

北京交通大学

本科毕业设计（论文）

基于图像特征向量的图像优选算法设计及系统的实现

**Design of optimize Image Algorithm based image
eigenvector and implementation of the system**

学 院：____软件学院____

专 业：____软件工程____

学生姓名：____XXX____

学 号：____XXX____

指导教师：____XXX____

北京交通大学

2016 年 5 月

学士论文版权使用授权书

本学士论文作者完全了解北京交通大学有关保留、使用学士论文的规定。特授权北京交通大学可以将学士论文的全部或部分内容编入有关数据库进行检索，提供阅览服务，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：

指导教师签名：

签字日期： 年 月 日

签字日期： 年 月 日

中文摘要

该项目来自于本人实习期间所在公司—北京喵喵物语科技有限公司的大数据背景电商导购项目“很赞”。近年来，随着以手机为主要代表的移动设备兴起和普及，移动设备上阅读逐渐成为人们生活中不可缺少的一部分。同时，在国民经济飞快发展的今天，人们开始追求讲究生活品质。导致以内容主导的消费模式开始逐渐取代过去的电子商务领域传统货架式购买。人们在 App 上阅读感兴趣的内容，发现心仪的商品，可以直接在 App 上即时完成交易，甚至不用离开正在阅读的页面，这就是“内容导购”从现在开始即将大行其道的场景。作者所在公司主导的项目“很赞”也面向内容导购方向。内容导购中，信息以信息流的个性化推荐直达用户痛点是其精髓。但是，在浩瀚的信息流中如何快速展示信息特点及其优质性，快速抓住用户眼球，是每个内容导购项目关注的问题。

本文的目标是通过开发一套图像优选算法及系统，从图文混排的内容中，找到画质精美、构图有张力且能快速吸引用户的图像作为内容的封面图展示在信息流中。以求增加内容的竞争力。本文基于图像特征的分析，调研优选了多种图像特征的提取方法，使用向量化的图像特征进行一定规则的排序优选，得到优选图像结果。

本文通过一套优选验证方案完成对优选结果的实际测试。通过对图像优选，提高了工作效率，节约了人力成本，提升了网站点击、转化等数据指标。该测试结果证明了图像优选算法及其系统的优异性能。

关键词：特征向量；图像优选；聚类算法；图像截取；surf 算法

ABSTRACT

The project comes from the company where I'm in practice. The company has a shopping guide project "henzan" based on Big Data Technology. In recent years, with the mobile phone rising fast and being popular, reading on mobile devices gradually become an indispensable part of life. Meanwhile, by the fast development of the national economy, people begin to pay attention to the pursuing quality of life. Which leads to the content-led consumption patterns began to replace the traditional the past shelves of formula purchased. People reading content of interest on the App, find favorite products, can be directly on App immediate completion of the transaction, even without leaving the page you are reading, which is "content shopping guide". "henzan" is one of the shopping guide focus on content shopping guide direction. Content shopping guide, the information to personalize the information flow. However, in the vast flow of information on how to quickly display the characteristics and quality of information to quickly seize the users' eye, is important and difficult.

Object of this paper is to develop a algorithm to choose prefer image from photo-text content, then make the image as the cover of the figure shows the contents in the stream. Based on the analysis of image features, making researches about preferring method for image feature extraction, collect image features to to sort and choose the best one as preferred image .

We test the results from the algorithm by a validation protocols. Through the preferred image algorithm, improve work efficiency, saving labor costs, improves the site clicks rate and conversions rate and so on. The test results prove the excellent performance of the image algorithm and the system.

KEYWORDS: feature vector; image prefer; clustering algorithm; image capture; surf algorithm

目 录

中文摘要.....	I
ABSTRACT	II
目 录.....	III
第一章 引言.....	1
1.1 研究背景和意义	1
1.2 研究现状	2
1.2.1 图像特征研究	2
1.2.2 图像优选研究	3
1.3 本文研究工作和内容安排	3
1.3.1 论文研究工作	3
1.3.2 论文内容安排	4
第二章 图像优选概述.....	6
2.1 图像优选的概念	6
2.2 图像优选的原则	6
2.2.1 画质	6
2.2.2 色彩和色彩协调	8
2.2.3 纹理	11
2.2.4 信息熵	11
2.2.5 内容语义	12
2.3 本章小结	12
第三章 基于图像特征向量的图像优选算法设计.....	13
3.1 图像特征选择和计算原则综述	13
3.2 色彩特征提取方法的优选	14
3.2.1 RGB 颜色模型	14
3.2.2 CMYK 颜色模型	14
3.2.3 HSL 颜色模型	15
3.3 纹理特征提取方法的优选	15
3.4 信息熵计算方法的优选	16
3.5 聚类算法介绍和优选	17
3.6 局部特征点检测方法研究	20
3.7 截图图像特征向量的图像优选算法逻辑设计	23
3.8 本章小结	25
第四章 基于图像特征向量的图像优选系统构建.....	26
4.1 图像优选系统模块定义	26
4.2 图像范围选择模块	27

4.3 图像筛选因子模块	28
4.4 图像优选算法模块	31
4.4.1 图像优选算法数据结构	31
4.4.2 图像优选算法的实现	32
4.5 图像截取模块	32
4.5.1 图像截取模块策略分析	32
4.5.2 图像截取模块的实现	33
4.6 方案验证	34
4.7 本章小结	37
第五章 总结与展望	38
5.1 总结	38
5.2 展望	38
参考文献	40
致 谢	42
附 录	43

引言

本章首先对课题的研究背景和意义进行了分析和介绍，对图像优选的概念进行了简单说明。并对图像特征领域的研究和图像优选方面的研究现状进行了简单介绍。最后，在论文研究工作介绍中，介绍了本文的研究框架和主要内容。在论文内容安排中，对各章内容和结论进行了简单介绍。

1.1 研究背景和意义

在过去的数十年间，随着计算机和互联网信息技术及其应用的飞速发展，网络购物深刻影响和改变了大多数消费者的消费方式。据调查到 2015 年 12 月，中国网民规模达 6.88 亿，互联网普及率达 50.3%；手机网民规模达 6.2 亿，占比提升至 90.1%^[1]，如表 1-1 中国网民规模和互联网普及率发展走势所示。国内无线网络覆盖成明显提升，网民 Wi-Fi 使用率达到 91.8%。网络环境的逐步完善和手机上网的快速普及，使得移动互联网应用的需求不断被激发。基础应用、商务交易、网络金融、网络娱乐、公共服务等个人应用发展日益丰富。其中，手机网上支付增长尤为迅速。截至 2015 年 12 月，手机网上支付用户规模达到 3.58 亿，增长率为 64.5%，网民使用手机网上支付的比例由 2014 年底的 39.0%提升至 57.7%。

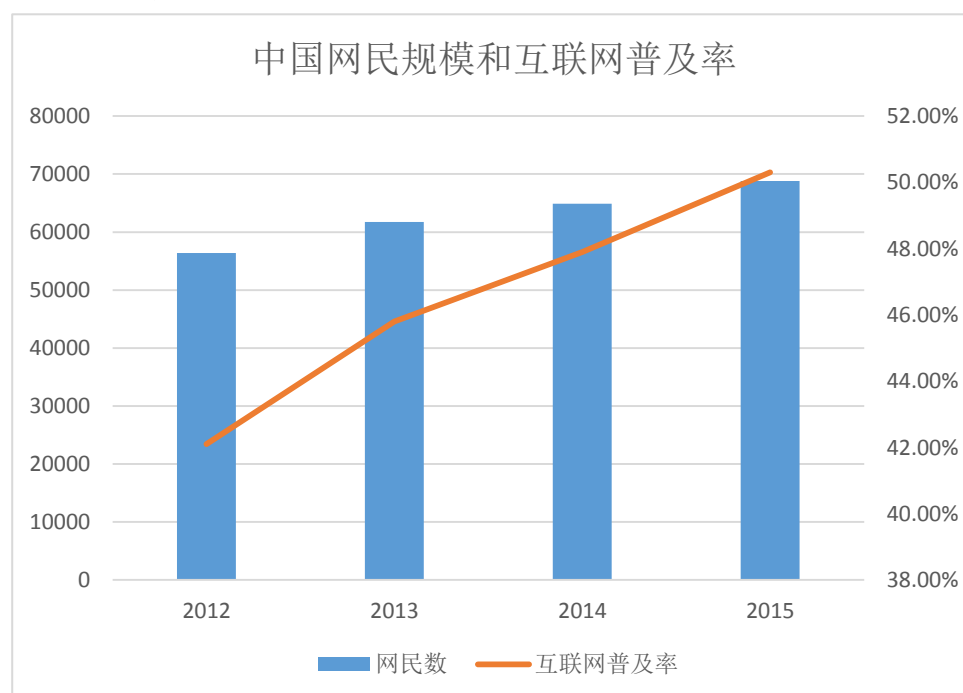


图 1-1 中国网民规模和互联网普及率发展走势

数据表明，互联网移动化和消费移动化是接下来互联网发展的大趋势。与此同时，网络购物深刻影响了我国广大消费者的消费行为，电子商务市场在互联网发展的大趋势下，也在悄然发生变革。在国民经济飞快发展的今天，人们开始追求讲究生活品质。导致以内容主导的消费模式开始逐渐取代过去的电子商务领域传统货架式购买。人们在 App 上阅读感兴趣的内容，发现心仪的商品，可以直接在 App 上即时完成交易，甚至不用离开正在阅读的页面，这就是“内容导购”兴起的场景。

内容导购以购物文章为主体，通过信息流的方式呈现。信息流中封面图、标题、摘要三大元素是用户接触信息的最直观、最便捷的方法。三大元素中，标题通过 UGC^[2] (User Generated Content) 产生，而封面图和摘要则要通过设计算法，通过算法计算得到。在信息流中，图像除了能直观传递信息外，优秀的图像还能吸引读者眼球。从商业角度上看，优秀的封面图可以有效提高用户对内容质量的第一印象，同时促进信息流中内容的点击率、转化率等多项指标。

在门户媒体作为互联网主力的时代，新闻内容的封面图使用最简单的编辑挑选填充的方法，这样的方法好处是可以人为控制封面图质量，缺点在于人工成本较高，封面图选择受个人因素影响高。在当前大数据爆发式增长的背景下，这种方法显然已经不再适用，为了满足公司期望的十万、百万量级内容数据处理，开发一套图像优选系统是必要也是重要的。

图像优选是一种利用图像识别将图像信息数据化，通过对数据的分析比较，采用一定的规则进行优选的技术。它利用了人工智能领域的图像处理相关技术，所处理的数据是图像。目的是对网络上抓取的内容（表现为图文混排的文章）中的图像进行数据化并采用算法进行优选。

1.2 研究现状

本节将从图像特征研究和图像优选研究两个角度介绍和本文相关领域研究现状。

1.2.1 图像特征研究

特征提取^[3]是图像处理和计算机视觉中的概念。是指使用计算机通过各种算法提取图像中的信息，根据这些信息判断图像中的点是否符合图像特征的要求。其结果是将图像上的点分到不同的子集中。这些点一般有孤立、线性、或者区域特性。

特征提取在图像处理过程中属于初级运算，也就是说只对图像进行了第一步处理。通过检测每个像素上的点来确定能否代表一个特征。在特征提取的算法应用中，一般检

查图像的特征区域。特征被检测到后，可以通过计算机从图像中抽离出代表特征的特征值，这个结果被称为特征描述。得到的和图像特征有关的特征值被称为图像特征向量。

图像特征是一种常用的研究方法，在图像识别、图像检索、特征分析、图像去噪等领域有广泛而且成熟的应用。

1.2.2 图像优选研究

图像优选是在图像识别的基础上根据一定的规则对图像指标进行判断选择的研究。该研究是基于图像识别进行的逻辑判断。由于优选标准和业务需求紧密结合导致优选标准没有统一性，使用哪张图最优，本身是一个主观判断的问题。而且它依赖于产品的设计逻辑，在很赞这个基于内容的电商导购产品背景下，也有不同的判断标准。如果以商品为主要逻辑，那么图像优选应该以商品图为首选原则；如果以商品体验信息为主要逻辑，那么图像优选应该以场景图为首选原则。不同策略下，优选的标准和策略是不同的。故图像优选是用算法的思路解决的工程问题。目前并没有完整成熟且通用的图像优选解决方案。

1.3 本文研究工作和内容安排

本节将从问题定义、方法介绍、研究框架、主要内容等方面介绍论文研究工作，并简述各章节论文内容具体安排。

1.3.1 论文研究工作

基于以上背景和现状分析，本文从购物类文章本身特点出发，结合电子商务领域对用户购物行为的研究^[4]，探讨如何通过算法和系统完成图像优选工作。本文定义“优选”包括两方面的考虑：一是图像在画质、色彩、纹理等方面的参数，效果主要依赖于人的肉眼识别和直观感受。二是图像截取和压缩。

所谓图像在画质、色彩、纹理等方面的参数以及人的肉眼识别和主观感受。从问题发现的背景和目的来说，优选图像是为了增加内容对用户的吸引力，让用户在第一时间接触信息流的时候，能快速从优质的封面图中得到文章要讲述的内容语义或者文章的特点，从而快速做出选择。感受本身就是主观的，但是一般来说，图像的画质、色彩、纹理等可量化的全局特性是可以计算的，并且优秀的图像在这些方面表现都很出色。所以可以通过量化计算来选择出表现优秀的图像参与优选。

所谓图像截取和压缩。是指在图像的存储和展示过程中，需要根据图像展示窗口比例进行相应的截取操作，但是这里的截取的目的要考虑到截取结果不能让图像的比例失调，所以需要考虑图像的局部特征等元素在图像质量中的影响，在不损害图像比例和内容语义的基础上，结合图像窗口的设计需求，完成图像的截取工作。图像的压缩则主要考虑图像存储和传输过程中的性能方面影响。

本文研究内容主要包括以下三部分：

1. 图像优选概述。

从业务需求角度阐述图像优选的概念，并从需求分析角度介绍图像优选的原则。

2. 基于图像特征向量的优选图像算法研究。

完成图像优选算法相关的解决方案优选和算法的逻辑设计，是本文的重点和难点。

3. 基于图像特征向量的优选图像系统研究和设计。

完成系统模块定义和各模块的需求分析和设计实现。

本文研究框架如图 1-1 所示

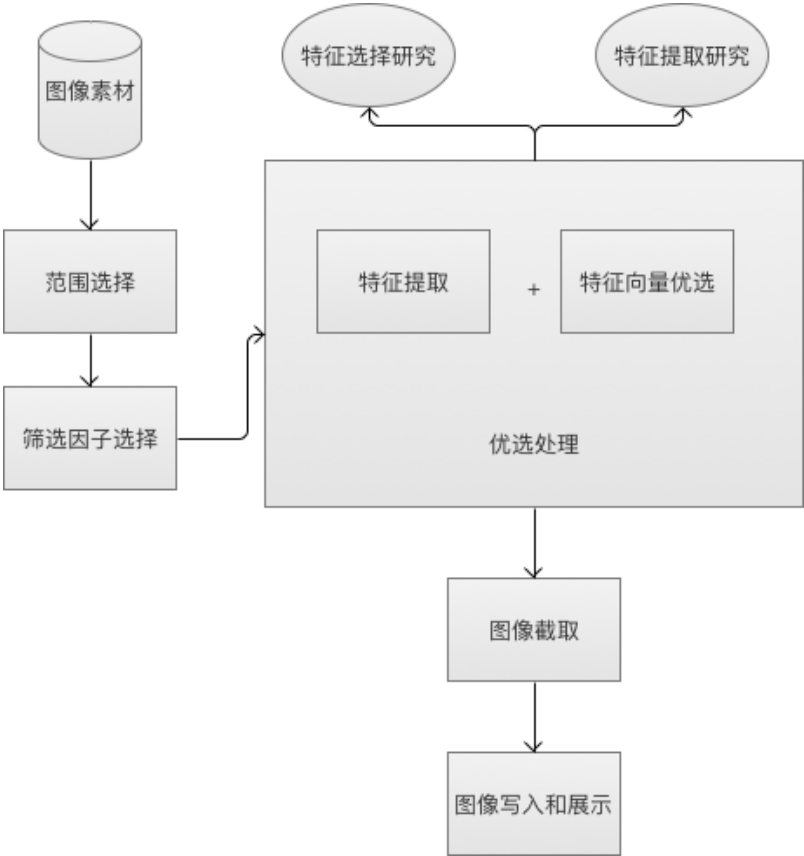


图 1-1 本文研究框架图

1.3.2 论文内容安排

本论文主要分为五个章节：

第一部分为引言部分，介绍了课题的研究背景、目的和意义，并对国内外研究现状进行说明，最后概括了本论文的研究内容和主要结构。

第二部分为图像优选概述，对题目中图像优选进行了详细的研究和说明。详细分析了图像优选的概念，和项目背景下图像优选的总体思路。并确定了图像优选的原则，对这些原则进行了分类说明。从而引出图像特征的概念。

