

分 类 号_____

学 号_____M200972460_____

学校代码_____10487_____

密 级 _____

华中科技大学

硕 士 学 论 文

外观专利搜索中图像特征提取
技术研究

学位申请人：周 挺

学 科 专 业：计算机应用技术

指 导 教 师：章 勤 教 授

答 辩 日 期：2012 年 2 月 5 日

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering**

**Research on Image Feature Extraction
Method for Design Patent Image Retrieval**

Candidate : Zhou Ting

Major : Computer Application Technology

Supervisor : Prof. Zhang Qin

Huazhong University of Science & Technology

Wuhan, Hubei 430074, P.R.China

Feb., 2012

独创性声明

本人声明所呈交的学位论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到，本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，即：学校有权保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权华中科技大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本论文属于 ☐ 保密，在_____年解密后适用本授权书。
☐ 不保密。

（请在以上方框内打“√”）

学位论文作者签名：

日期： 年 月 日

指导教师签名：

日期： 年 月 日

摘要

随着国家专利法的健全，人们法律意识的提高，外观专利的申请数量急剧增加，来自个人与企业的对外观专利图像的检索需求强烈。目前国内针对外观专利的检索系统主要是基于关键字的，而基于关键字检索外观专利的图像不能很好的解决用户的需求。

基于图像内容的外观专利搜索，从分析专利图像的特点入手，对这些特点进行研究，实现了外观专利图像的底层视觉特征与高层语义特征提取，在此基础上研究了基于图像特征的分类。专利图像底层视觉特征含有颜色特征与形状特征，颜色特征采用划分网格的色彩直方图的方式获取，形状特征采用金字塔化的梯度方向直方图的方式获取。在底层视觉特征中融入空间分布信息，使其更接近人类对物体的感知；图像的高层语义特征是为提高检索效果，缩小底层视觉特征与人类对图像理解之间的差距提出的。针对外观专利的图像的知识范围有限且语义定义具体，采用结合专利文档与图像获取图像语义信息，融合图像的底层视觉特征提取了专利图像的语义特征。在实现语义特征时采用样本空间来定义语义空间，采用融合外观特征与几何形状特征作为专利图像的底层视觉特征。其中语义空间的参考基是支持向量机在每个类上获取的最优分类面的集合，外观特征采用了基于局部区域的描述符，同时避免了局部特征的关键点检测部分，减少了提取特征的时间。

最后采用外观专利图像数据作为测试集合，对提取的图像特征在判别能力，提取时间等方面进行了详细的测试，测试结果表明这些特征能够满足外观专利图像检索的检索速度与效果要求。同时，把获取的语义特征与目前常用的效果很好的局部特征 SIFT 做了对比，语义特征的判别性明显优于局部特征 SIFT。

关键词：外观专利，图像检索，图像特征，局部特征，语义空间，语义特征

Abstract

With the improvement of the intelligent properties laws and the enhancement of human legality awareness, the number of applying patents is increasing dramatically from individuals and enterprise. Currently in domestic, the design patents retrieval systems are mainly based on text description, but those systems could not satisfy users' needs. Meanwhile Content-Based Image Retrieval (CBIR) systems are usually for general purpose, which works on images from Internet, So general CBIR systems are not practical for patents image retrieval.

By analysing the attributes of patent images, the low-level visual features and high semantic features are used for content-based patent image retrieval system. The research focus on image feature extracting method and discriminative ability. For the reasons above, this thesis focus on how to implement design patents images retrieval. The low-level visual features contains colour and shape features. Colour feature is obtained by dividing the images into small grids and then get colour histogram in grids. Shape feature is obtained by pyramid of histogram of orientation gradients. Spatial information is integrated into global features for closing to users' perception. Moreover, the knowledge of images used for applying patents is defined quite concrete and specific, which is good for aliasing semantic to images. By taking advantage of design patent images' semantics, in addition to low-level visual features, semantic features of patent images are used. For these special properties, a framework is proposed for patents images semantic features extraction. In implementing semantic features, semantic space is defined by selecting sample images and the space reference basis is acquired by the hyper-plane derived from support vector machine. The low-level visual features under semantic feature are integration of appearance and geometry shape features. Appearance features has taken advantage of local features, but avoided its key-points detecting part which is quite time-consuming, so this feature could meet the requirement of real-time applications.

Images features are evaluated in discriminative ability, feature's dimension and extraction time speed on testing images from Chinese design patent. The testing results demonstrated that these features can meet the requirements of retrieval on speed and performance. What's more, the semantic feature outweighs other features significantly including the state-of-art local features SIFT in discriminative ability.

Key words: Design Patents, CBIR, Image Features, Local Features, Semantic Space, Semantic Features

目 录

摘 要	I
Abstract	II
1 绪论	
1.1 研究背景与意义	(1)
1.2 国内外研究概况	(2)
1.3 研究内容	(5)
1.4 文章组织结构	(6)
2 外观专利图像搜索系统框架	
2.1 专利图像搜索系统整体设计	(8)
2.2 获取图像相似度设计	(10)
2.3 小结	(15)
3 外观专利图像的特征提取方法	
3.1 外观专利图像特点分析	(16)
3.2 外观专利图像的底层视觉特征提取方法研究	(17)
3.3 专利图像的高层语义特征提取方法研究	(25)
3.4 小结	(33)
4 功能测试与性能分析	
4.1 测试环境与数据	(35)
4.2 图像特征性能测试与分析	(37)
4.3 本章小结	(42)
5 总结与展望	
致 谢	(45)
参考文献	(47)

1 绪论

本章首先介绍了基于图像内容专利搜索的研究背景与意义，然后简要介绍了图像搜索的发展历程，着重对图像特征的提取方法进行了描述，接着给出了本文的研究内容，最后展示了本论文的组织结构。

1.1 研究背景与意义

随着国家专利法的完善与人们法律意识的增强，每天有大量的外观专利申请来自个人与企业。专利代理机构需要从庞大的数据库中找出类似的专利，用以审核新提交的专利，进而决定该专利是否可以申请成功。专利设计人员也需要查看专利库中的专利，用于改进设计或者防止重复设计。外观专利的主要载体是所提交的图像，文字起辅助作用，毫不相关的图像可能具有相同的文字描述，所以基于文字的搜索不能解决用于对专利审核与查重的需求。同时基于内容的通用搜索不能用于国内的专利系统中，基于以上原因，提出了基于图片内容的专利图像搜索系统。

基于图像内容的专利搜索属于图像搜索的领域，具有许多技术难点，也是学术界的研究热点。学术研究主要围绕如何提高用户满意度展开，难点主要集中于如何描述图像，使机器判别的相似度接近人类对图像的理解。图像搜索从输入关键字搜索发展到了以图搜图的时代。通用图像搜索引擎最近几年发展迅猛。以 Google 为代表的互联网公司，其通用图片搜索的效果越来越好。但是这些公司针对的数据集是互联网上的图片，对专利库中的图片效果并不好，部分原因是大多数互联网图片都预先处理过。而专利图片又具有其特殊的性质，并且大多数并没有在互联网上出现过，所以通用图像搜索并不能用于专利的搜索。

图像特征的发展经历了简单的颜色、纹理、形状等全局特征，基于区域分割的特征，到局部特征，再到目前接近人类对图像认知的语义特征。对应技术的图像搜索名字为基于图像内容的搜索 CBIR (Content-Based Image Retrieval)，基于区域特征的搜 RBIR (Region-Based Image Retrieval)，基于语义特征的搜索 SBIR (Semantic-Based Image Retrieval)。从此可见，特征发展正在逐步缩小机器视觉与人类视觉的差异。特征研究从开发复杂的底层特征转移到了与人类视觉更接近的高层语义特征上。图像特征描述符有如下评价标准：平移不变性，尺寸不变性，旋转不变性，光照不变性，视角不变性以及局部变化鲁棒性。但是，不同的应用对特征

要求会有所不同,所以需要根据专利图像数据的特点与检索应用需求,对专利图像的特征及其性质进行研究。

1.2 国内外研究概况

1.2.1 图像搜索技术研究概况

由于基于文本的图像搜索中,需要对图像进行文本标注,标注信息在部分应用中并不存在,人工标准需要耗费大量人力,并且标注信息主观性很强。鉴于基于文本搜索的主要问题,在 20 世纪 80 年代初期提出了基于图像内容的搜索 CBIR。基于图像内容搜索是首先提取图像的底层视觉特征,然后根据视觉特征向量的相似度来决定图像之间的相似度。Datta^[1]总结了 CBIR 的 3 个层次。第一层次是通过简单的图像底层视觉特征,如颜色、纹理、形状、空间分布等特征进行图像检索。第二个层次是通过识别提交图像的内容来进行图像检索,如查找含有人或建筑的图像。第三个层次是通过高层的知识,如根据图像描绘的事件、表达情感、显示的宗教内容等信息,进行图像检索。其中后两个层次与第一个层次的差别表现为第一个层次的检索缺乏语义。第三个层次的检索难度很大,目前很少见。目前的图像检索系统大都停留在第一、二个层面。图像的底层视觉特征是图像检索系统的基础。随着图像视觉特征研究的深入,图像搜索具有很大的发展,从早期的基于简单图像特征的搜索到基于区域特征的搜索,然后发展到基于语义特征的图像搜索。下面选择几个系统对其进行简要介绍,介绍的重点放在其所采用的特征上面。

1. PatSeek

PatSeek^[2]是针对美国国家专利数据库开发的专利搜索系统。PatSeek 提取的图像底层特征为边缘方向关联图 EOAC (Edge Orientation Auto-Correlogram), EOAC 特征提取方法是预先把线条归类如直线,双曲线,双折线,虚线等类型,然后获取图像内线条的类型,位置,与方向以及与其连接的边的信息,最后量化获取的所有线条信息得到图像的特征。相似度度量采用 L1 与 L2 距离度量方式。其针对的数据集合主要是工程制图或者线条图,这类图像并不具备颜色,梯度,纹理等信息,所以该系统并不适用于中国的外观专利。

2. PatMedia

PatMedia^[3]是希腊针对专利开发的一套系统。其数据库中的图像主要是二值图,

其采用分层自适应几何中心 AHGC (Adaptive Hierarchical Geometric Centroids^[4])来提取图像的几何形状特征, AHGC 主要把图像分为不同的图形, 如矩形, 圆形, 正方形等, 然后提取这些图形的特征, 如中心位置, 类型等, 最终得到整幅图像的特征。但是该方法只适应于由不同几何图形构成的图像, 其同样不适用于中国的外观专利。

3. PSEARCH

PSEACH^[5]是浙江大学的郑俊勇, 孙守迁等针对中国外观专利做的一个实验系统。其采用颜色、纹理、形状等图像底层特征, 并结合图像的语义信息进行专利图像检索。其语义获取方式是采用语义模板^[6]以及一定的语义生成规则获取的, 这种语义获取方式依赖于事先定义的语义规则, 并且语义模板不易于扩展。同时根据其文中的测试部分, 可以看出这种检索方式不易于使用。该试验系统并未公开, 对公众并不可用。

4. Google

通用图像搜索, 目前以 Google 的效果最好。Google 的图像搜索除了采用底层视觉特征, 同时也对图像提取了语义特征。其语义特征实现方式是通过模型预测的方式获取输入图像的语义信息, 把语义信息关联到文本标注, 然后根据文本标注在图像库中搜取相应的结果。由于 Google 的技术细节并未公开, 以下推测其使用的技术。其语义信息获取方式可能是根据输入图像获取图像库中最相似的几个图像, 根据最相似图像的文本标注信息来获得输入图像的标注信息, 在根据标注返回最终相似的图像。或者是以下方式, 建立关键字、底层视觉特征、图像内部语义关系网络的模型, 通过建立的模型来确定图像的语义, 最后展示相似的图像结果。Google 采用 Collaborative Filtering^[7]建立图像内部的相似度连接, 图像之间的相似度是通过用户的点击来确定的, 即图像之间相似是由广大互联网用户决定的。Google 图像搜索结果单页内显示大量的图片, 用户会从大量的图片中选择其认为相似的几张图像进行点击, 从而帮助 Google 建立了模型。Google 主要针对互联网上的图片, 一方面利用 HTML 上图像的文本信息对图像进行标注获取语义, 另一方面通过其拥有极大的用户来建立语义模型, 从而获取很好的效果, 但是专利数据库并不具备这个优势。其不能直接用于专利图片。

1.2.2 图像特征提取技术研究概况

图像特征包括底层视觉特征, 与高层语义特征。其中底层视觉特征是图像检索

系统的基础,即使是语义层次的检索,也需要底层视觉特特的支持。图像的底层视觉特征主要包括全局特征,如颜色、纹理、形状、空间分布等,局部特征。高层语义特征包括基本的语义特征,如人脸,建筑等的图像,抽象的高层语义特征,如图像表达的情感基调,宗教信息等。下面对各个特征做简单的介绍。

颜色特征的表达性最强,其在图像搜索中应用得最为广泛。颜色可以在不同的空间中表示,如 RGB, HSV, LUV 等空间, Ying^[8]指出在大多数情况下,颜色特征在不同空间中表现差别不大。常用的颜色特征有颜色直方图^[9], 颜色关联向量^[10], 颜色协防矩阵等。

纹理特征定义不如颜色特征明确,是一种更精细的特征。很多系统获取图像的梯度信息而并不使用纹理特征。纹理特征常用在医学图像处理,图像分类等系统中。常用的纹理特征提取方法有 Tamura texture features^[11], Gabor filtering^[12], wavelet transform^[13]等。

形状特征是图像的一种重要特征,现实世界的物体都具有一定的形状,也是人类感知物体的一个重要方面。形状特征主要分为基于区域的特征,如 Hu 不变矩, Zernike 矩等,基于轮廓的有傅里叶描述子,复数坐标等。基于轮廓的特征能更好的描述物体,其性能优于基于区域的特征,但是图像的轮廓往往不容易提取。目前常用的用于描述形状特征的方法是获取图像局部区域边缘分布来描述整体的图像轮廓,形状上下文^[14],方向梯度直方图^[15]。

空间分布在部分应用中是一种重要的特征,如汽车搜索中,大多数情况下车的轮子总在下面。空间位置可以简单的定义为上,下,左,右,居中等。在语义信息的获取中相对的空间关系要比绝对的空间位置更重要。常用的空间分布描述方法有区域中心及最小边界,划分网格与金字塔核^[16]等

