**

**学 位 类 别 工程硕士**

**专 业 名 称 计算机应用**

**研 究 方 向**

**论文提交时间**

注1：注明《国际十进分类法UDC》的类号。

**声 明**

本学位论文是我在导师的指导下取得的研究成果，尽我所知，在本学位论文中，除了加以标注和致谢的部分外，不包含其他人已经发表或公布过的研究成果，也不包含我为获得任何教育机构的学位或学历而使用过的材料。与我一同工作的同事对本学位论文做出的贡献均已在论文中作了明确的说明。

研究生签名： 年 月 日

**学位论文使用授权声明**

南京理工大学有权保存本学位论文的电子和纸质文档，可以借阅或上网公布本学位论文的部分或全部内容，可以向有关部门或机构送交并授权其保存、借阅或上网公布本学位论文的部分或全部内容。对于保密论文，按保密的有关规定和程序处理。

研究生签名： 年 月 日

# 摘 要

近年来，互联网的普及促进了电子商务的快速发展。人们在网络购物时，商品图像信息是购买与否的重要参考信息。然而目前商品图像并没有广泛的应用到商品搜索系统中，用户很难快速准确找到自己需要的商品。与此同时，智能手机的普及和发展带动了互联网的移动化浪潮。手机购物成为新的趋势，移动终端的交易份额逐年增加。如何在移动终端上实现图像搜索成为电子商务平台商品搜索智能化的迫切需求。本文对此展开了研究，主要工作有：

深入分析和研究了基于内容的图像检索的一些关键技术，主要包括图像主要低层特征（颜色、纹理、局部特征）的描述方法以及提高检索正确率（GrabCut）和加快检索效率（视觉词袋模型、倒排文档索引）的途径等。

结合服饰图像搜索的特殊要求，选择适宜的图像特征模型应用于服饰图像搜索系统。全局特征提取颜色信息和纹理信息，其中颜色信息使用量化过的HSV颜色分布直方图表示、纹理信息使用Uniform LBP特征表示。局部特征提取SIFT特征。SIFT特征借助视觉词袋技术进行描述压缩，并采用倒排文档索引加速搜索。同时利用GrabCut算法对图像进行前景提取减少背景的干扰。图像全局特征采用欧式距离进行相似度度量。使用TOP K算法对距离进行排序，得到相似度最高的图像。

详细介绍了基于Android的服饰图像搜索系统的设计和现实过程。基于Android的服饰图像搜索系统由客户端、服务器端、图像集和特征池三大部分组成。客户端基于Android实现，包括待搜索图像输入模块、搜索结果展示模块、其它辅助功能模块。服务器端基于Python Django实现，包括颜色特征提取、纹理特征提取、SIFT特征提取、视觉词、排序等模块。图像集和特征池部分主要包括网络爬虫模块及特征提取模块。

最后，对实现的服饰图像搜索系统进行了测试和实验。结果表明系统搜索准确率良好。

**关键词**：Android，基于内容的图像检索，LBP，SIFT，特征提取

# Abstract

- 4 -

In this dissertation, a research has been made on the surface hydrophilic modification of porous polyvinylidene fluoride(PVDF) membrane based on supramolecular chemistry process and its application in the treatment of low concentration oily wastewater.

**[Document Title]**

First, the mechanism of surface hydrophilic modification of PVDF membrane based on supramolecular

**Key word:**Android, CBIR, LBP, SIFT, Feature extraction

**目录**

[摘 要 - 3 -](#_Toc439064739)

[Abstract - 4 -](#_Toc439064740)

[1 绪论 7](#_Toc439064741)

[1.1 课题的研究背景和意义 7](#_Toc439064742)

[1.2 国内外研究现状 8](#_Toc439064743)

[1.3 论文的主要工作 9](#_Toc439064744)

[1.4 论文的组织结构 9](#_Toc439064745)

[2 基于内容的图像检索技术 10](#_Toc439064746)

[2.1 概述 10](#_Toc439064747)

[2.2 颜色特征 11](#_Toc439064749)

[2.3 纹理特征 12](#_Toc439064750)

[2.4 SIFT特征 14](#_Toc439064751)

[2.5 视觉词袋模型 17](#_Toc439064752)

[2.6 倒排文档索引 18](#_Toc439064753)

[2.6 图像分割技术 19](#_Toc439064754)

[3 Android平台相关技术 20](#_Toc439064755)

[3.1 Android平台的发展 20](#_Toc439064756)

[3.2 Android平台的组成 21](#_Toc439064757)

[3.2.1 Linux内核 22](#_Toc439064758)

[3.2.2 系统运行库 22](#_Toc439064759)

[3.2.3 应用程序框架 22](#_Toc439064760)

[3.2.4 应用程序 23](#_Toc439064761)

[3.3 Android应用程序开发技术 23](#_Toc439064762)

[3.3.1 Activity 23](#_Toc439064763)

[3.3.2 Service 25](#_Toc439064764)

[3.3.3 Broadcast 26](#_Toc439064765)

[3.3.4 ContentProvider 27](#_Toc439064766)

[4 基于Android平台的服饰图像搜索系统实现 28](#_Toc439064767)

[1.1 系统架构 28](#_Toc439064768)

[4.2 Android客户端实现 29](#_Toc439064769)

[4.2.1 逻辑架构 30](#_Toc439064770)

[4.2.2 图像输入模块 31](#_Toc439064771)

[4.2.3 结果展示模块 34](#_Toc439064772)

[4.2.4 其它辅助模块 37](#_Toc439064773)

[4.3 图像集及特征池的建立 39](#_Toc439064774)

[4.3.1 图像集的建立 39](#_Toc439064775)

[4.3.2 特征池的建立 41](#_Toc439064776)

[4.4 服务器端实现 41](#_Toc439064777)

[4.4.2 Django简介 41](#_Toc439064778)

[4.4.3 GrabCut模块 44](#_Toc439064779)

[4.4.4 颜色特征提取模块 45](#_Toc439064780)

[4.4.5 纹理特征提取模块 47](#_Toc439064781)

[4.4.6 SIFT特征提取模块 47](#_Toc439064782)

[4.4.7 视觉词模块 48](#_Toc439064783)

[4.4.8 排序模块 49](#_Toc439064784)

[5 系统测试与实验 50](#_Toc439064785)

[5.1 测试环境 50](#_Toc439064786)

[5.2 实验 51](#_Toc439064787)

[5.2.1 评价标准 51](#_Toc439064788)

[5.2.2 实验结果 52](#_Toc439064789)

[6 总结及展望 54](#_Toc439064790)

[5.1 论文总结 54](#_Toc439064791)

[5.2 研究展望 55](#_Toc439064792)

[致 谢 56](#_Toc439064793)

[参考文献 57](#_Toc439064794)

# 1 绪论

传统的服饰搜索通常采用关键词搜索技术对服饰信息进行整合，忽略了对电子商务而言非常重要的图片信息。另一方面，用户并不总能对商品的关键词进行准确描述，而随着智能手机的普及，随时随地用手机拍摄的服饰照片则对实物具有更直观、准确的描述。因此，将基于内容的图像检索技术应用于商品检索领域具有广阔的应用前景和推广价值。

## 课题的研究背景和意义

随着互联网前所未有的发展和智能手机的迅速普及，移动电子商务逐渐成为一个新兴的领域。网络购物移动应用日渐增多。

随着互联网前所未有的发展，网上购物已经成为一种流行与时尚。网上购物者的数量在不断的增长。购物网站也相继涌现、发展和壮大，比如淘宝网、天猫商城、京东、苏宁易购等等。和实体店购物不同的是，网络购物者都是利用计算机和网络来浏览所需要的商品。由于缺乏现实购物过程中的视觉和触觉体验，网络购物者很难挑选到合适的商品。因此，一个电子商务平台的商品搜索系统就显得尤为重要。

多数购物网站设有自己的商品索系统，像淘宝、拍拍网等大型专业购物网站依赖于自己开发的商品搜索引擎。一般的购物网站都是嵌入了百度、搜狗等搜索功能来进行搜索。由于技术的原因，大部分购物网站对于商品图片信息的检索都是基于标注的关键字来实现的。

图像信息在数据库中的检索方式有两种，即全文索引和关键词索引。其基本步骤就是在网页信息自动采集和标引作为搜索引擎的重要组成的基础上，建立全文索引和关键词索引。全文索引就是用图像所在网页的全部文字信息作为图像的注释，网页上任何文字信息都认为与图像相关，显然这种图像标注查全率高但是查准率却是很低；基于关键词的索引使用若干关键词来表示图像信息，这些关键词一般都是从卖家对商品的图像的描述中获得，相比之下查准率高，但是查全率低。

电子商务购物网站和一般网站相比具有其独特性，即为了直观呈现商品的信息，多使用图片作为信息载体。商品本身具有自己的特性，比如对时间的敏感性和对色彩的依赖性。这些因素导致图像数量的飞速增长和图像信息数量的爆炸。

采用上述两种检索方式存在着以下问题：

（1）商品销售方对商品信息分类方式的多样性。销售人员是网上信息的发布者，他们不可能按照有关分类法来进行信息的分类，而是根据自己的理解对图片文件加注标签，具有很强的主观性。因而必然会增加标引词的多样性，增加了检索的范围。

（2）消费者对商品信息标签理解的歧义性。即由于个人认识的差异，不同的人对待相同的标签可能就会有不同的理解。

（3）视觉依赖性强的商品类型划分不同。即对于这种视觉依赖性强的商品，不同的角度划分就会导致所属的类型不同。

（4）商品信息标注的不完全性。即商品的文本标注不能全部反映图像的内容。

由于上述各种主观因素的存在导致基于文本的信息检索不能够很好的满足消费者所需商品的检索匹配。用户通过文本搜索检索不到自己心仪的商品，这在一定程度上将会降低购物者的消费需求。如果有一种检索方式可以避免图像文字标注的主观性和差异性所带来的匹配不精确问题，而直接客观地从商品的图像来检索匹配，不仅可以提高检索的查准率和查全率，还可以提高搜索的时间效率。于是基于内容的图像检索在购物网站中的应用成为一种迫切的需要与必然趋势。

## 1.2 国内外研究现状

基于内容的图像检索逐渐成为图像理解和计算机视觉领域的热门研究课题，国内外的研究结构已经投入了大量人力物力开展了对该课题的广泛研究，并且研制了一些商业系统和实验系统。比较知名的基于内容的图像检索系统有IBM公司的QBIC系统、由哥伦比亚大学研究开发的VisualSEEK和WebSEEK系统、由美国Virage公司开发的Virage系统、由美国MIT媒体实验室开发的Photobook系统等。

同时近年来，计算机视觉技术和电子商务得到了迅猛的发展。基于图像外部特征描述的商品检索方式已经不能很好的满足用户的需求，因此，国内外的研究机构开始将基于内容的图像检索技术引入到了商品的检索和浏览中，其中比较有代表性的图像搜索引擎如QBIC、谷歌图像搜索以及移动端图像检索系统Google Goggles。

（1）QBIC

QBIC系统是由IBM Almaden研究中心开发的第一个商品化的基于内容的图像检索系统，它的系统框架、结构和技术对后来的图像检索系统有着深远的影响。QBIC系统支持基于例子图像、手绘略图、选择的颜色、纹理等查询，不仅支持图像检索，还支持视频、文本和语音等多种形式的信息检索。QBIC系统是少数几个考虑高维特征索引的系统。QBIC系统使用的颜色特征是颜色直方图。纹理特征采用粗糙度、对比度和方向性描述。形状特征包括面积、离心率、主轴方向和不变炬。颜色、纹理和形状均采用加权的欧氏距离比较。

（2）谷歌图像搜索系统

2011年6月份谷歌发布了以图搜图的图片搜索服务，谷歌以其强大的搜索功能而著称，以图搜图的搜索技术也相对比较成熟。谷歌图像搜索融合了图像分析、模式识别、人机交互等多种技术，提供关键字、视觉特征相结合的检索模式，提高了查询的准确性，提升了用户体验。其操作流程比较简单，用户通过选择查询图像的相关特征如颜色、纹理等进行检索。Google图像检索具有理想的检索效率，谷歌图像检索系统结合了图像特征和关键字对查询图像进行分析进而匹配出最相似的图片集合，提高了图像检索的准确率，优化搜索结果，改善了用户体验。

（3）Google Goggles

Google Goggles是谷歌推出的一款移动端的图像搜索工具，利用手机摄像头拍摄当地地标建筑、书籍封面、艺术作品、酒类标签以及产品商标等物体的照片后，软件就将会自动在Google上搜索相关信息，并予以识别显示。不过Google Goggles对查询图像的类别有一些限制，不能检索家具、服装之类的东西。

（4）拍立淘

拍立淘是阿里巴巴公司研发的商品图像搜索系统，目前应用于手机淘宝。拍立淘的领军人物是著名亚裔科学家华先胜。拍立淘是目前商业化最成功的图像搜索系统。它背靠淘宝数十亿级的商品图像，揉合了深度学习、大数据和海量用户反馈，取得很好的搜索效果。随着搜索结果相关性的提升，拍立淘的用户UV接连攀升。

## 1.3 论文的主要工作

本论文围绕着基于内容的图像检索系统，进行了如下工作：

阅读文献资料，研究基于内容的图像检索系统的特征提取方法及特征匹配方法。最终选择了HSV颜色直方图、Uniform LBP、SIFT作为本系统的特征提取方式。

研究Android平台的组成与特性，以及应用层开发关键技术。

搭建图像检索系统，实现拍照、结果展示功能。系统拍照后检索出相似的图片结果，并在用户界面展示给用户。

最后，测试图像检索系统，得出并分析实验数据。

## 1.4 论文的组织结构

本文共划分为六个章节，每章的主要内容如下所示。

第一章，绪论。主要介绍了系统上基于内容的图像检索研究的背景与意义，说明了当前国内外对基于内容图像检索的研究实践现状。给出了本论文的主要工作以及论文组织结构。

第二章，基于内容的图像检索技术。本章主要介绍了基于内容的图像检索原理，着重解析了HSV颜色直方图、Uniform LBP、SIFT这三种特征提取方式，并介绍了本文项目中使用到的GrabCut算法和视觉词袋模型。