第三部分为基于图像特征向量的优选图像算法研究和设计。介绍了图像优选原则对应的图像特征，以及这些特征的模型和提取方法。

第四部分为截图图像特征向量的优选图像系统的设计与实现。本章从图像范围选择模块、图像筛选因子模块、图像优选算法模块、图像截取算法模块四个部分，对基于图像特征向量的优选图像系统进行了详细的描述，最后通过实验分析对比了该系统在”优选“上的表现。

第五部分为总结与展望，总结了本文的研究内容，并对未来工作提出了展望。

第二章 图像优选概述

研究过程中有涉及到图像优选方法。图像优选是的关键在于“优选”，而优选是一个主观性极强的工程型问题。基于公司产品“以商品为核心，以内容为展示手段”的产品策略，接下来将对图像优选和图像特征进行分析介绍。

2.1 图像优选的概念

图像优选是指从一篇图文混排的文章中提取所有图像，通过算法计算出表现最优的图像，将此图像作为文章封面图展示到文章信息流列表中。

图像优选的总体思路是：

1. 挑选画质精美的图像
2. 挑选构图有张力、质感强的图像
3. 挑选摄影图像。不挑选背景单调、黑白的图像

通过上述 3 点描述，期望通过图像优选出能触动人类情感波动，甚至产生占有欲和购买欲。

2.2 图像优选的原则

图像优选原则是指以总体思路为本，列举所涉及的具体图像指标。以下将对这些指标进行分析介绍

2.2.1 画质

画质顾名思义就是画面质量。包括清晰度、锐度、色彩纯度、色彩平衡等多项指标^[5]。

清晰度指的是图像上各细部影纹和影纹边界的清晰程度，是评价图像质量的重要指标之一。这一点在图 2-1 和图 2-2 的值比截图的人脸部分可以清楚的看到。



图 2-1 佳能 EOS 600D 清晰度样片 100%截图



图 2-2 宾得 K-r 清晰度样片 100%截图

锐度，有时也叫“清晰度”。它是反映图像平面清晰度和图像边缘锐利程度的一个指标。如果将锐度调高，图像平面上的细节对比度也更高，看起来更清楚。比如，在高锐度的情况下，不但画面上人脸的皱纹、斑点更清楚，而且脸部肌肉的鼓起或凹下也可表现得栩栩如生。在另一种情况下，即垂直方向的深色或黑色线条，或黑白图像突变的

地方，在较高锐度的情况下，线条或黑白图像突变的交接处，其边缘更加锐利，整体画面显得更加清楚。因此，提高锐度，实际上也就是提高了清晰度，这是人们需要的、好的一面^[6]。这一点从硬币截图部分还可以看出对物体质感的表现。图 2-1 最好的还原了硬币的金属质感，而图 2-2 几乎难以分辨硬币材质

色彩纯度，是指色彩的鲜艳程度。有色相感的色彩，就有一定程度的鲜艳度。三原色可以组成任意色彩，在偶有色彩中，红、黄、蓝。原色的纯度最高，所谓色彩纯度是指原色在色彩中的百分比。

色彩平衡是控制图像的颜色分布的指标，可以让图像整体色彩达到一个可接受的平衡。在调整图像的颜色时，根据色彩平衡原理，主要通过补色的手段实现。

2.2.2 色彩和色彩协调

颜色可以分成两个大类：无彩色系、有彩色系。对于有彩色系的颜色而言，共性是具备三个基本特征：色相（Hue）、纯度（Chroma）、明度（Value）^[7]。

色相，顾名思义是色彩的质地面貌。也指特定波长的色光呈现的色彩感觉。对于每种颜色而言，都有独有的色相，这是他们区别于其他颜色的重要参数。

纯度指色彩的饱和度、纯净程度和鲜艳度。也就是一种色彩中所含该色素成分的多少。含的越多，纯度越高，含的越少，纯度越低。

明度是辨别色彩明暗的程度。

色彩协调理论是基于色轮的概念而形成的。色轮本质上是一个将所有的颜色按环形排列的圆。基本的颜色位于色轮上三个等间距的点。（通常是指红、绿、蓝三原色。）在色轮概念起源的印刷领域，通过混合 RGB 三原色，可以调出几乎所有其他的颜色。在现代印刷中，RGB 三原色被换成了 CMY（Cyan 青色，Magenta 洋红，Yellow 黄色），并加入了黑色来生成更暗的颜色，因此被称为 CMYK^[8]。理论上，所有的颜色都能在色轮中的某个位置上找到。并且不同颜色的混合可以得到新的颜色。

色轮用圆的形式表现颜色。离圆心越近，颜色越不纯；离色轮边缘越近，颜色的纯度和饱和度都越高。这是平面色轮的基础概念。

一般来说，颜色有五种协调的形式。

1. 互补色：互补色是最基础的协调形式。在色轮中互补色位于基本色的对面，与基本色呈 180 度夹角，故名互补色。事实上，除了邻近色，其他色彩协调方案都有参考互补色而形成自己的方案。图 2-3 使用艺术作品来展示了互补色的应用效果。



图 2-3 互补色

2. 分补色：在色轮中基本色的互补色的两侧区域。在实现基本色和其互补色之间的色彩协调的基础上得到了更大的色域。比起第 1 中互补色方案。给人更强烈的视觉对比体验。在实际实践中被证明，分布色是一种效果不错的方案。图 2-4 使用艺术作品来展示了分补色的应用效果。

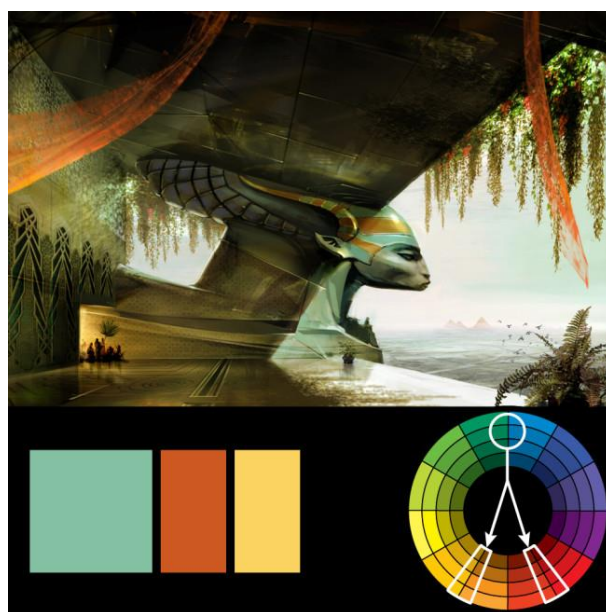


图 2-4 分补色

3. 对比色：也叫“三角色”。色轮中和基本色有 120 度的夹角，形成一个正三角形。正是因为对比色在色轮上具有相等的间距，色彩协调的同事色域的范围被扩大。故而这种色彩方案一般用于高光展示部分。图 2-5 使用艺术作品来展示了分补色的应用效果。



图 2-5 对比色

4. 邻近色：也叫类似色。是指在色轮中位于基本色两侧的颜色。综合使用这种邻近色可以达到一种相对不错的配合感觉。营造一种舒适的体验。但是问题在于邻近色的配色方案容易色彩单调，这是邻近色配色方案的不足之处。图 2-6 使用艺术作品来展示了邻近色的应用效果



图 2-6 邻近色

5. 四角色：四角色和三角色很相似，可以看做是在三角色的基础上增加了一个点。色轮上四个点之间的距离相等。不同与上述几种配色方案，四角色并没有使用色彩协调理论，却可以视作通过使用两组互补色从而达到色彩协调的结果。图 2-7 使用艺术作品来展示了四角色的应用效果。



图 2-7 四角色

通过介绍上联的 5 种色彩系统的基础配色方案，可以看出色彩协调、色彩平衡需要综合考虑的情况较多，这一部分特征的检测需要通过较为复杂的聚类算法实现，通过对图像色彩在 5 种色彩系统之间的分布关系给出图像的色彩平衡指标参数。关于聚类算法的具体内容参见 3.5

2.2.3 纹理

纹理分析是对图像灰度（浓淡）空间分布模式的提取和分析。纹理分析在遥感图像、X 射线照片、细胞图像判读和处理方面有广泛的应用^[9]。

计算机图形学中的纹理既包括通常意义上物体表面的纹理即使物体表面呈现凹凸不平的沟纹，同时也包括在物体的光滑表面上的彩色图案，通常我们更多地称之为花纹。纹理是一种重要的视觉线索，是图像中普遍存在而又难以描述的特征。^[10]

20 世纪 70 年代以来，纹理特征提取研究领域最有代表性的是 Haralick 对纹理的分析和理解以及提出的纹理特征提取方法，成为纹理研究的经典内容。其他还有自相关函数法、MRF 模型^[11]、Gibbs 模型^[12]等方法。本文中对纹理的研究将主要使用纹理特征标明图像质量。

2.2.4 信息熵

根据香农开创的现代信息论知识可以知道，图像能够传递的信息量可以用图像的信息熵来表示^[13]。对于数字图像，图像的构成是一些像素点。不同亮度的像素点占据图像

中的不同位置，所以图像表现出不同的形状。不同形状的图像表现出来的信息量是不同的，对于全黑或者全白的图像，可以看做必然事件，信息量为零。而当图像不同等级亮度的像素出现的概率相同时，图像信息量最大。

设 $\{F(x, y); x=1, 2, \dots, M; y=1, 2, \dots, N\}$ 是一个以色彩和灰度为参数的二维图像像素矩阵。若 $F(x, y)$ 表示 (x, y) 处像素的亮度，图像的像素亮度经量化计算后得到一个取值集合，这个集合的统计平均值可以用熵来表示，即图像的信息熵^[14]。

2.2.5 内容语义

语义分析是图像理解中的难点，和文本语义相比，存在语义理解和文本描述等方面的鸿沟。从课题背景出发，作者发现：在获取和分析图像的全局特征的基础上。通过灰度直方图这个全局特征的相似度判断^[15]，在第三章关于纹理特征的介绍中，可以知道：纹理是可以标明事物唯一的特征。在灰度共生矩阵中，共生矩阵用两个位置的像素的联合概率密度来定义，反映了亮度的分布特性和具有同样亮度或接近亮度的像素之间的位置分布特性，是有关图像亮度变化的二阶统计特征^[16]。所以根据图像的灰度共生矩阵中像素的概率密度分布特点的相似程度判断，可以简单的达到初级的内容语义理解的效果。

内容语义除了需要做到文章优选图像和商品图在语义上的高度近似，还需要考虑图像中的局部兴趣点描述和肖像画率。局部兴趣点描述的应用场景主要是用户对图像全局特征不感兴趣时，不再需要关注纹理、色彩、画质等全局特征，这是需要计算局部兴趣点，即可以呈现出图像中最有吸引力的局部特征。

2.3 本章小结

本章对图像优选的概念和总体思路进行了介绍。并从总体思路出发，选取并研究了可能对总体思路有影响的一些图像特征的原理和解决办法。在深入了解这些原理和解决办法后，为下一步针对特征向量的算法研究提供了基础。

第三章 基于图像特征向量的图像优选算法设计

在第二章介绍和论述图形优选概念和总体思路的基础上，对图像特征有了深刻的理解之后。本章将首先介绍在图像特征提取方面的工作，然后基于对图像特征的提取，介绍对图像优选算法的算法逻辑设计。这一章是全文的重点和难点。

3.1 图像特征选择和计算原则综述

在已知的画质、色彩、纹理、信息熵以及内容语义等图像优选原则中，可以发现每个原则对应一个或者数个特征，通过这些特征的综合表现可以决定图像在这些原则上的表现，所以首先对需要提取的特征进行归纳和总结，并在此基础上完成特征的提取，然后将其转换为数学模型中的一维向量。对这些一维向量，采用打分机制完成一个排序，排序的参数调优通过采用 100 个测试用例人工训练得到。这一部分是本文的重点也是难点。

总结作为排序因子的特征向量及计算原则：

1. PCB: 图片是否含有纯色背景色块(Pure Color Band)，过多的纯色块会使人不适。
2. RDH: 图片比例偏移值(Ratio Deviation Hold) 图片比例是否相比目标窗口偏移很大
3. HSL(s): 图片饱和度，参数在 0.3-0.8 之间时人感觉比较舒适。
4. HSL(I): 图片光照度，参数在 0.3-0.8 之间时人感觉比较舒适。
5. ASM:角二阶矩(Angular Second Moment)^[17]，图像灰度分布均匀程度和纹理粗细的一个标准，值越大，表明图像纹理细致且灰度分布均匀；值越小，表明图像纹理粗且灰度分布杂乱^[18]，属于图像全局特征之一。
6. CON:对比度(Contrast)，反映了图像清晰度、纹理沟纹深浅度。纹理沟纹越深，其对比度越大，视觉效果越清晰；反之，对比度小，则沟纹浅，效果模糊^[19]。灰度差即对比度大的像素对越多，这个值越大。属于图像全局特征之一。
7. COR:自相关(Correlation)^[20]，度量空间灰度共生矩阵元素在行或列方向上的相似程度，相关值大小反映了图像中局部灰度相关性。相关值大，表明矩阵元素均匀分布；相关值小，表明矩阵元素差异大。属于图像全局特征之一。
8. IDM:逆差距(Inverse Differential Moment)。反映图像纹理的同质性，度量图像纹理局部变化的多少。反映了纹理的清晰程度和规则程度，纹理清晰、值越大，表明纹理清晰、规律性强；值越小，表明纹理难以分辨，毫无规律。属于图像全局特征之一。

9. ENT:信息熵 (Entropy)。度量图像所含信息量, 衡量图像信息的复杂程度, 熵值越大, 代表图像信息越复杂; 熵值越小, 代表图像信息越简单, 属于图像全局特征之一。
10. IPD:兴趣点 (Interesting Point Descriptor) ^[21]。用于表达图像局部特征的细节, 在用户对全局特征不感兴趣时使用。
11. IPR:肖像画率。这个参数反映图像是否有前/后景的明显区分, IPR 越大, 区分越明显。

以上 11 个特征是基于第二章介绍的图像优选原则关注的画质、色彩、纹理、信息熵和内容语义等条件确定的。这些特征将作为参与图像优选的数据指标参与图像优选算法核心的排序逻辑。接下来在 3.2 到 3.6 部分从色彩、纹理、信息熵、局部特征点 (内容语义) 等综合角度介绍和优选相关特征的提取计算方法, 此外, 也从方案优选角度介绍了计算肖像画率和色彩平衡需要使用的聚类算法的优选。

3.2 色彩特征提取方法的优选

色彩特征提取主要涉及对纯色背景 (PCB) 的处理、图像饱和度、图像光照度处理。以下将对集中常见的颜色模型进行介绍和比较。

3.2.1 RGB 颜色模型

RGB(Red, Green, Blue)颜色空间是应用最广的颜色空间之一。常用语显示器系统, 显示器使用 R、G、B 参数值来驱动 R、G、B 电子枪发射电子, 然后电子分别激发荧光屏上的 R、G、B 三种颜色的荧光粉从而发出不同亮度的光线, 此外还可以通过相加混合产生各种颜色; 扫描仪也是通过吸收原稿经反射或透射而发送来的光线中的 R、G、B 成分, 并用它来表示原稿的颜色^[22]。RGB 色彩空间因其在设备制造方面的广泛应用被称为与设备相关的色彩空间。

3.2.2 CMYK 颜色模型

CMYK(Cyan, Magenta, Yellow)颜色空间应用于印刷工业, 印刷业通过青(C)、品(M)、黄(Y)三原色油墨的不同网点面积率的叠印来表现丰富多彩的颜色和阶调, 这便是三原色的 CMY 颜色空间。实际印刷中, 一般采用青 (C)、品(M)、黄(Y)、黑(BK)四种颜色印刷。当混合红绿蓝三原色时, 会产生白色, 但是当混合蓝绿色、紫红色和黄色三原色时会产生黑色。这种模型称之为 CMYK。CMYK 颜色空间应用于印刷业的特色, 使它在不同条件