图像的局部特征是基于关键点的图像视觉特征,能够更精细的描述图像。Hubel 与 Wiesel 对猫的视觉实验^[17]表明哺乳动物视觉是天然基于局部特征的。局部特征提取分 2 个阶段,第一个阶段是检测关键点,第二个阶段是获取兴趣区域描述符。关键点检测方法主要有基于边缘的,如 canny^[18] 算子, Marr-Hildreth^[19]等。基于角点的,如 Harris-Laplacian corner^[20]。基于区域的如 Laplacian of Gaussian^[21]。经过 Mikolajczyk 与 Schmid 的评测^[22]指出目前效果最好的局部特征描述符是类 SIFT^[23] (Scale Invariant Feature Transform)特征,是 Lowe^[24]在 1999 年首次提出并在 2004 年进行了改进。

图像语义特征是融合了人类对图像理解的高层图像特征,是用来弥补图像底层

视觉特征与图像丰富内容之间语义鸿沟。语义特征是目前学术研究热点，核心在于如何关联图像的语义信息。Vailaya^[25]等在城市和风景的图像数据集合上，采用朴素贝叶斯分类器^[26]对图像进行按等级分类，并结合边缘方向直方图与边缘方向相关量等图像的底层视觉特征获取了图像语义特征。Chen 与 Yu^[27]通过推测用户输入关键字的意图获取语义信息，并结合图像的底层视觉特征，获取图像的语义特征用于购物搜索。朱兴权^[28]等采用反馈的方式获取图像的语义信息，并结合图像的底层视觉特征得到图像语义特征用于图像检索。从以上关于图像语义特征的介绍可以看出，获取图像语义特征的前提是获取图像的语义信息。常采用的方式有 1. 通过识别图像中的物体获取图像的语义信息^[29]；2. 采用机器学习方式管理图像与查询关键词获取输入图像的语义信息^[30]；3. 通过建立语义模板^[31]获取输入图像的语义信息；4. 通过反馈，推测用户意图^[27]，从而获取语义信息；5. 融合互联网上的图像标注信息获取图像的语义信息^[32]。

尽管存在许多的图像特征提取方法，CBIR 系统还远达不到基于文本的搜索系统的性能，CBIR 系统远不能满足人们的要求。提高 CBIR 系统的用户满意度，一直是学术界研究的重点。

1.3 研究内容

基于图像内容的外观专利检索，是针对目前基于文本的专利搜索系统以及基于图像内容的通用检索系统不能满足广大用户的需求提出的。本文旨在通过分析专利图片的特点，找出合适的图像特征，并实现图像特征提取模块，应用于专利检索系统中，从而提高检索效果。

针对专利保护的不同方面，如专利会对图案的颜色，纹理，内部物体等保护，需要不同的图像特征对这些方面进行表达。专利图像中单张图片的背景单一，但是不同专利背景往往不一致，为减少图像特征受专利背景的影响，需要做基于区域的颜色特征提取。同一个专利一般会提供 6 个视图，使得专利图像很少受到视角的影响，但是部分专利图片会存在较大的旋转，需要图像特征能够对旋转保持不变的形状特征描述符。

本文利用专利中知识范围有限且语义定义清楚具体，提出了专利图像语义关联方法，并结合专利图像的特点，提出了专利的语义特征提取框架，并提取了专利图像的语义特征。对语义关联方法，以及语义空间定义方式以及语义空间参考基获取

方式等方面进行了研究。对语义特征的底层视觉特征，外观特征，形状特征等做了深入的研究。在提取语义特征的底层视觉特征时，采用了局部特征表示方法，但是避免了局部特征的关键点检测部分，从而提高了语义特征提取时间。

文章主要围绕图像特征展开研究，并对相似度获取及多特征融合等做了研究工作。目前有许多距离函数提出用于相似度获取，但是不同的距离函数有其适用范围。本文针对所采用的图像特征选取合适的距离函数，并融合不同特征的距离值获取图像之间的相似度，最后对相似度进行排序。

同时，为使图像特征提取模块与相似度获取模块易于系统使用，本文对特征代码以及距离函数接口进行定义与封装，使其能够使用其他语言开发的特征提取方法，提高了系统的可扩展性。

1.4 文章组织结构

第一章首先介绍了基于图像内容的专利搜索的研究背景和意义，接着简要介绍了国内外研究概况，主要包括国外专利搜索，Google 通用图像搜索系统，在介绍这些系统时偏重于介绍其采用的图像特征，对部分搜索系统介绍了其语义特征获取时关联语义的方式。然后说明本课题的研究内容是根据专利图像的特点，选取合适的特征描述。最后介绍了文章的组织结构。

第二章首先介绍外观专利图像搜索系统框架，并对其中重要的模块进行简要的描述，接着介绍图像搜索中图像之间相似度获取原理，然后对获取相似度的图像特征提取，距离匹配，特征融合等模块进行逐一说明。接着介绍图像特征，并指出影响图像特征的因素，并给出解决这些问题的技术方案。然后介绍了常用的距离匹配函数，与特征融合方式，其中融合方式采用线性融合方法，这一章为后续章节打下了基础。。

第三章从分析专利图像的特点入手，然后着重讲述专利图像的底层视觉特征提取方法。颜色，形状等特征对人类的感官影响较为强烈，同时也是专利保护的内容，首先讲解图像颜色，形状特征的提取方法，在提取这些图像特征时融合了空间分布信息。接着着重描述专利图像的语义特征提取方法。首先对图像的语义特征进行简单的介绍，接着给出语义特征提取框架，然后描述采用样本确定专利图像语义的方法，并讲解语义特征计算模型的获取方式。最后结合语义计算模型，外观专利图像底层视觉特征，介绍外观专利图像与特征的提取方法。

第四章详细首先介绍了测试特征的环境与测试数据，并说明了特征在系统中的使用方式，然后对语义特征在提取速度，判别性等方面进行测试，并对颜色，形状特征以及语义特征进行了检索效果展示。其中判别性测试通过与目前分类效果较好的特征进行对比，采用查准率与查全率进行说明。功能测试是各个特征在系统中的检索效果。

第五章对总结了全文并展望了以后的工作。

1. 交互模块

交互模块主要用于接收用户提交的待检索图像，把待检索图像发送给后台处理程序，并在后台处理完成时，返回结果给用户。

2. 图像特征提取模块

图像特征提取模块用于提取图像的特征，在后台提取专利图像库中图像的特征，建立专利图像特征库。同时该模块也提取前台提交的待检索图像的特征，用于后继的相似度获取模块。该模块是系统的核心模块，其所提取的图像特征决定了系统检索的效果，并且其提取时间对检索的时间也有很大影响。

3. 图像分类模块

图像分类模块用于对图像进行分类，后台根据图像的类别进行分类存放，前台提交的图像根据分类类别确定检索方式。

4. 索引模块

索引模块对专利图像的特征建立索引，该模块决定了系统的检索速度。

5. 相似度获取模块

相似度获取模块主要获取待检索图像的特征与专利图像特征库之间的相似度。该模块设计特征距离匹配，与多特征融合等关键技术。

6. 结果返回模块

结果返回模块主要用于对获取的图像相似度进行排序，然后把排序结果传给交互模块。

7. 反馈模块

反馈模块在查询结果不好时，用于接受用户的反馈，如对各特征权重进行调整，进行再次检索，知道系统给出较为满意的搜索结果。。

从以上介绍可以看出系统的核心模块在于图像分类模块，图像特征提取模块，相似度获取模块以及索引模块。其中系统的检索速度取决于索引模块与特征提取模块，系统的检索效果取决于图像特征提取模块与相似度获取模块。

2.1.2 系统数据处理流程

用户从提交图像到获取查询结果系统的处理流程如图 2.2 所示。系统首先获取待检索的图像然后提取图像的特征，并对该图像进行分类，根据待检索图像的分类标

号确定检索方式，如果待检索图像属于普通分类，则采用分类检索模式，否则采用全局检索模式。检索模式主要用于确定检索图像集合。确定检索范围后，然后在检索范围内获取带检索图像与库中图像的相似度，最后对相似度进行排序并输出结果。

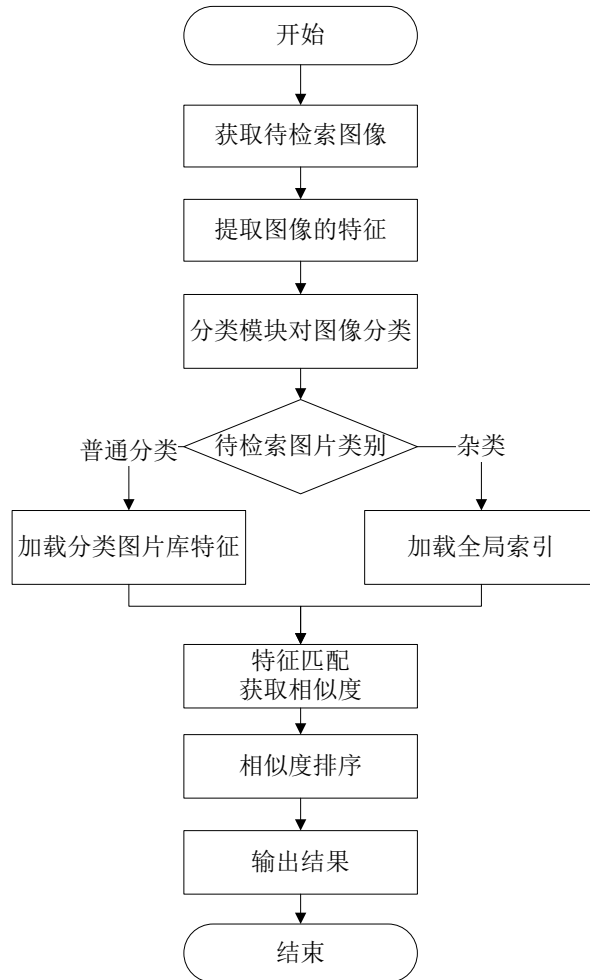


图 2.2 检索流程示意图

2.2 获取图像相似度设计

2.2.1 获取图像相似度流程

从上一节的分析可以看出，系统整个处理过程的目的在于获取待检索图像与库中图像的相似度。在基于图像内容的搜索中，图像之间相似度的获取不依赖于图像的文本标注，其是根据图像的像素信息来得到的。最直接的获取图像之间相似度的方法是逐个比较图像的所有像素，但是这种方法对大部分应用都不可行，因为图像

内像素之间的对应关系难以确定，并且基于这种方法的比对很容易受到各种因素如噪声，光照，摄像角度，旋转，图像尺寸等的干扰。因此，需要根据图像的像素得出图像特征描述符，使用图像特征描述符来代表图像，采用图像特征向量之间的相似度来代替图像之间的相似度。图 2.3 给出了获取两幅图像之间相似度的一般过程。首先要把图像文件转换成图像特征描述符，

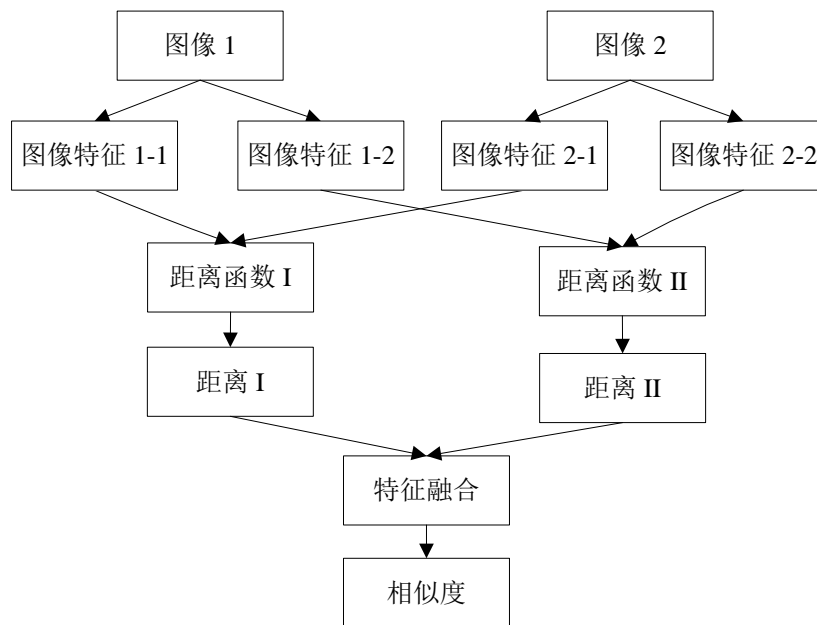


图 2.3 获取图像相似度示意图

获取图像之间的相似度首先需要采用图像特征提取算法提取图像的特征描述符，同一图像根据应用需求往往需要多个特征描述符对图像进行充分表达。然后根据图像特征的类型以及表示形式，采用合适的距离函数，如欧式距离，计算不同特征向量之间的距离。最后采用合适的融合方法，如线性融合，融合不同特征的距离得出最终的相似度。

从获取图像之间相似度的过程可以看出，有 3 个基本的步骤，其一是从图像到图像特征，即图像特征提取；其二是计算不同图像特征之间的距离，即特征向量匹配；其三是融合不同距离，即距离融合。在以上 3 个步骤中，最重要的是图像特征提取，其产生的图像特征描述符直接决定了获取的相似度是否可靠。

2.2.2 图像特征描述符的性质

数字图像大都是采用成像设备获取的现实世界物体的影像，其容易受到以下因

素的影像。如，由于图像摆放位置，角度等物理变化图像会发生相应的旋转，平移等变化，由于拍照时光照与焦距的不同，图像也会发生明暗度与视角的变化，同一张图像由于不同的缩放比例，会导致图像的尺寸不一致，同时在大多数情况下，图像可能具有噪声等。好的图像特征描述符要能够在以上情况发生时仍能保持强的表达性。因此，评价图像特征描述符的标准有以下几条：1.平移不变性；2.旋转不变性；3. 尺寸不变性；4.仿射不变性；5.光照不变性；6.能够容忍局部的微小变化。针对图像特征描述符的几条标准，目前都有其解决方案。下面针对几个图像重要的性质的解决方案进行表述。

