Saliency and Similarity Detection for Image Scene Analysis

Dissertation Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

in

Computer Science and Technology

by

Ming-Ming Cheng

Dissertation Supervisor : Professor Shi-Min Hu

Cooperate Supervisor : Quan-Zhan Zheng

March, 2012

图像内容的显著性与相似性研究

（申请清华大学工学博士学位论文）

培 养 单 位 ： 计算机科学与技术系

学　　科 ： 计算机科学与技术

研 究 生 ： 程 明 明

指 导 教 师 ： 胡 事 民 教 授

联 合 导 师 ： 郑 全 战 博 士

二〇一二年三月

摘 要

随着数码相机和智能手机的普及，以及微博、社交网络等传播媒介的快速发展，图像已经越来越广泛地融入并改变人类的生活方式。图像所承载的丰富知识和无尽乐趣为人们带来巨大方便的同时，也提出了极大的挑战。一方面，随着图像数量的爆炸式增长，如何通过计算机对图像进行有效地组织和快速地检索变得越来越重要；另一方面，虽然目前已经有很多图像编辑方法，但这些编辑方法大多直接操作底层图像元素，其编辑对象缺乏语义性。图像的智能组织和语义编辑需要对图像场景内容本身的理解。如何让计算机理解图像场景内容，以贴近人类感知、加工和存储信息的方式对图像内容进行分析和组织，是当前计算机视觉与计算机图形学领域的重要研究课题。

本文围绕图像场景内容的显著性与相似性分析及应用，研究了图像视觉显著性区域的检测与分割、群组图像视觉显著性区域的提取与检索、交互式的图像场景相似单元分析与编辑等重点问题。主要创新点包括：

1. 提出了一种基于全局对比度分析的图像视觉显著性区域检测算法。通过对图像区域间的全局对比度和空间相关性进行建模，该算法能够快速有效地检测并分割图像中的视觉显著性区域。在国际上现有最大 (含1000张图像) 的公开测试集上，该方法的检测结果优于已有方法，显著性区域分割结果的准确性从之前最好结果的正确率75%、召回率83% 提升到了正确率90%、召回率90% 。
2. 提出了一种利用公共显著性物体的形状与表观一致性，从一组相关、低质、不可靠的网络图像中鲁棒地提取群组显著性区域的方法，并建立了一个包含15000张图像的标注数据集来验证算法的可行性。实验结果表明，相对基于单张图像的方法，该方法的结果具有更高的正确率。基于区域提取结果的草图图像检索系统(SBIR)不但在正确率方面优于之前最好的 SBIR 系统，而且提供了额外的目标物体区域信息以支持更广泛的应用需求。
3. 提出了一种基于简单交互的相似物体快速检测与分析方法。该方法基于一种新颖的轮廓带图匹配算法快速准确地检测相似物体单元，并对检测结果进行进一步分析以建立包含物体级别对象及其相互关系的图像场景语义信息。这些场景信息使得一系列场景物体级别图像编辑应用得以实现。

关键词：显著性检测；感兴趣物体分割；图像检索；相似物体检测；图像编辑

Abstract

Ubiquity of digital cameras and smart-phones, especially the rapid development of social network and twitter, has resulted in an explosion of digital images in form of personal and internet photo-collections — leading to a revolution of our daily life. Although such images form a well-established communication medium for sharing experiences or blogging events, it also poses challenging research questions. On one hand, efficient image retrieval and effective organization remains unsolved and even more urgent. On the other hand, although for decades researchers have developed methods for editing such images, most of these methods work on the image at pixel or region level, ignoring the hard to infer underlying semantics of the scene. However, such intelligent organization and semantic editing requires understanding of the underlying image scene. Allowing the computers to understand image scene and organize these information in the attractive notation which accords with our human mental data representation, is an important research topic in computer vision and computer graphics. The major contributions of the paper are:

1. We introduce a regional contrast based saliency extraction algorithm, which simultaneously evaluates global contrast differences and spatial weighted coherence scores. The proposed algorithm is simple, efficient, and produces high quality saliency maps. Our algorithm consistently outperformed existing saliency detection methods, when evaluated using the largest (including 1000 images) publicly available dataset. The saliency segmentation results have much better accuracy (precision = 90%, recall = 90%) than previous best results (precision = 75%, recall = 83%) on this dataset.
2. We propose a group saliency approach for automatic localization and segmentation of salient objects from internet image collections with heterogeneous quality, using shape and appearance consistencies. A benchmark consisting of 15,000 images with ground-truth annotations is also introduced to evaluate the results. The evaluation demonstrates that group saliency consistently outperforms state-of-the-art saliency algorithm working on individual images. Such automatic segmentation results directly enable efficient shape-based image retrieval (SBIR). We experimentally show that the resulting SBIR system outperforms state-of-the-art SBIR

systems not only by yielding better retrieval rate, but also by producing additional object region information which enables a wide range of applications.

3. We present a novel framework where user scribbles are used to guide detection and analysis of similar objects. Our detection process, which is based on a novel boundary band method, robustly extracts the repetitions. Subsequent analysis of these repetitions allows us to get scene object level instances along with their mutual relations. We demonstrated how to use such scene information for a variety of object level tasks that are otherwise difficult to perform.

Key words: Saliency detection; object of interest segmentation; image retrieval; repetition detection; image editing

目 录

第 1 章　绪论

1.1　论文背景与意义

一图胜千言，图像是记录信息、传递思想和表达情感的重要媒介。随着手持式图像采集设备和智能手机等硬件设备的普及，人们获取图像的手段日益方便与灵活。近年来，由于社交网络、微博、网络相册等共享平台快速兴起，以及图像本身所具有的内容直观、获取容易、传播方便、表现力丰富等优势，数字图像迅速成为日常生活中最受欢迎的一种知识传播和信息共享的媒介。对数字图像的处理已经成为信息科学、工程学、生物学、医学、心理学甚至社会科学等领域的重要研究对象。图像处理已经为人类带来了巨大的经济和社会效益，成为了日常生活、科学研究和社会生产中不可缺少的有力工具。随着计算机处理能力的持续提高以及相关领域分析手段的不断发展，图像处理技术无论是从理论上还是实践上都有着巨大的发展潜力。