（纸张、油墨、工艺）下会有不同的结果，这种复杂的颜色模型也被成为与这杯有关的颜色空间。

3.2.3 HSL 颜色模型

HSL(Hue, Saturation, Lightness)颜色空间可以用台式机图形程序的颜色表示。HSL颜色空间的特点是：从人的视觉系统出发，以色调(Hue)、色彩饱和度(Saturation)和亮度(Lightness)描述色彩，用圆锥空间表示自己的颜色模型，虽然描述颜色空间的模型相当复杂，但却能把色调、亮度和饱和度的变化表现出来^[23]。这种色彩空间描述法对比上述介绍的其他颜色空间也是最为接近人类视觉特性的，同时，在计算机视觉处理中有大量关于 HSL 颜色空间的算法，可以大大简化分析和处理工作。

通过比较 RGB 颜色模型、CMYK 颜色模型和 HSL 颜色模型可以知道，HSL 颜色模型由于其更接近人类视觉特性同事在计算机视觉中有大量算法的特性，选择 HSL 颜色模型作为色彩特征提取的模型是理所当然的。提取 HSL 颜色模型特征的核心代码在附录中展示。

3.3 纹理特征提取方法的优选

纹理特征可以分为四种类型：统计型纹理特征、模型型纹理特征、信号处理型纹理特征和结构型纹理特征。这四种纹理特征提取方法中，信号处理型纹理特征主要从图像的变换域提取纹理特征，其他三种纹理提取方法则直接从图像域提取纹理特征。图 2-8 纹理特征提取分类对纹理特征提取方法进行了更详细的说明。

纹理特征提取对于计算机而言，其实就是提取刻画纹理的数字特征，灰度共生矩阵、局部区域的自相关函数、灰度分布以及灰度游程等统计量，是常见的数字纹理特征。

灰度共生矩阵是一种常用的图像纹理特征计算方法。图像的灰度共生矩阵（GLCM）已知被理论证明并且实验显示其在纹理分析中是一个很好的方法。共生矩阵定义为两个不同位置的像素联合概率密度，它很好的反映了亮度的分布特性和具有同样亮度或接近亮度的象素之间的位置分布特性，是关于图象亮度变化这个现象的二阶统计特征[8]。它是定义一组纹理特征的基础。

一幅图象的灰度共生矩阵能反映出图象灰度关于方向、相邻间隔、变化幅度的综合信息，它是分析图象的局部模式和它们排列规则的基础。

设 $f(x, y)$ 为一幅二维数字图象，其大小为 $M \times N$ ，灰度级别为 Ng ，则满足一定空间关系的灰度共生矩阵为

$$P(i, j) = \# \{ (x_1, y_1), (x_2, y_2) \in M \times N \mid f(x_1, y_1) = i, f(x_2, y_2) = j \}$$

其中 $\#(x)$ 表示集合 x 中的元素个数，显然 P 为 $N_g \times N_g$ 的矩阵，若 (x_1, y_1) 与 (x_2, y_2) 间距离为 d ,两者与坐标横轴的夹角为 θ ,可以得到各种间距及角度的灰度共生矩阵 $P(l, j, d, \theta)$ 。从共生矩阵可以导出排序因子“对比度 (CON)”“自相关 (COR)”“逆距差 (IDM)”三个参数的值。计算共生矩阵并导出参数作为排序因子的核心代码在附录中展示。

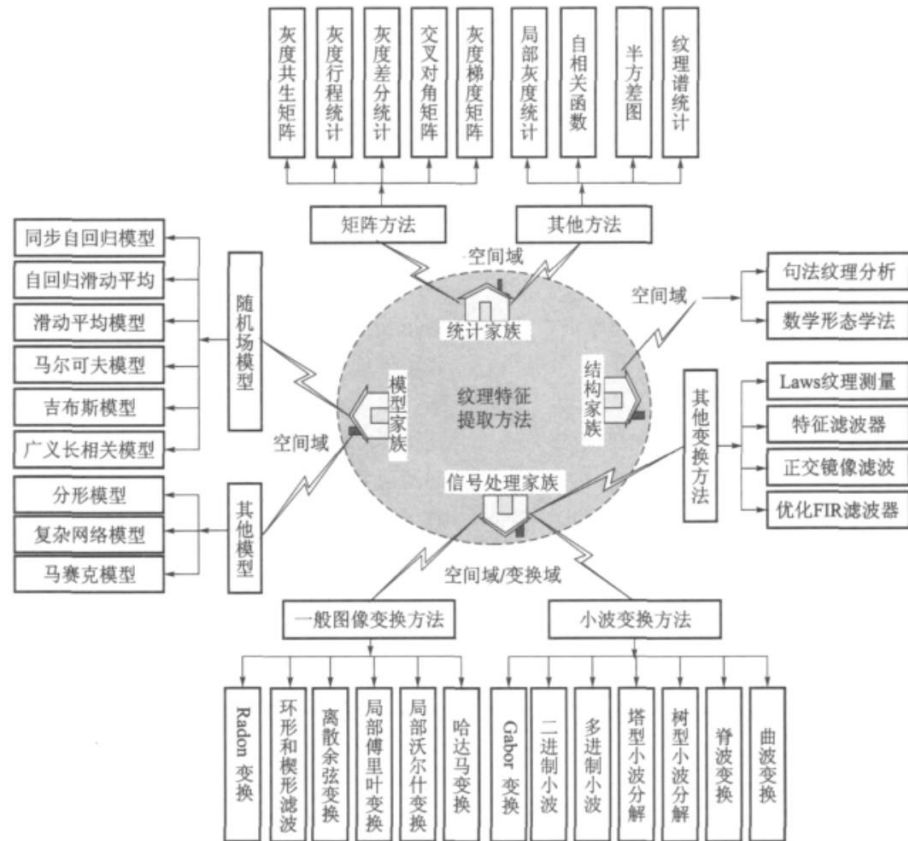


图 2-8 纹理特征提取分类

3.4 信息熵计算方法的优选

图像的信息熵，是一种二维的信息源，图像中不同位置的像素有不同的色彩和灰度代表熵的两个维度参数。像素的灰度用 x_i 表示， $p(x_i)$ 为各灰度级出现的概率，则熵的四种定义形式分别为^[24]：

1. 在假设各像素和各灰度统计独立的前提下，不考虑像素的几何位置，信息熵如式 3-1 所示：

$$H = - \sum_{i=1}^n p(x_i) \log(x_i) \quad (3-1)$$

2. 将图像的 $M \times N$ 个像素位置均看作一个陷阱，光子独立地投入陷阱中，投入的概率不同，每个像素会表现出亮度。设亮度与投入陷阱的光子数成正比，又设投射到整幅图像的光子数为 f_t ，则在 (m, l) 位置陷阱的光子数概率为式 3-2 所示：

$$p(f_{m,n}) = \frac{f_{m,n}}{f_t} = \frac{f_{m,n}}{\sum_{m=1}^M \sum_{n=1}^N f_{i,j}} \quad (3-2)$$

在不考虑 $f_{m,n}$ 间相关性的情况下，可以得到式 3-3（Frieden 定义的熵）：

$$H = - \sum_{m=1}^M \sum_{n=1}^N \left(\frac{f_{m,n}}{\sum_{m=1}^M \sum_{n=1}^N f_{i,j}} \right) \log \left(\frac{f_{m,n}}{\sum_{m=1}^M \sum_{n=1}^N f_{i,j}} \right) \quad (3-3)$$

3. 在考虑到不同像素之间的相关性的情况下，可以得到式 3-4（Pnatt 定义的熵）：

$$H(\bar{f}) = \sum_{q=1}^{MN} H[f_q | f_{q-1}, f_{q-2}, \dots, f_2, f_1] \quad (3-4)$$

4. 用差信号作为样本，在传输过程中会起到较好的效果，故可以用灰度差作为统计单元，从而建立灰度直方图。定义式 3-5：

$$p(\Delta x_i) = n(\Delta x_i) / MN \quad (3-5)$$

则差信号的熵为式 3-6 所示：

$$H_d = - \sum_{i=1}^{2^B} p(\Delta x_i) \log(\Delta x_i) \quad (3-6)$$

总结上述四种信息熵计算方法，式 3-4 和式 3-6 的计算复杂度高，过程繁琐，故在实际的图像处理过程中，基于图像熵的参数评价一般采用公式 3-1 和 3-2 或者其变形的形式。

3.5 聚类算法介绍和优选

聚类算法是做聚类分析的方法。聚类分析是指将对象的集合分组为由相似对象组成的数个类的过程。从计算机和统计学的角度看，聚类分析是通过数据建模来简化数据，寻找数据规律的一种方法。Everitt 在 1974 年关于聚类所下的定义^[25]：一个类簇内的实体是相似的，不同类簇的实体是不相似的；一个类簇是测试空间中点的会聚，同一类簇的任意两个点间的距离小于不同类簇的任意两个点间的距离；类簇可以描述为一个包

含密度相对较高的点集的多维空间中的连通区域，它们借助包含密度相对较低的点集的区域与其他区域(类簇)相分离

一般来说，聚类过程分为 5 步：

1. 数据准备
2. 特征选择
3. 特征提取
4. 聚类(或分组)
5. 聚类结果评估

目前关于聚类算法的研究成果很多，有多种成熟的聚类算法可以参考使用。下图(图 3-1 聚类算法分类示意图)列出当前有的聚类算法的分类：

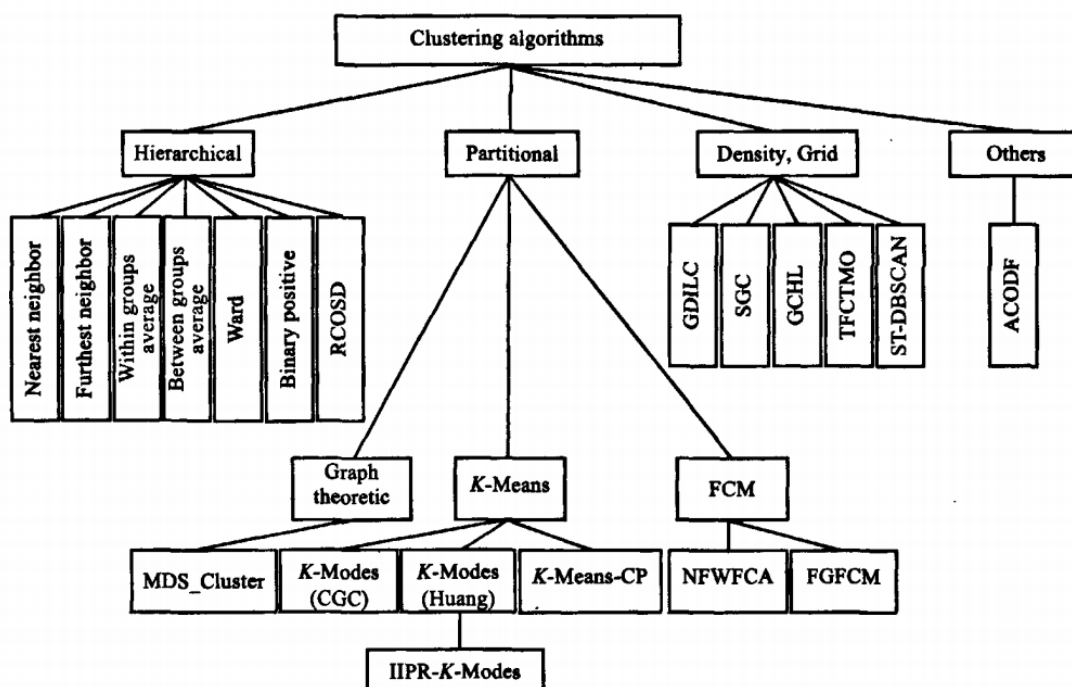


图 3-1 聚类算法分类示意图

从实现成本角度考虑，MacQueen 提出的 k-means 聚类算法本身思想比较简单，是迄今为止，最常用的聚类算法。所以接下来将对 k-means 算法进行研究。

k-means 算法主要思想是：在给定 K 值和 K 个初始类簇中心点的情况下，把每个点(亦即数据记录)分给离其最近的类簇中心点所代表的类簇中，所有点分配完毕之后，根据一个类簇内的所有点重新计算该类簇的中心点(取平均值)，然后再迭代的进行分配点和更新类簇中心点的步骤，直至类簇中心点的变化很小，或者达到指定的迭代次数^[26]。

K-Means 算法中，能否合理确定 K 值以及 K 个初始类簇中心点，对于聚类算法效果的会产生极大的影响。下面以五个二元正态高斯分布生成的类为测试用例，介绍算法。

1. 确定 K 个初始类簇中心点

确定 K 个初始类簇中心点，可以采取选择彼此距离尽可能远的点的方法：随机选择类簇中的一个点，将其作为初始类簇中心点，然后选择距离该点最远的点，将其作为新的初始类簇中心点，选择距离前两个点的最近距离最大的点，将其作为更新的初始类簇的中心点，以此类推，直至选出 K 个初始类簇中心点。

图 3-2 表示了五个二元正态高斯分布生成的类，图 3-3 是确定 K 个初始类簇中心点的测试结果

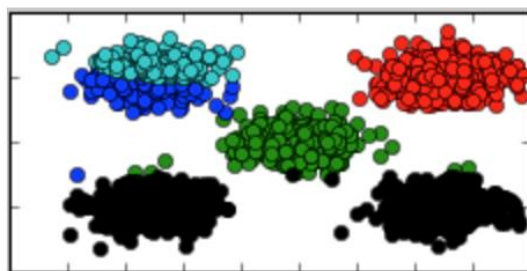


图 3-2 初始类簇结果

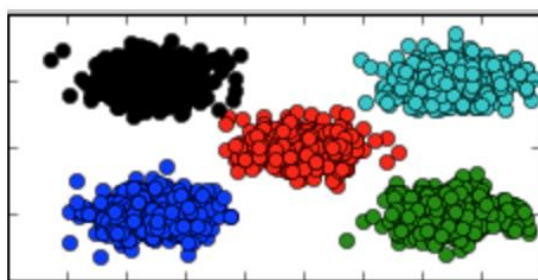


图 3-3 方法 1 聚类效果

2. K 值的确定

类簇内任意两点之间的最大距离即类簇的直径。类簇内各点到中心的距离最大值即类簇的半径。图 3-4 是 K 的取值从 2 到 9 时，聚类效果图；图 3-5 是 K 的取值从 2 到 9 时，类簇指标的效果图：

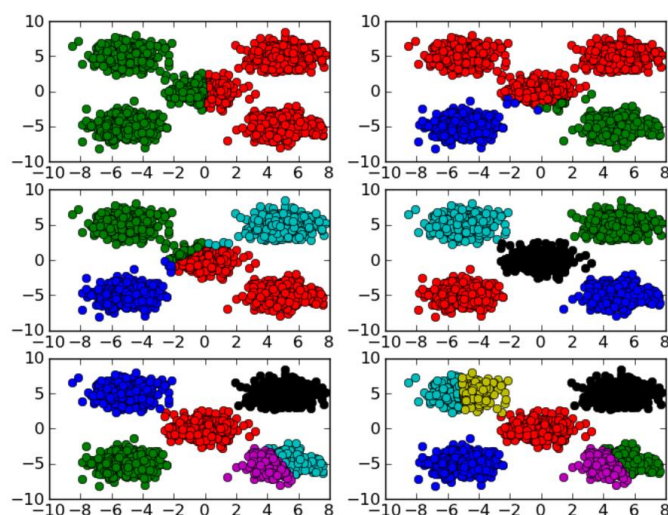


图 3-4 取值从 2 到 9 时，聚类效果图

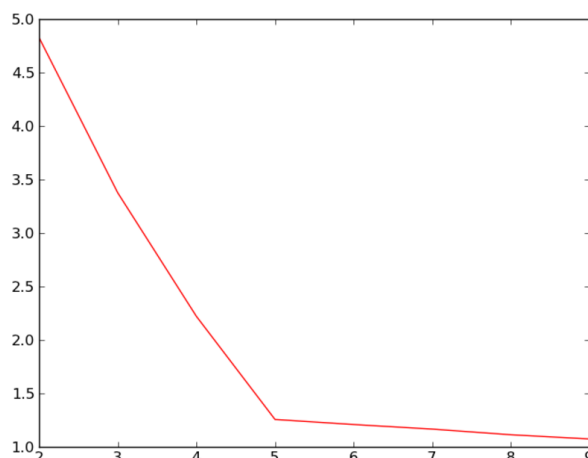


图 3-5 取值从 2 到 9 时，类簇指标的效果图

从上图中可以明显看到，当 K 取值 5 时，类簇指标的下降趋势最快，所以 K 的正确取值应该是 5。

以上是对 k-means 算法的简介。实际算法设计中，k-means 聚类算法主要用于计算肖像率和色彩平衡两个因子，然后具体核心算法见附录。

3.6 局部特征点检测方法研究

在上述的色彩、纹理等特征都是图像的全局特征。然而全局特征的局限在于，一方面，全局特征不能实现描述图像细节，会出现图像空间信息丢失的问题；另一方面，用户可能仅仅对图像中的某一部分感兴趣，此时全局特征就会失效。所以考虑图像的局部特征，可以弥补只考虑全局特征时，在图像内容分析上的不足导致的信息缺失。