1. 尺寸不变性

为应对图像尺寸变化引起的特征描述符不稳定的问题，一直困扰学术界多年，直到 1998 年 Lindeberg 提出的尺度空间理论^[33]，其首先在尺度空间中进行表示图像，然后在不同的尺度空间上提取图像的特征。图像 $I(x,y)$ ，采用转换到尺度空间的方法如公式 2.1 所示。

$$L(x, y; \delta) = G(\delta) \cdot I(x, y) \quad (2.1)$$

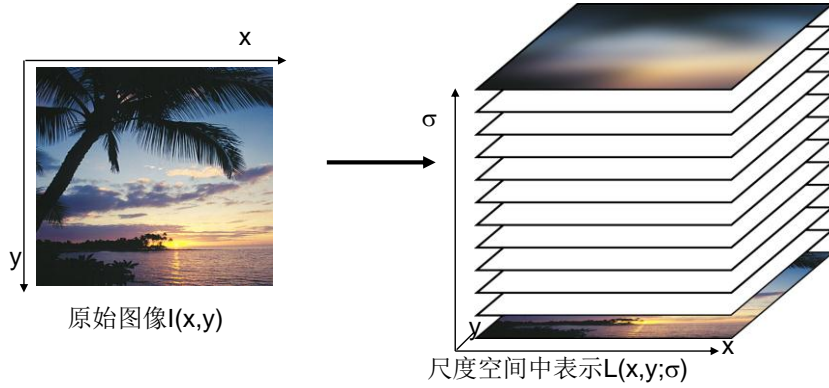


图 2.4 不同尺度空间中表示图像的示意图

其中 $G(\delta)$ 是高斯核函数，二维函数的高斯核函数如公式 2.2 所示，公式 2.1 中采用的高斯核函数参数 δ_x, δ_y 相等取为 δ 。

$$G(x, y; \sigma_x, \sigma_y) = \frac{1}{2\pi\sqrt{\sigma_x^2 + \sigma_y^2}} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right\} \quad (2.2)$$

图像在不同尺度空间上表示如图 2.4 所示，尺度空间由参数 δ 控制。

2. 仿射不变性

图像发生仿射变化时，其由于视角变化，导致图像的部分丢失或者多余的信息增加，使得提取出来的特征描述符很难保持一致，如图 2.5 所示是同一车由于拍照的角度不同发生的仿射变化。针对此类问题，提出的解决方案是从图像中采取小的图像片段，提取片段的特征，采用基于片段的特征进行比较，提取片段的方法如图 2.6 所示。如果视角变化过大，提取片段的方法也无法有效解决。



图 2.5 同一物体由于视角不同产生的图片

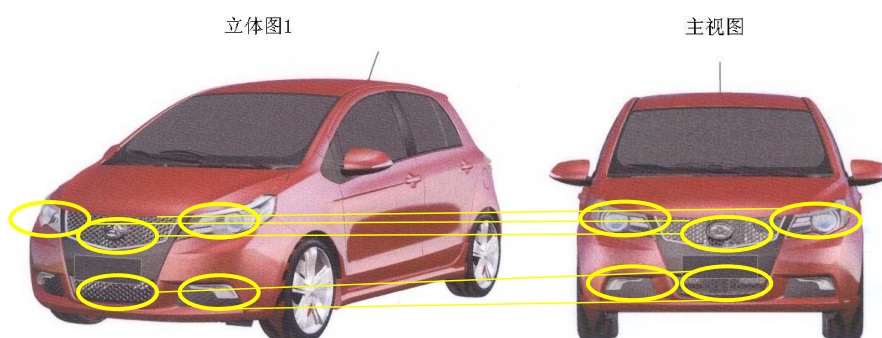


图 2.6 采用取 patch 方法解决视角变化

2.2.3 距离匹配与多特征融合

为了获取图像特征向量之间的相似度，需要计算特征向量之间的距离。图像的特征描述符一般表达图像颜色、纹理、形状、能量分布等信息。其常用的形式具有向量形式，直方图形式，Bag of Visual Words (BoG) 特征即词袋特征等。其中直方图是一种特殊的向量，其精度不会丢失。以向量形式表示的图像特征容易使用并且易于与其它特征融合。BoG 特征是模仿用词汇表达文本的方式，采用图像中关键点的特征来表示图像。不同的特征向量需要采用合适的距离函数，以便达到最佳的效果。基于向量表示形式的特征常用的距离函数有以下几种。

1. 闵可夫斯基(Minkowski)距离

Minkowski 距离是一组距离的定义,其可以看做是广义的欧式(Euclidean)距离与街道(Manhattan)距离。两个维度为 n 向量 u 与 w 之间的 Minkowski 距离定义如公式 2.3 所示。

$$d(u, w) = \left(\sum_{i=1}^n |u_i - w_i|^p \right)^{\frac{1}{p}} \quad (2.3)$$

其中 p 为正整数。

当 $p=1$ 时,其形式如公式 2.4 所示。

$$d(u, w) = \sum_{i=1}^n |u_i - w_i| \quad (2.4)$$

即街道(Manhattan)距离。

当 $p=2$ 时,其形式如公式 2.5 所示。

$$d(u, w) = \sum_{i=1}^n (u_i - w_i)^2 \quad (2.5)$$

即常用的欧式(Euclidean)距离。

当 p 区域无穷大时,其是契比雪夫(Chebyshev)距离。

2. 直方图相交

直方图相交距离匹配方式主要用于稀疏直方图特征之间的匹配,其计算比较快,定义如公式 2.6 所示。其中 u, w 为 n 维的特征向量。

$$d(u, w) = 1 - \sum_{i=1}^n \min(u_i, w_i) \quad (2.6)$$

3. JSD 距离

JSD(Jensen-Shannon divergence)距离常用于一般直方图特征之间的匹配,其定义如公式 2.7 所示,其中 u, w 为 n 维的特征向量。

$$d(u, w) = \sum_{i=1}^n \left(u_i \log\left(\frac{2u_i}{u_i + w_i}\right) + w_i \log\left(\frac{2w_i}{u_i + w_i}\right) \right) \quad (2.7)$$

4. 卡方(χ^2)距离

卡方距离在计算含有空间信息的,如金字塔化的图像特征,图像特征之间的距离比欧式距离,直方图相交距离性能要好。其定义如公式 2.8 所示,其中 u, w 为 n 维

的特征向量。

$$d_{\chi^2}(u, w) = \sum_{i=1}^n \frac{(u_i - w_i)^2}{u_i + w_i} \quad (2.8)$$

BoG 形式特征的距离匹配首先根据向量形式的距离函数计算相似单词元组，然后统计两幅图像之间相似单词的个数，采用相似单词的个数来确定图像之间的相似度。

在具体的应用中，根据需求往往需要多个图像特征对图像进行描述，不同的特征会产生不同的距离。为了得到最终图像之间的相似度，需要融合多个特征。融合多个特征的方式一般分为线性融合与非线性融合。常用的线性融合方式如公式 2.9 所示。

$$S = \sum_{i=1}^n (d_i w_i) \quad (2.9)$$

其中 d_i 是由不同特征计算出的距离， w_i 是对应的权重。

2.3 小结

本章首先介绍了系统的模块构成，并对特征提取，图像分类，相似度获取，相似度排序，索引等模块做了简单的描述，接着介绍了系统的数据处理流程，然后对获取图像相似度的过程的流程做了重点阐述。

本章还介绍了基于图像内容搜索的应用中获取图像相似度的一般方法，图像首先经过特征提取得到图像特征描述符，图像可能具有多个特征描述符，然后采用距离函数计算不同特征描述符的距离，最后融合不同的特征得出的距离得到图像之间的相似度。围绕获取图像相似度的流程，对图像特征的形式，不同的距离函数以及多个特征融合的方法进行了描述。对影响图像特征的因素做了简单的阐述，并介绍了常用的解决方案。

3 外观专利图像的特征提取方法

本章主要围绕提取外观专利图像的特征描述符展开论述。首先分析了专利图像的特点，获取表达专利图像所需要的特征以及性质，引出本章的重点。接着根据专利保护的图像颜色、形状等方面，阐述了对应的颜色特征、形状特征提取方法。最后描述了提取专利图像语义特征的方法，提出了语义特征提取框架并描述了专利图像的语义关联方式。

3.1 外观专利图像特点分析

有众多的图像特征可以使用，并且图像特征可以具有非常好的性质，但是并不是任何图像特征都适用于专利图像，并且由于代价的原因，并不一定要求专利图像特征达到所有的高标准。在具体应用中往往会在考虑众多因素，对图像特征的各个方面进行折中，从而选取合适的特征描述符，即图像特征与应用非常相关。下面分析专利图像的特点。

1. 外观专利图像的特点分析

专利图像是一类特殊的图像，其范围不如互联网上图像那么广泛，同时专利图像也不同于人脸图像，指纹等图像那样的单一。作为一种法律文档，专利图像需要说明保护的内容，也要保证提供的图像能够表达清楚并不存在歧义，这些是专利图像相对互联网上图像的优势。但是专利图像本身也具有多样性，不能算作传统意义上的垂直图像领域。

外观专利在申请时一般都会指定保护的方面，如请求保护外观设计的色彩，轮廓，内部图案等信息。图像的特征需要表达出专利保护的方面，这样才能达到用户的意图。

为了能够表达清楚外观专利的内容，专利所提交的图像需要提供各个视角的图像，并且提交的图像的尺寸大小均匀，明暗度适中。专利保护的都是现实世界存在的物体，所以其提供的图像一般都具备一定的形状。

从专利库中挑选出的专利图片样本如图 3.1 所示，通过该集合分析，得出专利图像更多的特点。图中左上角的图片其特征主要在于其内部物体，一朵花能够表达整张图像的特点，但是其并不需要外部形状对其进行描述，颜色也是次要的特征。最



图 3.1 专利图片样本

后 2 张图像都是领带，但是他们的存在一个较大的方向改变，其形状特征需要具有旋转不变性。同时部分图像，如衣服，因为摆放的原因，会在局部产生微小的变换，所以其特征需要能够容忍这些微小的变化，也能够容忍局部噪声产生的影响。

2. 表达专利图像所需要的特征及其性质

通过以上分析，可以看出需要从颜色，形状，外观等方面表达专利图像，即需要颜色特征，形状特征，外观特征，也要考虑图像的空间分布。同时专利图像表的语义范围有限，并且专利图像的语义定义很明确，所以可以提取专利图像的语义特征。

专利图像特征描述符的性质，专利图像会存在部分的旋转，所以要求对应的图像特征要具备旋转不变性。专利图像的位置会发生变化，所以要具备平移不变性。由于拍照的问题，图像特征要局部一定程度的尺度不变，同时也要能够容忍局部微小变化。但是特征描述符对视角变化，光照变化要求不高。

专利图像大都需要从局部对其进行把握，但是考虑到图像特征需要应用在图像检索中，但是局部特征提取的速度比较慢，所以不能直接在图像中使用局部特征。

3.2 外观专利图像的底层视觉特征提取方法研究

图像特征根据最终的特征描述符是否含有所有像素的信息分为全局特征与局部特征。全局特征主要包括颜色，纹理，形状等特征，其具有一定的通用性，但是区分能力不强。局部特征是采用图像中部分关键点周围像素区域的特征，来代表整幅图像的特征，这类图像特征的判别能力比全局特征强，但是其提取难度比较大。

3.2.1 专利图像的颜色特征

直接采用颜色直方图获取的颜色特征，其区分能力并不强。因其提取方式过于简单，完全忽略了图像内像素的空间分布，不同图像具有相同的颜色直方图如图 3.2(a) 与图 3.2(b) 所示，但是其视觉上不一样^[34]。外观专利提交的图像，一般具有一定的摆放方式，如提交的汽车图像，轮子总在下面。所以很有必要融入空间信息。

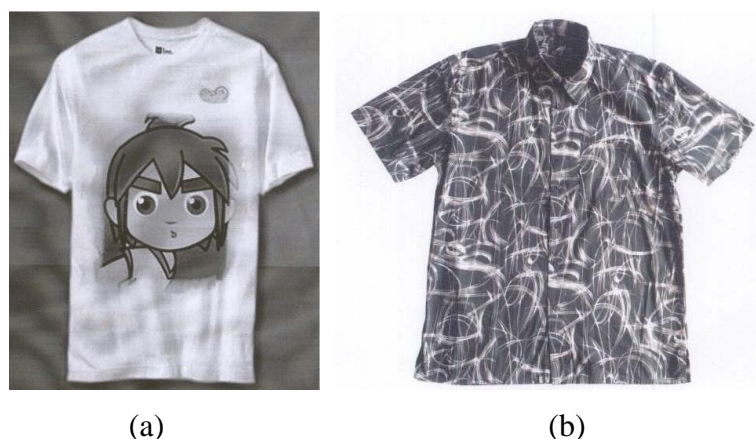


图 3.2 不同图像具有相同的颜色直方图

造成这种问题的根本原因是没有考虑图像内像素之间的关系，为解决由于空间信息引起的颜色特征区分能力不足的问题，采用划分网格的方式提取图像的颜色特征。划分网格是在计算机视觉中常采用的方式^[35]，最常用的是把图像划分成 4×4 个网格，图 3.3 给出了划分网格的示意图。



图 3.3 网格划分示意图 4×4

划分网格后，在网格区域内提取图像的颜色特征。由于划分网格后提取的颜色特征会造成图像特征维度急剧膨胀，所以要在网格数量与颜色直方图量化区间个数上做出调整，以达到合适的效果。

对于 RGB 颜色通道的彩色图像 I , 采用公式 3.1 量化其像素值。即每个通道取字节的最高 2 个 bit, 3 个通道共有 6 个 bit 的数值, 得到量化区间是 64, 即得出的颜色特征维度是 64。采用 4x4 网格划分后, 在每个方格的小区域内提取颜色直方图, 最终得出融合空间信息的颜色特征描述符维度为 $64 \times 16 = 1024$ 。

$$f(I) = \frac{B}{2^6} \times 2^4 + \frac{G}{2^6} \times 2^2 + \frac{R}{2^6} \times 2^0 \quad (3.1)$$

颜色还有其他众多特征, 如颜色矩, 颜色相关图等, 经过试验测试表明采用划分网格方式融入空间分布的颜色直方图的效果用于专利图像检索效果较好。

3.2.2 专利图像的形状特征

专利图像大都含有丰富的内部轮廓信息, 传统的基于轮廓的描述符, 只利用图像最外层的轮廓信息, 这样会导致损失很多有用的信息, 从而导致特征判别性低下。如图 3.4 具有相同轮廓的不同物体, 如果采用基于轮廓的描述符, 则不能区分这两者之间的区别, 所以形状特征必须考虑图像的内部轮廓信息。