图像在计算机中以像素为单位进行数字化存储。受到分析手段和计算机处理能力的制约，传统图像处理方法通常在像素、区块、信号等底层数据信息级别对

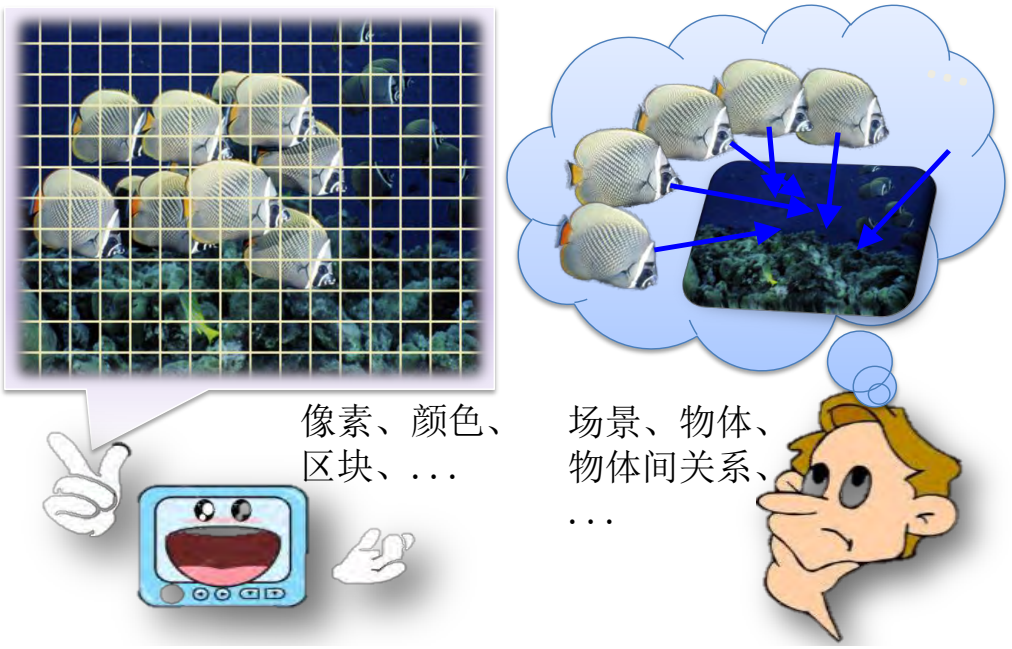


图 1.1　计算机与人类大脑关于图像的基本表示方式的差异：计算机通过像素、颜色等底层信息对图像进行数字化的表达和存储，而人类通过场景、物体、及物体间相互关系对图像进行认知和记忆。

图像进行处理。其主要研究内容包括：亮度变换、空间滤波、频域滤波、图像复原、相机配准、色彩变换、数据压缩、形态学操作等内容[1]。然而，人类对图像的理解是基于对相应的场景、物体以及这些物体之间相互关系的认知和记忆来进行的[2] (图 1.1给出了一个形象的示例)。因此，传统的像素或区块级别的图像编辑操作通常并不直接与人类所理解和认知的表达形式相对应，导致很多图像编辑工具需要复杂冗长的交互过程，仅能被有经验的专业人员所熟悉和掌握。近年来，计算机处理能力的持续增强为图像场景内容的深入分析和理解提供了坚实的硬件基础，同时也为更加智能的图像处理工具提供了强大的计算平台，让计算机像人一样理解图像中的高层语义对象(物体、空间关系、相互联系等)，从而提供自然、方便、直观、符合人类认知规律的图像处理工具，成为数字图像处理领域的重要发展趋势，值得深入研究。

下面简要介绍基于场景内容理解的图像处理领域中，几个重要研究方向上目前的主要研究成果，并讨论其中存在的主要问题，进而分析本文工作的定位和意义。

1.1.1 内容敏感图像处理

人类善于快速精确地判断图像中的显著性区域，解析重要的场景元素，并适应性地将注意力集中在那些对认知过程重要的区域。考虑到这种内容敏感的图像处理机制可以使得在后续的图像分析与合成过程中有倾向性的分配计算资源，自动精确地计算显著性区域变得非常迫切。对于图像内容显著性这个基本问题的研究也受到了认知心理学[3,4]、神经生物学[5,6]和计算机视觉[7,8]等多个领域研究者的广泛关注。

参考这种人类视觉机理，对不同图像内容进行自适应的差异化处理，已经在计算机视觉和图形学中的多个领域得到了非常广泛的应用。 Rutishauser等人利用显著性区域提取技术获得关于潜在目标物体的位置、大小、形状等信息，并成功地利用这些信息实现了无标注图像集合的无监督学习[9]。图像内容的显著性分析也可以被JPEG2000标准用来对图像中相对重要的区域进行更高质量的编码，以实现自适应压缩[10]。在最近几年被广泛研究的内容敏感图像缩放应用中，许多研究者利用显著性图来指导图像变形过程，并保持图像的视觉显著性区域与原始图像尽可能相似[11–14]。 Chen等人[15]利用显著性区域提取技术从海量互联网图像中选择与用户输入轮廓相近的图像及相关区域，并利用这些提取到的图像区域自动地合成具有真实感的图像，实现了用户草图到真实图像的转化。这种显著性区域提取的策略随后被用于多种图像编辑[16]、知识学习[17]和内容合成[18]等应用。

本文研究的一个重点就是如何准确高效地提取图像视觉显著性区域，为内容敏感的图像处理应用提供支持。虽然该领域的研究成果已经非常丰富，但是现有方法还不能很好地对显著性物体区域进行检测与分割。通过深入分析，作者设计了一种全新的基于区域对比度的视觉显著性检测算法。新算法在国际上现有最大的公开测试集上的检测结果明显优于近年来被广泛应用的各种代表性方法。该算法成果及相应代码公开至今不到一年时间，已被国内外400多个不同应用背景的研究或工程项目所使用。

1.1.2 基于内容的图像检索

随着智能手机等廉价的图像获取设备的普及和微博社交网络等方便共享平台的流行，每个人都成为了图像的生产者与发布者。例如，Facebook上已经分享了超过100亿张照片，腾讯用户平均每天上传的照片数也超过了1000万张。虽然高效的获取、存储和传输等技术使得人们具备了获取这些海量图像的潜力，但是要从这些共享的海量图像中通过浏览的方式获取用户期待的内容却无异于大海捞针。为了实现对海量图像数据的有效管理和组织，辅助用户快速、准确的找到期望的内容，基于查询的图像检索技术非常必要。用户查询方式通常以关键字等方式表达，而不是图像本身。传统方法利用图像的文件名、相应网页周围的伴随文本等网络元数据 (Metadata) 对图像进行检索。这种方法存在两大弊端：首先，从网页中自动提取的元数据本身可能并不是对图像内容的**准确**描述；另外，由于不同人的认识角度差异导致图像内容的文本标注不能**完全**反映图像内容本身。因此，基于内容的图像检索技术 (Content Based Image Retrieval, CBIR) 成为整个信息技术领域中最为重要的研究热点之一。

CBIR技术所面临的主要困难在于如何在用户查询和图像语义之间建立跨越感知鸿沟 (Sensory Gap) 和语义鸿沟 (Sematic Gap) 的桥梁[19,20]。其中，

(1) 感知鸿沟是指真实世界的物体和从该物体场景对应的图像中提取的描述信息之间的鸿沟；

(2) 语义鸿沟是指人们从视觉数据中所能提取到的信息和某个用户在特定情况下对相同数据的描述缺乏一致性。

图 1.2列出了解决这两大问题的几类主要方法。关于CBIR的详细信息可以参考Smeulders等人[19]和Datta等人[20] 的经典综述文献。在感知特征的描述方面，形状特征是人类理解物体的重要特征，是最强大的视觉线索[21]。认知心理学研究也表明，即使只有形状信息，人类也可以进行有效的物体识别[22]。在没有背景干扰的纯形状比较领域，相关研究已经相对成熟。即使对非常有挑战性的MPEG-7数据集，现有最先进的方法也能达到91.6%的检索率[23]。然而，由于