局部特征检测的解决思路是先对用户感兴趣的图像或者对象定位，然后通过局部描述子（IPD）对局部特征进行提取。SURF（Speeded-Up Robust Features）算法是一种用于特征点检测的快速鲁棒特征算法^[27]，和经典的特征点检测算法 Sift 算法相比，有更高的执行效率。这一点在实际测试过程中被认为是很重要的。

SURF 算法主要步骤：

1. 构建 Hessian 矩阵

假设函数 $f(x, y)$ ，图像某个像素点的 Hessian 矩阵如式 3-7 所示：

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (3-7)$$

由于特征点需要具备尺度无关性，所以需要在 Hessian 矩阵构造前，对其进行二阶高斯函数的滤波处理，得到式 3-8

$$H(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \quad (3-8)$$

最终需要的结果是原图像的变换图像，因此需要在变换图像上寻找特征点，然后通过将这个特征点映射到原图中。Herbert Bay 提出了近似公式替代的方法，公式 3-9 如下：

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (3-9)$$

2. 构建尺度空间

兴趣点需要在不同的尺度进行检测，不只是因为一致性搜索的比较需要在不同尺度的图片中进行。尺度空间通常以图像金字塔的形式呈现。图像多次被高斯平滑，然后下采样，以得到金字塔高层的图像。

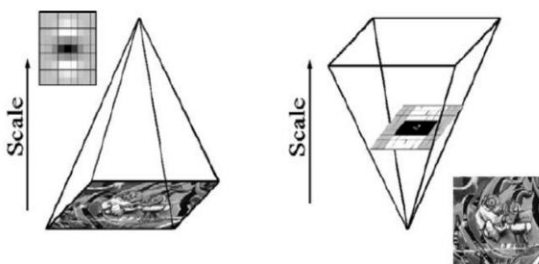


图 3-6 尺度空间金字塔模型

这里通过改变滤波器的大小达到改变尺度空间大小的效果。

3. 精确定位特征点

丢弃比预设极值小的取值，通过增加极值的方法使检测到的特征点数量减少，最后只有特征最强的点可以被检测到。检测过程是使用滤波器进行检测。下图 3-7 以 3*3 的滤波器为例，在滤波过程中，通过 26 个点的比较。标记为 x 的像素点特征在大于周围像素的特征值得时候，即可以确定这个像素点是这个区域的特征点。

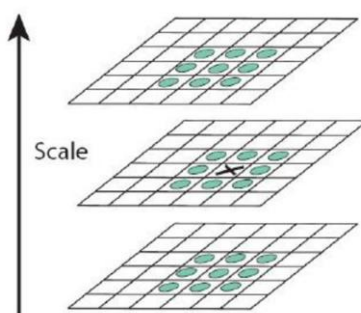


图 3-7 精确定位特征点示意图

4. 确定主方向

以特征点为中心，取半径 $6s$ 的领域内的点，计算这些点在 z 轴和 y 轴上的 Haar 小波响应，可以保证旋转不变性。然后以响应值为参数，按照靠近特征点的权重高，远离特征点的权重低的规则对参数进行赋以高斯权重系数。遍历整个圆形区域，选择最长矢量的方向，即可得到该特征的方向，这里矢量是取半径范围内的响应相加得到的新参数。通过逐个计算特征点的参数，可以得到所有特征点的主方向。

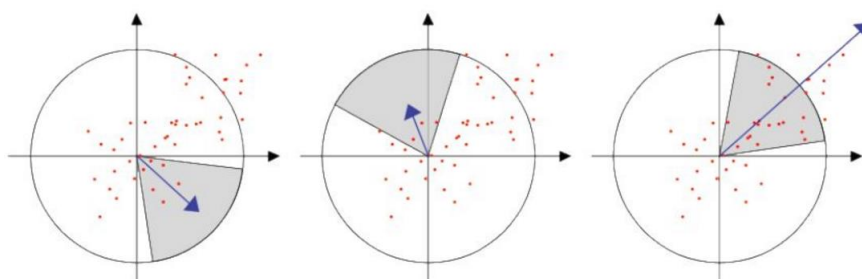


图 3-8 旋转确定主方向示意图

5. 生成特征点描述子

向着关键点方向将坐标轴旋转。如图 3-9 所示，得到旋转不变的结果。

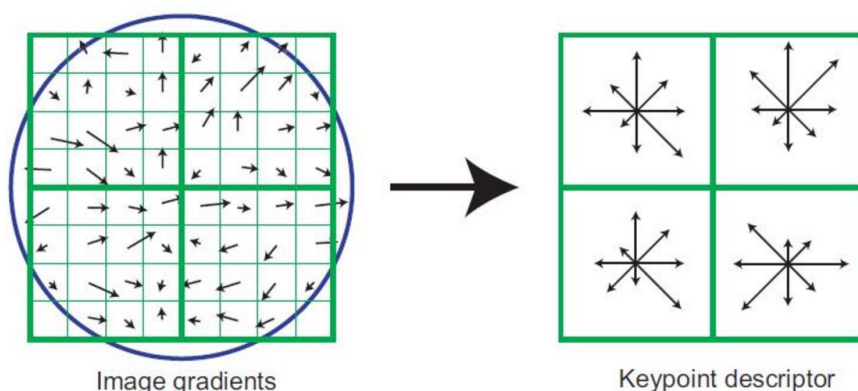


图 3-9 生成特征点示意图

接下来以关键点为中心取 8×8 的窗口。在图 3-9 中，左边图像表明当前关键点处于图像中央位置，图中每个小格表示关键点领域在尺度空间中所占的一个像素，通过公式可以求出每个像素的梯度方向和模值，使用箭头方向和箭头的长度代表上述两个参数，然后经过高斯窗口对这些参数进行加权计算，就可以将像素映射成向量，如右图所示，然后在 4×4 的小块上计算 8 各方向的梯度直方图，根据梯度方向上的叠加值，可以得到种子点。这种领域方向性信息的联合思想增强了算法噪声的抵抗能力，而且对于含有定位误差的特征匹配情况，也提供了一定的容错能力。

以上 5 步简单介绍了 SURF 算法的原理和实现。在使用 SURF 算法完成对兴趣点特征子的提取时，发现，对于图像而言，前景和背景的区别是否明显，也是图像表现的重要指标之一。这就是肖像画率（IPR）的重要作用。图 3-10 和 3-11 对比可以明显看出肖像画率对图像的影响。肖像画率的计算主要通过 SURF 算法和 k-means 聚类算法实现。具体核心代码参见附录。



图 3-10 肖像画原图



图 3-11 肖像画检测效果示意

3.7 截图图像特征向量的图像优选算法逻辑设计

在上面几个小节中，完成了对图像特征的提取方法优选，在此基础上可以进行图像优选算法的设计。

图像优选算法对于每篇图文混排的文章进行图像优选。故以存储一篇文章所有图像的列表作为一次最小运算。对于每张图像而言，需要从列表中读取到图像地址，并下载到本地计算。通过 HSL 算法、Haralick 算法、SURF 算法、聚类算法等优选出来的算法方案提取特征并将特征向量化，并且在提取过程中，按照图像筛选因子（参考第四章 4.3 小节）的规则将不符合规则的图像丢弃。待将一个列表中所有图像都完成特征提取和向量化之后，对特征向量按照规则（参考第四章 4.4 小节）进行排序。根据排序结果得到优选结果，这就是一个完整的最小运算过程。

图像优选算法逻辑设计如下图 3-12:

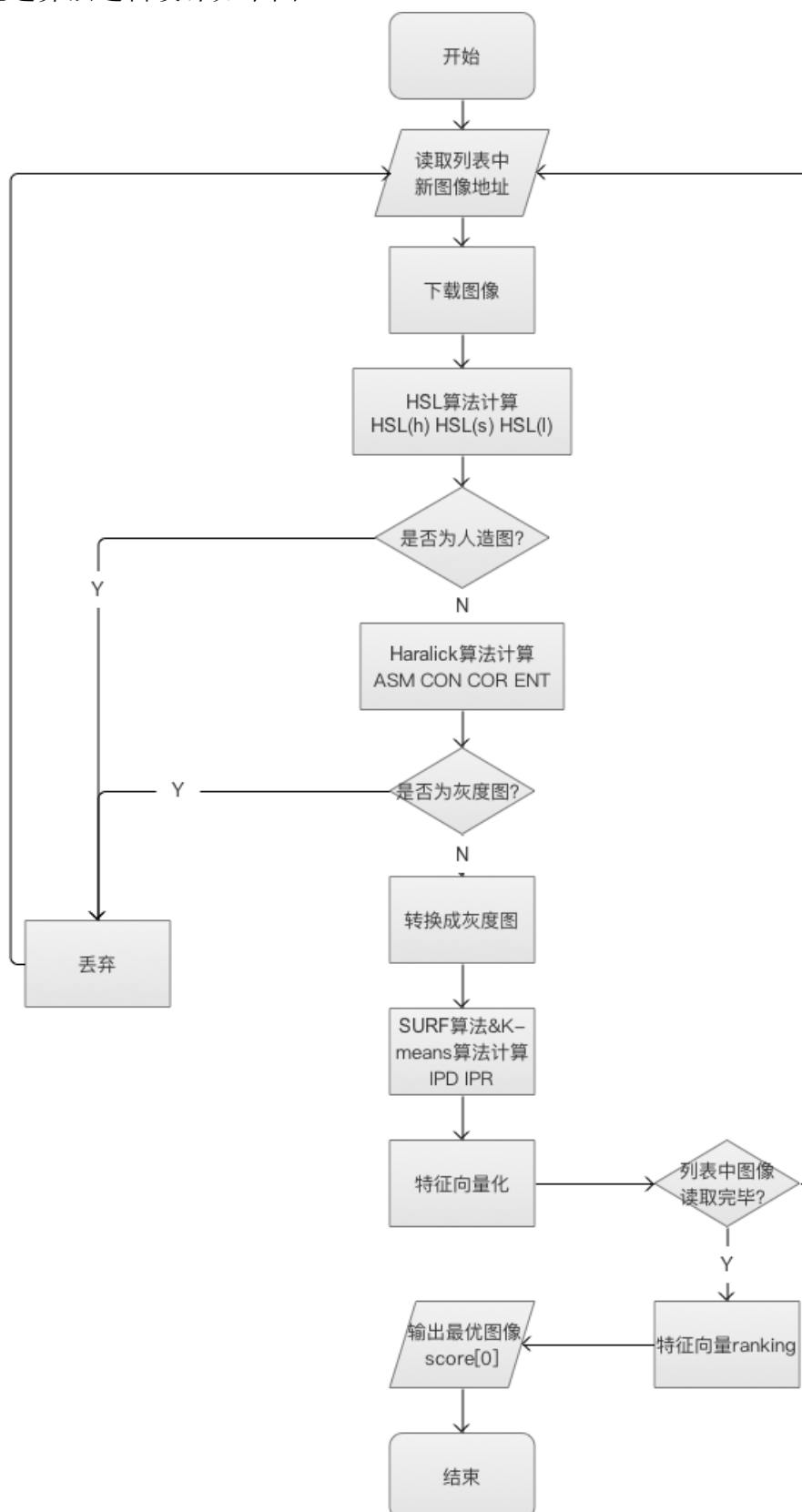


图 3-12 图像优选算法逻辑设计

3.8 本章小结

本章对色彩、纹理、信息熵等全局特征的原理、提取方法进行了介绍。同时对图像内容语义的两步处理：聚类分析和局部特征点检测的应用进行了研究和分析。色彩方面，通过对比三种经典的色彩模型，最终优选更符合人类视觉特征的 HSL 颜色模型作为色彩提取的依据，得到纯色、饱和度、光照度等参数指标。纹理方面，采用针对数字图像纹理处理中，成熟的灰度共生矩阵计算方法，根据计算结果，分析参数对比度、自相关和逆距差作为纹理特征。信息熵方面，提供了四种常用信息熵计算公式，最后选取式一作为信息熵计算方法。图像内容语义方面，首先可以简单的通过灰度直方图这个全局特征，根据灰度特征分布的文章图像和商品图像的灰度特征分布的相似程度进行打分；作为图像内容语义基本实现的补充，同时补充图像全局特征在图像局部上的作用缺失，引入局部特征点。通过 SURF 算法完成。最后，完成了图像优选算法的逻辑设计，算法的实现部分将在下一章完成。

第四章 基于图像特征向量的图像优选系统构建

在第三章综合研究并优选图像特征向量提取的基础上完成了图像优选算法的设计。对于一个完整的基于图像特征向量的图像优选系统来说，这是一个工程问题，所以解决问题的关键在于系统流程和业务要求的逻辑的梳理。接下来，将对图像优选系统的各个功能模块和实验结果进行说明和分析。

4.1 图像优选系统模块定义

图像优选系统功能流程说明：

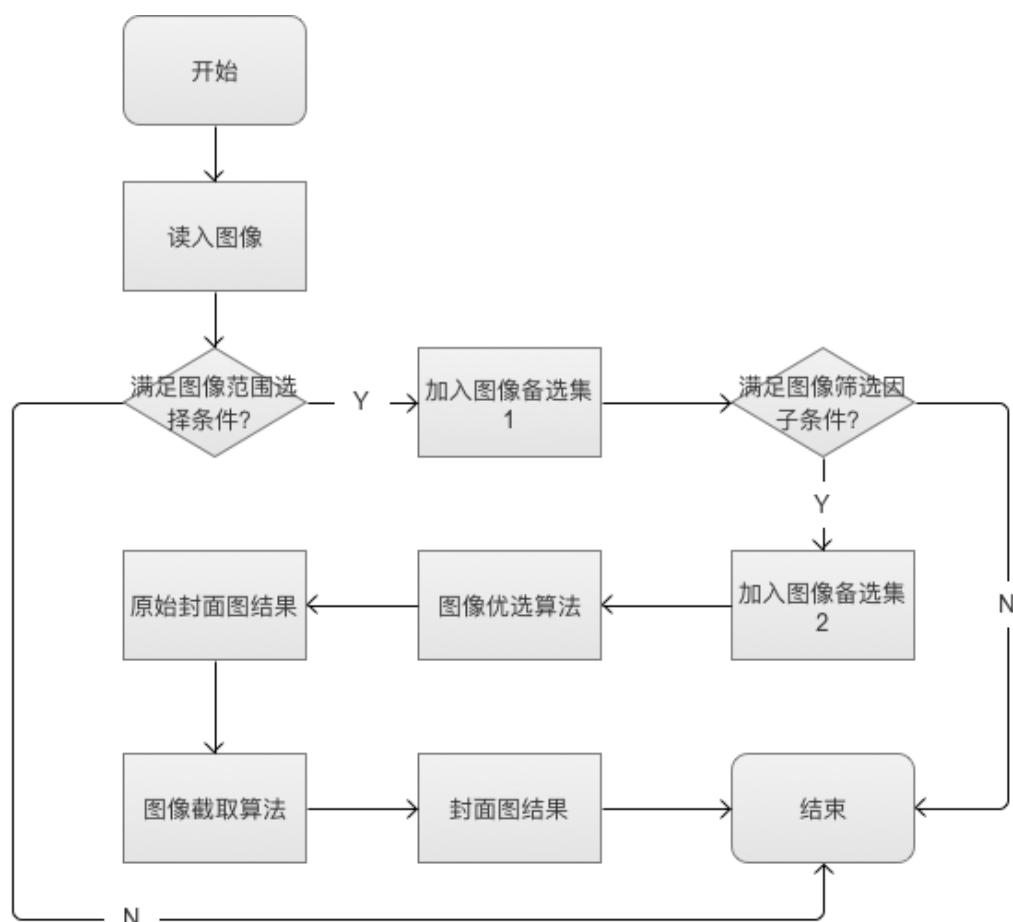


图 4-1 图像有选系统功能流程

Step1、开始，读入文章转码服务（是对文章文本进行初始化处理的一个服务）提供的图像列表

Step 2、判断并将满图图像范围选择条件的图像加入到图像备选集 1 列表，不满足条件则直接结束，计算结果为空。

- Step 3、读取图像备选集 1 列表，判断并将满足图像筛选因子条件的图像加入到图像备选集 2 列表，不满足条件则直接结束，计算结果为空。
- Step 4、使用图像优选算法计算出一张原始的封面图结果。
- Step 5、使用图像截取算法处理原始封面图结果。
- Step 6、得到封面图结果，结束。