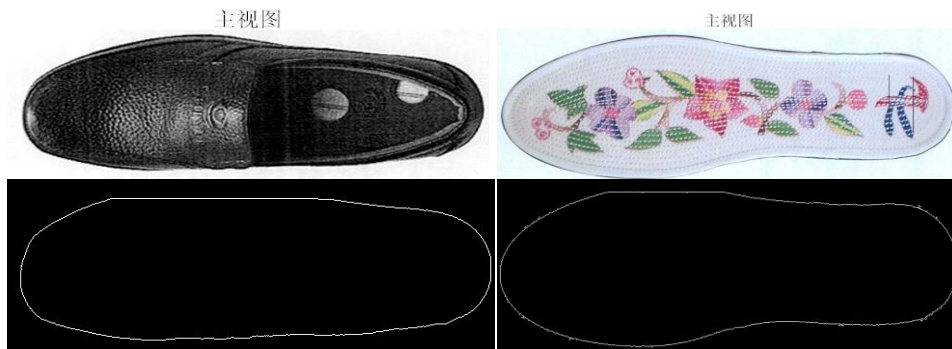


图 3.4 具有相同轮廓的不同物体

基于统计内部轮廓信息的形状特征有形状上下文, 边缘方向直方图, 梯度方向直方图^[15]HOG (Histogram of Orientation Gradients)等。

为了兼顾专利图像的特点, 得到更强表达能力的形状特征, 形状特征需要考虑图像的内部轮廓, 并且同时要融入空间信息。结合金字塔内核^[37], Bosch 与 Zisserman 对 HOG 进行了改进, 得到金字塔化的梯度方向直方图^[36] PHOG (Pyramid of Histogram of Orientation Gradients)。金字塔内核^[37]可以从粗糙到精细的描述图像信息, 其是对象分类与模式识别常采用的一种方式。与 HOG 一样, PHOG 也是提取像素点的梯度方向, 与其不一致的是采取的策略不同。PHOG 在预先划分的方格内进

行提取梯度方向，而 HOG 是在检测窗口中提取梯度方向，PHOG 划分网格的方式如图 3.5 所示。

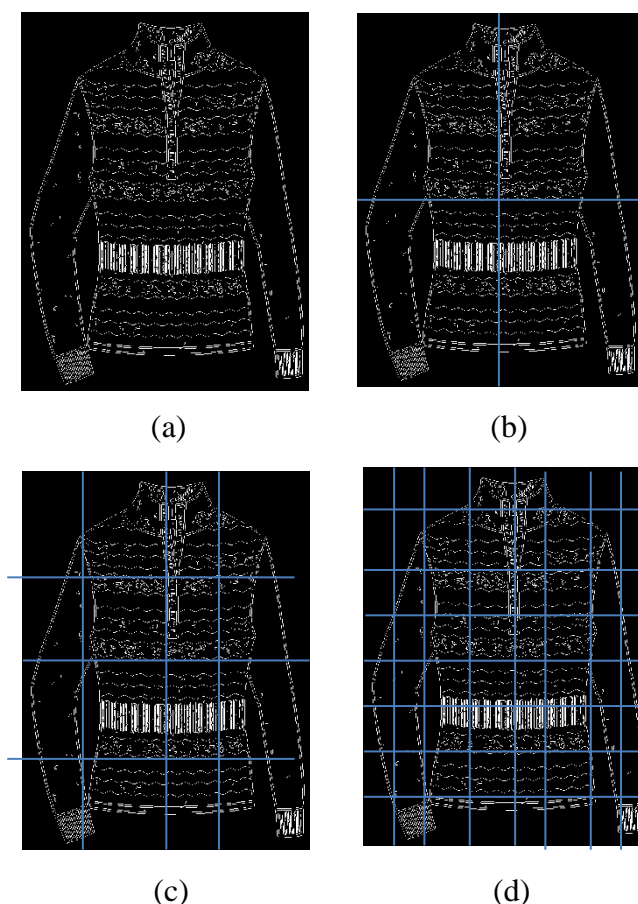


图 3.5 PHOG 划分网格方式

PHOG 形状特征描述符提取步骤如下：

- 1) 把输入图像转换成灰度图像 $Gray$;
- 2) 使用 canny 算子提取灰度图像 $Gray$ 的边缘 $Edge$ 。
- 3) 计算灰度图像 $Gray$ 在两个方向上的梯度 G_x, G_y ，并根据两个方向梯度计算综合梯度 $Gr = \sqrt{G_x^2 + G_y^2}$ 。
- 4) 根据上一步得到的方向梯度值，计算每个像素点的角度 $Angle = \text{atan}(G_y, G_x)$ ，并把每个像素点角度量化到 8 个区间。量化方法为 $f(Angle) = \frac{Angle \times k}{360}$ 。
- 5) 获取角度矩阵 $MatrixAngle$ 用于存放角度量化值，综合梯度矩阵 $MatrixGradient$ 用于存放综合梯度值。针对边缘图像 $Edge$ ，在边缘为 0 的点，

$MatrixAngle$ 与 $MatrixGradient$ 都为 0；不为 0 的点， $MatrixAngle$ 存放其角度量化后的值， $MatrixGradient$ 存放综合梯度值。

6) 求取方向梯度直方图。在网格范围内统计像素点的梯度幅值，相同的角度存放在相同的桶中。

7) 获取金字塔化的梯度方向直方图。如图 3.5 所示，划分网格单元个数依次为 $[2^{0 \times 2}, 2^{1 \times 2}, \dots, 2^{2 \times n}]$ ，每次在水平与垂直方向上增加一个划分线。每个网格中获取 K 个方向向量，则最终描述符大小为 $K \times \sum_{g=1}^n 2^{2g}$ 。

对应图 3.5 的划分网格获取的梯度方向直方图如图 3.6 所示，其中最大角度取 360 度，量化角度个数为 20。图 3.5(a)到图 3.5 (d)的图像的直方图与图 3.6(a)到图 3.6(d)一一对应，从直方图的分布可以看出，直方图特征是逐步求精的，直方图数量越来越多。

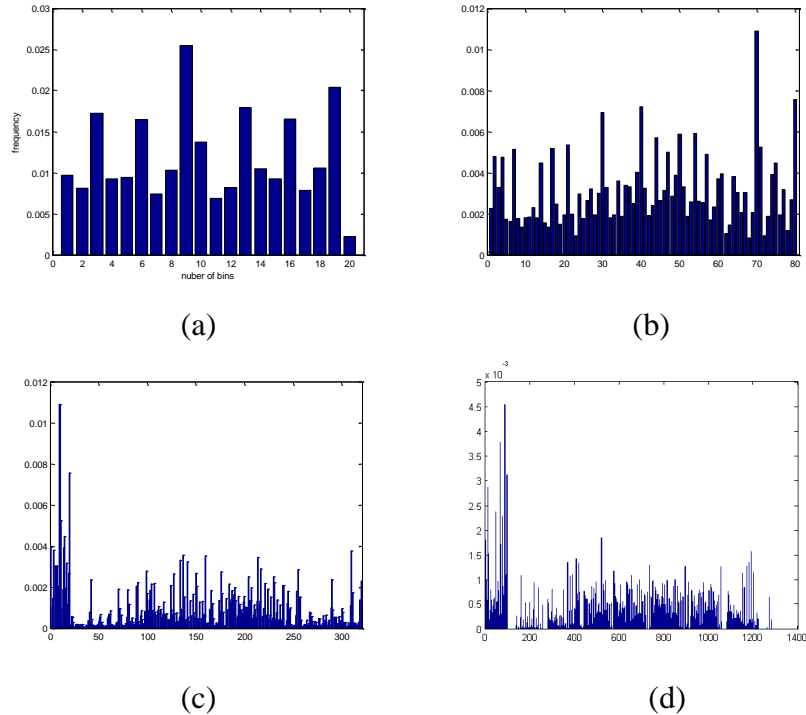


图 3.6 划分网格获取的梯度方向直方图

几何形状特征主要通过图像的梯度信息来反映图像的总形状，采用基于梯度强度分层的梯度方向直方图表示形状特征。与金字塔核心在图像平面上划分区域不同，基于对比度分成是按照图像的梯度强弱进行划分层次。如图 3.7 所示，图像的边缘是采用 canny 算子提取的，三个图像按照梯度强度划分，从左到右逐渐减弱。

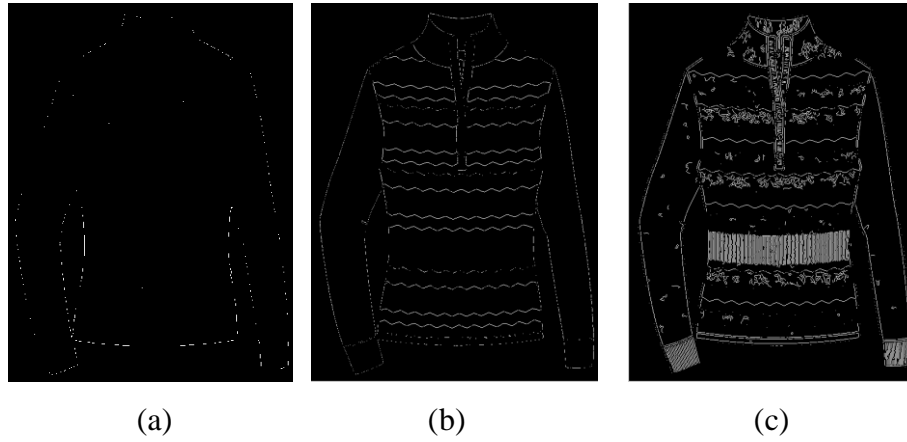


图 3.7 基于梯度强度分层方式

基于梯度强度分成的梯度直方图特征提取方式如下：

- 1) 获取图像 I 的灰度图，记为 $Gray$ ；
- 2) 计算灰度图像 $Gray$ 中每个像素在 x,y 方向上的梯度，记为 G_x, G_y ；
- 3) 根据像素的 x,y 方向上的梯度计算其综合梯度 $Gr = \sqrt{G_x \cdot G_x + G_y \cdot G_y}$ ，计算其角度 $Angle = \text{atan}(G_x, G_y)$ ；
- 4) 使用 $canny$ 算子提取图像 I 的边缘图像，记为 $Edge$ ；
- 5) 根据 Gr 的值把边缘分成 K 个层次，是每个层次非 0 的像素点个数相同；
- 6) 在每个层次统计梯度方向直方图。像素的角度 $Angle$ 在 0 到 360 度之间，分成 M 个区间， M 在 8 到 64 之间；
- 7) 收集说有层次的梯度方向直方图得到最终的几何形状特征，其维度 $dim = M \cdot K$ 。

提取的基于梯度强度分层的梯度方向直方图如图 3.8 所示，梯度方向采取无符号表示，划分到 20 个区间。3.8(a)s 所示图像是最终的特征直方图，3.8(b)，3.8(c)，3.8(d) 是对应图 3.7 的梯度方向直方图。

基于梯度强度分层的梯度方向直方图有如下优点：1.减少了噪声对特征的干扰，由图像内物体外轮廓产生的梯度强度比较大，其轮廓信息会在层次图中保留下来，而由噪声产生的梯度强度比较弱，其只在最后几副图像中有其信息，通过这种方式达到了减少噪声污染的目的。2.具有旋转不变性，从其提取过程很容易看出，基于梯

度强度分层的梯度方向直方图其本质是基于统计的，基于统计方法获取的特征能够保持旋转不变性。3.对局部变化稳健，提取直方图的过程中对边缘方向进行了量

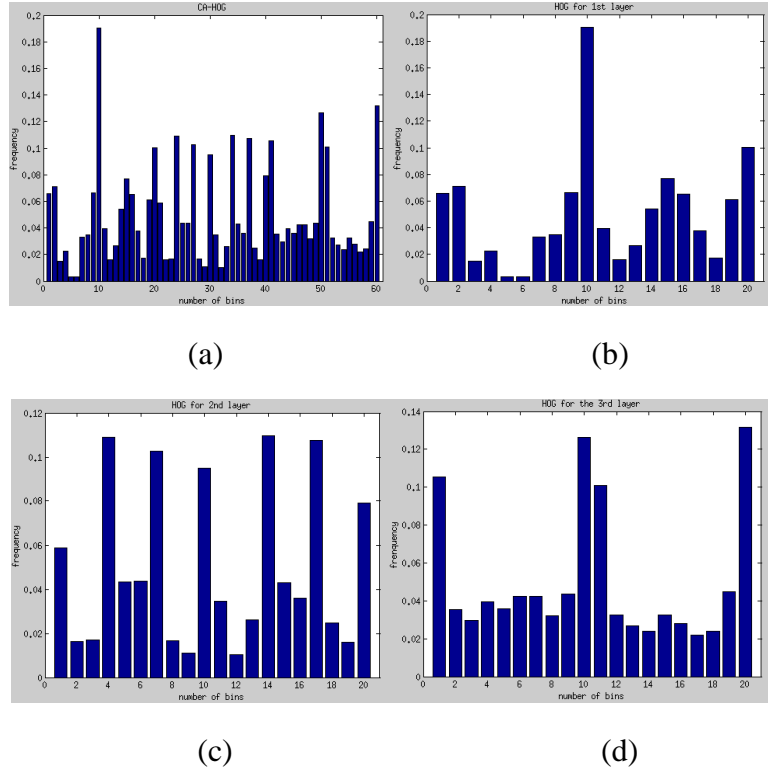


图 3.8 基于梯度强度分层的梯度方向直方图

化处理，局部微小变化会量化到同一个区间，不会影响到最终的特征。

3.2.3 基于局部特征的外观特征

图像的外观特征是表达人对图像综合感觉的一种特征。局部特征起源于生物视觉成像的启发，Hubel 和 Wiesel 对猫的视觉神经元实验表明生物视觉系统是基于局部特征的，因此局部特征的判别能力更加接近人类视觉的判别能力，所以采用局部特征来获取专利图像的外观特征。局部特征在模式识别，图像分类中使用广泛，取得了很好的效果。

局部特征提取主要包括两个步骤，其一是关键点检测，其二是描述关键点周围的兴趣区域。关键点检测需要在不同尺度空间内检测图像区域的极值点，这些极值点一般是边缘，斑点，拐角等信息，常用的关键点检测算子有 Harris-Laplacian corner^[38]，Laplacian of Gaussian^[33]。兴趣区域描述符是把关键点周围像素转成特征向量的一种方法，在众多局部特征描述符中，Mikolajczyk 与 Schmid 在 2005 年经过评测^[22]后指出，类 SIFT 描述符在分类与目标识别中表现最好。