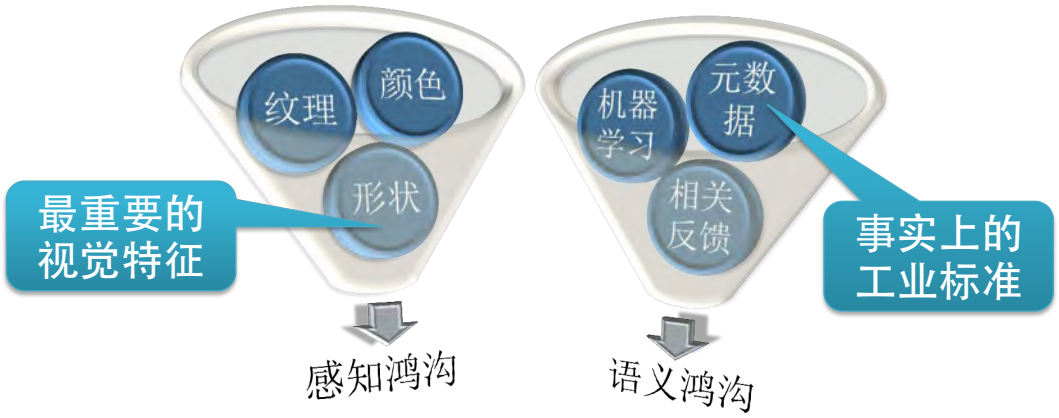


图 1.2　感知鸿沟和语义鸿沟是图像像检索所面临的最主要问题。其中形状信息是人类感知物体最重要的信息。很多时候，人们可以仅凭形状就能有效地识别目标[22]。元数据是建立图像和特定用户查询词输入之间语义桥梁的最重要工具，被广泛用于现有商业图像搜索引擎。

目标物体区域自动分割本身是一个非常困难的问题，可靠的形状信息难以获取，而经典形状比较方法非常容易受到背景轮廓线的干扰。因此，现有的CBIR系统中应用最广泛的还是颜色和纹理等容易提取的特征。颜色和纹理特征本身的描述能力比较有限，同样的颜色分布或者纹理特征经常与语义差异非常大的许多不同物体相对应。因而，利用网络元数据将用户查询关键字和图像内容进行对应依然是图像检索领域目前的工业标准，被广泛地用于各大商业引擎，如：Google图像 (http://images.google.com/)、 Baidu图像 (http://image.baidu.com/)、搜搜图像 (http://image.soso.com/)、 Flickr (http://www.flickr.com/)等。

本文同时利用物体形状特征和元数据信息，为用户提供基于草图的查询界面。这种查询方式相对颜色和纹理描述更加方便用户输入 (特别是在手机等触摸式设备普及的今天)。作者首先从同一个关键字对应的一组相关、低质、不可靠的网络图像集合所含的每个图像中提取视觉显著性物体区域。这些显著性区域信息排除了背景干扰，使得形状比较算法可以轻松地挑选出含有用户期待物体的图像。由于基于形状的检索并不破坏颜色和纹理等表观特征的多样性(diversity)，排序靠前的草图检索结果可以用于训练相应物体的表观模型(appearance model)，从而进一步改善显著性区域提取结果，并最终改善图像检索结果。由于目前尚且没有相关、低质、不可靠的网络图像标注集合，作者建立了一个包含15000张图像 (现有最大的单张图像显著性区域公开标注集合含有1000张图像) 的标注数据集来验证这种群组显著性检测与分割算法的有效性。实验结果表明，群组显著性分割结果明显优于单张图像处理的结果。基于这种自动分割结果的草图图像检索系统 (Sketch Based Image Retrieval, SBIR) 不但在正确率方面优于现有最好的SBIR系统[24]，而且提供了额外的目标物体区域以支持更广泛的应用需求。这种基于显著

性区域形状比较检索方法的早期版本(未利用表观统计特征改进分割)，已经在作者的早期项目 Sketch2Photo 中取的了较大的成功 ① ，并被后来的多个著名工作所借鉴[16–18]。可以预见，利用群组显著性对自动分割正确率的提升，将改善基于显著性物体区域的形状比较技术，也将在相关的图像检索、知识学习、照片合成、影视制作等领域发挥更大的作用。

1.1.3 图像场景建模与智能编辑

数十年来，对图像编辑方法的研究从未停止，但是现有的方法大多仅支持底层操作，通常被用于图像的修补和局部增强 [25–27]。近年来，各种各样的高层次图像编辑技术相继问世 [14,15,28–31]。这些技术允许用户通过简单的笔画来指定大范围、有意义的变化，将用户从复杂图像编辑过程中解放出来，并把繁琐的像素级操作留给底层算法完成。然而，这些方法中的大多数仅工作在图像的像素和区块级别，忽略了难以捕捉的潜在场景结构，并采用数据驱动的计算技术来产生看起来较为合理的编辑结果。

人类基于场景中物体之间的结构来组织和处理场景中的信息[2]。允许用户在场景物体级别对图像进行编辑操作是一种非常具有吸引力的概念，这种编辑方式与人类大脑中数据的表示形式相吻合。为了实现这种编辑，必须克服两个关键的难题：

(1) 图像是由散乱的像素点而非有意义的物体组成；

(2) 图像本身并不显式地含有物体在三维空间中排布的深度信息。

虽然自动的图像分割在图像处理和计算机视觉领域有着非常广泛的研究[32–34]，但是直到目前为止，对该问题的有效求解依然是一个难题[35]。大尺度的光照变化、阴影的存在、缺乏深度信息以及场景物体间相互的遮挡，都使得这个问题越发困难。适当的用户交互已经被证明是处理这个问题的一个强有力的途径[36–38]。

本文以含有相似结构物体的图像为载体，探索场景物体级别图像编辑应用。相似物体结构广泛地分布于各种人造的和自然的场景中，它们提供了关于类似数据的多份非局部的记录。如果能够正确地检测这种冗余性，就可以利用它来改进图像编辑，进行非局部的编辑操作[39,40]。这种冗余性提供了足够的信息来获取模板样式，以及刻画相似实例之间多样性的小形变。本文提出了一种用户交互算法

① Sketch2Photo 被法国政府参与组织的全球互联网行业论坛评为2009年全球十大最具创新性和可行性的发明之一；被英国《每日邮报》、德国《明镜周刊》等众多国际著名媒体专门撰文报道；获得了2011年中国计算机大会最佳研究成果奖；并成功转化为腾讯公司的商业产品“神笔小Q”(http://innov.labs.qq.com/official/sketch2photo/)。

框架来高效鲁棒地检测图像中相似模式的方法。虽然本方法并不恢复场景的真实语义信息，但是该方法通过提取图像中的相似结构以及它们之间相互的变形和层次关系，来达到对场景的部分理解。这促进了一系列直观高效的高层图像编辑应用，以及那些以其它方式难以完成的图像操作。基于这个算法框架，本文演示了许多新颖的图像处理应用。