第三章，Android平台相关技术。本章首先介绍了Android的发展历程和发展现状。然后阐述了Android平台的组成。最后解析了Android应用开发的关键技术。

第四章，系统实现。详细介绍了Android客户端、服饰图像集及特征池、服务器端的实现过程。认真阐述了工程工作中遇到的问题及解决办法。

第五章，实验结果与讨论。对实现的图像检索系统进行测试，分析检索效果。

第六章，研究总结及展望。在这一章将总结前面的研究及实践成果，讨论并展望下一步研究方向。

# 2 基于内容的图像检索技术

为了搭建上的图像检索系统，本论文首先研究了基于内容的图像检索原理。基于内容的图像检索包含图像特征抽取和特征匹配两大关键技术。在本章节中，会对基于内容的图像检索原理进行概述。在小节中介绍了本论文选用的特征描述符的原理，详细解析了的颜色、纹理信息提取及的执行与量化。

## 概述

随着互联网设备的发展以及图像采集设备的普及，使得快速、大量的采集、传输图片信息成为可能。由于一张图片往往难以用几个关键词准确形容，原来简单的基于文本的图像检索已经不能满足用户的需求。为了解决基于文本的图像检索中的诸多不足，人们想到了利用图像的内容进行检索。基于内容的图像检索简称应运而生。基于内容的图像检索是指由计算机自动提取图像的某些特征信息，并采用某种相似度的方法对这些特征进行匹配，并最终返回用户所需结果的技术。不同于基于文本的图像检索，基于内容的图像检索要直接从图片中提取特征值，而不需要标注关键字。这之后，基于内容的图像检索将所有图片特征值集合起来形成一个图片特征库。对于待检索图像，提取相同的特征值与图片特征库中的特征值逐一进行相似度测量，最后按照相似度的大小返回给用户，以方便用户的使用、查询。

相比于基于文本的图像检索，基于内容的图像检索具有以下三个特点。第一是，直接从图像提取特征值，突破了传统图像检索中基于文本关键词检索的局限。第二是，其本质是一种近似匹配技术，不同于常规的精确匹配方法，通过度量相似度，最后排列检索结果进行输出。第三是，图片特征提取通过计算机自动完成，减少了人工加注标示的工作量，并且消除了人主观性的干扰。基于内容的图像检索系统的工作流程示意图如图中所示。包含两个阶段：一是特征值预先提取阶段，二是图像搜索阶段。在特征值预先提取阶段，抽取图像数据库中的每一张图片的特征，存放在特征数据库中。在图像搜索阶段，利用同种特征提取手段提取待搜索图像的特征值，之后和特征数据库中的特征值逐一进行匹配，计算特征相似度，以获得最佳检索结果。

基于内容的图像检索选用的特征必须能够区别目标并且足够形容目标。基本上，实现一个成功的检索系统的关键就是选择一组特征，这组特征描述能力要强而且要独特以便利于区分判别。关于系统的类型，它们可以分为基于色彩特征的，基于纹理特征的以及基于形状特征的。仅用一种特征种类很难以达到理想的效果。很多图像检索技术都釆用提取多种图像特征放到一个直方图中的方法来进行图像特征抽取。

~~本文选用的特征提取方式为，这是一种将颜色信息和纹理信息结合于一个直方图的描述符。的优点是限制了描述符的大小，将每张图提取出的特征限制在字节。保证其仅需较少的时间就可以完成特征提取工作。~~

~~关于的具体内容将在中加以详细叙述。~~



## 颜色特征

在基于内容的图像检索中，图像的内容由图像的特征表示。图像的特征主要分为低层视觉特征和高层语义特征两大类。目前，受到计算机视觉、心理学、生物学等学科发展水平的制约，基于高层语义特征的图像检索技术还很不成熟。因此，目前基于内容的图像检索技术的研究热点是基于低层视觉特征的图像检索技术。

颜色特征具备旋转不变性和尺度不变性，而且相比其它特征容易获取，是描述一幅图像最简单有效的特征。因此颜色特征在基于低层视觉特征的图像检索中应用广泛。

RGB颜色空间是通过红绿蓝三原色来描述颜色的颜色空间，是最基本、最常用的颜色空间。数字图像一般均采用RGB颜色空间来表示。然而，RGB颜色空间的分量与亮度密切相关，即只要亮度改变，3个分量都会随之相应地改变。所以，RGB颜色空间适合于显示系统，却并不适合于图像处理。

HSV颜色空间是通过色调（Ｈ）、饱和度（S）、量度（V）３个分量来描述颜色的颜色空间，是一种面向视觉感知的颜色空间。其中，色调是指图像的主色，与混合光谱中主要光波长有关，如黄、绿、青代表不同的色调；饱和度是指色彩的深浅程度，与一定色调的纯度有关；亮度是指人眼感受到的光的明暗程度。HSV颜色模型与人眼的视觉特征比较接近，所以HSV颜色空间在图像颜色特征提取与分析中应用广泛。

从RGB颜色空间到HSV颜色空间的转换是一个计算简单的非线性变换。

RGB颜色空间到HSV颜色空间的转换公式如下所示：

 （2.1）

 （2.2）

  （2.3）

其中，，，， 。

为了计算的便捷，需要对颜色进行量化，即将颜色空间映射到一个给定的子集中，使其总体误差最小。

（1）根据人的视觉分辨能力，把色调 H 空间分成 8份，饱和度S空间分成3 份，亮度V空间分成3份。

 （2.4）

 （2.5）

 （2.6）

（2）根据色彩的不同范围和主观颜色感知进行量化

（3）构造一维特征矢量。按照以上的量化级，把各颜色分量合成为一维特征矢量：

 （2.7）

其中，和分别是S、V的量化级数，，。

公式x实际上为：

 （2.8）

这样，H,S,V三个分量在一维矢量上分布开来。W的取值范围是。计算 W获得72 bin一维直方图。

## 2.3 纹理特征

Castlema等人认为[2]：纹理是一种反映图像中一块区域的像素灰度级的空间分布属性，这种空间结构的固有属性可以通过邻域像素间的相关性刻画。常用的纹理分析方法有四种：统计分析方法、结构分析方法、模型分析方法和频谱分析方法。由于纹理特征提取计算成本过高，影响了图像纹理处理的实际应用，一些研究人员在图像纹理处理的计算简化上做了许多工作。Ojala 等人[15]提出了具有开创性的纹理特征提取方法，其计算简单可行，已经在许多研究领域内取得了实效。

局部二元模式（local binary pattern，简称LBP）是一种局部纹理描述算子。由于具有计算简单、对光照变化不敏感等优势，在纹理分类、人脸识别、医学图像处理等领域应用广泛。

基本LBP算子计算过程为：将中心像素点3\*3邻域内的8个像素点的灰度值分别与中心像素点的灰度值比较大小，根据大小关系对8个相邻像素点进行二值化。即如果相邻像素点大于中心像素点的灰度值，则将其置为1，否则置为0。然后按顺时针方向对这些值进行加权求和，得到该邻域的LBP特征值。

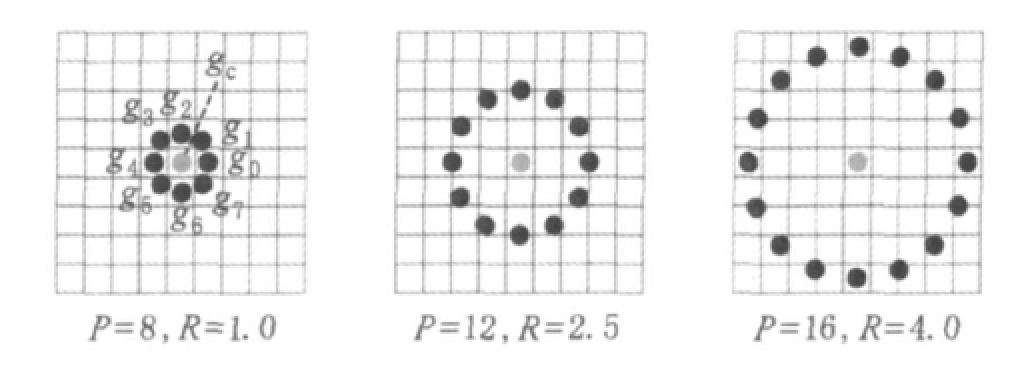
基本LBP算子具有无法提取大尺度纹理特征的局限性。为了提高LBP算子的有效性和完整性，Ojala等人对基本LBP算子进行了改进，LBP算子的计算不在局限于3\*3窗口内的相邻的8个像素点，而是给定一个采样半径R和采样点数P，在中心像素点为圆心、R为半径的圆周上等间隔的采样P个点。常用的P、R取值组合有P=8，R=1.0、P=12，R=2.5、P=16，R=4.0，如图x所示。

图2.1 常用的P、R取值组合示例

现在介绍P=8，R=1.0时改进LBP算子的计算过程，其它取值组合依次类推。设中心像素点的灰度值为，相邻的8个像素点的灰度值依次为，那么该中心像素点的LBP算子可表示如下：

 （2.9）

 （2.10）

LBP8,1.0有种不同的取值，如果我们采用LBP8,1.0特征直方图来表示一副图像的话，需要统计种类别，这个向量将是bin的，不便于计算。Ojala等人经过对大量纹理图像进行研究之后发现，如果将LBP算子二进制值首尾相连组成环，绝大多数的环至多存在2次0和1之间的跳变。以LBP8,1.0为例，00000000、01110000、01010000分别包含0、2、4次0和1之间的跳变。基于这样的统计规律，Ojala等人提出了均匀模式的LBP算子，即将至多存在2次0和1之间的跳变的LBP算子定义为均匀模式，分别归类。将存在2次以上0和1之间的跳变的LBP算子定义为混合模式，归为一类。压缩之后，LBP算子的种类可由原来的减少为。均匀模式占所有模式中的绝大多数，这样利用均匀模式和混合模式求统计直方图时，可以在不损失太多信息的情况下，而大大减少统计的类别。

本文纹理特征的提取使用均匀模式的LBP8,1.0，将原来的28种LBP特征压缩为8\*(8-1)+3=59种LBP特征，方便计算图像的LBP直方图。

## 2.4 SIFT特征

图像的高斯尺度空间

尺度空间理论的出现是为了模拟图像数据的多尺度特征。Koenderink与Lindeberg证明了高斯卷积核是唯一的线性尺度核。一副图像的尺度空间被定义为：

 （2.11）

其中为尺度可变高斯函数：

 （2.12）

式中：为尺度坐标；为尺度因子，决定图像的平滑程度，大尺度对应图像概貌特征（低分辨率），小尺度对应图像细节特征（高分辨率）。

SIFT特征

SIFT(Scale-Invariant Feature Tansform)描述子由Lowe于1999年提出，2004年总结完善。SIFT对旋转、尺度缩放、亮度变化具有不变性，对视角变化、仿射变换、噪声的容忍度较高。应用范围包含物体辨识、机器人地图感知与导航、影像缝合、3D模型建立、手势辨识、影像追踪和动作比对。SIFT特征提取主要分为4个步骤：