从上面的系统功能流程中，可以将图像优选系统划分为四个重要模块：图像范围选择模块、图像筛选因子模块、图像优选算法模块和图像截取模块。在 4.2 到 4.5 四个小节，将对这四个模块的需求分析和系统实现进行介绍。

4.2 图像范围选择模块

图文混排的购物文章，也就是元数据，图像的数量和质量（清晰度、大小等）是复杂多样的。这个特点决定了不可能将所有图像作为图像优选算法的元数据，防止过多的元数据在数据质量没有得到有效筛选时对图像优选算法在性能、效率上的影响。所以在通过图像优选系统的核心模块图像优选算法之前，需要对元数据进行清洗。这里对数据的清洗就是通过图像范围选择和图像筛选因子两步完成。

图像范围选择制定标准源自于对项目数据冷启动过程中精选的优质数据的研究。在项目的数据冷启动阶段，通过人工使用网络爬虫工具的方式，从不同来源网站抓取了 10000 篇购物类文章。这些文章涉及手机数码、汽车、户外运动、服饰鞋包、居家生活等不同领域。来源网站数目也在 100-150 个之间。在研究购物类文章图像选择范围特征时，选取了表 4-1 中的 10 个来源网站的 200 篇代表性购物文章。值得说明的是，一方面，这些购物文章是在很赞网站上阅读量超过 200（相对于未推广前的很赞网站所有文章而言是较高的阅读量），另一方面，感谢运营团队人工为这些购物文章挑选了封面图，便于算法和逻辑的研究。

表 4-1 图像范围确定样本

来源网站名称	文章类别	文章数量	人工选取图像 编号范围
什么值得买	综合	30	1-5
Chiphell	综合	30	2-5
汽车之家	汽车	20	1-3
Knewone	综合	20	1-3
手机中国	手机数码	10	2-5
Zealer	手机数码	10	1-end
易车网	汽车	20	1-3

虎扑识货	运动户外	20	1-end
骑行之家	运动户外	20	1-3
爱燃烧	运动户外	20	2-5

从上表数据可以知道，人工挑选购物文章中的优质图像作为文章封面图时，图像在前五张中基本可以选出，但是存在少部分购物网站的文章存在第一张图像不可取，同时，有一些内容质量普遍较高的网站中的文章图像质量接近，所有图像都可以作为备选图像。

在研究了 200 篇购物文章人工选取封面图的结果后，结合写作过程中的特点。可以得到结论：在文章首部和尾部是文章的黄金位置，一般情况下，这里的图像都是作者最想表达的。所以，图像范围选择的规则是选取“前四后一”，意思是对于每篇购物文章中的图像列表，选取从前到后 4 张通过筛选因子的图像和从后到前最少 1 张但不超过 2 张的图像，作为经过图像范围选择的图像，加入到图像备选集 1 中。

图像范围选择模块描述如下：

输入：原始图像列表 `proto_im`

输出：图像备选集 1 列表 `imlist`

#算法开始#

Step1: 遍历 `proto_im` 列表中图像地址并下载计算，同时记录图像顺序

Step2: 如果图像总数大于 5, 选取正序前 4 张图像和倒序第 1 张图像加入到 `imlist`;

如果图像总数小于等于 5, 将所有图像加入到 `imlist`。

#算法结束#

4.3 图像筛选因子模块

图像计算，也就是图像计算。在计算前，可以通过一些图像筛选因子的判断过滤元数据。图像筛选因子和图像范围选择的不同之处在于，图像筛选因子是根据图像特征的指标进行筛选；而图像范围选择是依据数据样本研究出来的图像选取标准。

图像筛选因子有以下 4 点：

1. 图像格式必须是 jpg、jpeg、png。

一般情况下，支持图文混排的富文本编辑器都只支持 jpg、jpeg、png 这三种图像格式，同时，从公司图床系统的支持情况来看，对这三种图像格式的支持表现极佳，而对其他格式则不予支持。所以，将图像格式作为图像筛选因子。

2. 图像宽高比和目标截图窗口比例不能相差过大。

经过图像优选系统选择出来的图像最终会作为文章封面图展示在推荐文章信息流、商品文章信息流等不同位置。虽然在不同位置的图像宽高不尽相同，但

是不同位置的图像宽高比在设计的时候是近似的。如表 4-2。图像宽高比和目标截图窗口比例不能相差过大，否则会导致经过图像优选系统的封面图结果在展示的时候出现比例失调的问题。

表 4-2 目标窗口宽高数据

目标截图窗口	宽（px）	高（px）	宽：高
推荐信息流	400	280	10： 7
同类文章	232	156	21： 13
商品购物文章	400	280	10： 7
登录页推荐文章	400	280	10： 7

根据上面表格可以看出，目标截图窗口的高宽比大约是 $7/10=0.7$ 。

3. 图像不能是灰度图。

灰度图像是图像中每个像素均只采取一个颜色的图像，俗称为黑白图。这种图像特点是色调较冷，且色彩单一，信息量也较色彩图少^[28]。图像优选系统选取优质的文章封面图，文章是和商品关联的，选取优质文章封面图的主要目的就是展示文章信息，吸引用户，期望触动用户情感波动。显然，灰度图和目的是不一致甚至背道而驰的。图 4-3 和图 4-4 同一个文章封面图采用灰度图和非灰度图的对比。



图 4-3 非灰度图



图 4-4 图 4-3 的灰度图

灰度图的计算方法，可以使用第三部分介绍的纹理特征计算时的灰度直方图计算结果。

4. 图像不能是“人造图”。

人造图是定义人工通过计算机软件制造或者加工过的图像。这类图像的特点是图像中心区域含有纯色块或者纯色块的颜色过多，或者颜色过少，或者补色过多。图 4-5 和图 4-6 说明了人造图现象。



图 4-5 人造图——纯色块



图 4-7 人造图——补色过多

人眼的色彩识别结构在同时看到三原色（Primary Color）时感到比较舒适。在看到两种原色或者两种原色的混合色的时候，呈现在人眼中的周围环境或者人眼中的盲点就会导致出现幻觉，得到第三种颜色。所以，将图像不能为“人造图”作为图像筛选因子条件之一。通过第三部分介绍的纯色背景（PCB）的提取方法 HSL 颜色模型可以判断图像是否为“人造图”。

以上所有 4 点筛选因子为布尔值判别条件。表 4-8 说明了判断结果和相应处理。

表 4-8 布尔值判断筛选因子结果示意

筛选因子	图像格式	图像比例	灰度图判断	人造图判断
图像格式	✓✓	✓✓	✓✓	✓✓
图像比例	✓✗	✓✓	✓✓	✓✓
灰度图判断	✓✗	✓✗	✓✓	✓✓
人造图判断	✓✗	✓✗	✓✗	✓✓

在上面表格中，将四个筛选因子作为判断条件，用✓和✗分别表示布尔值的 True 和 False。当且仅当四个筛选因子都通过时，图像才符合筛选因子条件。

图像筛选因子模块的算法描述如下：

输入：图像备选集 1——imlist

输出：图像备选集 2——imdata

#算法开始#

Step1: 遍历 imlist 中所有图像地址并下载计算

Step2: 判断图像格式。当图像格式为 jpg、jpeg、png 其中一种时，进行下一步；否则，开始计算下一张图像。

Step3: 判断图像比例，图像宽 imw，图像高 imh 满足条件：

$\text{abs}((0.7-1.0 \cdot \text{imh}/\text{imw})/0.7) \leq 0.45$ ，进行下一步；否则，开始计算下一张图像。

Step4: 判断灰度图。调用 isgrey 函数，返回结果为 False，进行下一步；返回 True，开始计算下一张图像。

Step5: 判断人造图。调用 `man_made` 函数, 返回结果为 `Flase`, 将图像存入 `imdata`; 返回 `True`, 开始计算下一张图像。

#算法结束#

4. 4 图像优选算法模块

进过图像范围选择和图像筛选因子选择两个判别条件后, 图像优选算法有了表现比较好的元数据, 接下来应该使用第三章介绍设计的基于图像特征向量的优选算法对图像备选集 2 中的图像进行优选。

4.4.1 图像优选算法数据结构

本系统使用 python 语言开发。由于 python 语言的数据结构中, 列表提供了多种操作方法便于列表的插入、查询、排序、复制等操作, 同时还支持列表当做队列、堆栈使用, 以及嵌套计算二维矩阵等特性。所以结合算法中对图像的存储要求和对特征向量的存储要求, 关键部分均使用列表作为算法存储计算的数据结构。其他部分涉及一些变量, 可以参考附录 A 中具体实现。

图像地址存储在列表 `proto_im`、`imlist`、`imdata` 中。图像特征提取结果的向量表示存储在列表 `feature` 中。特征 `feature` 列表和特征的映射关系如下:

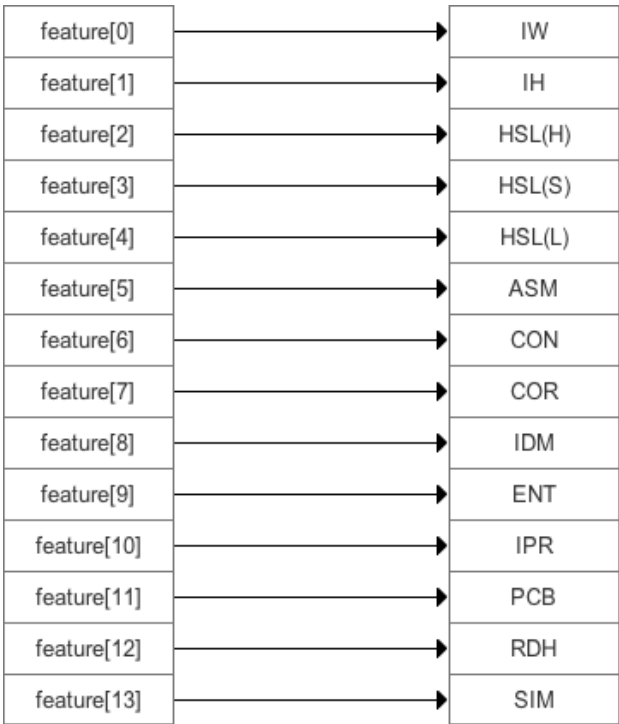


图 4-8 特征数据结构映射关系

4.4.2 图像优选算法的实现

图像优选算法具体步骤描述如下：

输入：图像备选集 2 中的图像列表 imdata。

输出：优选图像结果 scores[0]。

#算法开始#

Step1: 开始收集图像特征，并将特征向量化。

Step2: 将图像转化为灰度图，便于矩阵计算。

Step3: 使用 surf 算法计算局部兴趣点（IPD）和肖像画率（IPR）两个特征。

Step4: 使用 Haralick 算法计算全局特征，得到角二阶距（ASM）、对比度（CON）、自相关（COR）、逆距差（IDM）、熵（ENT）等全局特征。

Step5: 计算图像宽高比例，得到图像比例偏移值（RDH）。

Step6: 计算 HSL 颜色模型，得到纯色（PCB）、图像饱和度、图像光照图等全局特征。

Step7: 将上述所有特征向量化。

Step8: 使用上述向量化的特征：IW、IH、HSL(H)、HSL(S)、HSL(L)、ASM、CON、COR、IDM、ENT、IPR、PCB、PDH、SIM 总计 13 个特征向量进行打分计算，打分加权规则是（4-1）：

$$\text{score} = \text{CON} * \text{ENT} * (0.0001 + \text{IPR}) * \text{SIM} * (\text{HSL}(s) + \text{HSL}(l) + \text{ASM} + \text{COR} + \text{IDM} + \text{RDH}) \quad (4-1)$$

Step9: 选择 score[0] 作为优选结果输出。

#算法结束#

4.5 图像截取模块

图像截取模块是对图像优选结果进行进一步处理的模块。本模块将从截取策略分析和算法实现两部分进行介绍。

4.5.1 图像截取模块策略分析

在使用基于图像特征向量的图像优选算法完成图像截取步骤后，得到了可以作为封面图的优选图像结果。虽然在不同文章中得到的图像优选结果宽高比是接近 10:7，但是具体的宽高值是不同的。而且对于一些特定的图像，用户更关注的是局部特征。也就是说，当图像局部特征表现十分优秀时，是应该取图像局部作为封面图，而非整个图像。如下图 4-10 和 4-11 中，图 4-10 是封面图优选结果，但是这篇主要内容是将键盘的文

章中，应该更多将表达聚焦在键盘以及键盘配件上。图 4-10 虽然选出了键盘相关图像，但是图像本身比较模糊且途中还有鼠标、盒子等其他物体，所以这张图像在局部特征表现上效果不佳。经过处理的图像图 4-11 则对原图进行了图像截取和细节参数调整，得到一张基于原图像的新图像，这张新图像在清晰度、局部特征表达方面的表现就显得更好了。



图 4-10 优选图像结果



图 4-11 截图处理后结果

从上面分析中，可以看出，对图像进行截取处理是必要的。那么参考 twitter 的图像截取策略，决定图像截取的策略是：尽量对图像核心区域进行特写，截出有窗口既视觉的图像结果。

4.5.2 图像截取模块的实现

根据上述策略，具体的图像截取操作分以下几步：

- 1 根据计算出的兴趣点描述子（IPD）结果，按照 scale 大小取前 25%的点，计算包围边界（bounding）。
- 2 在 1 中计算结果的基础上向外夸张 40%，作为最终的边界值，进行图像截取
- 3 根据目标窗口比例（默认 400*280），进行截取和缩放。其中缩放采用 Antialias 模式^[29]，10 级 jpeg 压缩等级进行存储，以期尽可能保留画质水准。

图像截取算法具体步骤如下：

输入：图像优选算法结果 score[0]

输出：图像

#算法开始#

Step1: 计算边界值。调用边界值计算函数 crop 计算 score[0], 得到图像边界值 region

Step2: 截取图像。根据边界值结果 region，截图图像并完成缩放。

Step3: 将图像处理结果 cover_selector(list)写入数据库中封面图字段 henzan_cover

#算法结束#

4.6 方案验证

图像优选这个目标本身就是主观的，它深度依赖于产品策略，如果产品策略以商品为主，文章为辅，则应该以商品图为优选第一原则；如果产品策略以购物体验为主，商品为辅，则应该以场景图为优选第一原则。不同的策略下，各项特征的参数权重是不同的。所以图像优选系统是一个深度基于实践的工程型项目。图像优选算法及其系统的研究和设计背景是以商品策略为背景开始的，所以最终实验结果的“优选”判定标准是下列4点，其中第1、4点最重要：

- 1、图像是否以商品图为第一优选原则
- 2、图像质量（肉眼感觉）在备选图像中是否为最优
- 3、图像局部截取是否定位到准确的兴趣点
- 4、优选图像结果对网站信息流的点击率、转化率等指标有明显作用

方案验证的数据来源是项目数据冷启动阶段，积累的不同来源网站多种类型的10000篇购物信息有关文章（图文混排），以这些真实的购物信息文章为数据来源。数据选择方法方面，从10000个数据来源中综合考虑不同商品分类、随机算去200个数据，保证每个分类下覆盖15-20个数据。