1.2 研究目标与主要贡献

在巨大的应用需求推动下，对随手可得的海量图像进行有效地组织、快速地检索、智能地编辑显得尤为重要。本文的主要研究目标包括：快速、智能的图像处理技术；自然、方便的图像检索方法；直观、高效的图像编辑体验。由于这些智能的图像处理手段无一不依赖于对图像场景内容的语义理解，同时考虑到场景内容分析本身所具有的复杂性，本文以图像的显著性和相似性分析为出发点对图像场景内容进行分析 (三个主要工作之间的关系见图 1.3)，提出了解决图像智能处理关键问题的一系列算法：

(1) 基于全局对比度分析，提出了一种图像视觉显著性区域快速鲁棒的检测方法。通过对图像区域间的全局对比度和空间相关性进行建模，该算法能够快速有效地检测并分割图像中的视觉显著性区域。在含有显著性区域像素级别精度精确标注的国际上现有最大 (含1000张图像) 公开测试集上，该方法的检测结果优于已有方法，其显著性区域分割结果的正确率与召回率 (正确率=90%，召回率=90%) 相对之前的最好结果 (正确率=75%，召回率=83%) 也有大幅提升。为了验证该算法在更大数据集上的可扩展性，作者进一步建立了一个同样含有像素级别精度的基准数据集THUS10000，其数据量是之前最大数据集的10倍。在此数据集上将本文算法与国际上现有的15种方法进行了进一步比较，实验结果再次验证了该算法的鲁棒性。

(2) 利用公共显著性物体的形状与表观一致性，提出了一种从相关、低质、不可靠的网络图像集合中鲁棒地提取群组显著性 (Group Saliency) 区域的方法。由于目前尚且没有关于大规模相关、低质、不可靠网络图像集合显著性区域提取的公开测试集，为了验证算法的可行性，作者利用5类关键词从Flickr上自动下载了15000张图像并对其进行像素精度的标注，并将其命名为THUR15000测试集。该数据集将和THUS10000数据集一起被公开以方便后续的研究工作。在THUR15000数据集上的实验结果表明，本文提出的群组显著性区域提取方法明显优于现有最好的基于单张图像的提取方法。基于这种自动分割结果的草图图像检索系统 (SBIR) 不但在正确率方面优于之

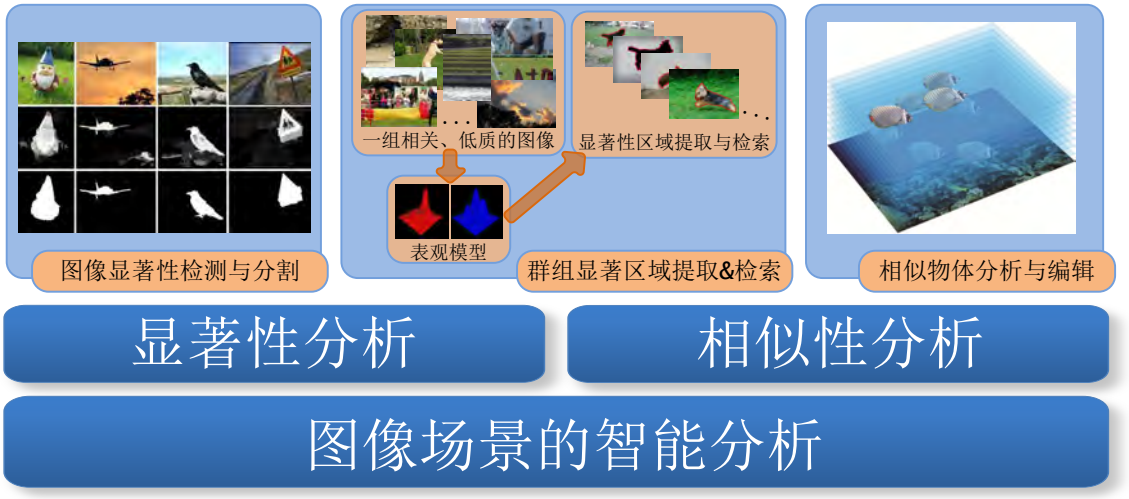


图 1.3　本文主要工作之间相互关系的示意图。

前最好的SBIR系统，而且提供了额外的目标物体区域信息以支持更广泛的应用需求。

(3) 提出了一种基于简单交互的相似物体快速检测与分析方法。该方法基于一种新颖的轮廓带图匹配算法快速准确地检测相似物体单元，并对检测结果进行稠密对应关系计算、遮挡部分补全、层次关系估计以及透明度抠图，以建立包含物体级别对象及其相互关系的图像场景语义信息。这些场景信息使得一系列场景物体级别图像编辑应用得以实现，包括以下几大类：图像重排、编辑扩散、变形传播以及物体替换。所有这些应用都只需非常简单的用户输入就能完成。

1.3　本文的组织结构

本文各部分的内容：第1章介绍本文的课题背景、研究意义，并简述研究内容和贡献；第2章详细介绍基于全局对比度的图像显著性区域检测，包括基于直方图的方法和基于区域的方法；第3章介绍本文提出的基于相似性分析的群组图像显著性区域提取与检索方法；第4章详细讨论图像中相似结构的检测与分析，以及利用这些分析结果的多种物体级别图像编辑应用；第5章总结本文并讨论进一步的研究方向。

第 2 章　基于全局对比度的图像视觉显著性区域检测与分割

本章主要研究基于全局对比度的图像视觉显著性区域检测与自动分割问题。第 2.1节中介绍背景知识、研究动机和解决方法概要； 第 2.2节给出与本章方法紧密联系的相关工作概述； 第 2.3节和第 2.4节分别介绍视觉显著性区域检测的两种高效算法：基于直方图的全局对比度分析方法和基于区域的全局对比度算法；第 2.5节给出了视觉显著性区域的自动提取算法。 第 2.6给出实验验证和结果分析。最后，第 2.7节对全章进行总结。

2.1　引言

2.1.1　背景知识

在漫长的进化过程中，人类获得了实时理解复杂场景的能力。人类是如何利用神经元硬件来实现这一复杂的认知过程呢? 心理物理学和生理学的研究者们通过对认知反应时间和信号在生物途径中的传输速度进行深入的研究和推理，得到了关于人类视觉注意选择机制的一整套认识，形成了视觉显著性理论。该理论认为：人类视觉系统只详细处理图像的某些局部，而对图像的其余部分几乎视而不见[41,42]。这种被视觉系统详细处理的局部区域通常被称为显著性区域 (saliency region)、重要性区域 (important region)、感兴趣区域 (region of interest)。静态图像中，能够引起人类视觉注意的信号刺激主要包括：新异刺激、较强刺激和人所期待的刺激。据此，以认知心理学家Triesman和Gelade[41] 与神经生物学家Koch和Ullman[42]为代表的研究者们从人类视觉认知的角度将视觉注意机制分为两个阶段：

(1) 快速的、下意识的、自底向上的、数据驱动的显著性提取；

(2) 以及慢速的、任务相关的、自顶向下的、目标驱动的显著性提取。

2.1.1.1　自底向上的图像显著性检测

自底向上的图像显著性区域检测算法主要关注由刺激信号本身所引起的视觉注意，这种方法由数据驱动，与目标任务无关，因而通常更加快速。由于各种显著性检测方法非常多，这里主要介绍最有影响力的和最近几年来在重要计算机视觉会议或期刊上发表的工作。