构建高斯差分尺度空间（DOG scale-space）

高斯差分尺度空间利用不同尺度高斯差分核与图像卷积生成。是两个相邻尺度图像的差。

 （2.13）

如图x所示。

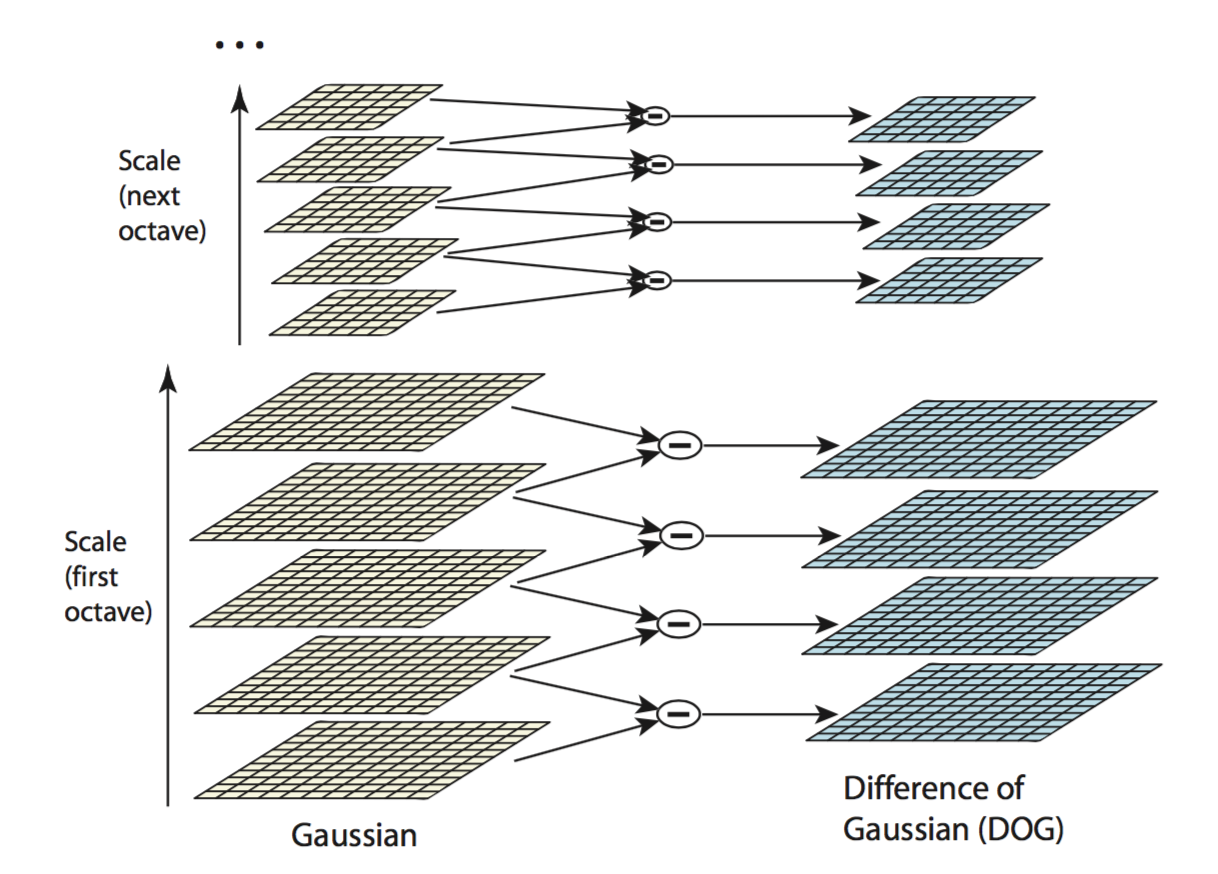


图2.2 构建高斯差分尺度空间（DOG scale-space）

关键点定位

如果一个像素在DOG尺度空间本层以及上下两层的领域中具有最大值或最小值时，则认为该像素是图像在该尺度下的一个关键点。如图x所示。

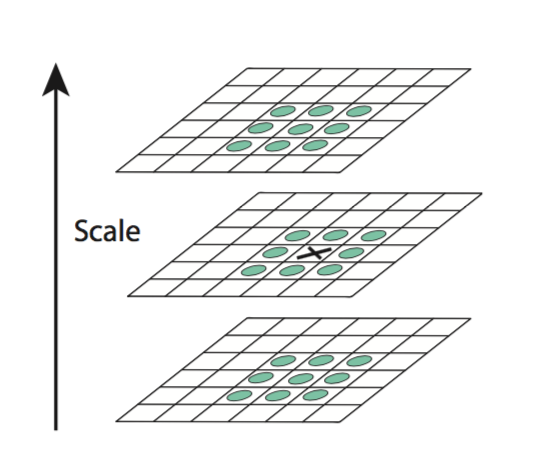


图2.3 DoG尺度空间关键点定位

中间的检测点和它同尺度的8个相邻点和上下相邻尺度对应的9×2个相邻点共26个相邻点进行比较，以确保在尺度空间和二维图像空间都检测到极值点。

关键点方向确定

基于关键点邻域像素的梯度分布特性，分配给每个关键点一个或多个方向参数。后续的对图像的操作均相对于关键点的方向、尺度和位置进行变换，从而提供对于这些变换的不变性。

  （2.14）

 （2.15）

公式x、x分别为处梯度的模值和方向公式。其中所用的尺度为每个关键点各自所在的尺度。

在实际计算时，我们在以关键点为中心的邻域窗口内采样，并用直方图统计邻域像素的梯度方向。梯度直方图的范围是0～360度，其中每45度一个柱，总共8个柱, 或者每10度一个柱，总共36个柱。Lowe在论文中建议使用高斯函数对直方图进行平滑，减少突变的影响。直方图的峰值代表该关键点处邻域梯度的主方向，用作该关键点的方向。图x是采用7个柱来统计邻域像素梯度方向并确定主方向的示例。

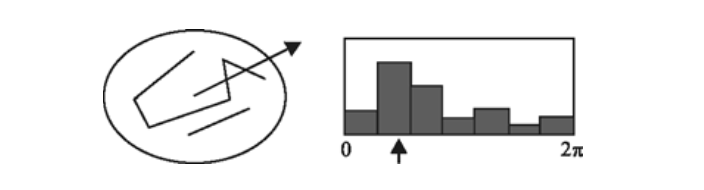


图2.4 确定关键点方向

关键点描述子的生成

首先将坐标轴旋转为关键点的方向，以确保旋转不变性。以关键点为中心取8×8的窗口。

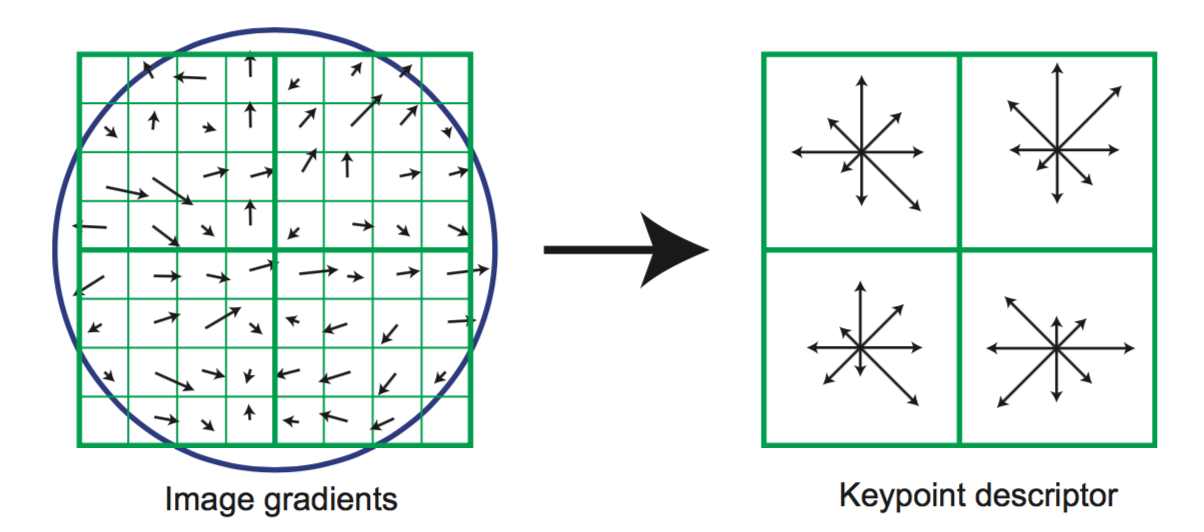


图2.5关键点描述子的生成

图左部分的中央为当前关键点的位置，每个小格代表关键点邻域所在尺度空间的一个像素，利用公式x求得每个像素的梯度幅值，利用公式ｘ求得每个像素的梯度方向，箭头方向代表该像素的梯度方向，箭头长度代表该像素的梯度模值，然后用高斯窗口对其进行加权运算。图中圈代表高斯加权的范围（越靠近关键点的像素梯度方向贡献越大）。然后在每4×4的小块上计算8个方向的梯度方向直方图，绘制每个梯度方向的累加值，即可形成一个种子点，如图右部分示。此图中一个关键点由2×2共4个种子点组成，每个种子点有8个方向信息。这种邻域方向性信息联合的思想增强了算法抗噪声的能力，同时对于含有定位误差的特征匹配也提供了较好的容错性。

计算关键点周围的16×16的窗口中每一个像素的梯度，并使用高斯下降函数降低远离中心的权重。这样就可以对每个关键点形成一个4×4×8=128维的描述子。将这一向量归一化之后，可进一步去除了光照的影响。

## 2.5 视觉词袋模型

词袋（Bag of words）[x]是文档分析领域统计关键词出现频率的一项技术，被广泛应用于信息检索和文本分类。视觉词袋模型是词袋模型在计算机视觉领域的应用。它将图像映射为视觉单词的分布直方图，既保存了图像的局部特征又压缩了图像的描述。视觉词汇的使用使得在大规模图像检索上面有效的提高了检索效率[]。除了用在图像检索领域外，视觉词汇在图像分类上面也有应用[]。

在文本信息检索系统中，一般采用如下标准处理过程：

步骤一：首先将文档解析成一系列单词，即进行分词处理。

步骤二：每个词由其单词主干替代，例如，“ate”、和“eating”都替换成单词主干“eat”。

步骤三：蹄选掉一些及其常见的词，如“for”、“at”等，这些词对文档的可区分性很差，因为太常见了。

步骤四：对于剩余的每个不同词汇赋予一个唯一的标识符，该文档用每个词出现的频率得到一个频率直方图向量。

步骤五：对于每个词都有一个权重系数，不同的系统规定的权值系数取值方法不一样，在做检索查询时返回与查询向量最接近的文档。

视觉词袋实现过程如下：

步骤一：特征提取。提取图像集中每张图片的局部特征，每张图片提取结果是多个局部特征描述符。

步骤二：聚类生成视觉词典。步骤一种得到的所有特征构成一个特征空间，利用K-means进行聚类。设是一组观测值序列，其中，每个观察值都是一个d维向量。利用K-means将这n个观察值划分到k个序列中，其中是的均值。

 （2.16）

步骤三：将图像用词典表示。将图像中的局部特征映射到距离最近的视觉词上，统计图像中各个视觉词的频率，得到分布直方图，最终用直方图K维向量表示图像。其中，K为词典中包含的视觉词数量。局部特征之间的相似度用欧式距离度量。同文本信息检索过程没有考虑到词汇的顺序，只考虑词汇出现频率一样，视觉词汇也没考虑到视觉特征的空间分布信息，只考虑视觉词汇的出现的频率。

在上述聚类过程中，选择的聚类数目也就是视觉词汇表的大小要适中，如果词汇表过大了，则产生的视觉特征就会非常精细，一些可能正确的匹配会检测不到，此外词汇表过大对于时间和空间的开销也变大了。而词汇表过小则会将差别比较大的视觉特征通过量化映射到了同一个特征，从而导致本该错误的匹配结果被当作是正确的，系统的正确性得不到保证。

## 2.6 倒排文档索引

图2.6 倒排文档索引结构

本文采用文本信息检索里面的倒排文档索引结构对图像特征数据进行索引，索引结构如图x所示。在上述图中，每个视觉词汇表项都记录了包含此视觉词的图像编号以及此视觉词在该图像中出现的次数。使用了文档倒排索引结构后，在进行检索操作时，可以将检索过程看成是一个投票的过程。即在查询图像中提取的每个特征描述符量化到相应的视觉词汇，然后对每个视觉词汇对其所索引的所有图像进行投票。借鉴于文本搜索处理的技术还可以通过不同的视觉词汇的不同TF-IDF（Term frequency-Inverse document Frequency）值对图像中的视觉词汇进行加权投票，最后按照得分高低排序选择出检索的图像。TF-IDF的计算方法如下：

 （2.17）

Nid表示的是视觉词汇i在图像d中出现的次数，Nd表示的是该图像中的视觉词汇数目，N表示的视觉词汇表的大小，Ni表示的词汇i在整个图像库中出现的次数，log(N/Ni)其中的是在创建视觉词袋模型的时候计算好的。

## 2.6 图像分割技术

图像分割是图像工程中目标检测、特征提取和参数测量的基础，是图像分析、模式识别、计算机视觉领域的关键问题之一，它使得高层次的图像理解成为可能。

利用基于能量最小化框架的图割理论进行图像分割已成为近年来的一个研究热点。它的优势包括它的全局最优求解能力以及结合了多种知识的统一图像分割框架。在此基础上，针对不同应用场景，人们提出了多种变种分割方法[1-6]。

在Interactive Graph Cuts算法[x]基础上，Rother等人提出了GrabCut算法。改进的内容包括：（1）采用高斯混合模型（Gaussian Mixture Model，GMM）替代灰度直方图，支持彩色图像分割；（2）在GMM参数估计过程中，采用多次迭代算法替代一次最小估计；（3）算法采用非完全标号（incomplete labelling）的方式，降低了用户的交互工作量。

图割理论

图像分割即把像素标为前景/背景，是典型的二元标号问题。首先构造一个能量函数，用于计算像素的标号值。之后借助网络流理论，把标号问题转化为最大流/最小割问题解决。

设为一无向图，是定义在边集上的容量函数：，则无向图及其边集上的容量函数构成一个网络，记作，其中，是网络的源点，是网络的汇点。借助最小化能量函数把顶点集划分为两个顶点集，，分别与源点和汇点相连，(,)。

 （2.18）

其中，为的一个标号，；为数据项，用来衡量和所观察到的数据的不一致性；为光滑项，用来衡量非分片光滑的程度；为相互作用的相邻顶点对。

网络中的顶点对应图像中的像素，网络边上的容量对应像素特征之间的差异或者相似度，即用网络表示一个图像。图像分割能量函数的最小值对应网络的最小割。根据最大流/最小割定理，网络的最大流与最小割是等价的。最终，能量函数的最小化问题转化为网络的最大流问题。

GrabCut算法

GrabCut算法将图像分割问题定义为：对图像的每个像素点，求其标号值，其中，1代表前景；0代表背景。代表前景/背景的概率密度模型。借助图割理论，图像分割问题可表示为

 （2.19）

GrabCut算法基本步骤包括：首先，由用户在前景周围画一个矩形，通过“非完全标号”的方式来标定图像的背景区域（Trimap Background）和未知区域（Trimap Unknown），矩形框外为背景区域，矩形框内为未知区域。将未知区域内的像素标注为1，背景区域内的像素标注为0，目标区域（Trimap Foreground）设为空。利用用户标定的背景区域和未知区域分别初始化前景/背景GMM。然后将把未知区域划分为前景/背景两类，对新划分的前景/背景像素进行切割，更新GMM。在新的GMM参数下继续对未知区域进行划分，迭代直至满足收敛条件，最终确定GMM参数。利用最终的GMM对未知区域进行一次切割，最终得到前景图像。

# Android平台相关技术

2008年7月23日，Android 1.0正式发布，Android 1.0是Android的一个商业版本。2015年10月5日，Android 6.0正式发布。目前，Android已发展出Android Wear、Android TV和Android Auto等分支平台。其中，Android Wear专为智能手表等可穿戴设备设计；Android TV专为智能电视设计；Android Auto专为汽车设计。Android系统一直处于快速发展之中。2013年7月3日，Google宣布，全世界激活的Android设备已经超过10亿台。2015年7月，Google宣布，全世界访问网络的智能手机和平板中59.1% 搭载了Android系统，中国市场的比例为68.3%，美国市场的比例是40.78%。