下面表4-12是实验结果，由于篇幅限制，只展示部分。

表 4-12 实验结果（部分）

编号	人工选取结果	图像优选系统选取结果
1		
2		

3		
4		
5		
6		
7		



从上面部分对比结果中，可以直观且清晰的感觉到图像优选算法在挑选和商品相关的优质图像作为文章封面图这个问题上的优异表现。

依据前面提出的 4 点判断标准，前 3 点的判断主要以人的感觉来进行，从上面部分对比结果中，已经可以基本验证，并且根据测试完成 200 个数据的图像优选结果，在运营团队内部进行了“封面图满意度调查”，根据运营团队内部问卷结果，认定图像优选算法及其系统在封面图挑选准确率上高达 95.6%。第 4 点其实是根据网站日常访问 2 万人的行为，做出对图像优选结果的判断，是大量数据样本对实验结果的证明。图 4-12 和图 4-13 是使用图像优选系统前后最主要的网站推荐页用户行为数据指标，数据由谷歌分析提供。

表 4-13 使用图像有选系统前后很赞数据指标对比

数据指标	使用前	使用后
网站 7 日均访问 IP	130000	240000
推荐页 7 日均访问 IP	40000	100000
详情页 7 日均访问 IP	6400	43000
推荐页到详情页 7 日均转化率	0.16	0.43

由上表数据可知，在使用图像优选系统之后，网站访问量得到明显提升，这里主要原因是通过喵喵折插件上的精选文章推荐导流。推荐页访问量在网站访问量中占比明显提高，推荐页访问量到详情页访问量的转化率也明显提高。结合后两点，可以认定该系统缺失起到了他优选图像作为封面图的效果，而表现优秀的封面图成功吸引了用户注意力，激发了用户点击阅读的欲望。除了这些表现在网站数据指标上的好处，该系统性能在单线程下可以实现 5 分钟/篇的计算速度，每天可完成 7200 此计算。完全满足当前计

算要求，大大节约了运营人工选取封面图的时间成本，降低了人工选取封面图的主观因素，在大量数据下，综合表现更加优秀。

4.7 本章小结

在第三章设计并实现了基于图像特征向量的优选图像算法的基础上。为了能够满足公司业务需求，本章提出了将该算法工程化的基于图像特征向量的优选图像系统的设计与实现。包括数据处理的图像范围选择模块和图像筛选因子模块，核心的图像优选算法模块，以及为了更好的展示封面图效果的图像截取模块。

经试验证明了该系统在图像优选作为封面图操作上的效果和性能都大大优于人工处理的结果，同时使用该系统后，对网站各项数据指标有显著提升，为公司业务带来增长。

第五章 总结与展望

5.1 总结

基于内容的导购服务已经成为电子商务即将成为电子商务行业下一个爆发点，从阅读出发到购物的购物模式将改变人类的消费习惯。而在内容数据处理过程中，优质图像选择作为封面图展示在信息流，能否打动用户，成为重要的业务问题。目前为止，还没有专门针对购物信息中的图像进行优选的算法和设计，同时，这种高度出于业务而设计的系统需要和产品的设计逻辑高度吻合。不同策略下，系统对于优选的定义是不同的。而本文正是在此背景下，对图像优选算法及其系统进行研究。探讨如何以挑选画质精美、构图有张力、摄影质感强、能触动人类情感波动的图像作为优选策略。

本文主要做了以下几个方面的工作：

- 1、从人类认知和感知的角度出发，确定了以画质、色彩、纹理、信息熵、内容语义为优选原则，并对这些原则对应的模型和特征进行了了解和学习，为后续特征提取和算法研究打下基础。
- 2、提出了基于图像特征向量的优选图像算法。在研究数据样本的基础和第一步从认知角度提出的优选原则的基础上，提出了参与优选的图像特征。并对这些特征的提取方法进行了研究和概述。
- 3、提出了基于图像的特征向量的优选图像系统的设计与实现。在完成基于图像特征向量的优选图像算法的基础上，提出将优选图像工程化。并且确定了数据前期准备工作：图像范围选择、图像筛选因子；确定了核心的图像优选算法的流程、设计与实现；确定了图像后期处理的图像截取算法原则、设计步骤。经实验证明，图像优选系统在实现优选的同时，还达到节约了人工成本，提高了网站数据指标的效果。

5.2 展望

本文提出算法的思路是从图像的外在表现，从而进一步联系到图像特征。通过打分机制处理图像特征的向量化数据，完成图像的优选。相对于当前业务要求来说，达到满意的效果，但是仍有一些不足和需要继续改进的地方，未来作者将从以下几个方面继续研究：

- 1、算法性能在时间效率上还有改善的空间。

2、本文以图像特征为出发点，在内容语义特征方面。只是通过灰度直方图特征进行简单的判断实现。如何通过更好的内容语义识别算法，选择更加符合内容语义，能选出文章想要表达的商品特性的图像，是未来需要努力研究的方向。

参考文献

- [1] CNNIC. 第 37 次中国互联网络发展状况统计报告. 2016. 2
- [2] 赵宇翔, 范哲, 朱庆华等. 用户生成内容 (UGC) 概念解析及研究进展[J]. 中国图书馆学报, 2012, (5): 68-81. DOI:10.3969/j.issn.1001-8867. 2012. 05. 007.
- [3] 倪国强, 刘琼. 多源图像配准技术分析与展望[J]. 光电工程, 2004, 31(9): 1-6. DOI:10.3969/j.issn.1003-501X. 2004. 09. 001.
- [4] 袁贺, 卢姗, 汪明峰等. 电子商务环境对消费者传统购物行为的影响[J]. 商业研究, 2011, (2): 211-216. DOI:10.3969/j.issn.1001-148X. 2011. 02. 039.
- [5] 百度百科. 2016. 画质.
<http://baike.baidu.com/link?url=j-NKJnKlIPwUqIc0R9tp0FIIHmiqmwndo6bNsQJwk1d88vsnkXrnHWYA3jLJfLibQ6-QaGGBPKp5P-1Dbvl4Hq>
- [6] 百度百科. 2016. 锐度 <http://baike.baidu.com/view/1500619.htm>
- [7] 百度百科. 2016. 色彩 <http://baike.baidu.com/subview/37967/6055194.htm>
- [8] 黄小花. RGB 与 CMYK 色彩模式[J]. 信息通信, 2012, (6): 21-22. DOI:10.3969/j.issn.1673-1131. 2012. 06. 012.
- [9] 薄华, 马缚龙, 焦李成等. 图像纹理的灰度共生矩阵计算问题的分析[J]. 电子学报, 2006, 34(1): 155-158, 134. DOI:10.3321/j.issn:0372-2112. 2006. 01. 032.
- [10] 刘丽, 匡纲要. 图像纹理特征提取方法综述[J]. 中国图象图形学报 A, 2009, 14(4): 622-635.
- [11] 陆明俊, 王润生. 基于 MRF 模型的可靠的图像分割[J]. 电子学报, 1999, 27(2): 87-89. DOI:10.3321/j.issn:0372-2112. 1999. 02. 023.
- [12] 马晓川, 侯朝焕, 赵荣椿等. 一种新的纹理图像 Gibbs 模型[J]. 计算机学报, 1998, 21(z1): 303-307. DOI:10.3321/j.issn:0254-4164. 1998. z1. 052.
- [13] 廖大麟. 随机事件的不确定性或信息量的度量——信息熵[J]. 毕节学院学报, 2006, 24(4): 35-38. DOI:10.3969/j.issn.1673-7059. 2006. 04. 009.
- [14] 周丽华, 周纪勤, 柏正尧等. 基于图像信息熵的商标图像检索[J]. 计算机应用, 2000, 20(2): 21-23.
- [15] 闻京. 基于 Matlab 图像的文字区域检测新方法[J]. 网络与信息, 2011, (1): 42-43. DOI:10.3969/j.issn.1008-0252. 2011. 01. 028.
- [16] 薄华, 马缚龙, 焦李成等. 图像纹理的灰度共生矩阵计算问题的分析[J]. 电子学报, 2006, 34(1): 155-158, 134. DOI:10.3321/j.issn:0372-2112. 2006. 01. 032.
- [17] 任国贞, 江涛. 基于灰度共生矩阵的纹理提取方法研究[J]. 计算机应用与软件, 2014, (11): 190-192, 325. DOI:10.3969/j.issn.1000-386x. 2014. 11. 047.
- [18] Pantic, I., Pantic, S., Paunovic, J. et al. Nuclear entropy, angular second moment, variance and texture correlation of thymus cortical and medullar lymphocytes: Grey level co-occurrence matrix analysis[J]. Anais da Academia Brasileira de Ciencias, 2013, 85(3): 1063-1072.
- [19] 殷松峰, 曹良才, 杨华等. 提高夜视融合目标可探测性的颜色对比度增强方法[J]. 红外与毫米波学报, 2009, 28(4): 281-284. DOI:10.3321/j.issn:1001-9014. 2009. 04. 010.
- [20] 陈国金, 朱妙芬, 张克松等. 图像调焦过程的清晰度评价函数研究[J]. 数据采集与处理, 2009, 24(2): 165-169. DOI:10.3969/j.issn.1004-9037. 2009. 02. 008.

- [21] 张良, 王海丽, 吴仁彪等. 基于改进局部不变特征的兴趣点匹配[J]. 电子与信息学报, 2009, 31(11): 2620-2625.
- [22] 黄小花. RGB 与 CMYK 色彩模式[J]. 信息通信, 2012, (6): 21-22. DOI:10.3969/j.issn.1673-1131.2012.06.012.
- [23] 魏宝刚, 鲁东明, 潘云鹤等. 多颜色空间上的交互式图像分割[J]. 计算机学报, 2001, 24(7): 770-775. DOI:10.3321/j.issn:0254-4164.2001.07.016.
- [24] 朱铮涛, 黎绍发, 陈华平等. 基于图像熵的自动聚焦函数研究[J]. 光学精密工程, 2004, 12(5): 537-542. DOI:10.3321/j.issn:1004-924X.2004.05.014.
- [25] 孙吉贵, 刘杰, 赵连宇等. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48-61. DOI:10.3724/SP.J.1001.2008.00048.
- [26] Marjan Abdeyazdan. Data clustering based on hybrid K-harmonic means and modifier imperialist competitive algorithm[J]. Journal of supercomputing, 2014, 68(2): 574-598.
- [27] 王君本, 卢选民, 贺兆等. 一种基于快速鲁棒特征的图像匹配算法[J]. 计算机工程与科学, 2011, 33(2): 112-117. DOI:10.3969/j.issn.1007-130X.2011.02.022.
- [28] 林开颜, 吴军辉, 徐立鸿等. 彩色图像分割方法综述[J]. 中国图象图形学报, 2005, 10(1): 1-10. DOI:10.3969/j.issn.1006-8961.2005.01.001.
- [29] Li-Qian Ma, Kun Xu. Efficient antialiased edit propagation for images and videos. Computers & graphics, 2012, 36(8)

致 谢

在论文完成之际，我要由衷感谢北京交通大学软件学院四年对我的培养。同时，还要感谢喵喵物语科技有限公司同事对项目的支持和帮助。特别感谢企业指导老师高扬在算法和工程化方面的指导。感谢 xxx 老师在毕业设计开题和论文撰写过程中的细心指导。

附 录

附录 A 翻译文献

原文：Speeded-Up Robust Features (SURF)

3. Interest point detection

Our approach for interest point detection uses a very basic Hessian matrix approximation. This lends itself to the use of integral images as made popular by Viola and Jones [41], which reduces the computation time drastically. Integral images fit in the more general framework of boxlets, as proposed by Simard et al. [38].

3.1. Integral images

In order to make the article more self-contained, we briefly discuss the concept of integral images. They allow for fast computation of box type convolution filters. The entry of an integral image I_{Σ} at a location \mathbf{x} represents the sum of all pixels in the input image I within a rectangular region formed by the origin and \mathbf{x} .

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

Once the integral image has been computed, it takes three additions to calculate the sum of the intensities over any upright, rectangular area (see Fig. 1). Hence, the calculation time is independent of its size. This is important in our approach, as we use big filter sizes.

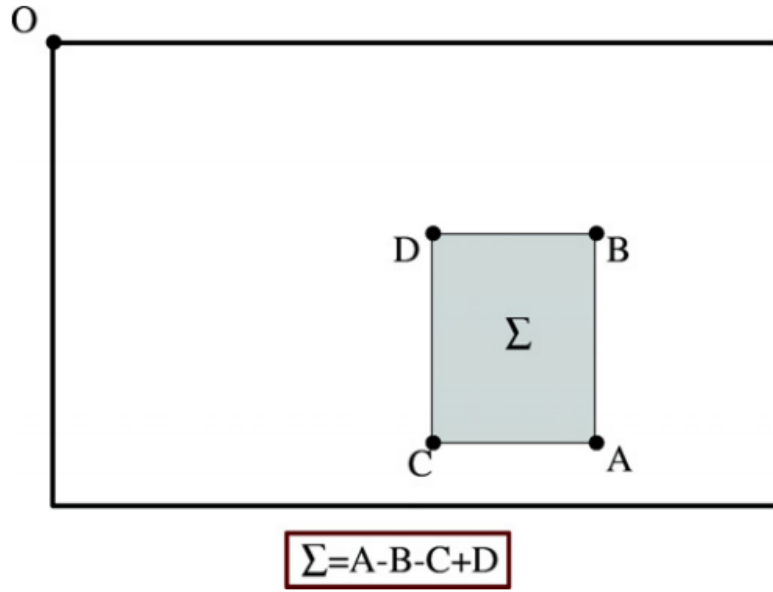


Fig. 1. Using integral images, it takes only three additions and four memory accesses to calculate the sum of intensities inside a rectangular region of any size.

3.2. Hessian matrix-based interest points

We base our detector on the Hessian matrix because of its good performance in accuracy. More precisely, we detect blob-like structures at locations where the determinant is maximum. In contrast to the Hessian-Laplace detector by Mikolajczyk and Schmid [26], we rely on the determinant of the Hessian also for the scale selection, as done by Lindeberg [21]. Given a point $\mathbf{x} = (x, y)$ in an image I , the Hessian matrix $\mathcal{H}(\mathbf{x}; r)$ in \mathbf{x} at scale r is defined as follows

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}$$

where $L_{xx}(\mathbf{x}; r)$ is the convolution of the Gaussian second order derivative $\partial^2 \mathcal{G} / \partial x^2$ with the image I in point \mathbf{x}

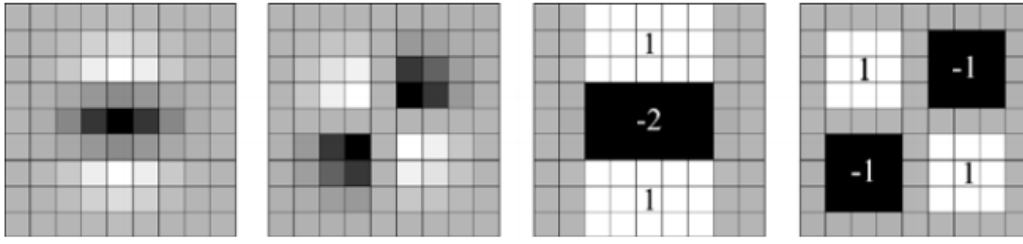


Fig. 2. Left to right: The (discretised and cropped) Gaussian second order partial derivative in y - (L_{yy}) and xy -direction (L_{xy}), respectively; our approximation for the second order Gaussian partial derivative in y - (D_{yy}) and xy -direction (D_{xy}). The grey regions are equal to zero.

Gaussians are optimal for scale-space analysis [19,20], but in practice they have to be discretised and cropped (Fig. 2, left half). This leads to a loss in repeatability under image rotations around odd multiples of $\pi/4$. This weakness holds for Hessian-based detectors in general. Fig. 3 shows the repeatability rate of two detectors based on the Hessian matrix for pure image rotation.