1998年，美国加州理工大学的Laurent Itti等人[7]提出了一种基于生物启发模型[42] (biologically plausible architecture) 和特征整合理论[44] (feature integration theory) 的视觉显著性计算方法。该方法首先利用线性滤波器对输入图像进行分解，以获取颜色、亮度、方向等特征的特征图。然后，在每个特征图内部，不同位置按照赢者通吃 (winner take all) 的准则进行显著性的竞争，以便使得局部最为突出的位置得以保留。最后，所有的特征图被按照一种纯粹自下而上的方式组合成一个主显著性图，这个主显著性图包含了视觉场景内的局部显著性。在灵长类动物中，这样的显著性图被认为是存储在后顶叶皮层[45] 以及丘脑后结节[46]之中。这个框架提供了一个大规模并行化地快速选择少量感兴趣图像位置的机制，以便降低更为耗时的物体识别过程所需的计算。这个基于生理学原理的可计算模型是视觉显著性检测领域最为经典的一个早期工作，被大量地应用于快速物体识别、内容敏感图像编辑等领域。

2003年，微软亚洲研究院的Yu-Fei Ma和Hong-Jiang Zhang提出了一种基于图像局部对比度分析的方法，来产生显著性图。进而通过模拟人类的感知方式，采用模糊增长的策略从显著性图中提取感兴趣的物体区域。该显著性检测算法框架支持三个层次上的视觉显著性：显著性视角、显著性区域和显著性点。

2007年，上海交通大学的Xiaodi Hou和Liqing Zhang[47] 提出了一种基于谱残留的方法来检测图像的视觉显著性。该方法通过分析输入图像的对数谱 (log-spectrum)，从图像的频域提取谱残留 (spectral residual)，并在对应的空间域构建显著性图。

同年，美国加州理工大学的Jonathan Harel等人提出了一种基于图结构的视觉显著性检测算法。该方法首先为每个图像位置提取特征向量，然后利用这些特征向量构建动标图 (activation maps)，最后对这些动标图进行归一化以突出和其它动标图相兼容的显著性组合。

2008年，加州大学圣地亚哥分校的Lingyun Zhang等人提出了一种基于贝叶斯理论和自然图像统计特征的方法来检测视觉显著性[48]。除了从待处理图像中获得自下而上的显著性特征以外，该方法还利用了大量自然图像中的统计特征来得到自上而下的显著性目标的统计信息。同年，Achanta等人利用不同尺度上的对比度测定滤波器从亮度和色彩这种底层信息中获取和原图具有同样分辨率的显著性图[49]。

2009年，Achanta等人[8]提出了一种基于频率调整的算法来获取多个尺度的显著性区域检测结果； Neil D. B. Bruce和John K. Tsotsos[50]提出了一种基于信息论的方法对显著性、视觉注意、和视觉搜索过程进行建模； Hae Jong Seo和Peyman Milanfar[51]提出了一种基于自相似的视觉显著性检测方法。

2010年，Esa Rahtu等人[52]利用统计模型从亮度、颜色、运动等特征的局部特征对比度中提取显著性区域； 2011年，Naila Murray等人利用滤波器卷积、中心邻域机制和空间合并的方法获取图像的显著性区域；同年，Lijuan Duan等人提出了一种基于空间加权差异性的显著性区域检测算法。

综上所述，关于如何从图像中自动地检测显著性区域这个计算机视觉基本问题的研究已经非常广泛。然而，以上所有模型都没有同时使用空间加权和全局对比度，而本章方法正是利用这种机制在现有最大的公开测试集上获得了正确率明显优于以上方法的结果，同时计算效率也比以上绝大部分方法更高。

2.1.1.2　自顶向下的图像显著性检测

自顶向下的显著性区域检测方法通常是根据具体任务对自底向上的检测结果进行尺度、方位、大小、形状、特征数目、阈值、类型、组合参数、描述形式等进行一定的调整而实现的[53]。相应的显著性图建立通常采用两种方法。

第一种方法是依据检测任务，人工设计显著性区域的检测模型。1986年，John Canny[54]提出了一种边缘检测的计算模型。为了找到最优的边缘检测算法， Canny首先定义了最优边缘检测的含义，其中包括最优检测准则、最优定位准则和检测点与边缘点一一对应约束，然后使用变分法来寻找满足这些功能的函数，进而得到了一个鲁棒的边缘检测算法。同年，J. Brian Burns[55]提出了一种直线检测算法。该方法将像素分为梯度方向相似的支持区域，然后利用支持区域的相关信息来确定直线的参数性质。 1998年，萨里大学的Farzin Mokhtarian 和 Riku Suomela[56] 提出了一种基于尺度空间曲率表达的图像角点检测方法。该方法首先利用Canny算法从原图中提取边缘，然后通过边缘的局部曲率绝对值最大值来定义角点。在检测过程中，先在较高尺度检测角点，然后细化到较低尺度以改进位置信息。这些方法的共性在于角点、边缘、直线等检测目标都是根据具体任务，人工建立的显著性区域检测模型。这些计算模型的建立需要设计者对目标的特殊结构和性质有着明确清晰的理解，并将这种理解反映到模型本身之中。

第二种方法是根据任务示例或样本，自动建立显著性区域模型。由于目标任务的可能形态成千上万，对各种目标任务的全面综述已经超出了本文的范畴，这里仅给出几种典型的例子。 2004年，微软研究院的Paul Viola和三菱电机研究实验室的Michael J. Jones[57] 提出了一种从人脸标注库中学习快速检测模型的算法。该方法主要包含三个部分：利用积分图快速地计算检测器所需的特征；利用AdaBoost从大量候选特征集中寻找一小部分数目的关键特征；通过瀑布模型快速地排除背景区域。 2005年，法国INRIA的Navneet Dalal 和 Bill Triggs[58] 提

出了一种基于梯度方向直方图 (Histograms of Oriented Gradient, HOG)特征的人体检测算法。该算法深入地研究了HOG构建过程中，细尺度梯度、细粒度方向划分、相关大尺度空间划分和重叠特征区域内高质量局部对比度归一化等因素对检测结果的影响，并最终构建了高质量的人体检测系统。 2007年，微软亚洲研究院的Tie Liu等人[59]提出了一种基于条件随机场的学习模型来优化三种不同显著性图的组合系数。这三种不同的显著性图分别是：反映局部显著性的多尺度对比度 (multi-scale contrast) 、反映区域显著性的中心邻域直方图 (center-surround histogram) 和反映全局显著性的颜色空间分布 (color spatial distribution)。示例样本中含有关于目标物体的大量指导信息，能够有效地指导算法过程对潜在显著性物体进行正确的判断。但是，不同的目标物体往往需要不同的物体特征、大量的训练样本和不同的学习方法，因而不能得到更具普遍意义的显著性物体区域。

2.1.2　研究动机

人类的这种视觉选择机制对于计算机识别和处理图像具有四个方面的重要促进意义：

(1) 高效性：视觉显著性区域的准确判断有助于极大地提升物体检测和跟踪等重要计算机视觉任务的运行效率。传统的物体检测算法大多采用滑动窗口 (sliding window) 机制来分析图像中每一个位置、方向、尺度下的区域是否是待检测物体。由于不同位置、方向、尺度的组合经常有上千万之多，其计算效率受到了很大的制约。如果能够有效地分析视觉显著性区域，就可以让计算机像人类一样重点地对少数几个区域进行判断，从而大大提升检测和跟踪效率。