## 3.1 Android平台的发展

安迪·鲁宾（Andy Rubin）被称为“Android之父”。他于2003年创立了Android科技公司（Android Inc.）。2005年7月11日，Android科技公司被Google收购，成为Google的全资子公司。安迪·鲁宾加入Google，继续基于Linux内核的Android系统的研发。2007年11月5日，Google牵头成立了开放手持设备联盟（OHA，Open Handset Alliance），最初的成员有Broadcom、HTC、Intel、LG、Marvell、Samsung（三星）等。开放手持设备联盟成立的同一天也推出了第一部搭载Android系统的智能手机。小米、华为、索尼和ARM等也于2008年12月9日加入OHA。2007年，AOSP（Android Open Source Project）项目启动，它是Android源码开发更新项目。众所周知，Android系统是开源的，代码公开免费。其中，Android的大部分源码以Apache开源条款2.0发布，而Linux内核部分继承GPL v2开源。Android发展至今，比较重要的版本包括：Cupcake (1.5)、Donut (1.6)、Eclair (2.0–2.1)、Froyo (2.2–2.2.3)、Gingerbread (2.3–2.3.7)、Honeycomb (3.0–3.2.6)、Ice Cream Sandwich (4.0–4.0.4)、Jelly Bean (4.1–4.3.1)、KitKat (4.4–4.4.4, 4.4W–4.4W.2)、Lollipop (5.0–5.1.1)、Marshmallow (6.0–6.0.1)。

图3.1 Android系统架构

## 3.2 Android平台的组成

如图3.1所示，Android系统大致可以分为五层，由上到下依次为应用程序层、应用程序框架层、函数库层、Android运行层和Linux内核层。各层各司其职，上层使用下层提供的服务，屏蔽本层及下层的差异，为上层提供统一的服务。分层结构具有高内聚、低耦合的优点。

### 3.2.1 Linux内核

Linux核心层是Android的最底层，也是Android的基础。Andriod依赖Linux2.6内核提供包括进程管理、内存管理、网络、硬件驱动等核心系统服务。同时，Linux核心层也是硬件和软件之间的抽象层，它屏蔽硬件差异，为上层提供统一服务。

### 3.2.2 系统运行库

Android运行层的核心是Dalvik虚拟机。在Android系统中，每一个Android应用程序对应一个Dalvik虚拟机。Dalvik虚拟机在设计上也使得一台移动设备可以高效的运行多个它的实例。Dalvik虚拟机可执行文件格式是dex。dex格式是专为Dalvik虚拟机设计的压缩格式，适合内存和处理器速度有限的移动设备。开发Android应用时，Java程序通过编译生成.class文件，还需要通过SDK中提供的dx工具转化为dex文件才能在虚拟机上执行。一个dex文件通常会包含多个class。Dalvik虚拟机依赖下层的Linux内核来实现线程和内存管理等功能。

### 3.2.3 应用程序框架

应用程序框架层是与Android开发者联系最紧密的一层，它为Android开发提供了一套服务，主要有：

（1）视图（View）

Android应用框架层为开发者提供了文本视图（TextView）、按钮（Button）、列表视图（ListView）等基本视图，也提供了线性布局（LinearLayout）、相对布局（RelativeLayout）、帧布局（FrameLayout）等布局来组织基本视图。视图以树状结构组织起来构成我们所见到的各式各样的UI界面。

（2）内容提供者（Content Providers）

内容提供者实现了一种Android应用程序之间的数据共享机制。应用程序可以访问、修改其他应用程序共享的数据，也可以共享自己的数据。

（3）资源管理器（Resource Manager）

在Android开发中，界面可以通过布局文件（xml文件）来定义。布局文件即是这里所说的资源的一种。资源是相对代码而言的，可以说Android应用程序是有代码和资源组成的。除了上述的布局文件，资源还包括图标、字符串等等。资源管理器统一管理这些资源，为开发者提供一致的资源获取接口。

（4）通知管理器（Notification Manager）

在Android开发中，经常有在状态栏通知中心显示通知的需求。通知管理器是应用程序框架层为Android开发提供的服务，Android开发者只需调用通知管理器的API即可实现通知推送的功能。

### 3.2.4 应用程序

应用程序层在Android系统的最上层，面向用户提供服务。应用程序层主要包括系统应用和第三方应用。我们平常使用的拨号、消息、计算器等应用均位于此层。

## 3.3 Android应用程序开发技术

Android SDK为Android开发者提供了四大应用程序组件，是Android开发不可或缺的强大工具，分别是Activity、Service、Broadcast、ContentProvider。

### 3.3.1 Activity

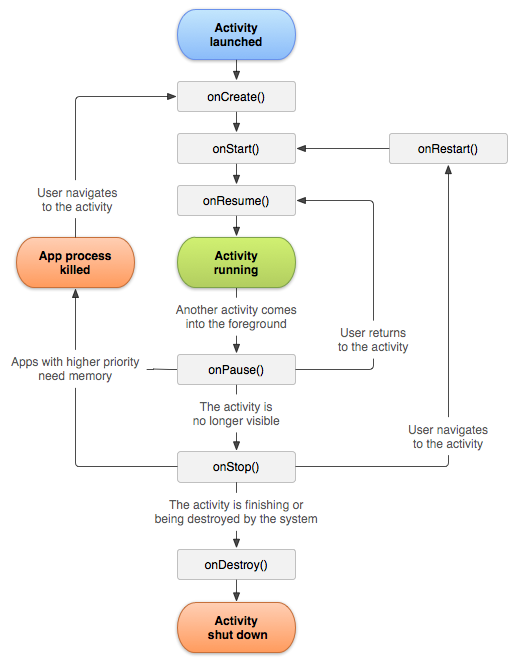
Activity称为活动，是应用程序的表示层，每个程序包括一个或多个Activity。应用程序的每个屏幕显示都通过继承和扩展Activity基类实现。在Android中，不同应用程序的Activity可以相互调用，共同协作完成一个任务。这种协作是通过任务栈机制实现的。任务栈机制将来自不同应用程序的Activity放在同一个任务栈中，对于用户来说，好像一个应用程序完成了这个任务。任务栈本身仅是Activity的组织结构，无法设置参数。任务栈的特性通过Activity的启动参数来确定。根据不同的启动参数，可以把Activity的启动模式分为四种，分别是Standard、SingleTop、SingleTask和SingleInstance。

（1）Standard。Standard模式为默认模式，以Standard模式启动一个Activity，总是创建一个新的Activity实例，并把这个示例置于当前任务栈栈顶。

（2）SingleTop。以SingleTop模式启动一个Activity时，将首先判断当前任务栈栈顶是不是该Activity的实例，如果不是，接下来的操作与Standard模式相同；如果是，调用栈顶Activity实例的onNewIntent方法，而不再创建一个新的实例。

（3）SingleTask。以SingleTask模式启动一个Activity时，将首先判断当前任务栈中是否存在该Activity的实例，如果不存在，接下来的操作与Standard模式相同；如果存在，调用Activity实例的onNewIntent方法，并销毁其之上的Activity，让该Activity实例居于任务栈栈顶，接受用户响应，而不再创建一个新的实例。

（4）SingleInstance。以SingleInstance模式启动一个Activity时，将首先查找系统中是否存在该Activity的实例，如果不存在，创建一个新的任务栈并创建一个新的Activity实例置于新的任务栈的栈底，并且保证此任务栈自始至终只有这一个Activity；如果存在，调用Activity实例的onNewIntent方法。

图3.2 Activity生命周期

Activity从创建到销毁可能经历多种状态，在不同状态之间切换时，将调用不同的回调方法（CallBack），如图3.2所示。Activity生命周期中包含四种状态，分别是活动状态、暂停状态、停止状态和非活动状态。

（1）活动（Active/Running）状态。处于屏幕前台，可以响应用户操作的Activity处于活动状态。此时，它处于任务栈的栈顶。同一时刻至多只有一个Activity 处于活动状态。

（2）暂停（Paused）状态。对用户仍可见但失去焦点，无法响应用户操作的Activity处于暂停状态。比如，一个Activity被另一个Activity部分遮挡时，它处于暂停状态。一个Activity透明或者尺寸小于屏幕尺寸，会部分遮挡启动它的Activity。暂停的Activity仍处于存活状态，它保留着所有的状态和成员信息。Activity的UI视图也保持着与窗口管理器（WindowsManager）的连接。仅当系统资源极度紧张时，暂停状态的Activity才可能会被系统销毁以回收资源。

（3）停止（Stopped）状态。被另一个Activity完全遮挡的Activity处于停止状态。停止的Activity仍然保留着所有的状态和成员信息。但Activity的UI视图已与窗口管理器（WindowsManager）断开了连接。当系统资源紧张时，停止状态的Activity往往会被系统销毁以回收资源。

（4）非活动（Dead）状态。尚未被启动、已经被终止、已经被系统销毁的Activity处于非活动的状态。非活动的Activity不保存任何状态或者信息。

### 3.3.2 Service

图3.3 Service生命周期

### 3.3.3 BroadcastReceiver

在Android中，广播分为两种，系统广播和自定义广播。常用的系统广播包括启动广播、低电量广播、网络可用广播等等。自定义广播可以用于APP内部通信。自定义广播

### 3.3.4 ContentProvider

ContentProvider为存储和获取数据提供了统一的接口。无论应用程序是否启动，其它应用程序都可以通过接口来操作该应用程序的内部数据，包括增加数据、删除数据、修改数据、查询数据等。一般来说ContentProvider是单例模式的，当多个应用程序通过ContentResolver操作ContentProvider的数据时，ContentResolver调用的数据将会委托给同一个ContentProvider处理。

# 4 基于Android平台的服饰图像搜索系统实现

基于上文所述的相关理论和研究成果，本文设计并实现了一个基于Android平台的服饰图像搜索系统。本章将详细介绍系统的架构设计及各个模块功能的实现过程。

本论文实现的服饰图像搜索系统共包含以下这几个模块：搜索图像输入模块：基于内容的图像搜索需要用户输入搜索图片，而不是输入关键词。本文实现的图像搜索系统支持两种输入图片的方法。一种是直接利用手机进行拍照，自己实现拍照功能，而非调用系统的拍照服务。用户拍照后将照片以临时文件的形式存储，当做输入图像。第二种方式是直接读入手机当中的图片进行检索。本文实现了一个自己的图片选择器，用户可以选择自己想要搜索的图片。这些模块的具体实现过程将在到小节中进行详细描述。

Android客户端开发使用Android Studio集成开发环境。服饰图像数据库及特征池的建立以及服务器端开发使用Pycharm集成开发环境。

## 系统架构

本文系统由三个部分组成，一是Android客户端；二是图像集及特征池；三是服务器端。系统架构如图4.1所示。搜索流程如下：移动端以拍照或从图库选图的方式获得待搜索图片并上传至服务器；服务器提取待搜索图片特征，计算与特征池中特征距离，得到相似度最高的图像列表返回给客户端；移动端得到相似图像列表并向用户展示，用户可以收藏、分享、购买这些服饰；特征池的建立是离线的，首先使用网络爬虫爬取各大电商平台的服饰图像建立服饰图像集，然后提取特征建立特征池。



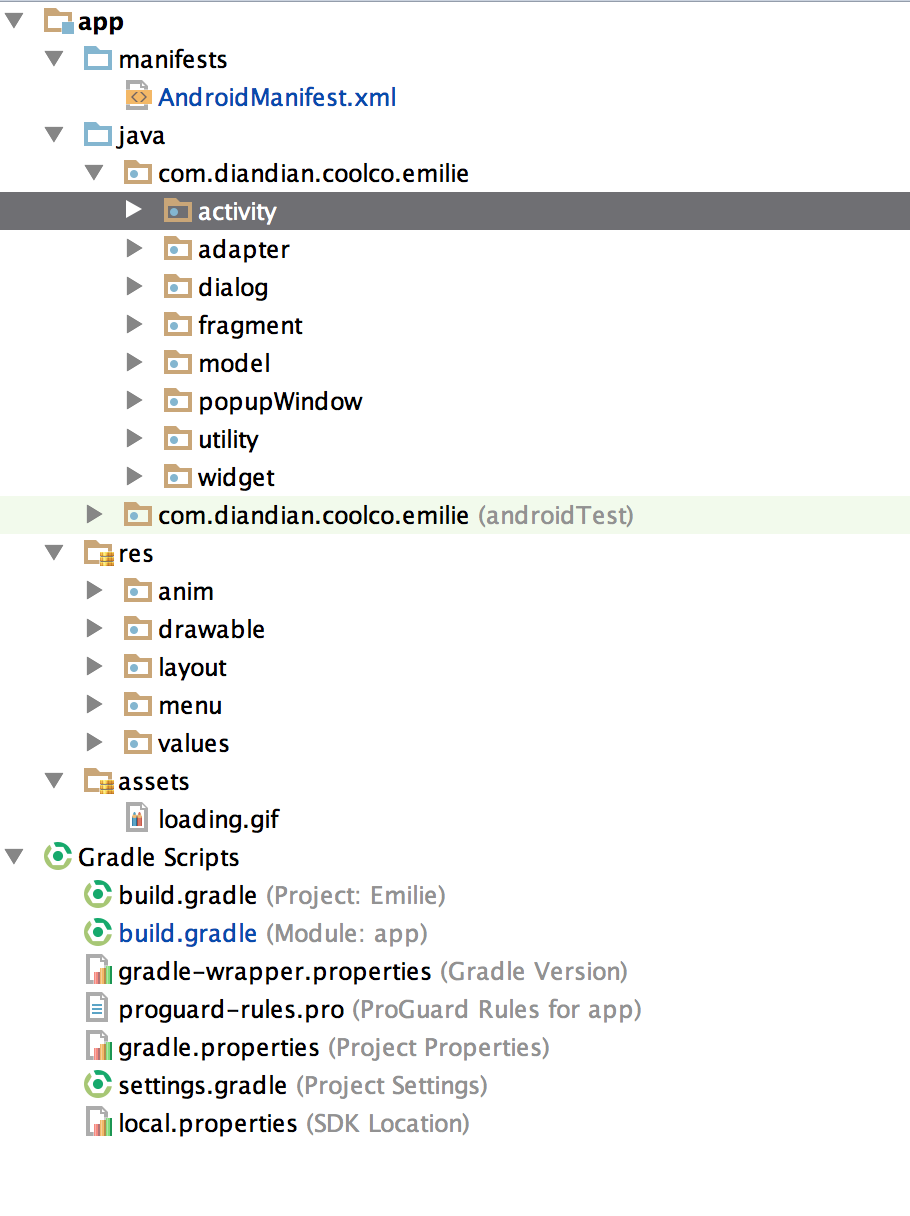
图4.1 系统架构

## 4.2 Android客户端实现

Android客户端主要由查询图像的获取和搜索结果的展示的两个模块组成。查询图像的获取有两种方式，一是通过摄像头拍照得到；二是通过手机图库选择得到。搜索结果有两种展示模式，一是概要展示模式，将结果图片缩略图以瀑布流的形式展示给用户；二是详情展示模式，将结果图片原图以翻页的形式展示给用户。在详情展示模式下，支持收藏、分享、购买等操作，也可以查看服饰品牌、价格、材质等相关信息。除查询图像的获取和搜索结果的展示的两大主要模块之外，本文的Android客户端还包括新手引导模块、收藏模块、用户模块、通用模块等辅助模块。功能结构如图4.2所示。