SIFT 特征可以很好从局部对图像进行描述，但是 SIFT 特征的提取速度非常慢，

限制了其在图像检索中的使用。所以需要对 SIFT 进行改进，鉴于 SIFT 特征耗时的主要部分在关键点检测上，这里使用密集采样的 SIFT 特征，即 DSIFT^[39]特征描述符，DSIFT 在不同的空间尺度中直接描述图像的局部特征，从而保持了原始 SIFT 特征尺度不变优点。图 3.9 给出了 DSIFT 采样方式与 SIFT 采样方式的区别。左边的图像是原始 SIFT 采用 DoG 算子检测出的区域示意图，右边是 DSIFT 密集采样的区域示意图。图中 SIFT 采用的 DoG 算子的参数如下，关键点过滤阈值为 0.0067，边缘边界值为 10。DSIFT 的参数是尺寸大小为 20，采样区域步长为 25。

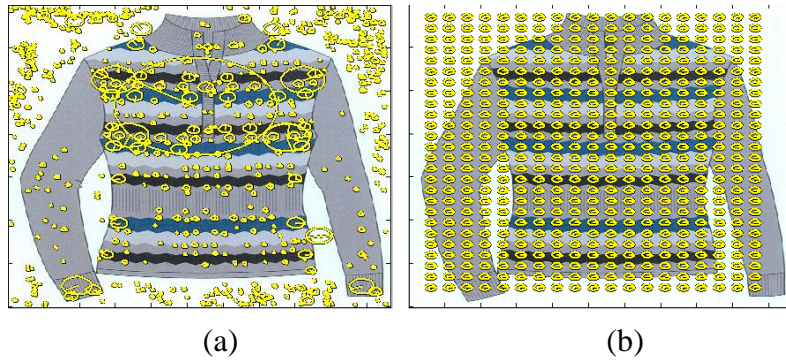


图 3.9 SIFT 与 DSIFT 采样效果对比图

图像的 DSIFT 特征仍然是 bag of words 形式的，为了便于使用需要对其进行量化处理，量化方法如图 3.10 所示，首先建立量化码本，然后根据码本得到量化的外观特征描述符。经过量化后得到的外观特征描述符是简单的向量形式，其维数与码本大小一致，并且所有图像的特征的维数相同。

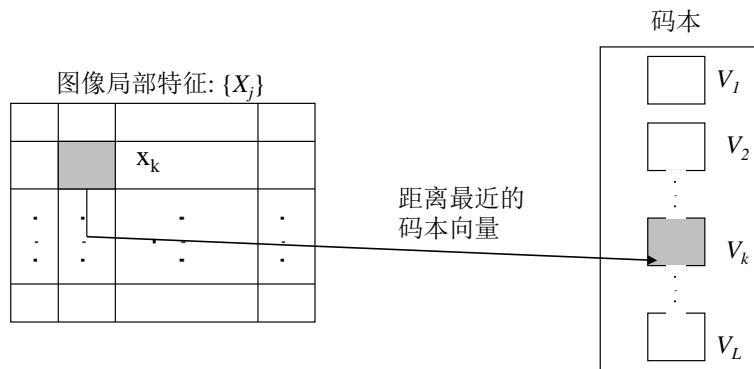


图 3.10 局部特征量化方式

建立图像 DSIFT 特征的码本，其步骤如下：

1) 选取部分图像集合，提取集合中图像的 DSIFT 特征，特征表示为 bag of words 形式。

2) 对图像集合的 bag of words 形式的特征使用 KMEANS 进行聚类, 得到 K 个聚类中心。

3) 使用 KDTREE 对上一步的聚类中心进行聚类, 聚类中心的索引。该步骤是为了加快查询的速度。

根据码本量化图像 DSIFT 特征, 得到量化的外观特征描述符:

1) 对图像的所有 DSIFT 特征, 使用公式 $d(X_j, V_k) = \frac{1}{n} \sum_{i=1}^n (X_j(i) - V_k(i))^2$ 在码本的 KDTREE 索引上进行一次查询, 其中 X_j 是图像的 words 形式的特征, V_k 是码本中的 words, 得到一组距离值, 选取最小距离的标号, 保存所有标号。

2) 经过上一步图像的所有 DSIFT 都一一对应一个标号, 然后按照码本的顺序统计各个标号出现的次数, 保存为向量, 即得到量化的外观特征描述符, 其维度 $dim=K$ 。

为得到描述性更强的外观特征描述符, 融合空间分布信息, 如图 3.11 所示, 把图像在两个方向分成 4×4 个区域, 在小区域中获取外观特征的描述符, 收集小区域的外观特征描述符, 得到整个图像的特征描述符, 其维度 $dim=16 \times K$ 。

3.3 专利图像的高层语义特征提取方法研究

传统的图像检索采用图像的底层视觉特征之间的相似性来代替图像之间的相似

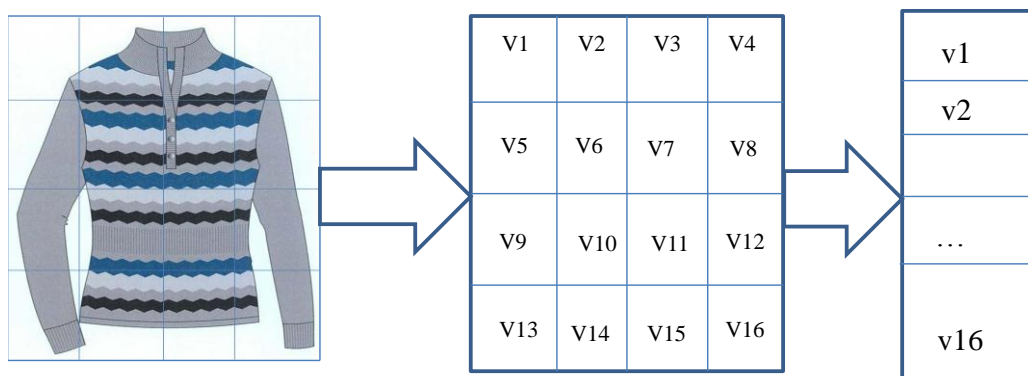


图 3.11 划分区域得到分区域的外观特征描述符

性, 这类系统忽略了人类对图像的理解, 其既不知道用户的意图也不知道图像之间的语义关系, 即忽略了图像底层视觉特征与高层语义之间语义鸿沟的存在, 所以其性能不理想。语义特征是为缩小语义鸿沟而提出的, 是图像的高层特征。图像的高层语义特征与底层视觉特征的区别是, 语义特征融合了人类对图像的理解, 同时高层语义特征提取难点在于如何获取图像的语义信息, 并把获取的语义信息关联到图像的底层视觉特征。

3.3.1 专利图像语义特征获取流程

1. 语义特征提取框架

为方便表述，首先给出语义特征提取框架，如图 3.12 所示。从图中可以看出，框架的核心部分是语义特征模型。图中水平方向指出了语义特征计算模型的建立过程，其是采用 **SVM** 分类器在人工挑选的分类样本上经过学习训练得出的，其融合了图像的语义空间与图像的底层视觉特征。

这个框架适用于图像数据库的语义空间有限，并且容易划分的情况。对于观察者来说，图像的语义空间有限就是图像所表达的知识范围有限。在图像知识范围有限并且容易划分的情况下，可以针对细分的知识范围，挑选出样本图像，采用样本图像的语义空间代替图像数据库的语义空间。为使图像的语义信息与图像的底层视觉特征进行关联，采用同一个知识范围内的样本图像都带上人工确定的语义标签。

图像的语义特征是图像对观察者所传递的知识，其在语义空间中表示为空间中的点。图 3.12 中垂直方向给出了图片语义特征的获取过程。首先提取图像的底层视觉特征，然后把图像的底层视觉特征带入语义特征模型，语义特征模型根据事先定

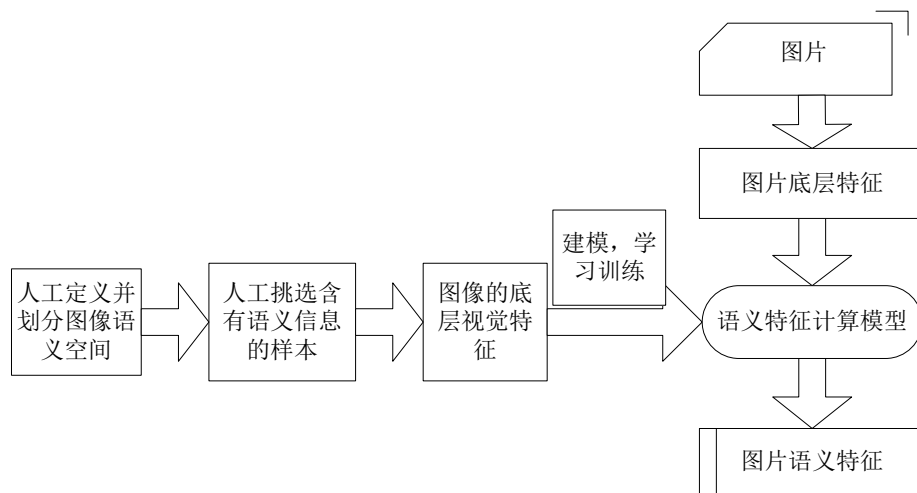


图 3.12 专利图像语义特征提取框架

义的规则产生语义特征。

2. 描述专利图像的底层视觉特征

图像的底层视觉特征在图像的语义特征中占有重要的地位，语义特征融合了图像的高层语义知识与图像的底层视觉特征，其底层视觉特征要具有很强的判别性。

针对专利图像的性质，大部分图像可以根据其几何形状进行区分，少部分图像需要根据内部物体的外观进行区分，需要外观特征的图像如图 3.13 所示，其内部是几朵花，其需要外观特征进行描述。为了与定义明确的语义空间对应，图像的底层视觉特征融合外观特征与几何形状特征。



图 3.13 需要外观特征的图像

外观特征能够很好的区分内部物体比较复杂的图像，大多数图像都可以采用形状进行区分。单独的外观特征与形状特征并不能完全区分出不同的专利图像，所以需要融合两个特征得到区分性更强的底层视觉特征。

采用公式 3.2 对两个特征融合,该公式是线性融合方式，其中 α 与 β 是形状与外观特征的权重， X, Y 分别代表两张不同的图像， X_{app} 代表图像的外观特征， X_{shp} 代表图像的形状特征。

$$K(x, y) = \alpha K_A(X_{app}, y_{app}) + \beta K_s(X_{shp}, y_{shp}) \quad (3.2)$$

该公式属于 Mercer^[40]核心，通过调整参数 α 与 β 的权重值获取最终的融合特征，经过测试，调整参数 α 为 2， β 为 1，加强外观特征的权重，达到较好的分类效果。

3.3.2 定义专利图像语义空间

由于外观专利在同一个大类别中的知识范围有限，所以专利图像的语义空间采用人工事先划分。划分专利图像语义空间前首先要得到专利图像的语义信息，然后再确定划分方式，最后确定语义空间的参考基。

1. 专利图像语义信息的获取方式

得到语义特征的一个关键步骤是获取图像的语义信息。外观专利保护的对象是图像中的物体，并且图像的物体承载了图像的语义。外观专利具有人工定义类别，

其按照层次定义，分为大类，小类信息，所以专利的图像知识非常明确且范围由限，并且专利作为一种法律文件决定了其图像语义不具有多义性。所以专利图像的语义信息容易定义。下面从专利文档分类与专利图像内物体两个方面讲述获取专利图像的语义获取方式。

1) 专利文档

根据中华人民共和国专利法^[41]的规定，申请外观专利应当写明使用该外观专利设计的产品及所属的类别。其大类主要有：01 类食品，02 类服装和服饰用品，12 类运输或提升工具等 31 类。同时专利分类表中也确定了小类信息，如 02 类服装与服饰用品含有的小类如下：02-01 内衣、女内衣、妇女紧身胸衣、乳罩睡衣；02-02 服装；02-03 帽子；02-04 鞋、短袜和长筒靴袜等 8 个小类。

2) 专利图像内物体

专利图像是外观专利的主题，是专利保护的内容。并且图像内的物体单一，可以通过特定的特征对其进行识别与分类，如汽车可以通过形状与局部特征进行区分，衣服可以通过形状与外观进行区分，然后辅助人工添加的语义标签，可以获取图像的类别，由于类别涵盖了人工定义的语义信息，并且同一类图像具有隐含相似的图像特征。通过建立图像底层特征与高层语义之间的模型，该模型可以对任意图像进行预测，确定该图像所属的类别。

从以上分析可以看出，从专利文档中可以得出图像的粗粒度语义信息。从图像底层特征入手获取图像更精细的语义信息，所以从专利文档的归类与图像的底层特征入手可以确定专利图像的语义信息。

2. 定义专利图像的语义空间

专利图像语义空间可以通过小样本空间进行代替。从上一节分析可以看出，不同大类定义比较清晰，可以用来对语义进行粗粒度划分。但是其小类信息定义比较抽象，难以与底层图像关联，不便于机器处理。因此，需要根据具体的图像定义明确的语义。为了得到专利图像的语义，采用人工挑选样本的方法来确定样本空间。

确定专利图像的语义空间具体方式如下：首先根据外观专利提供的大类对图像进行一次粗糙分类；然后在大类中参考小类信息，根据图像内物体的轮廓，外观进行精细分类；再根据精细的分类，人工挑选出具有代表性的样本。挑选样本时需要注意，所有类别的样本对于观察者来说，要能够通过底层视觉特征进行区分。例如

02 大类总共分为 53 个类别，每个类别选出 45 张样本图片，语义空间由 53 个类的图片隐式表达。

3. 获取语义空间参考基

语义空间可以通过一组参考基进行表达，空间内的任意点可以通过该参考基进行表达。语义空间的参考基通过 SVM 分类器在样本的底层视觉特征上训练学习得到，是每个类别分类面的集合。支持向量机常用于分类中，其分类主要任务是获取两个类别之间的分类面。其分类原理如下所述。