(2) 智能性：视觉显著性区域的准确判断可以使得许多图像编辑技术对图像内容进行自适应的处理，以达到智能编辑的目的。例如，为了适应不同长宽比显示设备的显示需求，内容敏感的图像缩放技术[43]利用显著性图将缩放过程中不可避免的扭曲分布到相对不重要(或者说不显著)的区域，以保护重要区域尽量不被扭曲。

(3) 集约性：利用视觉显著性区域的判断结果，可以在图像压缩过程中对显著性区域予以优先编码。这样既能节省存储空间和传输带宽，又能有效地保持图像中重要区域的视觉质量[10]。

(4) 准确性：图像显著性区域的准确判断可以有效地避免无关区域对待检测识别的物体区域特征提取的干扰，从而提高物体检测、识别等重要计算机视觉任务的正确率[9]。

参考人类视觉注意机理，在没有任何关于相关场景内容的先验假设和知识的情况

下对图像的视觉显著性进行自动地检测，能够使得对图像进行空间自适应的处理成为可能。因而，鲁棒的图像显著性估计是许多计算机视觉和计算机图形学算法的重要基础，其中包括：图像分割、物体识别、自适应压缩、内容敏感的图像缩放等。

2.1.3　解决方法概要

本章基于图像的对比度分析来进行自底向上、数据驱动的图像显著性检测(见图 2.1)。人们普遍认为，为了优先响应高对比度的刺激，人类的大脑皮质细胞在它们的接受域可能是**硬编码**(*hard wired*)的[94]。在一个稍有不同的领域，这种机制也被用于生成伪装图像(camouflage images) [95]。这种伪装图像生成的基本原理是利用高对比度的伪造信号来有效地隐藏嵌入的图像元素，以使得显著性的估计变得具有挑战性。然而在自然场景中，这种杂乱性并不常见。基于以下观察，本章提出了一种提取高分辨率显著性图的全局对比度分析方法：

(1) 基于全局对比度的方法倾向于将大范围的目标和周围环境分离开。这种方法要优于那些通常只在轮廓附近产生较高显著性的局部对比度方法。全局的考虑可以为图像中相似的区域赋予相近的显著性的值，并且可以均匀地突出整个目标。

(2) 一个区域的显著性，主要是由它和周围区域的对比度决定，相距很远的区域起的作用较小。

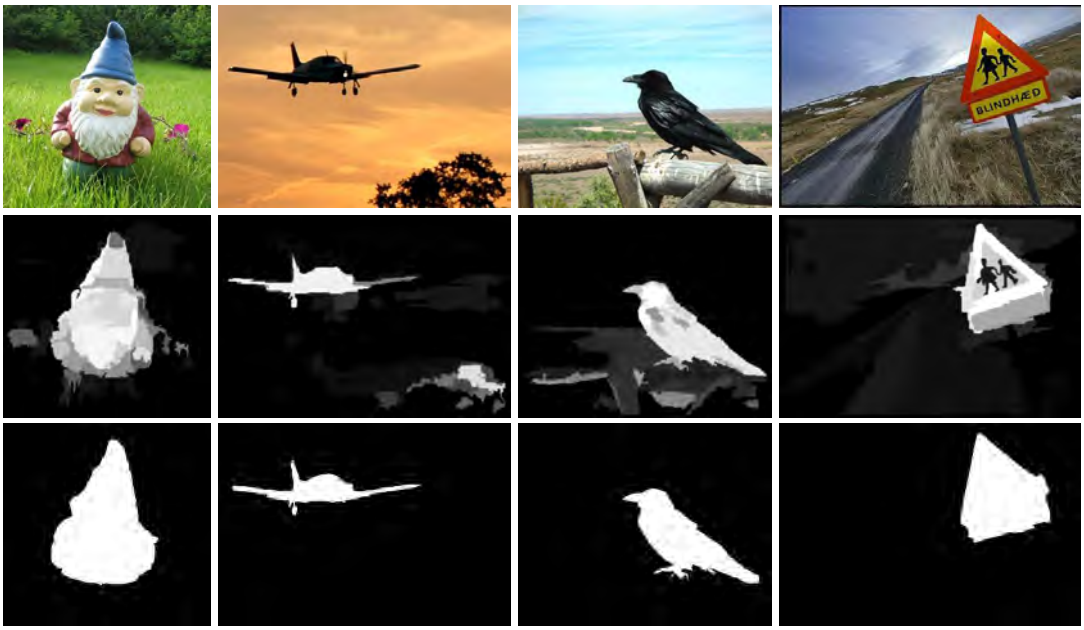


图 2.1　给定输入图像(上)，本章通过全局对比度分析得到高分辨率的视觉显著性图(中)。这种视觉显著性图可以进一步被用来获取感兴趣物体区域(下)。

(3) 人们在拍摄照片的时候，主要的物体通常更靠近图像的内部区域，远离图像边缘(见Jiang等人的工作[96]以及相关的引文)。

(4) 为了能够适应大规模图像集处理和高效的图像检索与分类的应用需求，显著图检测算法应该具有简单快速的特点。

本章提出了一种**基于直方图对比度的方法** (*Histogram based Contrast*, HC方法) 来检测显著性。 HC方法依据与其它像素的色彩差异来计算像素的显著性值，并用以产生具有全分辨率的显著性图像。本章用基于直方图的方法来进行高效处理，并用色彩空间的平滑操作来控制量化的缺陷。

作为HC方法的改进，作者结合空间关系提出了RC方法。首先将图像分割成区域，再为每个区域分配显著性值，从而形成**基于区域对比度的** (*Region based Contrast*, RC方法)显著性图。区域的显著性值由全局对比度值计算得到，其中，全局对比度值由当前区域相对于其它区域的对比度以及空间距离来计算。这种方法更好地反映了图像区域关系与显著性确定之间的联系。

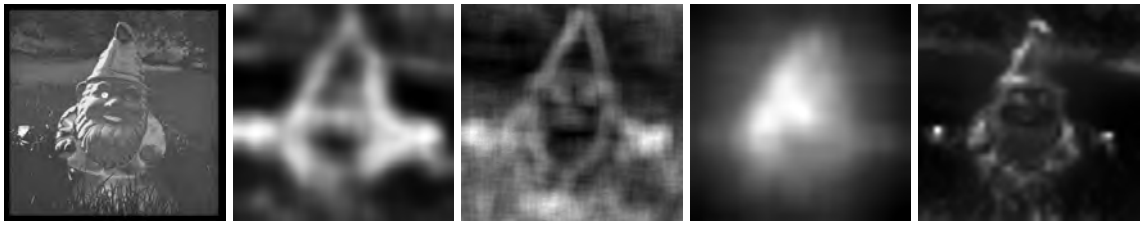
对于很多计算机视觉和计算机图形学应用来说，能够从静态图像中自动地提取感兴趣的区域有着重大意义。许多研究者致力于减少这一过程中的用户交互量，其中， GrabCut技术成功地将用户交互减少到仅需在目标物体外围选中一个矩形区域，该技术是一种迭代的优化能量方程技术，在此能量方程中同时考虑了纹理和边缘信息。因此，本章提出了一种迭代的GrabCut方法并将其与显著性检测算法相结合，得到了一种性能明显优于现有最先进的感兴趣区域自动提取方法的新算法。