图4.2 Android客户端功能结构

硬件要求：摄像头；系统要求：Android 4.0 以上；权限要求：访问摄像头权限（CAMERA）、写外部存储权限（WRITE\_EXTERNAL\_STORAGE）、访问网络状态权限（ACCESS\_NETWORK\_STATE）、访问网络权限（INTERNET）、挂载外部文件系统权限（MOUNT\_UNMOUNT\_FILESYSTEM）



### 4.2.1 逻辑架构

在逻辑上，本文APP可划分为四个层级，分别为界面层、核心层、接口层、模型层。界面层负责UI展示；核心层负责处理业务逻辑；接口层负责与服务器进行交互；模型层定义所有的数据模型。如图4.3所示，界面层依赖核心层、模型层，核心层依赖接口层、模型层，接口层依赖模型层，模型层不依赖任何层级。在构建项目，依照模型层、接口层、核心层、界面层的顺序依次构建。



图4.3 Android客户端逻辑结构

模型层横跨所有层级，定义模型类，封装项目用到的数据。在本文项目中，主要定义了Image类，封装服饰图像的图像地址、购买地址、品牌、价格、材质等信息。

接口层的主要的功能是封装网络请求。接口层首先将请求封装好发送给服务器，然后将网络访问得到的数据转化为模型层定义的数据类型返回给上层的核心层。在本文项目中，客户端和服务器以json格式交互数据。发送请求使用HttpURLConnection系统类。请求结果处理上，本文利用Gson库将服务器返回的json格式服饰图像信息直接映射成为Image类型。

核心层介于接口层和界面层之间，主要处理业务逻辑。向上，向层界面层提供待展示的数据。向下，调用接口层获取服务器的数据。由于网络访问是异步的，此层暴露给上层界面层的API都有一个CallBack参数。界面层向核心层请求数据的时候，核心层会立即返回，在后台调用接口层向网络层请求数据，在得到数据并处理后通过CallBack返回给界面层。CallBack模式的引入可以避免网络请求阻塞UI线程造成的ANR(Application Not Responding)错误。

界面层处于最上层，负责UI的展示。根据不同类型划分，主要分为以下几个包：activity、adapter、fragment、util、view。其中activity、adapter、fragment各自都有一个基类，做统一的处理，比如定义了一些共用的常量、对象和方法等。

### 4.2.2 图像输入模块

客户端支持拍照搜索，从本机图片中选取图片搜索两种搜索方式。点击拍照搜索将跳转到自定义的拍照页面进行拍照，点击选图搜索将跳转到自定义图片选择页面进行选图。尽管图像来源不同，拍照和选图都将返回一个图片的绝对路径，供后续搜索模块使用。

|  |  |
| --- | --- |
| 拍照搜索1 | 图库 |
| 图4.4 拍照 | 图4.5 图片选择器 |

#### （1）拍照

拍照功能可以调用系统服务完成。本文中的客户端采用自定义页面进行拍照的原因主要有两个。一是项目需要在拍照预览界面加矩形框，提醒用户将搜索目标置入矩形框内，进而提高搜索准确度。而系统服务并不支持界面的定制。二是不同版本的Android系统拍照界面也不同，一些版本的系统拍照界面十分简陋。。考虑到了用户体验的一致性以及界面风格的统一，本文项目的拍照功能由自己实现，而非调用系统服务。拍照页面如图4.4所示。

页面背景为摄像头的实时预览（preview）。在实时预览上层有一个矩形框，框内透明，框外蒙着一层灰色，和矩形框上部的文字一起提醒用户将搜索目标置入矩形框内然后再拍照。矩形框和蒙层由一个自定义视图（Custom View）实现，这个自定义视图不借助Android SDK中的控件，显示效果完全由自己绘制。

页面底部为拍照按钮。点击拍照按钮进行拍照。然后播放转场动画（Transition Animation）——拍照按钮向下坠落然后消失，取消、确定、重拍按钮向上升起然后出现。转场动画的存在让客户端的页面跳转更加友好，没有动画的过渡，用户会对界面的变化感到困惑（come from nowhere）。拍照后将出现取消、确定、重拍三个按钮。如果对拍照结果满意，点击确定按钮，跳转到搜索结果展示页面。如果对拍照结果不满意，点击重拍按钮。点击取消，跳转到搜索主界面。

如果点击了确定按钮，客户端将立即跳转到搜索结果展示页面，在此页面展示进度条（Progress Bar）提醒用户等待，避免出现应用程序无响应错误（ANR，Application Not Responding）。点击了确定按钮，拍照页面将把拍照所得图像做旋转处理并存储为临时图像文件并得到临时图像文件的路径的工作交给整个应用程序共用的线程池去执行，执行结束后通过EventBus异步的通知搜索结果展示页面。数据处理和页面跳转相互分离。之所以先跳转到搜索结果展示页面再通过广播把临时图像文件路径传递给搜索结果展示页面是为了减少一次进度条的展示。直接的做法是在拍照页面展示一次进度条，让用户等待APP存储临时图像文件，得到临时图像文件的路径后，夹带路径信息，跳转到搜索结果展示页面，展示一次进度条，让用户等待服务器返回搜索结果。但是这样处理，将连续展示两次进度条，而且两次之间用户没有任何操作，严重影响用户体验。通过上述的方法，可以很好的合并进度条，提升用户体验。

#### （2）图片选择器

图像选择也可以通过调用系统服务实现。之所以项目中选择自己实现图像选择器主要原因有两个。一是，Android系统图像选择服务以文件夹的形式组织图片，选中一张图片至少需要两步（选中相册、选中图片），并且选择过程中涉及界面的跳转，不够简单便捷。项目中的自定义图像选择器把图像文件夹选择和图像选择置于同一页面中，图像文件夹的选择以底部弹窗（Popup Window）的形式出现。自定义图像选择器还增加了所有图像这一逻辑文件夹，并按时间倒序排列图像，所有图像这一逻辑文件夹也是图像选择器的默认文件夹，所以，用户打开自定义的图像选择器就可以看到最近拍摄、保存的所有照片，点击即可选择目标图片。

图片的展示采用GridView布局。Android设备内存有限，在展示大量图像的时候容易导致OOM（Out of Memory）错误。造成OOM的主要原因是图像文件在展示之前需要解压（decode）为位图（Bitmap），而位图非常占用内存，在展示大量图片的时候容易导致OOM。针对这一问题，本文项目做了如下优化：

1．使用缩略图代替原图。在图像选择器页面中，每张图片只占一格，其显示尺寸远远小于其实际尺寸。所以在解压图片的时候，可以一定的采样率去解压，然后进行伸缩变换，得到的位图的尺寸将与显示的尺寸的匹配，而占用内存远远小于将图片直接解压做占用的内存。

2．缓存缩略图。用户在选择图片的时候，经常会上下滑动。如果采取图片被滑出屏幕时被回收（recycle），滑入屏幕时重新解压（decode）的策略，有两个弊端。一是，解压是高计算量操作，频繁解压将大量消耗计算资源。二是，用户在上下滑动的时候，大量的图片被滑出屏幕和滑入屏幕，大量的位图将被创建和回收，导致频繁的GC（垃圾回收），将造成内存抖动，进而导致界面卡顿。

本文项目中，利用LRU策略缓存图片缩率位图，在现实的图片时，将首先到缓存查找对应的位图。如果命中，直接显示。如果未命中，从原图片文件中采样解压、缩放得到与显示尺寸匹配的位图，显示并加入缓存。图片在刚刚被滑出屏幕又被滑入屏幕时，不会被回收，也不需要重新从原图片文件中解压得到，很好的解决了上文描述的两个问题。

#### （3）图像裁剪

用户拍照后，APP参照拍照预览上的矩形框自动对拍照所得图像进行裁剪。使用图片选择器选择图片后，跳转到裁剪页面进行裁剪。裁剪页面如图所示。用户可以拖动矩形框的边缘改变矩形框的大小，拖动其它位置改变矩形框的位置，直到矩形框将服饰框住。点击确定即可预览图片。在搜索列表展示页，也有裁剪页面的入口，方便用户再次搜索。

|  |  |
| --- | --- |
|  |  |
| 图4.6 用户引导 | 图4.7 裁剪图像 |

### 4.2.3 网络模块

Android客户端的网络模块主要包括上传查询图像文件和解析服务器返回数据两个部分。Android客户端通过网络模块上传查询图像文件，服务器将返回相似服饰信息数据。相似服饰信息中最重要的信息是相似服饰图像的URL。图像搜索时，首先返回的相似服饰图像的URL。在搜索结果展示模块，客户端再通过此URL去加载相似服饰图像。这么做的目的一是可以加快响应速度，提升用户体验。二是可以实现惰性加载，节省流量。一次搜索可能产生50个相似服饰图像结果，而用户并不一定会把它们全部浏览完。搜索时返回相似服饰图像的URL，用户在浏览搜索结果时，浏览到哪一件相似服饰时再去下载该相似服饰的图像可以减少下载不必要的图像而浪费流量。

（1）上传查询图像文件

本文Android客户端使用HTTP POST方法上传查询图像文件。在实现上主要借助Android SDK中的HttpUrlConnection类，HttpUrlConnection实现了HTTP协议，可以用于发送请求，获得响应。发送请求时，首先设置报文头中的“Content-Type”属性为 “multipart/form-data”，然后将位图（Bitmap）转化为字节数组，然后填充到请求报文体中，发送报文，实现文件上传。关键代码如下：

...

URL url = new URL(actionUrl);

con = (HttpURLConnection) url.openConnection();

/\* 允许Input、Output，不使用Cache \*/

con.setDoInput(true);

con.setDoOutput(true);

con.setUseCaches(false);

/\* 设置传送的method=POST \*/

con.setRequestMethod("POST");

/\* setRequestProperty \*/

con.setRequestProperty("Connection", "Keep-Alive");

con.setRequestProperty("Charset", "UTF-8");

con.setRequestProperty("Content-Type",

"multipart/form-data;boundary=" + boundary);

ds = new DataOutputStream(con.getOutputStream());

ds.write(jsonSB.toString().getBytes());

if (imageNames != null && bitmaps != null) {

for (int i = 0; i < imageNames.size(); i++) {

String picName = imageNames.get(i);

Bitmap bitmap = bitmaps.get(i);

ds.writeBytes(twoHyphens + boundary + end);

ds.writeBytes("Content-Disposition: form-data; "

+ "name=\"upload\_image\";filename=\"" + picName + "\""

+ end);

ds.writeBytes(end);

ds.write(bitmap2Bytes(bitmap));

ds.writeBytes(end);

ds.writeBytes(twoHyphens + boundary + twoHyphens + end);

}

}

ds.flush();

/\* 取得Response内容 \*/

...