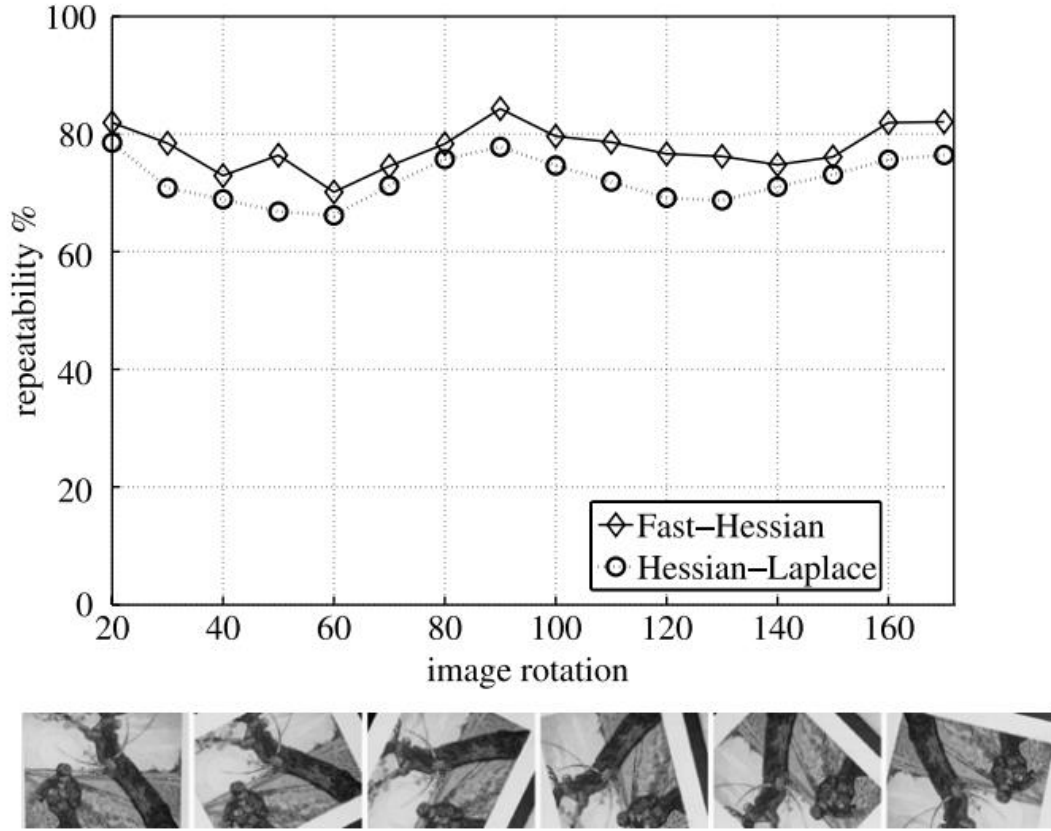


Fig. 3. Top: Repeatability score for image rotation of up to 180° . Hessian-based detectors have in general a lower repeatability score for angles around odd multiples of $\frac{\pi}{4}$. Bottom: Sample images from the sequence that was used. Fast-Hessian is the more accurate version of our detector (FH-15), as explained in Section 3.3.

The repeatability attains a maximum around multiples of $\pi/2$. This is due to the square shape of the filter. Nevertheless, the detectors still perform well, and the slight decrease in performance does not outweigh the advantage of fast convolutions brought by the discretization and cropping. As real filters are non-ideal in any case, and given Lowe's success with his LoG approximations, we push the approximation for the Hessian matrix even further with box filters (in the right half of Fig. 2). These approximate second order Gaussian derivatives and can be evaluated at a very low computational cost using integral images. The calculation time therefore is independent of the filter size. As shown in Section 5 and Fig. 3, the performance is comparable or better than with the discretised and cropped Gaussians. The box filters in Fig. 2 are approximations of a Gaussian with $\sigma = 1, 2$ and represent the lowest scale (i.e. highest spatial resolution) for computing the blob response maps. We will denote them by D_{xx} , D_{yy} , and D_{xy} . The weights applied to the rectangular regions are kept simple for computational efficiency. This yields

$$\det(\mathcal{H}_{\text{approx}}) = D_{xx}D_{yy} - (wD_{xy})^2$$

The relative weight of the filter responses is used to balance the expression for the Hessian's determinant. This is needed for the energy conservation between the Gaussian kernels and the approximated Gaussian kernels,

$$w = \frac{|L_{xy}(1.2)|_F |D_{yy}(9)|_F}{|L_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.912... \simeq 0.9$$

where $|x|_F$ is the Frobenius norm. Notice that for theoretical correctness, the weighting changes depending on the scale. In practice, we keep this factor constant, as this did not have a significant impact on the results in our experiments.

Furthermore, the filter responses are normalised with respect to their size. This guarantees a constant Frobenius norm for any filter size, an important aspect for the scale space analysis as discussed in the next section. The approximated determinant of the Hessian represents the blob response in the image at location x . These responses are stored in a blob response map over different scales, and local maxima are detected as explained in Section 3.4.

3.3. Scale space representation

Interest points need to be found at different scales, not least because the search of correspondences often requires their comparison in images where they are seen at different scales. Scale spaces are usually implemented as an image pyramid. The images are repeatedly smoothed with a Gaussian and then sub-sampled in order to achieve a higher level of the pyramid. Lowe [24] subtracts these pyramid layers in order to get the DoG (Difference of Gaussians) images where edges and blobs can be found.

Due to the use of box filters and integral images, we do not have to iteratively apply the same filter to the output of a previously filtered layer, but instead can apply box filters of any size at exactly the same speed directly on the original image and even in parallel (although the latter is not exploited here). Therefore, the scale space is analysed by up-scaling the filter size rather than iteratively reducing the image size, Fig. 4. The output of the 9×9 filter, introduced in previous section, is considered as the initial scale layer, to which we will refer as scale s . $1:2$ (approximating Gaussian derivatives with $r : 1:2$). The following layers are obtained by filtering the image with gradually bigger masks, taking into account the discrete nature of integral images and the specific structure of our filters.

Note that our main motivation for this type of sampling is its computational efficiency. Furthermore, as we do not have to downsample the image, there is no aliasing. On the downside, box filters preserve high-frequency components that can get lost in zoomed-out variants of the same scene, which can limit scale-invariance. This was however not noticeable in our experiments.

The scale space is divided into octaves. An octave represents a series of filter response maps obtained by convolving the same input image with a filter of increasing size. In total, an octave encompasses a scaling factor of 2 (which implies that one needs to more than double the filter size, see below). Each octave is subdivided into a constant number of scale levels. Due to the discrete nature of integral images, the minimum scale difference between two subsequent scales depends on the length l_0 of the positive or negative lobes of the partial second order derivative in the direction of derivation (x or y), which is set to a third of the filter size length. For the 9×9 filter, this length l_0 is 3. For two successive levels, we must increase this size by

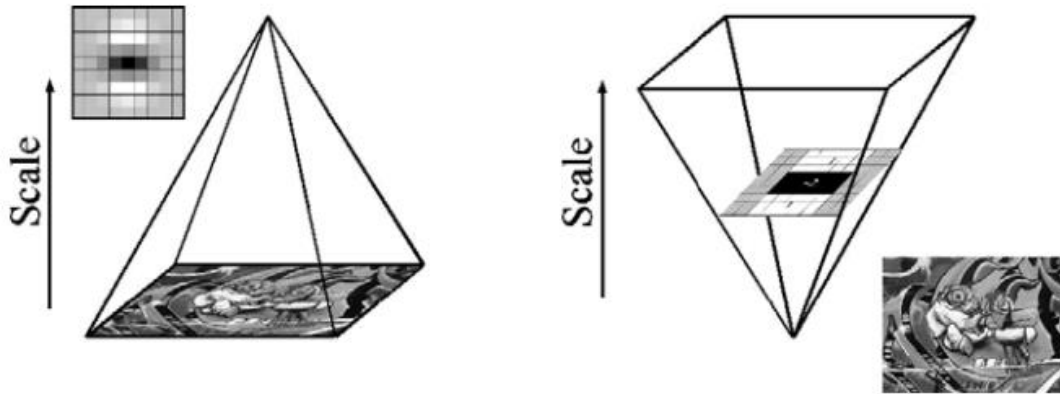


Fig. 4. Instead of iteratively reducing the image size (left), the use of integral images allows the up-scaling of the filter at constant cost (right).

a minimum of 2 pixels (1 pixel on every side) in order to keep the size uneven and thus ensure the presence of the central pixel. This results in a total increase of the mask size by 6 pixels (see Fig. 5). Note that for dimensions different from (e.g. the width of the central band for the vertical filter in Fig. 5), rescaling the mask introduces roundingoff errors. However, since these errors are typically much smaller than, this is an acceptable approximation.

The construction of the scale space starts with the filter, which calculates the blob response of the image for the smallest scale. Then, filters with sizes 15×15 , 21×21 , and 27×27 are applied, by which even more than a scale change of two has been achieved. But this is needed, as a 3D non-maximum suppression is applied both spatially and over the

neighbouring scales. Hence, the first and last Hessian response maps in the stack cannot contain such maxima themselves, as they are used for reasons of comparison only. Therefore, after interpolation, see Section 3.4, the smallest possible scale is $r = 1: 6 \cdot 1: 212.9$ corresponding to a filter size of 12×12 , and the highest to $r = 3: 2 \cdot 1: 224.9$. For more details, we refer to [2]. Similar considerations hold for the other octaves. For each new octave, the filter size increase is doubled (going from 6–12 to 24–48). At the same time, the sampling intervals for the extraction of the interest points can be doubled as well for every new octave. This reduces the computation time and the loss in accuracy is comparable to the image sub-sampling of the traditional approaches. The filter sizes for the second octave are 15, 27, 39, 51. A third octave is computed with the filter sizes 27, 51, 75, 99 and, if the original image size is still larger than the corresponding filter sizes, the scale space analysis is performed for a fourth octave,

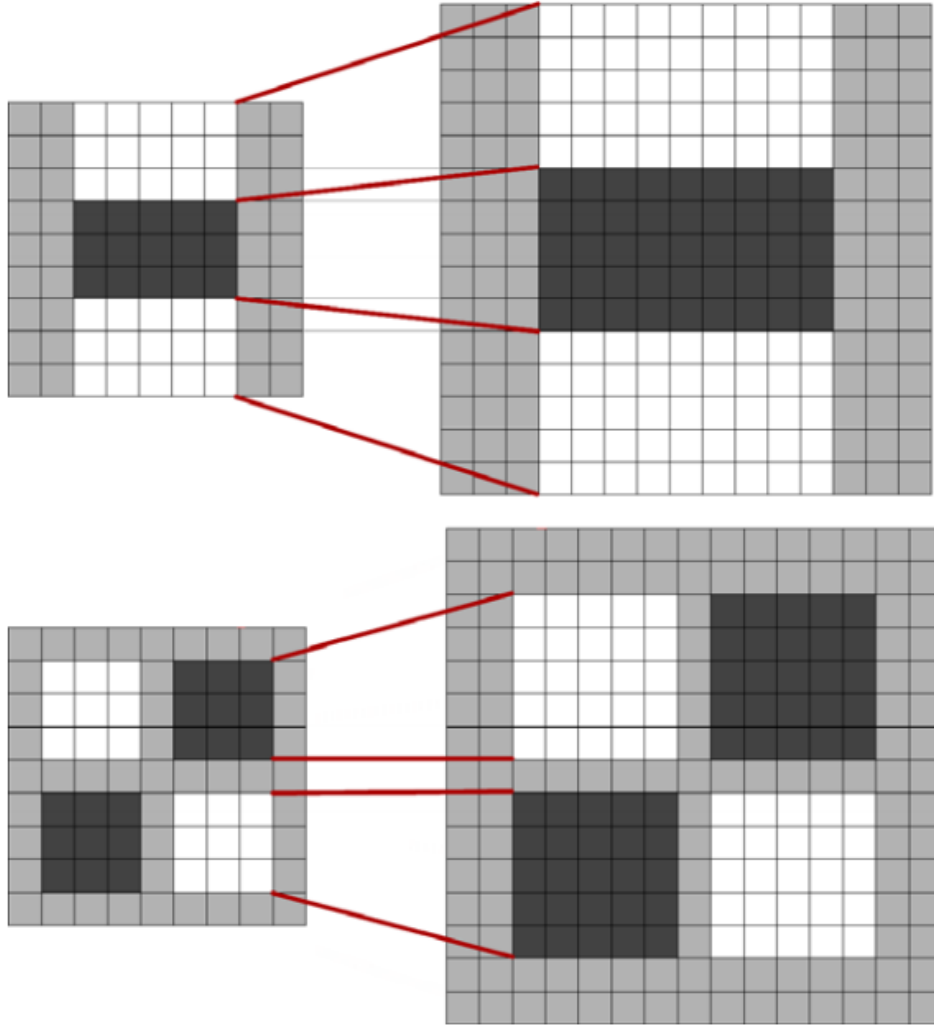


Fig. 5. Filters D_{yy} (top) and D_{xy} (bottom) for two successive scale levels (9×9 and 15×15). The length of the dark lobe can only be increased by an even number of pixels in order to guarantee the presence of a central pixel (top).

using the filter sizes 51, 99, 147, and 195. Fig. 6 gives an overview of the filter sizes for the first three octaves. Further octaves can be computed in a similar way. In typical scalespace analysis however, the number of detected interest points per octave decays very quickly, cf. Fig. 7.

The large scale changes, especially between the first filters within these octaves (from 9 to 15 is a change of 1.7), renders the sampling of scales quite crude. Therefore, we have also implemented a scale space with a finer sampling of the scales. This computes the integral image on the image up-scaled by a factor of 2, and then starts the first octave by filtering with a filter of size 15. Additional filter sizes are 21, 27, 33, and 39. Then a second octave starts, again using filters which now increase their sizes by 12 pixels, after which a third and fourth

octave follow. Now the scale change between the first two filters is only 1.4 (21/15). The lowest scale for the accurate version that can be detected through quadratic interpolation is 1:2.

As the Frobenius norm remains constant for our filters at any size, they are already scale normalised, and no further weighting of the filter response is required, for more information on that topic, see [22].

3.4. Interest point localisation

In order to localise interest points in the image and over scales, a non-maximum suppression in a neighbourhood is applied. Specifically, we use a fast variant introduced by Neubeck and Van Gool [33]. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space with the method proposed by Brown and Lowe [5]. Scale space interpolation is especially important in our case, as the difference in scale between the first layers of every octave is relatively large. Fig. 8 shows an example of the detected interest points using our ‘Fast-Hessian’ detector.

翻译：Speeded-Up Robust Features (SURF)

兴趣点检测

我们的兴趣点检测方法使用了非常基本的 Hessian 矩阵近似，使用由 Viola 和 Jones 推广的积分图，这将大大减少计算量。Simard 等指出，积分图适用于大量箱式小波的一般框架。

积分图

为了让这篇文章更完备，我们简要讨论一下积分图的概念。它使得箱型卷积滤波速度更快。积分图 $I_{\Sigma}(\mathbf{x})$ 在位置 $\mathbf{x} = (x, y)$ 处代表 I 在 \mathbf{x} 之前的矩形区域像素值的和。

$$I_{\Sigma}(\mathbf{x}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

一旦积分图计算好了，计算任何垂直矩形区域的密度只需要三个加法，如图 1。因此，计算时间将不依赖于大小。这在我们的方法中是很重要的，因为我们使用大尺寸滤波器。

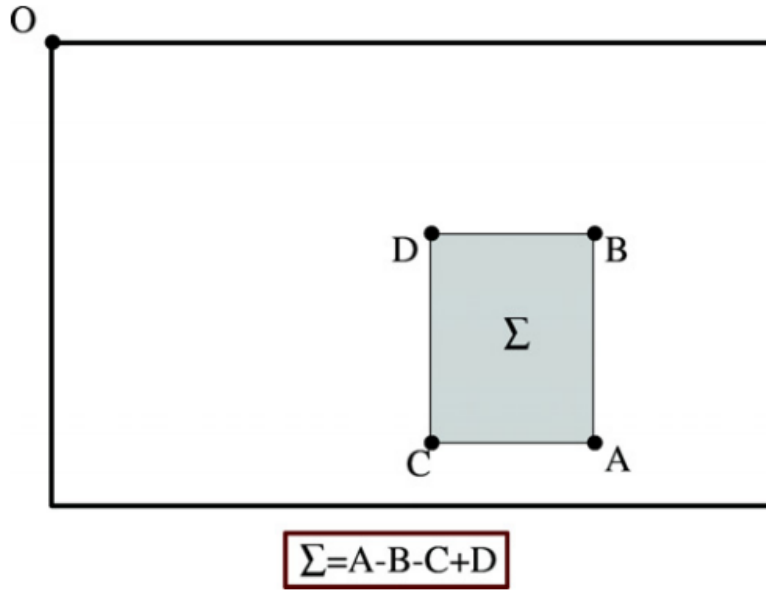


Fig. 1. Using integral images, it takes only three additions and four memory accesses to calculate the sum of intensities inside a rectangular region of any size.