作者在国际上现有最大 (含1000张图像) 的公开基准数据集上广泛地测试了本方法的结果，并且与现有最先进的的一组显著性图像提取方法 [7,8,47,49,65,97–99] 的结果，以及人工标注的参考数据进行了比较 ①。为了更好地验证本章方法在更大的数据集上的可推广性，作者建立了一个包含10000张图像的具有像素精度人工标注的参考数据集 (见第 2.6.1.2节)，该数据集比之前最大的同类数据集[8]大了一个数量级。作者在这个数据集上进行了广泛地测试，并把本章算法的结果和15个现有最先进的方法 [7,8,47–50,52,96,98–104] 的结果进行了比较 ②。该算法和之前已有的方法相比，在正确率和召回率方面都有明显的提高。总地来说，和HC方法相比，RC方法的正确率和召回率更高，但是计算量也更高。作者还将提取出的显著性图像应用于固定阈值分割、显著性区域自动提取、内容敏感的图像缩放以及非真实感渲染等应用。

① 对Achanta数据集[8]中的1000幅测试图用以上方法得到的所有结果、算法代码以及原型软件可以从本方法的项目主页上获取: http://cg.cs.tsinghua.edu.cn/people/~cmm/saliency/。

② 所列15种方法在所有10000张图像上的测试结果同样可以在作者的项目主页上找到。

(a) 输入图像　(b) IT方法[7]　(c) AIM方法[50]　(d) IM方法[103]　(e) MSS方法[102]



(f) SEG方法[52]　(g) SeR方法[101]　(h) SUN方法[48]　(i) SWD方法[104]　(j) GB方法[100]

(k) SR方法[47]　(l) AC方法[49]　(m) CA方法[99]　(n) FT方法[8]　(o) LC方法[98]

(p) CB方法[96]　(q) HC方法　(r) RC方法　(s) RCC　(t) 用户标注

图 2.2　通过各种现有最先进的方法计算得到的视觉显著性图 (b-p)，以及本章提出的HC方法(q) 和RC方法(r). 基于RC方法检测结果的感兴趣区域自动分割结果(s)非常接近用户标注(t)。大部分其它方法的结果使边缘部分更显著，或者具有较低的分辨率。图 2.8-2.10以及作者的项目主页中包含更多的例子。

2.2　相关工作

本章主要关注自底向上的显著性检测相关的文献。这类方法或是基于生物学原理的，或是纯计算的，或两者兼顾。这些方法利用亮度、颜色、边缘等底层特征属性来确定图像某个区域和它周围的对比度。本章宽泛地把这些算法分类为局部的和全局的两大类。

2.2.1　基于局部对比度的方法

基于局部对比度的方法利用图像区域相对于(一个小的)局部邻域的稀缺度来检测显著性区域。在Koch和Ullman[42]提出的非常有影响力的生物启发模型基础上, Itti等人[7]定义了图像的显著性，此定义利用了多尺度图像特征的中心-周围的差异来得到。 Ma和Zhang[65] 提出了另一种局部对比度分析方法来产生显著性图像，并用模糊增长模型对其进行扩展。 Harel等人[97]通过将Itti等人的特征图归一化来突出显著部分，并且可以和其它显著图像结合。 Liu等人[105]通过将高斯图像金字塔的对比度线性结合，提出了多尺度对比度。最近，Goferman等人[99] 同时对局部底层线索、全局考虑、视觉组织规则以及表层特征进行建模来突出显著性物体。这些利用局部对比度的方法倾向于在边缘部分产生高显著性值，而不是均匀地突出整个物体(见图 2.2)。

2.2.2　基于全局对比度的方法

基于全局对比度的显著性区域计算方法用一个区域和整个图像的对比度来计算显著性值。 Zhai和Shah[98]定义了基于某个像素和其余像素点对比度的像素级显著性。然而，出于计算效率方面的考虑，该方法仅用亮度信息，忽略了其它颜色通道的显著性线索。 Achanta 等人[8]提出了频率调谐方法，此方法用某个像素和整个图像的平均色的色差来直接定义显著性值。但是，该方法只考虑了一阶平均颜色，不足以分析复杂多变的自然图像。 图 2.8-2.10和 图 2.13展示了这种方法定性和定量的缺点。此外，这些方法忽略了图像各部分间的空间关系，而这个因素对可靠的显著性检测来说可能是至关重要的(见第 2.6节)。

2.2.3　感兴趣物体的自动分割

显著性图被大量用于无监督的物体分割(unsupervised object segmentation)：Ma和Zhang[65]利用模糊区域增长的方法从他们的显著性图中得到矩形的显著性区域。 Ko和Nam[106]利用图像区域特征来训练一个用于检测显著性区域的支持向量机(support vector machine)。这些显著性区域被进一步聚类得到感兴趣物体。近期，Achanta等人[8]对均值漂移分割 (mean-shift segmentation)区域内的显著性值进行平均，并选择平均显著性值高于某个阈值的区域作为感兴趣物体区域。实验中，他们选择这个图像平均显著性值的两倍作为这个阈值。作者借鉴并发展了经典的GrabCut[63]方法，通过迭代地进行GrabCut，并在每次迭代之后更新前景、背景和区域位置的信息，使得新方法能够自适应地根据已有显著性区域的信息更新其内部的表观模型，改进了其对噪声输入的鲁棒性。这种鲁棒性使得作者可以利

用显著性图检测结果自动地初始化感兴趣区域自动分割方法。在现有最大的公开测试集以及本章中构建的数据集(见第 2.6.1)上的分割结果显示了本算法在正确率方面相对传统方法的明显优势(第 2.6节)。

本章提出的视觉显著性检测方法及初步结果公开后 ①， Jiang等人[96]提出了一种类似的方法，该方法同样利用区域间的对比度对图像显著性进行建模。在分割阶段，他们的方法借鉴了本章方法的若干技术特性，该方法采用了迭代的GraphCut和基于直方图的外观描述模型，并同样在每步迭代之后收缩和扩展Trimap。由于本章所采用的GrabCut本身是一种利用GraphCut和高斯混合(Gaussian Mixture Model, GMM)外观描述模型的方法， Jiang等人的方法和本章所述的方法具有很高的相似性。然而，本章所述的方法和该方法具有两大不同点：

(1) 与本章提出的方法不同，Jiang等人[96] 利用不同的分割参数得到不同尺度的区域，随后对基于这些不同尺度分割结果的显著性图进行组合。

(2) Jiang等人利用形状先验对感兴趣物体分割进行改进。

由于直接邻域的范围和分割有紧密联系，该方法对基于不同尺度分割结果的显著性图进行组合对可靠的显著性图估计很有必要。然而，本章提出的基于全局对比度的方法对图像分割并不敏感。不出所料，利用人类摄影习惯的空间先验可以进一步改进本章提出的显著性物体区域检测与分割。然而，在一些没有显示的空间先验信息的应用中，作者更推荐选择不用这种强的先验，因为这种强先验在一些应用中有可能产生有偏差的结果 (例如，在自动监视系统中)。