如图 3.14 所示，图中是支持向量机在二维空间划分数据的方式。图中实心点与空心点分别代表不同的训练样本， H 为两类数据的分类线， $H1$, $H2$ 是过各类样本中与分类线 H 距离最小的且平行的直线， $H1$ 与 $H2$ 之间的距离是两类分类间隔，支持向量机需要找到一个分类线能把两类无误的分开，并且还要使两类的分类空隙最大。在高维空间中最优分类线既是最优分类超平面，对于超平面获取超平面的法向量称为支持向量。设线性可分样本集为 $(x_i, y_i), i=1, \dots, n, x \in R^d, y \in \{+1, -1\}$, y 类别标号， d 是维度数目。则 d 维空间中线性判别函数的一般形式为 $g(x) = wx + b$, 分类面方程如式 3.3 所示：

$$wx + b = 0 \quad (3.3)$$

判别函数最后进行归一化，达到 $|g(x)| \geq 1$ 。

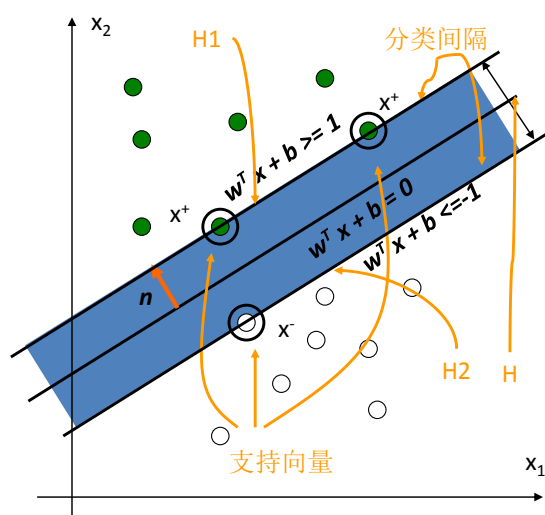


图 3.14 二维空间中支持向量机原理示意图

专利图像的一个大类可以划分为多个小的类别，每个类别隐含了一个参考基，

该参考基通过 SVM 学习得出，是每个类别的最优分类超平面。因此，每个分类通过 SVM 获取的超平面可以作为语义空间的参考基。

3.3.3 获取专利图像语义特征计算模型

在得到了关联语义信息，并且提取了图像的底层视觉特征后，就可以获取图像语义特征计算模型。语义特征计算模型是获取语义特征的核心部分，其融合了底层视觉特征与图像的语义信息，能把图像的底层视觉特征转换为图像的高层语义特征。

语义特征计算模型是计算图像底层视觉特征与语义空间参考基之间距离的计算器。获取语义特征计算模型示意图如图 3.15 所示。

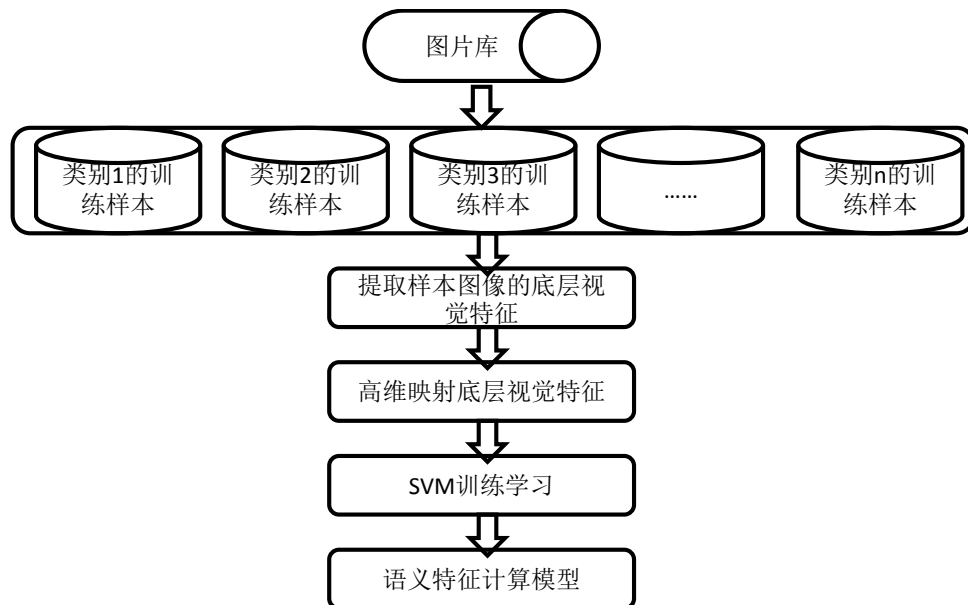


图 3.15 获取语义特征计算模型示意图

获取语义特征计算模型步骤为：

1) 人工分类并挑选出样本图像。首先根据专利图像的轮廓把图像分成多个小类，每个类别挑选出 45 张代表性的样本图像，其中 30 张用于模型训练，15 张用于测试。

2) 提取样本图像的底层视觉特征，DSIFT 外观特征描述符与基于梯度强度分层的形状特征描述符。其中在形成 DSIFT 码本时 KMEANS 聚成 600 个聚类中心，得到的量化后的外观特征描述符维数为 $600 \times 16 = 9600$ ；提取基于梯度强度分层的梯度方向直方图时层次划分为 3 层，保持每个层次的像素个数基本相同，像素的角度采用

无符号的形式，并把像素的角度量化到 20 个区间中，形状特征的维数为 $20 \times 3 = 60$ 。并采用公式 3.1 对这 2 个特征进行融合，为保持均衡，两个权重都设置为 1。

3) 获取图像底层特征在高维空间的向量。为使底层特征线性可分，在进行训练前需要把图像的底层特征映射到高维空间中。采用卡方核函数做高维映射，得到高维向量采用的阶码是 2，即得到映射后的向量维数是原始向量维数的 3 倍。

4) 采用线性核函数的支持向量机获取每个分类的判别式。训练数据设置为一对多的方式以利用 SVM 来进行多分类，即该类的样本数据为正样本，其它所有类别的样本设置为负样本。每个类得到的判别式为 $wx + b$ ，其中 w 是 SVM 训练出的超平面的支持向量， b 是偏移量， x 是图像底层特征对应的高维度向量。

5) 获取语义特征模型。收集所有类别的判别式，获取语义特征模型。

图像的特征在进行训练之前需要把其映射到高维空间中，以保证其线性可分。采用的映射核函数是卡方核^[42]，如公式 3.4 所示：

$$K_{\chi^2}(x, y) = 1 - \sum_{i=1}^d \frac{(x_i - y_i)^2}{\frac{1}{2}(x_i + y_i)} \quad (3.4)$$

其中 x_i, y_i 为原始向量， $K_{\chi^2}(x, y)$ 是映射后的向量，原始向量的维度为 d 。卡方核函数属于同类映射核函数，设原始向量 $x \in R$ ，则映射后的向量为 $\Psi(x) \in R^{2n+1}$ ，得到的映射后的向量为 $\langle \Psi(x) \rangle$ ，其维度为 $d(2n+1)$ 。

3.3.4 获取专利图像的语义特征

1. 获取语义特征

语义特征是语义空间中的一个点，其可以通过底层视觉特征经过语义特征计算模型转换得到。语义特征模型在专利语义空间中是每个类判别式的集合，把图像底层特征对应的高维空间向量带入判别式 $wx + b$ ，得到的值向量即为图像的语义特征。

图像在语义空间中的表示示意图如图3.16所示，平行四边形代表各个分类的超平面，椭圆是图像在高维空间中的超向量，带箭头的线是图像的超向量到各个超平面的距离，虚线是到各个超平面的最大距离。

在分类应用程序中，只使用最大距离，表明该图像最可能属于对应的类，分类模型用来判别分类类别，如图3.16中用虚线表示。由于支持向量机获取了每个类别的超

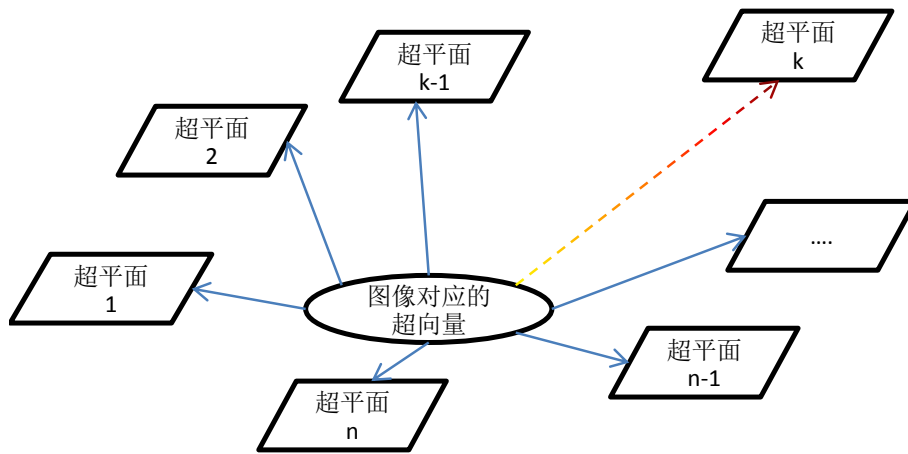


图 3.16 图像在语义空间中的表示示意图

平面的法向量，这些法向量形成了高维空间的一组参考基。图像在高维空间中的超向量可以通过该参考基进行表示，从而得到图像的语义特征。

获取图像语义特征的具体步骤如下：

- 1) 提取图像的底层视觉特征，外观特征描述符与形状特征描述符，得到的融合特征为 T 。
- 2) 使用卡方核函数对图像的底层特征 T 进行高维度映射，得到高维度向量 V 。
- 3) 把得到的高维向量 V 带入语义特征模型中，对每个分类判别式 $wx+b$,得到一个值，每个类分别获取一个值得到一组向量，即为图像的语义特征。

获取的图像语义特征的维数与人工确定的分类数目有关，在02类中，其维度是53维。该维度大大低于图像的底层视觉特征的维度，如外观特征描述符为9600维，而形状特征为60维。并且得到的特征是简单的向量形式，非常适合在图像检索中使用。

在分类应用程序中，只使用最大距离，表明该图像最可能属于对应的类，分类模型用来判别分类类别，如图3.16中用虚线表示。由于支持向量机获取了每个类别的超平面的法向量，这些法向量形成了高维空间的一组参考基。图像在高维空间中的超向量可以通过该参考基进行表示，从而得到图像的语义特征。

获取图像语义特征的具体步骤如下：

- 1) 提取图像的底层视觉特征，外观特征描述符与形状特征描述符，得到的融合特征为 T 。
- 2) 使用卡方核函数对图像的底层特征 T 进行高维度映射，得到高维度向量 V 。

3) 把得到的高维向量 \mathbf{V} 带入语义特征模型中, 对每个分类判别式 $w\mathbf{x}+b$, 得到一个值, 每个类分别获取一个值得到一组向量, 即为图像的语义特征。

获取的图像语义特征的维数与人工确定的分类数目有关, 在02类中, 其维度是53维。该维度大大低于图像的底层视觉特征的维度, 如外观特征描述符为9600维, 而形状特征为60维。并且得到的特征是简单的向量形式, 非常适合在图像检索中使用。

2. 语义特征的距离比对方法

不同图像的语义特征在语义空间中表现为不同的点, 并且语义特征中的元素可能含有负值, 表明该向量与对应类别的非相关距离, 所以采用欧式距离, 如公式3.5所示, 进行比对。式子中 u, w 表示分别两张不同图像的语义特征。

$$d(u, w) = \sum_{i=1}^n (u_i - w_i)^2 \quad (3.5)$$

3.4 小结

本章围绕提取图像的特征展开论述。首先分析了专利图像的特点, 得出描述专利图像所需要的图像与其性质, 为后续的图像特征的论述做好了铺垫。然后描述了专利图像底层视觉特征提取放, 最后讲述了专利图像的语义特征提取方法。

在讲述提取图像底层视觉特征的过程中也详细描述了空间信息融合方法, 如划分网格方法, 金字塔内核方法。如颜色特征采用分区域的颜色特征提取方式, 其把图像等分成 n 个小区域, 然后在每个小区域内提取颜色直方图, 汇集所有小区域的颜色直方图得到最终的颜色特征。形状特征通过局部形状信息来获取图像的形状特征, 最后利用金字塔核心分层次获取图像的梯度方向直方图得到最终的形状特征。这两个特征都融合了空间分布, 其性能优于原始的特征。

对于底层视觉特征, 全局特征如颜色、形状等特征具有一定的通用性, 但是其判别能力不足, 局部特征虽然判别能力高, 但是其特征维数往往比较高, 并且其表示形式为 Bag of Words 形式不便于在图像检索中使用, 也不便于与其它特征融合, 因此提出了专利图像的语义特征。

获取语义特征最关键步骤是得到语义特征模型, 语义特征模型是结合图像的语义信息与图像的底层视觉特征, 经过支持向量机训练得到的。该模型可以把图像的底层特征转换为语义特征。最后指出了图像语义特征的优势, 其判别能力取决于图

像的底层视觉特征与语义融合方式，其中语义信息经过半监督学习训练在一定程度上可以增加其判别性，并且得出的语义特征是向量形式，其维度低，方便使用，易于与其它特征融合。所采用的语义特征提取框架容易扩展，可以采用其它的底层视觉特征，并可以用于知识范围明确的垂直图像检索系统中。

4 功能测试与性能分析

本章首先介绍了测试环境与测试数据，接着介绍了提供的特征接口。接着对图像特征在特征提取时间，判别性等方面进行了详细的测试，指出这些特征用于图像检索的优势。其中判别性测试采用查准率与查全率的统计方法。最后展示了图像特征用于检索系统中的效果。

4.1 测试环境与数据

4.1.1 图像特征提取测试环境

1. 测试的硬件环境

表4.1显示了测试图像颜色特征，形状特征与语义特征提取时间的机器配置情况。

表 4.1 特征提取算法测试机器的配置

硬件配置	CPU	Intel Core i7, 2.67GHz
	内存	6 GB
	硬盘	1TB
	网卡	Realtek RTL8186/8111 Gigabit
操作系统	Linux	Redhat AS5.5