3.2 基于 Hessian 矩阵的兴趣点检测

我们使用 Hessian 矩阵是因其良好的精确度。确切的说，我们检测行列式最大的类似斑块的结构。相对于 Mikolajczyk 和 Schmid 的 Hessian-Laplace 检测子，我们同样使用 Hessian 矩阵来选择尺度，这一点 Lindeberg 已经做过。

对于一个图像中的点 $\mathbf{x} = (x, y)$ ，尺度为 σ 的 Hessian 矩阵 $\mathcal{H}(\mathbf{x}, \sigma)$ 定义为：

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}$$

其中 $L_{xx}(\mathbf{x}, \sigma)$ 为点 \mathbf{x} 处的高斯二阶倒数 $\frac{\partial^2}{\partial x^2} g(\sigma)$ 的卷积。 $L_{xy}(\mathbf{x}, \sigma)$ 和 $L_{yy}(\mathbf{x}, \sigma)$ 与之类似。

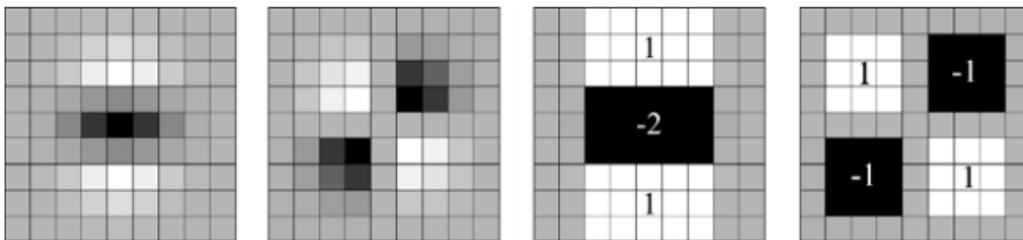


Fig. 2. Left to right: The (discretised and cropped) Gaussian second order partial derivative in y - (L_{yy}) and xy -direction (L_{xy}), respectively; our approximation for the second order Gaussian partial derivative in y - (D_{yy}) and xy -direction (D_{xy}). The grey regions are equal to zero.

高斯对于尺度分析是最优的，但在实际应用中需要离散化和裁剪（图 2）。这使其在 $\pi/4$ 的奇数倍旋转时的可重复性有损失。这是基于 Hessian 检测子的通病。图 3 显示了两种基于 Hessian 的检测子对于单纯图像旋转的可重复性。可重复性在 $\pi/2$ 的整数倍时达到最大，这是因为滤波器是方形的。然而，检测子的表现还是不错的，性能的微小减少并不比离散和裁剪带来的快速卷积的优点重要。实际的滤波器不可能是理想的，考虑到 Lowe 在 LoG 近似方法上的成功，我们进一步近似了 Hessian 矩阵。这种高斯二阶倒数近似可以利用积分图，大大减少计算复杂度。计算时间也不依赖于滤波器大小。第五部分和图 3 显示，其性能与离散裁剪的高斯法相当或更好。

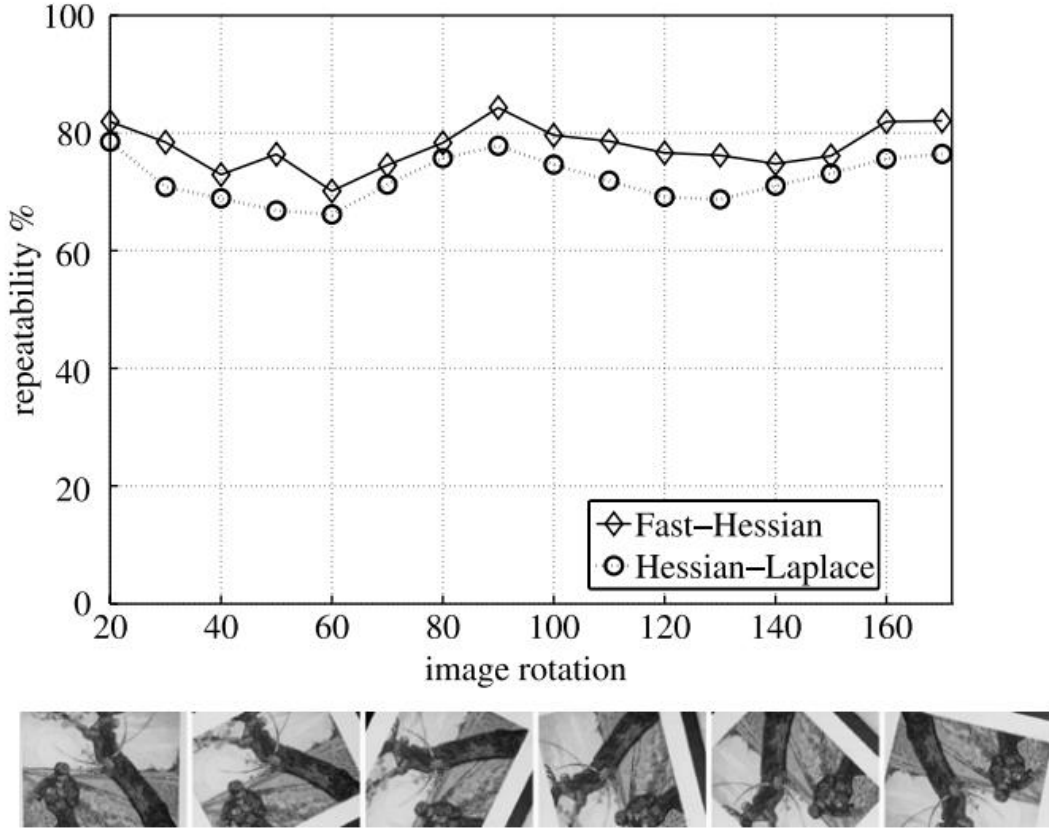


Fig. 3. Top: Repeatability score for image rotation of up to 180° . Hessian-based detectors have in general a lower repeatability score for angles around odd multiples of $\frac{\pi}{4}$. Bottom: Sample images from the sequence that was used. Fast-Hessian is the more accurate version of our detector (FH-15), as explained in Section 3.3.

图 2 中的 9×9 方形滤波器是 $\sigma = 1.2$ 的高斯滤波器的近似，这也是计算斑块响应图的最小的尺度（最大的空间分辨率）。我们将其标注为 D_{xx} 、 D_{yy} 和 D_{xy} 。为了计算简单，我们将方形区域的权重设为固定值。因此：

$$\det(\mathcal{H}_{\text{approx}}) = D_{xx}D_{yy} - (wD_{xy})^2$$

滤波器响应的相关权重 w 是为了平衡 Hessian 行列式的表达式。这是为了保持高斯核和近似高斯核的能量守恒。

$$w = \frac{|L_{xy}(1.2)|_F |D_{yy}(9)|_F}{|L_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.912... \simeq 0.9$$

其中 $|X|_F$ 是 F 范数。为了理论上的正确性，权重需要根据尺度变化。实际中，我们设为常数，因为它对实验结果并没有重要影响。

此外，滤波器响应根据尺度进行了归一化，这保证了对于不同尺寸有等量的 F 范数，这是上一节中讨论的，对于尺度分析很重要的一方面。

近似的 Hessian 行列式代表了图像在 X 点处的斑块响应。这些响应保存在不同尺度的斑块响应图中，3.4 节中将会检测局部最大值。

3.3 尺度空间表示

兴趣点需要在不同的尺度进行检测，不只是因为一致性搜索的比较需要在不同尺度的图片中进行。尺度空间通常以图像金字塔的形式呈现。图像多次被高斯平滑，然后下采样，以得到金字塔高层的图像。Lowe 将金字塔不同层次的图像相减以得到 DoG (Difference of Gaussians) 图像。

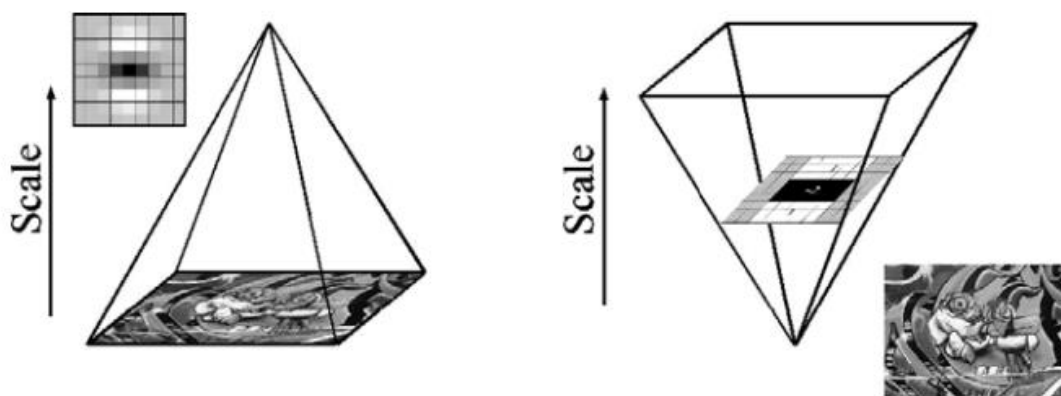


Fig. 4. Instead of iteratively reducing the image size (left), the use of integral images allows the up-scaling of the filter at constant cost (right).

因为用了积分图，我们不需要将同样的滤波器迭代应用在上次输出的图层上，而是使用任意尺度的方形滤波器，这样不同层可以有同样的速度，甚至可以并行执行（虽然还未在本文使用）。因此，尺度空间是使用放大的滤波器尺度，而不是迭代减小图像大小。图 4 是上节介绍的 9×9 滤波器输出，这是初始尺度层，我们记为尺度 $s = 1.2$ （近似于 $\sigma = 1.2$ 的高斯导数）。接下来的图层通过使用逐渐增大的滤波器得到，也利用积分图的离散特性和我们滤波器的特殊结构。

我们这样采样的动机是为了减小计算量。此外，因为我们不需要下采样，因此也不需要对齐。在下层，方形滤波器保存放大后会消失的高频部分，这会限制尺度不变性。然而这在我们的实验中并不明显。

尺度空间被划分为组，一组代表一系列滤波器响应图，通过对同样的图像使用不断增大的滤波器卷积得到。总体上，一组包含尺度因子为 2（这意味着滤波器大小需要至少增加一倍）。一个组又被分为固定数量的子尺度层。因为积分图的离散型，两个子层

间的最小尺度差依赖于 l_0 ，二阶偏导在倒数方向（x 或 y）的突出块的长度，这个值被设置为滤波器大小的三分之一。对于 9×9 滤波器，这个值是 3。对于两个连续层，尺度至少增加 2 像素（一边一个），以保持大小为偶数，确保中心点的存在。这使得总的覆盖面积增加了 6 像素（如图 5）。另外，由于 l_0 对于面积的差异性（例如，垂直滤波器中间段的宽度），滤波器大小的调整带来了舍入误差。然后，由于这些误差通常小于 l_0 ，这是可以接受的近似。

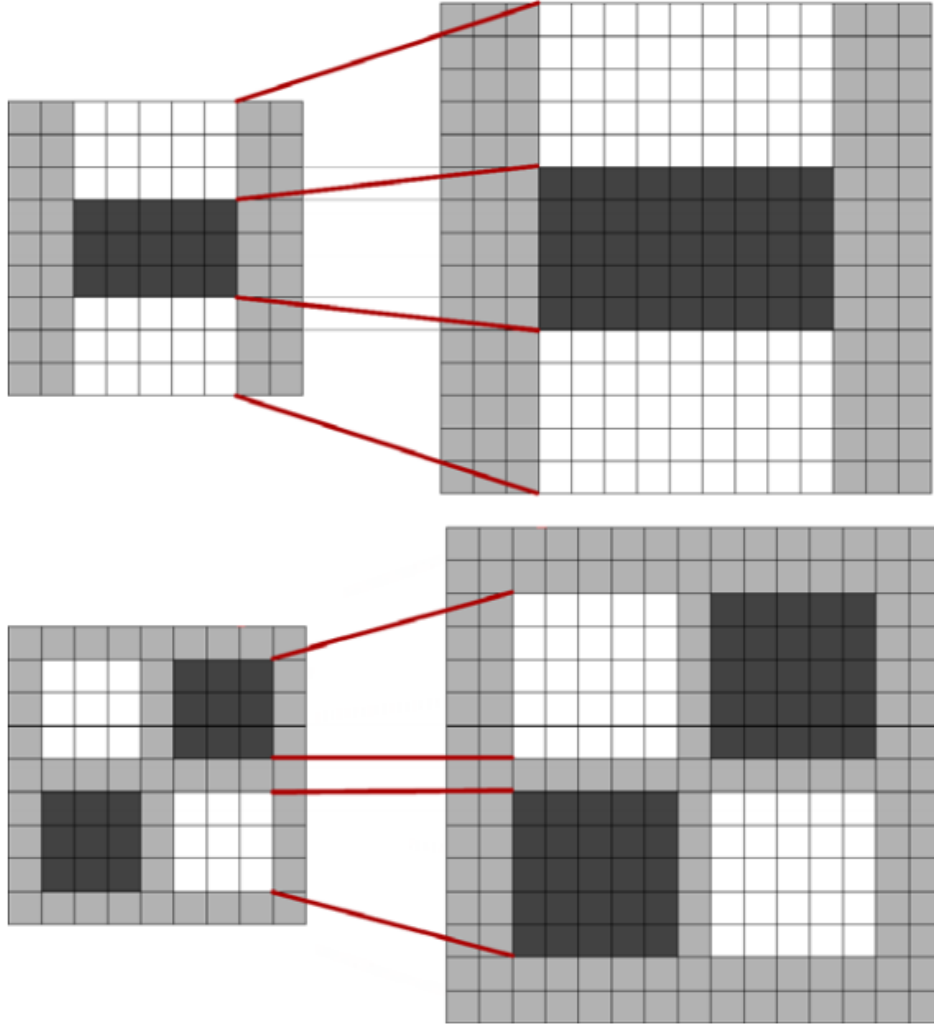


Fig. 5. Filters D_{yy} (top) and D_{xy} (bottom) for two successive scale levels (9×9 and 15×15). The length of the dark lobe can only be increased by an even number of pixels in order to guarantee the presence of a central pixel (top).

尺度空间的构建从 9×9 滤波器开始，计算图像的斑块响应，作为最小尺度层。然后，滤波器增至 15×15 ， 21×21 ，和 27×27 ，尺度变化了两倍多。但这是必须的，因为 3D 非最大化抑制不仅要在本空间内，还要在临尺度内进行。因此，第一和最后一层 Hessian 响应图本身并不能包含最大值，他们只是用来同其他层做比较的。所以，如 3.4 节所述，插之后可能的最小尺度是 $\sigma = 1.6 = 1.2 \cdot (12/9)$ ，对应的滤波器大小是 12×12 ，最大到 $\sigma = 3.2 = 1.2 \cdot (24/9)$ 。

其他组也类似。对每个新组，滤波器大小增倍（从 6-12 到 24-48）。同时，每个新组提取兴趣点的采样间隔加倍。相对于传统下采样方法，这将减少计算量和精确度损失。第二组的滤波器大小是 15,27,39,51，第三组是 27,51,75,99，如果原始图片还大于对应滤波器，尺度空间还需要第四组分析，滤波器大小为 51,99,147, 195。图 6 前三组滤波器大小的总图。之后的组可用同样的方法计算。然而在典型的尺度分析中，每组检测到的兴趣点会急速减少。

较大的尺度变化，特别是每组的第一个滤波器（从 9 到 15，1.7 倍），提供的尺度采样相当粗糙。因此，我们同样实现了一种较好的尺度采样的尺度空间。在放大两倍的图像的积分图上计算，第一组的第一个滤波器大小为 15，之后的大小是 21,27,33,39。第二组开始，滤波器大小按 12 递增。第三和第四组按同样方法。如此，第一个滤波器变化率就是 1.4（21/15）。这种精确版本的，可用二次插值检测的最小尺度是 $s = (1.2 * (18/9)) / 2 = 1.2$ 。因为任何尺度滤波器的 F 范数都是常数，因此本身就是尺度标准化的，不需要再对滤波响应加权，更多信息参见文献 22。

3.4 兴趣点定位

为了在图像及不同尺度中定位兴趣点，我们用了 $3 \times 3 \times 3$ 邻域非最大化抑制。具体的，我们用了 Neubeck and Van Gool 方法的快速变种。Hessian 矩阵行列式的最大值在尺度和图像空间被插值，依据 Brown and Lowe 提出的方法。

尺度空间在我们的方法中尤其重要，因为每组第一个滤波器的尺度差异比较大。图 8 显示了用我们的快速 Hessian 检测子检测到的兴趣点。