（2）解析服务器返回的相似服饰数据

服务器返回的数据使用JSON格式。JSON是目前应用非常广泛的数据交换格式，它具有易于阅读，信息比例高等优点。并且各个语言对它的支持都很好，封装和解析都十分方便。在本文中，服务器端使用Python标准模块json对数据进行封装和解析，Android客户端使用Google的Gson开源库对数据进行封装和解析。

图 服务器返回JSON数据示例

图 客户端相似服饰图像模型类

使用Gson可以方便的实现Java对象和JSON数据之间的相互转换，并且速度较快。客户端接受到服务端的响应之后，从响应中提取响应字符串。响应字符串是一个JSON数组。以JSON数组和预先定义好的相似服饰图像模型类为输入，Gson即可将JSON数组转换为相似服饰图像模型类的对象数组。服务器返回的JSON数据样例如图所示，包含的键（Key）有相似服饰图像下载地址（downloadUrl）、服饰品牌（brand）、服饰名称（productName）、服饰货号（productId）、服饰材质（material）、服饰价格（price）、服饰优惠价格（specialPrice）、服饰购买地址（buyUrl）。相似服饰图像模型类如图所示，相似服饰图像类的成员变量与服务器返回的JSON对象的键一一对应，以便使用Gson进行转换。

### 4.2.3 结果展示模块

#### （1）搜索结果列表展示

搜索结果列表展示页主要功能是让用户上下滑动快速浏览系统搜索结果，并选择感兴趣的图片点击跳转到搜索结果详细展示页。此外，页面还提供继续裁剪待搜索图片、搜索结果数量统计、搜索耗时统计等辅助功能。搜索结果以瀑布流的形式分两列展示给用户，从上到下相似度依次降低。搜索结果图片的加载使用开源的图像加载组件Fresco。Fresco 是一个强大的图片加载组件。

解压后的图片，即Android中的Bitmap，占用大量的内存。大的内存占用势必引发更加频繁的GC。在5.0以下，GC将会显著地引发界面卡顿。在5.0以下系统，Fresco将图片放到一个特别的内存区域。当然，在图片不显示的时候，占用的内存会自动被释放。这会使得APP更加流畅，减少因图片内存占用而引发的OOM。

Fresco的设计遵循MVC模式。DraweeView继承自View，负责图片的显示，相当于MVC中的V（View）。DraweeHierarchy用于组织和维护最终绘制和呈现的图片，比如上文代码中占位图片和按下蒙层图片的设置均通过DraweeHierarchy实现，相当于MVC中的M（Model）。DraweeController负责和image loader交互（默认是Fresco中的image pipeline），实现对所要加载的图片的更多控制，比如，图片加载失败情况下单击重新加载。相当于MVC中的C（Controller）。服务器返回的相似图片URL交给Fresco去加载，用户点击感兴趣的图片跳转到图片详情页面。

|  |  |
| --- | --- |
| 裁剪后的搜索结果 | 结果详情 |
| 图4.6 搜索结果列表 | 图4.7 搜索结果详情 |

#### （2）搜索结果详细展示模块

图片详情页面全屏展示图片。为了给用户沉浸式体验，默认情况下，顶部ActionBar（工具条）不可见。单击图片可显示ActionBar。ActionBar的更多（overflow）菜单中包括收藏、分享、保存到相册、去购买、查看服饰信息等菜单项，如图4.7所示。

收藏操作将把当前图片URL及对应的服饰信息保存到本地数据库中，之后可在收藏页面查看。具体实现见下文收藏模块。

保存到相册功能是把服饰图片保存到本地相册。

去购买功能给用户提供了购买的渠道。购买地址是在使用网络爬虫抓取服饰图片时一并抓取并保存的。用户点击去购买菜单，APP将打开相应的购买网页，如图4.8所示。打开网页操作在APP内嵌的浏览器完成，并没有调用系统浏览器或第三方浏览器。内嵌浏览器保证了APP在各个Android版本的手机上使用体验一致。本文项目中实现的内嵌浏览器还提供了复制链接、用浏览器打开链接和分享三个辅助功能。其中，分享的内容并不仅限于服饰的购买地址，还包括APP相关信息。内嵌浏览器基于Android SDK中的标准组件WebView实现。Android从4.4起提供基于Chromium实现的WebView。此前WebView基于WebKit实现。WebKit提供网页解析、布局和绘制以及JS运行等基础功能。Chromium在WebKit基础上为WebView提供进程、线程和渲染等基础构架。因此基于Chromium实现的WebView更好地提供了网页浏览功能。

点击查看服饰信息菜单将弹出对话框（Dialog）用于展示服饰信息，如图4.9所示。展示的服饰信息包括服饰品牌、服饰价格等信息。

|  |  |
| --- | --- |
| 去购买 | 查看服饰详情 |
| 图4.8 购买服饰 | 图4.9 查看服饰信息 |

分享操作的实现的基础是系统的分享功能。Android有着完善的应用间通信机制，其与Unix的开发式工具链哲学一脉相承，是一个完整的体系。在 Android 中，所有的应用都向系统注册其所持有的组件和每个组件能够处理的任务。这样系统对于每个应用的每个组件能完成何种任务都了如指掌。当一个应用需要启动其它组件完成任务时，只需向系统发送一个消息，即可调用系统中可以完成该项任务的组件，而应用本身无需关心具体哪个应用的哪个组件最终处理了这个消息。这样跨应用的内容分享变得很方便。在本文项目中，要分享的是纯文本，其中包含了图片的URL以及服饰相关信息。

收藏、分享、查看服饰信息等命令除了通过点击相应菜单项触发以外，还可以由手势（gesture）触发。收藏通过下拉触发，（与常见的下拉刷新操作类似），如图4.11所示。如果已收藏，下拉则取消收藏，添加收藏与取消收藏均有文字提示。分享通过上拉触发，如图4.10所示。查看服饰信息通过双击触发。双击显示服饰信息的过程为3D翻页动画，感觉上服饰信息就在图片背后。翻页后，显示服饰信息，图片在背面。再次双击，再次翻页，图片回到正面，服饰信息被隐藏在背面。这种交互方式源于生活经验，也易于理解。手势操作的优势是全局触发，相比菜单项必须点击某一矩形区域而言，手势更加方便。除此之外，众所周知移动设备屏幕尺寸有限，使用手势代替菜单可以使界面简洁，释放大量视觉空间。并且，手势操作被用户接受并熟悉后，会变成用户的直觉化操作，高效而便捷。

在图片详情页面，用户可左右滑动来查看上一张图片和下一张图片。左右滑动浏览大图的实现使用的是Android SDK中的标准组件ViewPager。

|  |  |
| --- | --- |
| 上拉分享 | 收藏 |
| 图4.10搜索结果详情页上拉进行快速分享 | 图4.11 搜索结果详情页下拉快速收藏服饰 |

### 4.2.4 其它辅助模块

#### （1）新手引导模块

上文中提到，本项目客户端中较多的使用了手势命令，为此本项目客户端也配备了新手引导模块。在用户第一次打开相关页面的时候，利用动画引导用户相关手势操作。引导动画相比文字、静态图片更加形象，最大程度的降低了用户的学习成本。在第一次打开相关页面的时候引导该页面可以使用的手势操作，可以把操作学习分散到各个实际场景，更有利于用户接受和学习。引导动画的实现使用Android中的属性动画（ValueAnimator）。用SharedPrefrence记录用户是不是第一次打开此页面。

#### （2）收藏模块

本文项目客户端支持用户收藏感兴趣的服饰。用户收藏服饰时，APP将服饰信息存入本地数据库。用户浏览自己的收藏时，APP从本地数据库读取数据并显示。本地数据库使用SQLite。SQLite是一个进程内的库，实现了自给自足的、无服务器的、零配置的、事务性的 SQL 数据库引擎。与其他数据库一样，SQLite 引擎不是一个独立的进程，可以按应用程序需求进行静态或动态连接。SQLite 直接访问其存储文件。客户端使用ORMLite简化数据库操作。ORMLite是一个轻量级的Java对象关系映射持久层框架。支持包括 MySQL、Postgres、Microsoft SQL Server、H2、Derby、HSQLDB和SQLite等在内的数据库。提供灵活的QueryBuilder来构建复杂的数据查询以及强大的抽象DAO（Database Access Object）类。

|  |  |
| --- | --- |
| 上拉分享 | 我的收藏 |
| 图4.12引导页面 | 图4.13 收藏页面 |

#### （3）用户模块

用户模块支持用户设置自己的头像和昵称。系统将统计用户的搜索次数、收藏次数，并由此计算成长值（积分）。系统将按照成长值为用户颁发勋章，鼓励用户使用APP，也让APP的使用具有一定趣味性。

#### （4）通用模块

通用模块包括“欢迎页”、“用户协议”、“常见问题”、“告诉朋友”、“意见反馈”、“版本升级”、“给我们评分”、“开源许可”、“关于我们”。其中，“欢迎页”用于宣传介绍APP的核心特色功能，在APP首次打开时也将显示；“用户协议”声明用户需要遵守的行为规范，列举禁止事项（比如反向编译本客户端）；“常见问题”将列举用户反馈最多的问题并作出解答，常见问题页面是一个静态网页，由内嵌浏览器打开，易于时常更新；“告诉朋友”具有分享APP的作用；“意见反馈”收集用户的意见和建议，指导APP功能改进、体验提升；“版本升级”即APP内部升级；“给我们评分”提供用户到应用商店为APP打分的途径，提高APP的传播率和影响力；“开源许可”声明APP开发中使用第三方开源库，本文APP使用了EventBus、Fresco、StaggerGridView三个开源库；“关于我们”介绍系统开发成员。通用模块的存在使得APP功能更加完整，为后期的推广以及商用打下坚实的基础。

|  |  |
| --- | --- |
| 用户信息 | 设置 |
| 图4.14用户页面 | 图4.15 设置页面 |

## 4.3 图像集及特征池的建立

服饰图像集的建立和服饰图像特征池的建立是离线的。

### 4.3.1 图像集的建立

本文系统采用的服饰图像来自国内主流电商网站（唯品会、亚马逊等），通过网络爬虫程序得到。网络爬虫不仅爬取服饰图像本身，也爬取服饰相关信息，比如品牌、价格、材质、货号。图4.16给出图像集的示例图像。

图4.16 图像集示例图片

网络爬虫程序使用Python实现，借助pyquery库。pyquery利用lxml实现快速的xml和html操作。pyquery支持在xml文件上做类似jquery查询，API设计和jquery非常相似，可以非常方便的定位HTML标签并提取该标签的属性值，比如输入参数“div.cat-item-pic”可以定位到class为cat-item-pic的div。这意味着，利用pyquery可以精准抓取服饰图像，而不是将服饰购买页上的所有图像都抓取下来之后过滤到非服饰图像。相比使用正则表达式在网页源码文本搜索感兴趣的字符串，借助pyquery实现网络爬虫更加简洁高效。

本文利用网络爬虫建立的图像集共包含36960张图片，包括1100 × 1390、234× 295两种尺寸，大小为5941MB，如表4.1所示。

表4.1本文项目使用的图像集信息

|  |  |
| --- | --- |
| 图像数量 | 36960 |
| 图像尺寸 | 1100 × 1390、234× 295 |
| 图像集大小 | 5941MB |

网络爬虫程序爬行到一件服饰的购买页时，不仅抓取服饰的图片，服饰的品牌、商品名、货号、材质、价格、优惠价格以及当前购买页的地址都将一并被抓取下来。服饰的相关文本信息以JSON的格式存放在文本中。本文爬取服饰文本信息示例如图所示。

图 网络爬虫采集服饰信息示例

本文共爬取19161件服饰的文本信息，文本信息大小为10.3MB，如表所示。

表 网络爬虫采集服饰信息规模

|  |  |
| --- | --- |
| 服饰数量 | 19161 |
| 服饰信息大小 | 10.3MB |

### 4.3.2 特征池的建立

服饰图像特征池具体是指存储服饰图像特征的数据库，本文系统使用MySQL数据库。特征提取使用Python实现，借助OpenCV库。OpenCV通过BSD协议开源。OpenCV旨在高效计算，并专注于实时应用。底层基于C/C++，OpenCV可以利用多核进行并行计算。如果启用OpenCL，OpenCV可以借助各式各样的计算平台的硬件加速。

## 4.4 服务器端实现

服务器端主要功能是特征距离的计算与排序。服务器用Python语言实现，并部署在Django架构之上。

服务器层在整体架构中占有至关重要的地位。本文系统的服务器端分为两个层次，上层是Web服务层，下层是业务核心层。

Web 服务层的功能是：接收客户端的请求、调用业务核心层为查询图像提取特征、读取数据库中存储的图像特征、计算相似度并排序、为用户返回查询结果。Web服务模块看似处理的事务很多，实际上，关键的部分均是调用系统核心部分的接口实现的，比如为查询图像提取特征、查询图像与目标图像的相似度度量。

业务核心层处于服务器端的下层，供Web服务层调用，功能相对独立，与其他模块无耦合。这个“核心部分”主要包括：

### 4.4.2 Django简介

Python是一种面向对象、直译式的计算机程序语言，具有近二十年的发展历史。它包含了一组功能完备的标准库，能够轻松完成很多常见的任务。它的语法简单，与其它大多数程序设计语言使用大括号不一样，它使用缩进来定义语句块。Python经常被用于Web开发。比如，通过mod\_wsgi模块，Apache可以运行用Python编写的Web程序。使用Python语言编写的Gunicorn作为Web服务器，也能够运行Python语言编写的Web程序。Python定义了WSGI标准应用接口来协调Http服务器与基于Python的Web程序之间的沟通。一些Web框架，如Django、Pyramid、TurboGears、Tornado、web2py、Zope、Flask等，可以让程序员轻松地开发和管理复杂的Web程序。Python对于各种网络协议的支持很完善，因此经常被用于编写服务器软件、网络蠕虫。第三方库Twisted支持异步在线编写程序和多数标准的网络协议（包含客户端和服务器），并且提供了多种工具，被广泛用于编写高性能的服务器软件。另有gevent这个流行的第三方库，同样能够支持高性能高并发的网络开发。

本文系统服务器采用Python Django架构。Django是基于Python的开源Web框架。它有部署快速、安全性高、伸缩性强三个主要特点。

Django以MTV（Model-Template-View）模型架构，其中Model为数据模型，与数据库操作相关；Template为模板系统，用于数据的格式化显示；View为控制器，负责调用Model、Template进行业务逻辑的处理，是MTV模型的调度中枢。

Django作为Web架构提供的功能有：

* Application的可插入（Plug-in）管理
* ORM（对象关系映射）

使用Django开发Web应用，只需用Python定义数据模型即可，数据库将由类映射得到，数据库操作也都有相应的API，不必自己写SQL语句（如果需要，也可以自己写SQL语句）。

* URL分发

在Django中，URL的分发通过正则表达式实现，简单高效。

* 模板系统

模板系统很好的分隔了数据与显示。Django的模板系统是可扩展的，能很好的满足个性化需求。

* Cache系统

Django为开发者提供了各种粒度的缓存策略，方便开发者快速开发。

图4.18 Django架构图

Python应用程序的基本单元是Module（模块），一个Python应用程序由一系列的Module组成。Module是一个Python文件，包含Python对象定义和Python语句。每个Module（模块）有自己命名空间，并可以被别的Module（模块）导入并使用。Module（模块）实现了Python语言的解耦和复用。

服务端核心Module有Grabcut Module、颜色特征提取Module、纹理特征（Uniform LBP）提取Module、SIFT特征提取Module、视觉词Module、距离计算Module、排序Module。