2. 测试的特征

测试的几个特征有颜色特征，如彩色直方图，划分网格的颜色直方图；形状特征，金字塔化的梯度方向直方图；语义特征与局部特征。彩色直方图采取的参数如下：采用RGB三色通道，每个通道去字节的高3个bit位，特征维度为512；划分网格的颜色直方图参数如下：采用RGB三色通道，每个通道取字节的高2位，网格划分成4x4个单元；金字塔化的梯度方向直方图参数如下：梯度方向采用无符号的360度表示，角度量化的区间数目为8，金字塔化的层次数为3。在后续章节从特征提取时间与判别能力两个方面对各个特征进行测试，其中对语义特征与SIFT进行了对比，然后对各个特征的检索效果进行了展示。表4.2给出了测试特征的基本信息。在说明一列中介绍了特征的获取方式，其对应了维度的来源，其表示形式大都是向量的形式，直方图是精度不损失的向量形式。

表 4.2 测试的特征基本信息

特征		表示形式	维度	说明
颜色特征	彩色直方图	直方图	512	每个通道取高3bit位，共 $2^9=512$ 维
	划分网格的颜色直方图	直方图	1024	4x4网格，每个通道取高2个bit，共 $4 \times 4 \times 2^6=1024$ 维
形状特征	金字塔化的梯度方向直方图(PHOG)	向量	680	角度为无符号360度，角度量化区间为8，金字塔分3个层次，共 $8 \times \sum_{i=0}^3 2^{2^i} = 680$
局部特征	普通SIFT特征	BoW	Nx12 8	区域分成4x4领域，量化角度到8个区间，得到维度4x4x8=128
	密集采样的DSIFT特征	BoW	Nx12 8	同上
	划分网格后量化的局部特征	向量	9600	码本数目为600，网格划分成4x4共16个网格区间，得到向量维度16x600=9600
语义特征	语义特征	向量	N	专利在语义上分为N个类，每个语义分类形成一个向量

4.1.2 测试数据与特征接口

1. 测试数据

系统测试数据采用中国外观专利02类服装与服饰用品，选用六视图中的左视图与立体图，图像数据总量为63, 350。用于对图像特征的性能测试采用的数据为总数据的子集，数量为13,309，其中对于语义特征2634张图像用于进行模型训练，10,674张用于测试。

2. 特征接口

图像特征需要用在检索系统中，为保持系统易扩展与方便使用。对提供的图像特征接口具有一定的要求。系统接受3中形式的特征：向量形式，直方图形式，bag of visual words形式。向量特征一般存储为双精度型，直方图形式的特征为避免精度损失，存储为长整型。其中划分网格的直方图属于直方图特征，金字塔化的直方图属

于向量特征。各个特征的具体格式定义如下。

向量形式的特征，具有3个字段：i.向量特征关键字；ii. 特征维度数目；iii. 数据。

直方图形式特征，具有5个字段：i.直方图特征关键字；ii.直方图个数；iii.直方图量化的区间数目；iv.单个直方图总的统计值；v.单个直方图中元素值。

Bag of Visual Words，具有4个字段：i. BoW特征关键字； ii. Visual words个数；iii. Visual words的维度； iv.按行排列的Visual words.

各个特征的具体参数以注释的形式放在特征文件中，系统接受二进制与文本文件存储的特征，其中以二进制存储的特征不接受特征参数的注释行。

4.2 图像特征性能测试与分析

4.2.1 测试方法

根据图像检索系统对图像特征要求，分别对图像特征的运行时间，特征向量维度，判别性等方面进行了测试。对于通用的特征，如颜色特征，因为其判别能力不足，只进行检索效果展示。而对判别能力稍强的形状特征，语义特征，对其进行了性能对比测试。

测试运行时间时采用标准的lena.jpg图像进行测试，，原始尺寸是512x512维，经过转换到600x600, 512x512, 500x500, 400x400, 300x300, 200x200尺寸对图像的运行时间进行测试。对特征向量维度与特征向量的形式采用表格进行描述。判别性测试使用人工挑选的图像集合，对每张图像进行标注，根据返回结果的数目在10%, 20%, 30%, 50%, 70%, 90%时进行查全率与查准率统计。

查准率与查全率是图像特征性能测试常用的指标。查准率是指在一次检索中，返回的相关结果与返回结果总量的比例，查全率是指返回的相关结果数量与检索集合中相关结果总量的比例。定义如公式4.1，4.2所示。其中 N 表示测试的数据集合记录的总数目， R 表示测试的数据集合中相关记录的总数量， I_R 表示返回结果中相关结果的数目， I_N 表示返回结果的数目。

$$\text{查准率} = \frac{I_R}{I_N} \quad (4.1)$$

$$\text{查全率} = \frac{I_R}{R} \quad (4.2)$$

4.2.2 图像特征测试与分析

图像检索是一类实时性的应用，其既要保证检索速度也要保证检索的效果。其检索速度受特征提取时间与特征维度数目影响较大。其检索效果取决于图像特征的判别能力。

1. 特征提取时间

图4.1给出了颜色特征，形状特征的提取时间与图像尺寸的关系，为保证提取时间具备可重复性，测试采用标准的lena图像进行测试。

从图4.1可看出各个特征与图像的尺寸成正比例关系。划分区域的颜色特征提取与原始的提取方法在时间上相差不明显。采用金字塔化的梯度方向直方图表示的形状特征，其运行时间稍低于颜色特征，与其实实现有关。图像的全局特征因为对图像进行一次扫描，所以其时间差别并不明显，可以达到实时性应用需求。

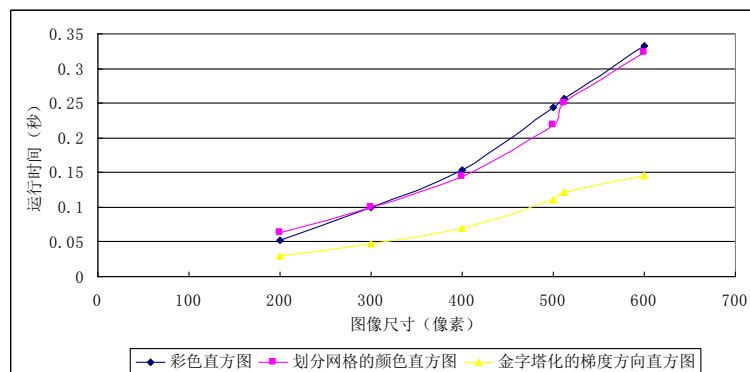


图 4.1 颜色与形状特征与图像尺寸的关系

语义特征底层采用了DSIFT特征与CA-HOG特征，其提取时间与这两者的速度密切相关，另外语义特征需要加载模型，故其提取时间会多于DSIFT特征与CA-HOG特征，而CA-HOG是全局特征，其提取时间相对DSIFT可以忽略，语义特征、DSIFT特征提取时间与图像尺寸的关系如图4.2所示。

2. 特征判别能力评测与分析

语义特征与SIFT特征的查准率如图4.3所示，语义特征与SIFT特征的查全率如图4.4所示。颜色特征，形状等特征的判别性非常低，所以不对其进行判别性测试，在下一节进行效果展示。从图中可以看出语义特征的判别性远远高于局部特征。选用的对比特征是在分类与对象识别中表现比较好的SIFT特征，提取SIFT特征的关键点参数为：关键点过滤阈值为0.0067，边缘边界值为10。

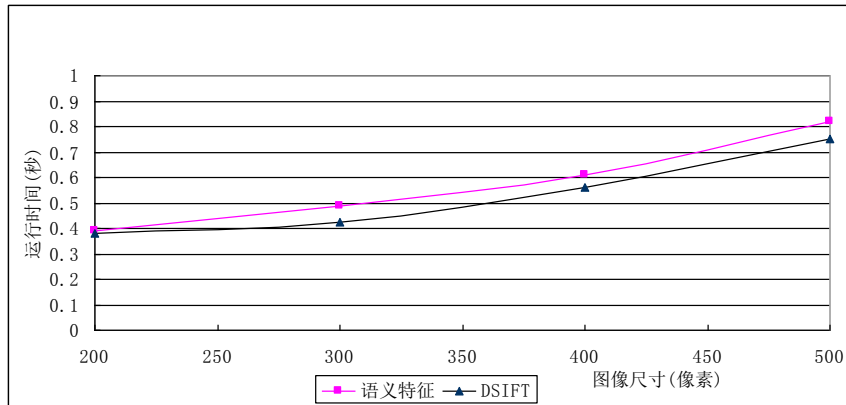


图 4.2 语义特征、DSIFT 特征提取时间与图像尺寸的关系

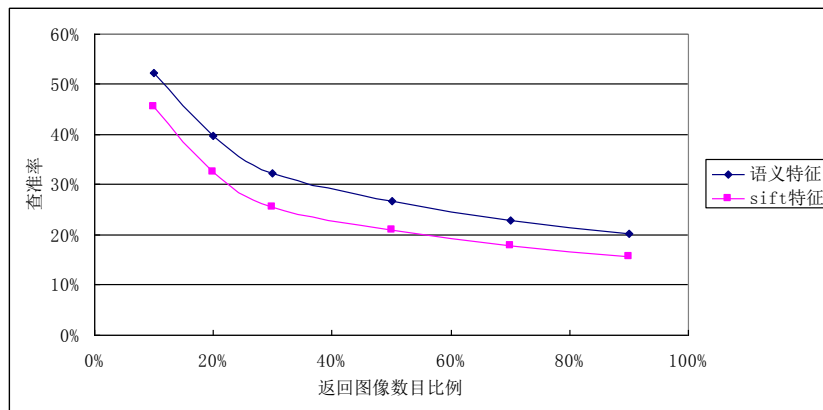


图 4.3 语义特征与 SIFT 特征的查准率

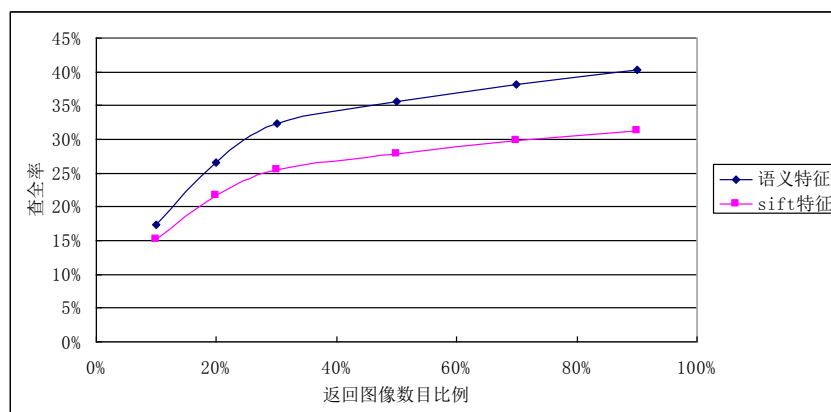


图 4.4 语义特征与 SIFT 特征的查全率

4.2.3 系统的检索效果

1. 颜色特征效果展示

颜色特征采用检索效果进行展示，上传测试图像如图 4.5所示，采用普通颜色直方图的检索效果如图 4.6所示，采用划分网格的颜色特征的检索效果如图 4.7所示。



图 4.5 上传测试图像



图 4.6 采用普通颜色直方图的检索效果



图 4.7 采用划分网格的颜色特征的检索效果

2. 形状特征效果展示

测试形状特征的图像如图 4.8所示，其来自互联网。PHOG形状特征的检索效果



图 4.8 测试形状特征的图像

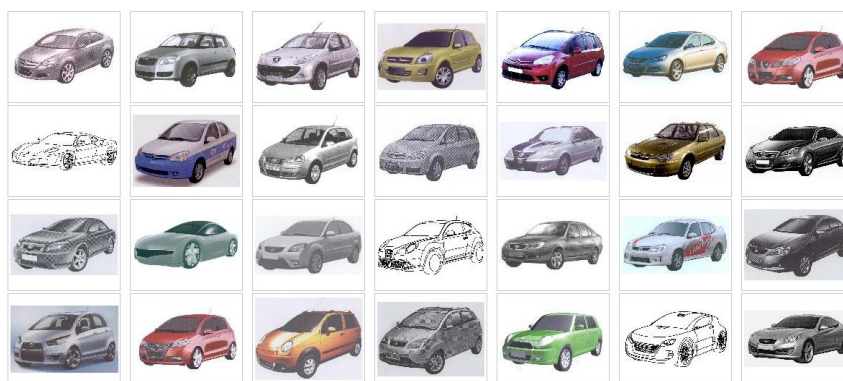


图 4.9 PHOG形状特征的检索效果

如图 4.9 所示。从图中可以看出融合了空间信息的检索结果所有车辆都向一个方向，这会导致形状特征漏掉一些拍照方向不同的图像。

3. 语义特征检索效果展示

测试语义特征的图像如图 4.8，其检索效果如图4.10所示。从图中可看出语义特征的检索结果存在放置方向不同的图像。

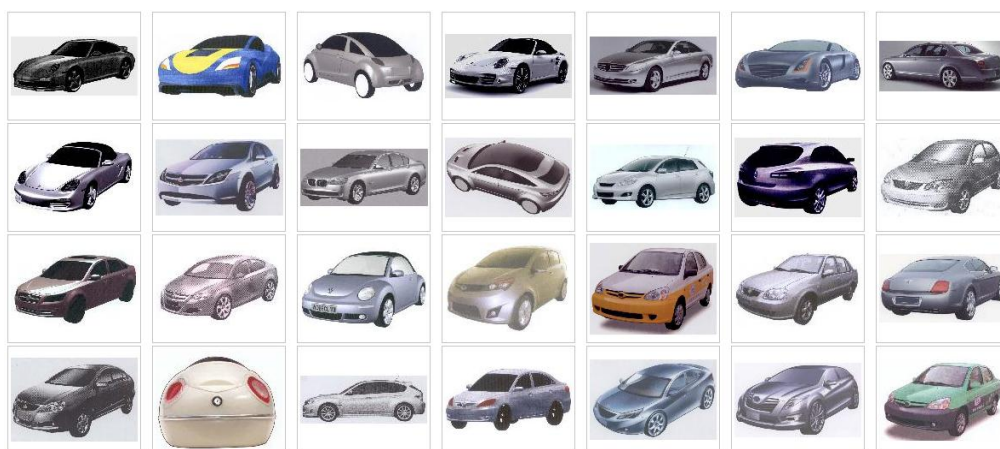


图 4.10 语义特征的检索效果

4.3 本章小结

本章围绕对各种图像特征的测试展开论述，首先介绍了测试环境与数据，并对图像特征在系统中的使用方式进行了具体的介绍，系统提供了良好的接口以利于特征的使用与扩展。

本着围绕影响检索效果的两个因素，检索速度与检索效果对图像的颜色特征，形状特征，局部特征以及语义特征进行了详细的测试。从测试结果可以看出语义特征在特征维度，判别性等方面性能优于目前常用的性能较好的局部特征SIFT。同时其提取速度也可以达到实时性要求。

本章最后以直观的形式展示了图像特征在系统中的检索效果。从检索效果可看出融合空间信息的颜色特征要优于普通的颜色直方图，而形状特征可能漏掉部分旋转的图像，而语义特征能够容忍图像的旋转，这也验证了第3章对这些特征的论述。

5 总结与展望

随着国家知识产权法律的健全以及人们法律意识的增强，人们对专利检索的要求越来越高。传统的基于文本的搜索无法满足人们的需求，更加方便的以图搜图即基于图像内容的专利搜索应运而生。基于图像内容的搜索要达到良好的用户体验，其必须在检索速度与检索质量两个方面达到较好的性能。然而这两个方面都与检索系统所采用的图像特征密切相关，本文针对专利图像搜索中的图像特征提取方法进行了研究。现在讲全文的工作总结如下：

1. 分析了外观专利图像的特点并研究了对应的解决方案。影响专利图像的因素有局部噪声，旋转等，但是专利很少受视角，光照影响小。根据这些因素得出外观专利图像的特征描述符需要对旋转，局部变化等因素保持不变，而对视角，光照变化要求不高。

2. 研究了适用于专利图像的底层视觉特征。专利图像具有受光照角度因素影响较小且尺寸变化不大等特点，根据专利图像的特点，提出并实现了适用于专利图像检索的底层视觉特征。同时也考虑到人类对图像的颜色，形状等信息比较敏感，也对图像的颜色与形状特征进行了研究，并在这些特征中采用划分子区域与划分金字塔等方式融入了空间分布的信息，从而提高了这些特征的性能。研究了局部特征的检测符与描述方式。局部特征被认为是目前最符合人类视觉感知物体的方式，文中对关键点检测原理进行了深入分析，并指出了局部特征在检索系统中不易于扩展与使用等局限性。

3. 研究了针对专利图像的语义特征提取方法。图像的底层视觉特征与人类对图像的认知存在很大的差距，为了解决这个问题，引入了语义特征。文中提出了语义特征提取框架，该框架很容易扩展到其他底层视觉特征上，并且根据该框架获取的语义特征性能较好，其维度低，判别能力强，最重要的是其带有语义信息，缩小了语义鸿沟。语义关联是提取语义特征最重要的环节，文中分析了常用的关联语义的方式，并针对专利图像提出了其合适的语义关联方式。本文还提出了采用样本确定语义空间定义方法，采用样本图像定义语义空间，避免了文本表述能力的不足，并且根据图像内容定义图像语义空间的方式易于实现与推广。

4. 评测了图像特征的性能及其检索效果。对图像特征在维度数目，提取时间，

判别性等方面进行了测试，并验证了图像语义特征在特征维度与判别性等方面的优势，同时本文提出的语义特征也可以满足实时性需求。

本人通过研究生阶段对图像特征的研究，对基于图像内容的搜索提出以下几点，作为参考与完善的地方：

1. 图像底层视觉特征的发展日趋成熟，尺度空间理论，图像信息挖掘也逐渐完善。图像特征的研究重点可以从研究更复杂的图像特征算法转向针对具体应用的特征融合方式，以及语义特征。

2. 图像的语义特征最重要的方面是含有图像的高层语义。更多的研究可以转向如何关联图像语义，语义关联方式可以从引导用户建立图像库的内在关系，建立语义关联模型等方面做深入研究。

致 谢

转眼间，两年半的研究生学习生活就过去了。华中科技大学服务计算技术与系统教育部重点实验室/集群与网格计算湖北省重点实验室的科研氛围浓厚，生活节奏催人向上。在实验室度过的两年使我的视野得以展开，学术水平得以提高，同时我也结识了许多良师益友，这两年的生活是我人生的宝贵财富。

在我的论文完成之际，首先衷心感谢我的导师章勤教授，两年前很荣幸地成为章老师的学生。章老师治学态度严谨，工作积极认真，为我以后的工作树立了榜样。同时在章老师的严格要求下，我学到了很多知识，包括学术上的与生活上的，这为我今后的工作与生活奠定了坚实的基础。通过点点滴滴，我对章老师的敬佩之情也是随着时间流逝而逐渐增加。非常感谢章老师这两年对我的悉心指导，愿章老师身体健康，家庭美满幸福。

同时我还要感谢我的指导老师郑然副教授。郑老师平易近人，关心学生，对事情与人都很有耐心，是学习生活的好朋友。在与郑老师的交流中，我学到很多做研究，做工程项目的方法。她也经常提醒我们项目进展，使我时时刻刻感受到敦促与鞭策。项目在郑老师的建议下，得以顺利完成，在此衷心感谢郑然老师，祝郑然老师工作顺利，合家欢乐。

深深感谢我们实验室主任金海教授。金老师阅历丰富，知识渊博，学术洞察力敏锐，让我感到钦佩。金老师时刻走在学术的最前沿，给予实验室师生最正确的指导，使我们受益颇多。通过对金老师回复我邮件的观察，金老师每天都在凌晨以后才入睡，金老师的积极工作态度使我更加感到敬佩。在此，向金老师致以崇高的敬意与衷心的感谢，同时也提醒金老师注意身体，祝金老师身体健康，带领实验室走得更高更远。

感谢韩宗芬教授。对韩老师最大的印象就是韩老师的脸上一直洋溢着幸福温暖的笑容，为实验室增加了几分家的韵味，韩老师为我树立了生活上的好榜样，祝韩老师天天开心，健康到永远。感谢李胜利教授，虽然李老师接触的比较少，但在开题时感受到李老师对学生的深切关怀，他对实验室同学的开题方向提出非常中肯的意见，在此对李老师表示深深的感谢，祝李老师身体健康，事事顺心。

在实验室里，我能时刻感受到各位老师的关心，以及各位老师们营造的积极向

华中科技大学硕士学位论文

上的氛围，在此，感谢吴松，廖小飞，袁平鹏，邹德清，石宣化，谢夏，邵志远，邵志远，喻之斌，丁晓峰，陈汉华，马晓静等老师，感谢他们为实验室的科研工作与氛围付出的艰辛努力，感谢他们给我在生活学习上的指导。

感谢实验室的刘英书、王小兰、赵娜、耿聪等几位老师。她们为实验室创造了良好的工作环境，为大家妥善安排日常的学习生活，在此表示真诚的感谢！

感谢项目组的朱磊博士，在朱博士的带领下，我们解决了很多技术难题。同时感谢项目组的郭明瑞，黎帅，与他们在一个项目组工作多了很多的乐趣。感谢曾经同组的博士生李波，冯晓文，同组的师兄汪聪，江武，文石磊，他们在我刚开始研究生生活时，给予了很大的帮助。感谢同组的师弟邓巍，陈龙，孙浩川，姚传威给项目组带来了活泼的气氛。

感谢我的宿舍好友曹杰，曾敬翔，程伟，王希斌，刘谱，与你们在一起交流使我受益颇多。感谢我们在一起的点点滴滴，丰富了我的人生经历。

特别感谢我亲爱的父母，感谢他们的养育之恩。父母对我的恩情难以言表，每次看到他们的白发都有说不清的感觉，祝愿父母身体健康，幸福长寿。

参考文献

- [1] R. Datta, J. Li, J. Z. Wang. Content-Based Image Retrieval-A Survey on the Approaches and Trends of the New Age. In workshop of Multimedia Information Retrieval, 2005, 593-603
- [2] F. Mahmoudi, J. Shanbehzadeh, A. Masoud, et al. Image Retrieval based on Shape Similarity by Edge Orientation Autocorrelogram. In Journal of Pattern Recognition, 2003, 36(8): 1725-1736
- [3] S. Vrochidis, S. Papadopoulos, A. Moumtzidou, et al. Towards content-based patent image retrieval: A framework perspective. In World Patent Information, 2010, 32(2): 94-106
- [4] M. Yang, G. Qiu, J. Huang, et al. Near-Duplicate Image Recognition and Content-based Image Retrieval using Adaptive Hierarchical Geometric Centroids. In International Conference on Pattern Recognition, 2006, 4(8): 958-961
- [5] 郑俊勇. 基于内容的外观专利检索系统研究: [硕士学位论文]. 杭州: 浙江大学图书馆, 2005
- [6] Y. Liu, S. Lin. Well-Defined Semantic Templates for Pornographic Images Identification. In IEEE International Conference on Information Science and Engineering, 2009, 1507-1510
- [7] K. U. Barthel, S. Muller, D. Backstein, et al. Image Retrieval using Collaborative Filtering and Visual Navigation. In Posters of ACM International Conference on Computer Graphics and Interactive Techniques, 2010
- [8] Y. liu, D. S. Zhang, G. Lu, et al. A survey of content-based image retrieval with high-level semantics. In the Journal of the Pattern Recognition, 2007, 40(4): 262-282
- [9] R. Brunelli, O. Mich. Histograms Analysis for Image Retrieval. Pattern Recognition, 2001, 34(8): 1625-1637
- [10] P. Balamurugan, R. Rajesh. Guided Classification using CCV and Wavelet Features for Greenery and Non-Greenery Images. In IEEE International Conference on Computational Cybernetics, 2008, 27-29
- [11] D. P. Dogra, K. Tripathy, A. K. Majumdar, et al. A comparative study on texture features used for segmentation of images rich in texture. In IEEE International Conference on Signal and Image Processing Applications, 2009, 336-339

- [12] C. Caeleanu, D. S. Huang, V. Gui, et al. Interest Operator versus Gabor Filtering for Facial Imagery Classification. In *Journal of Pattern Recognition*, 2007, 28(8): 950-956
- [13] J. Z. Wang, J. Li, G. Wiederhold. SIMPLIcity: semantics-sensitive integrated matching for picture libraries. In *IEEE Transaction on. Pattern Analysis and Machine Intelligence*, 2001, 23(9): 947-963
- [14] S. Belongie, J. Malik, J. Puzicha. Matching shapes. In *Proceedings of the eighth IEEE International Conference on Computer Vision*, 2001, 1(8): 252-461
- [15] N. Dalal, B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, 1(7): 886-893
- [16] S. Lazebnik, C. Schmid, J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, 2(3): 2169-2178
- [17] D. H. Hubel, T. N. Wiesel. Receptive Fields, Binocular Interaction and Functional Architecture In The Cat's Visual Cortex. In *the Journal of Physiology*, 1962, 160(7): 106-154
- [18] J. Canny. A computational approach to edge detection. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1986, 8(6): 679-698
- [19] 段瑞玲, 李庆祥, 李玉和. 图像边缘检测方法研究综述. *光学技术*, 2005, 31(3): 22-27
- [20] K. Mikolajczyk, C. Schmid. Indexing based on scale invariant interest points. In *Proceedings of eighth IEEE International Conference of Computer Vision*, 2001, 1(8): 525-531
- [21] T. Lindeberg. Scale-space theory: A basic tool for analyzing structures at different scales. In *Journal of Applied Statistics*, 1994, 21(2): 224-270
- [22] K. Mikolajczyk, C. Schmid. A performance evaluation of local descriptors. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2005, 27(10): 1615-1630
- [23] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of IEEE International Conference on Computer Vision*, 1999, 2(7): 1150-1157
- [24] D. G. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, 2004, 2(7): 91-110
- [25] A. Vailaya, A. K. Jain, H. J. Zhang. On Image Classification: City vs. Landscapes. In

- Pattern Recognition, 1998, 31(12): 1921-1936
- [26] D. Lowd, P. Domingos. Naïve Bayes Models for Probability Estimation. In ACM International Conference on Machine Learning, 2005, 529-536
 - [27] Y. Chen, N. Yu, B. Luo, X. W. Chen. iLike: Integrating Visual and Textual Features for Vertical Search. In Proceedings of ACM International Conference on Multimedia, 2010, 211-130
 - [28] 朱兴全, 张洪江, 刘文印等. iFind: 一个结合语义和视觉特征的图像相关反馈检索系统. 计算机学报. 2002(7): 681-685
 - [29] V. Mezaris, I. Kompatsiaris, M.G. Strintzis. An ontology approach to object-based image retrieval. In Proceedings of the International Conference on Image Processing, 2003, 3(5):11-14
 - [30] Y. Li, L.G. Shapiro, J. Bilmes. A generative/discriminative learning algorithm for image classification. In Proceedings of International Conference on Computer Vision, 2005, 2(10): 1605-1612
 - [31] Y. Liu, D. S. Zhang, G. J. Lu, et al. Integrating Semantic Templates with Decision Tree for Image Semantic Learning. In Multimedia Modeling, 2007, 185-195
 - [32] D. Cai, X. He, Z. Li, et al. Hierarchical clustering of WWW image search results using visual, textual and link information. In Proceedings of the ACM International Conference on Multimedia, 2004, 952-959
 - [33] T. Lindeberg. Feature detection with automatic scale selection. In International Journal of Computer Vision, 1998, 30(2): 77--116
 - [34] X. Y. Wang, J. F. Wu, H. Y. Yang. Robust image retrieval based on color histogram of local feature regions. In Multimed Tools Application, 2010, 49(2): 323-345
 - [35] A. Torralba, K. P. Murphy, W. T. Freeman, et al. Context-based vision system for place and object recognition. In Proceedings of IEEE International Conference on Computer Vision, 2003, 1(9): 273-280
 - [36] A. Bosch, A. Zisserman, X. Munoz. Representing shape with a spatial pyramid kernel. In Proceedings of sixth ACM International Conference on Image and Video Retrieval, 2006, 401-408
 - [37] S. Lazebnik, C. Schmid, J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, 2169-2178
 - [38] K. Mikolajczyk, C. Schmid. Scale and affine invariant interest point detectors. In

- International Journal of Computer Vision, 2004, 60 (1): 63–86
- [39] A. Bosch, A. Zisserman, X. Muoz. Scene classification via plsa. In Proceedings of European Conference on Computer Vision, 2006
- [40] J. C. Ferreira, V. A. Menegatto. Eigenvalues of Integral Operators Defined by Smooth Positive Definite Kernels. In Integral Equation and Operator Theory, 2009, 64(1): 61-81
- [41] 专利分类表, [online]. Available: <http://www.lawtime.cn/info/zhuanli/waiguansjzl/>
- [42] A. Vedaldi, A. Zisserman. Efficient additive kernels via explicit feature maps. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2010, 3539-3546