OpenCV是一个开源的可以跨平台运行的计算机视觉库，包含了许多图像处理和计算机视觉方面的通用算法。其中的源代码都是开源免费的代码，因此可以用于科研人员的研究领域，也可以用于商业领域。整个OpenCV包含五百多个函数，其中涉及到计算机视觉的各个领域，包括人机互动、物体识别以及机器视觉等等。

### 4.4.3 GrabCut模块

服饰图像往往包含其它无关元素，比如人脸、背景等。而我们只对图片中的服饰部分感兴趣，无关元素的存在将降低图像特征提取的效率并影响搜索的准确度。本文项目选择使用交互式图像分割方法—GrabCut提取服饰图像的前景，减少背景服饰图像特征提取的干扰。GrabCut是预处理操作，利用openCV中的方法对图像集中的所有图像进行图像分割，一定程度上过滤背景，将前景图像保存。后续的特征提取均在前景图像上进行。前景图像和源图像一一对应，前景图像用于图像搜索，源图像用于返回给客户端。OpenCV 中GrabCut初始化有两种方式：一是使用前景背景标记分别标记一部分前景像素和背景像素；二是使用一个矩形框标记未知区域和背景区域。本文项目使用第二种初始化方式。电商平台的服饰图像一般由专业摄像师拍摄，它们的构图基本一致。通过观察与实验发现，服饰图像左右两次各有20%的留白（背景），上下分别有25%的留白（模特脸部与脖子）和15%的留白（模特腿部）。矩形框的尺寸位置根据服饰图像的特点以及实验经验设置。矩形框的尺寸由图像尺寸计算获得，长为图像长度的60%，宽为图像宽度的60%，矩形框水平居于图像中间，垂直上，距离图像上边缘距离为图像长度的25%。本文项目中应用的GrabCut仅粗略的过滤了背景，在一定程度上提高搜索准确度。关键代码如下：

for filename in os.listdir(workDir):

img = cv2.imread(os.path.join(workDir, filename))

if img is None:

continue

cv2.grabCut(img, mask, rect, bgdModel, fgdModel, 1, cv2.GC\_INIT\_WITH\_RECT)

mask2 = np.where((mask == 2) | (mask == 0), 0, 1).astype('uint8')

img = img \* mask2[:, :, np.newaxis]

cv2.imwrite(os.path.join(destDir, filename), img)

|  |  |
| --- | --- |
| \\Mac\Home\Documents\Emilie\dress\103_0.jpg | \\Mac\Home\Documents\Emilie\103_0_cutted.jpg |
| 图4.19 原始图像 | 图4.20 前景图像 |

如图所示，使用GrabCut处理图像可以在一定程度上过滤模特脸部以及背景。但是本文项目中的GrabCut算法初始化矩形参数是根据经验设置，仅适用于本文项目图像集中的图像，通用性较差。

在项目中，使用GrabCut对图像集进行预处理，后续特征提取均在处理后的图像上进行。预处理操作是离线进行的，尺寸为1100\*1390的图像平均每张耗时37.95s，尺寸为234\*295的图像平均每张耗时0.16s。

### 4.4.4 颜色特征提取模块

本文项目在进行颜色特征提取时使用了第三方库PIL和标准库colorsys。PIL：Python Imaging Library，已经是Python平台事实上的图像处理标准库了。PIL功能非常强大，但API却非常简单易用。colorsys的主要功能是进行颜色空间的转换，在本项目中被用于把数字图像的RGB颜色空间转换为HSV颜色空间。依据上文理论，对h、s、v分别进行量化（compress），量化后H[0,7]、S[0,2]、V[0,2]。按照量化级数，把各个颜色分量合成为一维特征矢量，公式为。一维特征矢量W的范围是[0,71]，统计其分布直方图，然后进行归一化，即可得到图像的72维颜色特征向量。关键代码如下：

wl = [0]\*72

for count, (r, g, b) in img.getcolors(img.size[0]\*img.size[1]):

h, s, v = colorsys.rgb\_to\_hsv(r/255.0, g/255.0, b/255.0)

h = int(h\*360)

H, S, V = compress\_hsv(h, s, v)

W = 9\*H + 3\*S + V

assert W < 72

wl[W] = wl[W] + count

wl = normalize(wl)

表4.2 颜色直方图示例

|  |  |
| --- | --- |
| 原始图像 | 颜色直方图 |
| Z:\170_0.jpgZ:\129_0.jpgZ:\189_0.jpg | \\Mac\Home\Documents\cf_129_0.png\\Mac\Home\Documents\cf_189.png\\Mac\Home\Documents\cf_170.png |

表4.2展示了三张原始图片以及它们的颜色直方图，直观上来看，第一张示例图片与第二张示例图片颜色更加接近，而与第三张示例图片差距较大。从表中可以看到第一张示例图片的颜色直方图与第二张图片的颜色直方图形状更加相似，而与第三张示例图片相差较大。通过计算，第一张示例图片的颜色特征与第二张图片的颜色特征之间的欧式距离的平方为0.0455；第一张与第三张之间的值为0.4332。从示例中得知，本文选取的颜色特征模型可以较好的区分图像之间的颜色差异。

利用上述代码对图像集中的36960张图片进行颜色特征提取，总计耗时33.21分钟（1992.88s），平均每张图片耗时52.24ms。

### 4.4.5 纹理特征提取模块

本文项目的纹理特征采用R=1、P=8的Uniform LBP特征。特征提取算法如下所示：定义均匀模式LBP特征向量映射表，将均匀模式LBP特征向量映射为一个序号值。读取图像并转化为灰度图。遍历每个像素点，针对每个像素点执行以下操作。将该像素点与邻域内8个像素点进行比较，若周围像素点灰度值大于中心像素，则该像素点标记为1，否则标记为0。从而从关键点邻域内得到八维二值向量pattern[]。计算pattern中包含的跳变（由0变1或者由1变0）次数variations。如果variations大于2，表示该特征是非均匀模式特征。如果variations小于等于2，表示该特征是均匀模式特征，通过查表得到其序号值。统计均匀模式特征向量和非均匀模式特征向量的分布，归一化后得到Uniform LBP特征向量。关键代码如下：

for x in range(0, len(img)):

for y in range(0, len(img[0])):

center = img[x,y]

pixels = []

for point in range(1, P + 1):

r = x + R \* math.cos(2 \* math.pi \* point / P)

c = y - R \* math.sin(2 \* math.pi \* point / P)

if r < 0 or c < 0:

pixels.append(0)

continue

if int(r) == r:

if int(c) != c:

c1 = int(c)

c2 = math.ceil(c)

w1 = (c2 - c) / (c2 - c1)

w2 = (c - c1) / (c2 - c1)

pixels.append(int((w1 \* get\_pixel\_else\_0(img, int(r), int(c)) + \

w2 \* get\_pixel\_else\_0(img, int(r), math.ceil(c))) / (w1 + w2)))

else:

pixels.append(get\_pixel\_else\_0(img, int(r), int(c)))

elif int(c) == c:

r1 = int(r)

r2 = math.ceil(r)

w1 = (r2 - r) / (r2 - r1)

w2 = (r - r1) / (r2 - r1)

pixels.append((w1 \* get\_pixel\_else\_0(img, int(r), int(c)) + \

w2 \* get\_pixel\_else\_0(img, math.ceil(r), int(c))) / (w1 + w2))

else:

pixels.append(bilinear\_interpolation(r, c, img))

values = thresholded(center, pixels)

variations = find\_variations(values)

transformed\_img.itemset((x,y), 255)

if variations <= 2:

res = 0

variating\_blocks += 1

for a in range(0, len(values)):

res += values[a] \* 2 \*\* a

transformed\_img.itemset((x,y), res)

pixel\_values.add(res)

else:

unassigned.append((x,y))

|  |  |
| --- | --- |
| Z:\139.jpg | \\Mac\Home\Documents\nlbp_139.png |
| 图4.21 原始图像 | 图4.22 均匀模式LBP特征直方图 |

### 4.4.6 SIFT特征提取模块

本文项目中的SIFT特征提取使用openCV库中方法。关键代码如下：

img = cv2.imread(img\_path)

sift = cv2.SIFT()

kp, des = sift.detectAndCompute(img, None)

|  |  |
| --- | --- |
| 图4.23 原始图像\\Mac\Home\Desktop\8_1.jpg | 图4.24 标注了SIFT特征点的图像\\Mac\Home\Desktop\8_2_kp.jpg |

提取SIFT特征平均数量以及平均耗时如表4.3所示。

表4.3 SIFT特征提取性能

|  |  |  |
| --- | --- | --- |
| 图像尺寸 | 每张图片平均特征数量 | 每张图片特征提取平均耗时 |
| 1100\*1390 | 4502 | 0.7238s |
| 579\*830 | 1098 | 0.1765s |
| 234\*295 | 244 | 0.0356s |

### 4.4.7 视觉词模块

为了保证搜索的速度和效率，本文项目采用视觉词技术压缩图像局部特征（SIFT特征）描述。对图像集中的每一张图片进行特征提取，总共提取得到8316000个特征向量。然后由这些特征向量构成特征空间，采用均值聚类（KMeans）方法进行聚类，建立视觉词汇表。聚类中心即为视觉词，聚类中心的集合即为视觉单词本。本文选取的聚类中心个数为1000。

本文首先分别对图像集中的图片提取SIFT特征并保存为临时文件。然后依次读取SIFT特征文件得到特征空间，之后利用KMeans算法对所有SIFT特征进行聚类。聚类操作使用openCV中的方法。终止条件设置为迭代100次或者距离小于10。关键代码如下：

k = 1000

attempts = 100

epsilon = 10.0

criteria = (cv2.TERM\_CRITERIA\_EPS + cv2.TERM\_CRITERIA\_MAX\_ITER, attempts, epsilon)

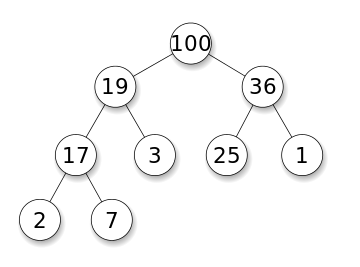
compactness, labels, centers = cv2.kmeans(des, k, criteria, attempts, cv2.KMEANS\_RANDOM\_CENTERS)

函数返回值中，compactness是每个特征向量与其中心之间距离的和；labels是标记数组，值的范围是[0,1,…,9999]，表示每个特征向量所属的中心序号；centers是聚类得到的中心特征向量数组。

### 4.4.8倒排索引模块

### 4.4.8 排序模块

本文项目中的排序问题是典型的Top K问题，给定N（N>>10000）个特征间距离，求出其中前K个最小的距离。由于输入为大量数据，而只需要求得前K个最小值，因此对整个输入数据进行保存并排序是不必要的。

图4.25 大顶堆

本文项目利用大顶堆数据结构来处理该问题。大顶堆如图4.25所示，每个非叶子节点的数值，一定不小于其孩子节点的数值。这样可以使用含有K个节点的大顶堆来保存K个目前的最小值（当前根节点是其中的最大值）。每次有数据输入时，先于根节点比较，如果不小于根节点，则舍弃该数据；如果小于根节点，用该数据替代根节点，并进行大顶堆调整。数据输入完时，大顶堆保存的即是N个数据中前K个最小值。然后使用快速排序（qsort）对这K个数据进行排序，从而得到从小到大有序的前K个最小距离。

Python标准库中包含heapq 模块（Module），本文项目中的排序模块依赖其中的堆操作实现TOP K算法，关键代码如下：

heap = images[:k]

heapq.heapify(heap)

for i in xrange(k, len(images)):

if images[i].dist < heap[0].dist:

biggest = heapq.heapreplace(heap, images[i])

top\_k\_images = sorted(heap, key=lambda x:x.dist)

本文对上述TOP K排序算法和常规快速排序算法进行了对比实验。实验过程如下：随机生成一个长度为36960（图像集中的图像数量）的数组，分别使用上述TOP K排序算法和快速排序算法对数组进行排序，TOP K排序算法中的参数K取值50，然后重复进行100次实验，计算两种排序算法的平均耗时。实验结果如下：上述TOP K排序算法平均耗时4.10ms；快速排序平均耗时10.12ms。实验表明上述TOP K排序算法相比快速排序性能有很大提升。

# 5 系统测试与实验

本论文中使用来衡量检索结果，即对前个检索结果中的正确率对检索结果进行分析。求分别使用、、词汇树的检索方法各自对不同类型商品的平均准确率，和使用和组合、使用和词汇树组合对不同类型商品的平均准确率。得到在检索结果中前张前张图片中的平均准确率如下图所示：

## 5.1 测试环境

众所周知，Android设备厂商众多，Android设备系统版本、屏幕分辨率各式各样。本文中Android客户端在设计和实现上很好的解决了Android的碎片化问题，实现了对主流机型的适配。真机适配测试结果如表5.1所示。

表5.1: 真机适配测试结果

|  |  |  |  |
| --- | --- | --- | --- |
| Android手机型号 | Android系统版本 | 分辨率 | 测试结果 |
| 三星GalaxyS3 | Android 4.3 | 1280\*720 | UI显示及功能均正常 |
| 华为荣耀3C | Android 4.4 | 1280\*720 | UI显示及功能均正常 |
| SnoyXperia C6603 | Android 4.1 | 1920\*1080 | UI显示及功能均正常 |
| 摩托罗拉Moto G | Android 5.0 | 1280\*720 | UI显示及功能均正常 |

系统测试时，手机终端选择使用摩托罗拉Moto G 。摩托罗拉Moto G 的硬件配置如表5.2所示。

表5.2: 摩托罗拉Moto G 硬件配置

|  |  |
| --- | --- |
| 系统配置 | 硬件参数 |
| Operating System | Android 5.0 |
| CPU | Qualcomm Snapdragon 400 四核 1.2GHz |
| GPU | Adreno 305 |
| RAM | 1GB |
| 屏幕大小 | 5.0英寸 |
| 屏幕分辨率 | 1280\*720 |

表5.3: 服务器硬件配置

|  |  |
| --- | --- |
| 系统配置 | 硬件参数 |
| Operating System | Windows 7 |
| CPU | Intel(R) Xeon(R) CPU E5-2650 v2 2.6GHz 2.59GHz(双处理器) |
| RAM | 8GB |
| 硬盘大小 | 2TB |

## 5.2 实验

### 5.2.1 评价标准

为了评价各种图像特征提取算法的性能，需要有一个好的评价标准，对于图像检索的不同应用中，适合该应用的特征提取算法和相应的相似性度量准则也不相同。图像检索的结果判定具有很强的主观性，同时也没有一个统一的图像库用来测试。目前对于图像检索系统的性能评价还没有一个比较好的统一标准，可以借鉴文本信息检索中所使用的一些度量标准，下面介绍一些常用的性能评价指标。

（1）正确率与检索率：

正确率（precision）和检索率（recall）是应用最广泛的一种评价标准，正确率的含义是指一次查询结果中相关图像数目占返回图像数目的比例，而检索率则是指在查询结果中的相关图像数占图像库中与查询图像相关的图像数目的比例，其中还包含查询结果中的相关的图像和不在查询结果中的相关图像。

正确率数学公式可以表示为：

 （5.1）

检索率数学公式可以表示为：

 （5.2）

其中A代表结果集中的相关图像， A加B代表结果中的图像集合，A加C代表相关图像集合。从上面可以看出，正确率是反映了一次检索的准确性，而检索率反映的是该次检索的全面性。

（2）前N个结果的正确率与检索率：

前N个结果的正确率就是在返回的前N个结果中正确的比例，定义R为某一相关的图像集合，（）为查询图像，其返回的N个结果记为，则前N个结果的正确率表示为：

 （5.3）

 （5.4）

选取K个测试样例进行评价，得到的平均正确率可表示为：

 （5.5）

前N个结果的检索率可以表示为：

 （5.6）

同样对于K个测试样例可以得到的平均检索率表示为：

 （5.7）

### 实验结果

本文实验采用K个测试样例的平均正确率作为评价标准。其中，N有5种取值，分别为5、10、20、30、50，K取值为10，图像有关与否有人工主观判定。本文还对GrabCut效果做了对比实验，测试GrabCut对平均正确率的影响。本文图像集中图像数量较多，人工标注所有图像彼此是否相关难以实现，因此本文实验结果并不包含检索率。

1. 颜色特征

图5.1是系统利用颜色特征进行搜索时的平均正确率。其中，Origin表示查询图像以及图像集中的未经过GrabCut过滤背景时的平均正确率，GrabCut表示查询图像以及图像集中的经过GrabCut过滤背景时的平均正确率。由图可知，利用颜色特征进行搜索时，平均正确率良好。使用Grabcut过滤背景，平均正确率有所提升。

图5.1 使用颜色特征搜索平均正确率

1. 纹理特征

图5.2 使用纹理特征搜索平均正确率

图5.2是系统利用纹理特征进行搜索时的平均正确率。其中，Origin表示查询图像以及图像集中的未经过GrabCut过滤背景时的平均正确率，GrabCut表示查询图像以及图像集中的经过GrabCut过滤背景时的平均正确率。由图可知，利用纹理特征进行搜索时，平均正确率良好。使用Grabcut过滤背景，平均正确率有所提升。

（3）SIFT特征

图5.3 使用SIFT特征搜索平均正确率

图5.3是系统利用SIFT特征进行搜索时的平均正确率。其中，Origin表示查询图像以及图像集中的未经过GrabCut过滤背景时的平均正确率，GrabCut表示查询图像以及图像集中的经过GrabCut过滤背景时的平均正确率。由图可知，利用SIFT特征进行搜索时，平均正确率良好。使用Grabcut过滤背景，平均正确率有所提升。

# 6 总结及展望

## 6.1 论文总结

随着智能手机的普及和电子商务的兴起，手机购物越来越流行。面对海量的商品，传统关键字匹配的商品搜索方式逐渐显现出其局限性。而图片作为人们购买时重要的参考信息包含的信息量巨大。近年来，在移动端商品搜索中加入图片作为输入受到了越来越多的关注。

本文首先介绍了基于内容的图像检索技术常用的特征，包括颜色特征、纹理特征、局部特征。着重介绍了HSV颜色空间下颜色分布直方图作为颜色特征、Uniform LBP特征作为纹理特征、SIFT特征作为局部特征的优势及其提取算法。并简单介绍了视觉词袋模型、GrabCut图像分割算法。视觉词袋模型应用于SIFT特征上，可以有效的压缩图像的描述，提供检索的速度和效率。GrabCut图像分割算法可以在一定程度上过滤掉背景，排除干扰信息，提高检索准确度。

Android平台以其开放的优势，在各个领域都得到了广泛的应用。本文实现的服饰图像搜索系统客户端基于Android平台开发。本文介绍了Android平台的发展、Android平台的组成以及Android应用程序开发关键技术。

基于以上理论和技术，本文实现了服饰图像搜索系统的Android客户端。Android客户端共分为界面层、核心层、接口层、模型层四个逻辑层次，图像输入模块、结果展示模块、其它辅助模块三个功能模块。并对实现过程做了详细的阐述。

本文也实现了服饰图像搜索系统的服务器端。对文中介绍的颜色特征提取算法、Uniform LBP特征提取算法、TOP K算法使用Python进行了实现。GrabCut图像分割、SIFT特征提取、SIFT特征聚类则借助于OpenCV库。然后对各个模块进行整合，搭建完整可用的服务器平台，为客户端提供稳定的服务，并对实现过程做了详细的阐述。

## 6.2 研究展望

本文实现的系统还有一些缺陷。本文项目使用的图像集仅有36960张图像，远远少于电商平台的服饰图像数量。因此，图片以普通方式存储，特征之间距离的计算和排序操作计算量也并不大。而在实际应用场景中，图片文件是海量的。海量图片的存储与管理是图像搜索系统面临的挑战。由于图像集数量有限，本文图像搜索算法在海量图像集上的搜索正确率以及响应时间不能被很好的估计。本文实验采用平均正确率作为搜索性能评价标准。而在平均正确率的计算中，搜索结果图片与查询图片是否相似依靠人工主观判断。首先，人工判断使得实验过程非常耗时；其次，主观差异使得计算得出的平均准确率并不可靠。如果存在已经标注图像之间相似度的图像集，相似度不再需要主观判断，那么实验过程即可实现自动化，计算得出的平均准确率也将更加可靠。本文项目使用GrabCut算法进行背景过滤，GrabCut是基于人机交互的，需要人工标注一定区域为目标或背景。在本文中，针对自己的图像集，依赖其中的图像构图具有规律性，用一个矩形框标注了图像的前景与背景。这种做法适应性差，效果不好。在这种应用背景下，使用无监督图像分割技术进行背景过滤更加恰当。

Android应用程序开发中，由于移动设备电量、内存等硬件资源有限，性能优化受到广泛关注。在处理、展示图像时，性能优化的要求尤为迫切。同样，网络数据传输的安全性也值得注意，如今网络环境越来越复杂，客户端与服务器的交互安全需要每个系统开发人员关注。响应速度也是衡量用户体验的一个重要指标。网络传输大量图片的技术还需要进一步优化。

# 致 谢

不知不觉中，论文工作已经接近尾声。这篇论文的完成与一群人的帮助密不可分。

首先要感谢我的导师赵春霞教授。感谢她研究生阶段一直以来的关心与指导。

还要特别感谢李旻先老师。这篇论文从选题目、做实验到写文章都在他的悉心指导下完成。不仅在学术上帮助我成长，李旻先老师也教会我很多做人做事的道理。每次与李旻先老师谈话都有很大的收获。我为能被这样一位优秀的老师指导而感到非常幸运。同时，我要感谢南京理工大学。

感谢任良成、张洪伟、张懿鑫三位好友。

感谢朱凌峰、张金康、李宁三位室友，是他们见证了我的成长！

最后要感谢我的爸爸、妈妈、哥哥、姐姐还有外甥女。虽然我的妈妈不知道论文为何物，但她听说我要写五六十页，她就为我骄傲，感谢她一直无条件的爱我。感谢我的爸爸，感谢他一直努力工作，让我可以不为钱发愁，有一个安稳小康的家。感谢我的哥哥，一直默默的关心我的，遇到事情喜欢挡在我的前面的哥哥（虽然有些事情，我觉得我可以处理好的）。感谢我的姐姐，永远把我当小孩儿看的姐姐，感谢这么多年来给我的这么多的红包。感谢我的小外甥女，她马上两周岁了，感谢她学会的第三个词就是舅舅，感谢她给我们带来那么多的快乐。感谢他们，因为有他们，我感到幸福。

# 参考文献

1. Psaraftis H Interactive graph cuts for optimal boundary&region segmentation of objects in ND images Computer Vision 14 2001
2. Michael R.Swihart, Jason D, Papastavrou Exploiting Smart-Phone USB Connectivity For Fun And Profit European Journal of Operational Research 114 2010 447-464
3. Teodorovic D, Pavkovic G Privilege escalation attacks on Android Transportation Planning and Technology 16 2011 261-273
4. Minkoff A A Stealthy and Context-Aware Sound Trojan for Smartphones Operations Research 8 2010 1279-1295
5. 曹健 图像目标的表示与识别 机械工业出版社 2011
6. 吴小珍，李表奎 Android 手机的轻量级访问控制 计算机应用研究 5 2014 176-178
7. 郭耀煌，谢秉磊 Linux 内核完全注释 机械工业出版社 4 2004 114-115
8. 陈壁峰，陆昊娟，黄樟灿 Linux 内核的分析及应用 西安邮电学院学报 3 2002 46-48
9. 蒋腾飞 深入理解 Linux 内核 东南大学出版社 2013
10. 郑海虹 一种基于图的交互式目标分割算法 计算机工程与应用 4 2006
11. 李晓东，王东 人体图像中周边物品检测分类技术研究 南京邮电大学 1 2014 172-175
12. 李军，郭耀煌 面向购物搜索的目标提取算法研究及系统实现 西南交通大学 2001
13. 常朝稳，徐江科 终端行为可信评估及其访问控制方法研究 小型微型计算机系统 3 2014 493-499
14. 兆凤久 交互式服饰图像检索研究及系统实现 西南交通大学 2 2013 1
15. 章权 云计算及安全分析 华南理工大学学报 8 2011 109-112
16. Van Hemert J I, La Poutre J A Parallel Problem Solvers Nature VIII 2004 690-699
17. Parfaits H Compressed histogram of gradients a low bit rate feature descfiptor Transportation Science 17 1983 351-360
18. 李宁，邹彤，孙德宝 基于内容的图像检索综述 基于内容的图像检索综述 4 1999 130-135
19. Bodin L D, Sexton T R Distinctive image features from scale—invariant keypoints Working Paper MS/S 1982
20. 陆琳 基于文本的与基于内容的图像检索技术比较研究 南京航空航天大学 2009
21. 郭建红 基于内容图像检索中的对象提取与检索算法研究 北京交通大学 2013
22. THANGIAH S, NYGARDK, JUELL P G Towards optimal bag—of-features for object categorization and semantic video retrieval Proceedings of the seventh Conference on Artificial Intelligence Applications 1991 322-325
23. JOE L, ROGER L Query by image and video content: The QBIC system Proceedings of Fifth International Conference on Genetic Algorithms 1993 452-459
24. Martin Desrochers, Jacques Desrosiers, Marius Solomon Towards optimal bag—of-features for object categorization and semantic video retrieval Operations Research 40(2) 1992 342-354
25. Diaz A Through the Google goggles:Sociopolitical bias in search engine design WebSearch.SpringerBerlinHeidelberg 40 1992 342-355
26. Van der Bruggen, Lenstra L J , Schuur P C Variable-depth search for the single vehicle pickup and delivery problem with time windows Transporation Science 27 1993 298-311
27. 霍住震，张磊 基于 CIFS 协议的存储加密代理设计与实现 第 24 次全国计算机安全学术交流会 5 2009 64-66
28. 叶耀华，朱晓梅，陈霖 理解 Android 上的安全性 ：利用沙箱、应用程序签名和权限增强应用程序安全性 系统工程理论与实践 3 2009 110-112
29. 施朝春，王旭 隐蔽信道研究 计算机工程与应用 34 2009 21-24
30. 马华伟 用 Android 开发手机应用 程序员 2009
31. R. Feris and L. Davis. Image ranking and retrieval based on multi-attribute queries. In CVPR, 2011. 6
32. Z. L. Guangcan Liu and Y. Y. Robust subspace segmentation by low-rank rep- resentation. In ICML, 2010. 6
33. B. Hasan and D. Hogg. Segmentation using deformable spatial priors with application to clothing. In BMVC, 2010. 2
34. C. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In CVPR, 2009. 3
35. Z. Lin, M. Chen, L. Wu, and Y. Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. Arxiv preprint, 2010. 6
36. S. Pan and Q. Yang. A survey on transfer learning. TKDE, 2010. 3
37. Z. Song, M. Wang, X. Hua, and S. Yan. Predicting occupation via human clothing and contexts. In ICCV, 2011. 3, 5
38. N. Wang and H. Ai. Who blocks who: Simultaneous clothing segmentation forgrouping images. In ICCV, 2011. 2
39. X. Wang and T. Zhang. Clothes search in consumer photos via color matching and attribute learning. In ACM MM, 2011. 3
40. J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. TPAMI, 2009. 5
41. M. Yang and K. Yu. Real-time clothing recognition in surveillance videos. In ICIP, 2011. 2, 3
42. Y. Yang and D. Ramanan. Articulated pose estimation with flexible mixtures-of-parts. In CVPR, 2011. 2, 3, 4