2.3　基于直方图的全局对比度分析

人类的生物视觉系统对于高对比度的视觉信号非常敏感。根据这一观察，作者基于输入图像的颜色统计特征提出了基于直方图对比度 (Histogram Contrast, HC)的图像像素显著性值检测方法。具体来说，一个像素的显著性值用它和图像中其它像素颜色的对比度来定义。例如，图像I中像素I_k的显著性值定义为

$$S(I_k) = \sum_{\forall I_i \in I} D(I_k, I_i), \tag{2-1}$$

其中 $D(I_k, I_i)$ 为像素I_k 和像素I_i 在$L^*a^*b^*$空间的颜色距离度量 (参考[98])。 式 (2-1)可以按照像素顺序展开为

$$S(I_k) = D(I_k, I_1) + D(I_k, I_2) + \cdots + D(I_k, I_N), \tag{2-2}$$

① 该工作的初步结果发表在CVPR 2011 [107]。

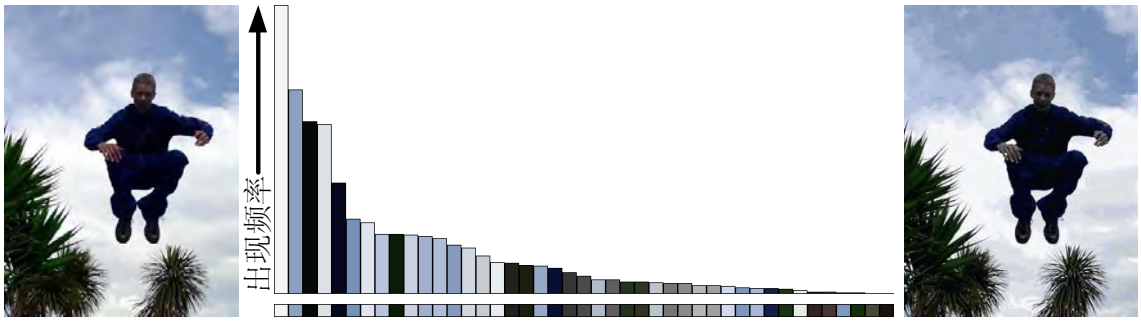


图 2.3　给定一个输入图像(左)，本方法计算其颜色直方图(中)。直方图中每一个bin所对应的颜色在下方的横条中显示。量化后的图像(右)仅仅使用43个直方图bin的色彩，且仍然保留了显著性检测所需的足够的视觉质量。

其中 N 为图像 I 的像素数。可以看到，由于忽略了空间关系，在这种定义下具有相同颜色值的像素具有相同的显著性值。因此可以对式 (2-2)进行重排，使得具有相同颜色值c_j的像素归到一起，得到每个颜色的显著性值

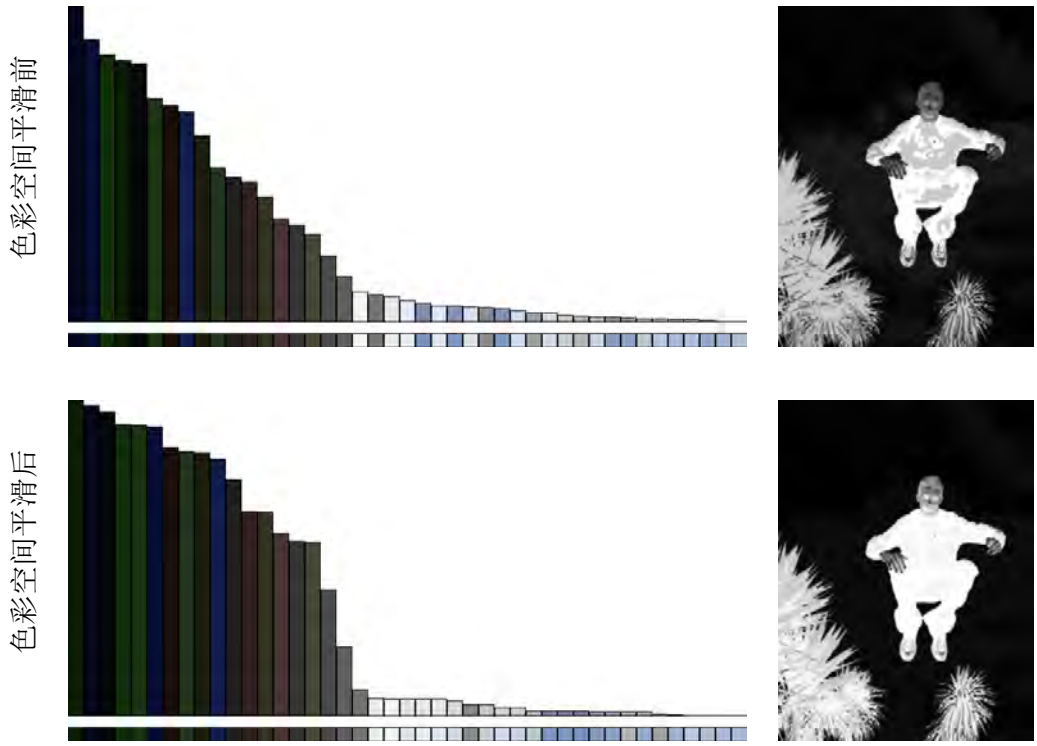
$$S(I_k) = S(c_l) = \sum_{j=1}^{n} f_j D(c_l, c_j), \tag{2-3}$$

其中像素I_k的颜色值为c_l，n为图像中所含的颜色总数， f_j为c_j在图像I中出现的概率 (由于整个图像的显著性值最终会被归一化到[0, 1]之间，由式 (2-2)到式 (2-3)过程中的归一化系数被忽略)。

2.3.1　基于直方图的算法加速

用式 (2-1) 来朴素地计算图像每个像素显著性值的方法的时间复杂度为$O(N^2)$。对于中等大小的图像来说，计算代价就已经非常高。上述公式的等价表示式 (2-3)花费$O(N) + O(n^2)$ 的时间。如果 $O(n^2) \leq O(N)$，时间复杂度就可以改进到$O(N)$。因此，加速的关键在于减少图像像素颜色的总数。然而，真彩色空间包含256^3种可能的颜色，比图像的像素总数多很多。

为了减少颜色数量n，Zhai和Shah[98] 仅用了图像色彩的亮度通道。这种方法得到的$n^2 = 256^2 = 65536$(通常情况下，$256^2 \ll N$)。然而，他们的方法的缺陷在于忽略了颜色信息的可区别性。在这个的工作中，作者用全色彩空间来代替仅用亮度的方法。为了减少需要考虑的颜色数目，本方法先将每个通道的颜色量化得到12个不同的值，这会将颜色数量减少到$12^3 = 1728$。考虑到自然图像中的颜色只占据整个色彩空间很小的一部分，可以将出现频率较小的颜色丢掉以减少色彩数目。通过选择高频出现的颜色，并确保这些颜色覆盖95%以上的像素，本方法通常可以做到将颜色数目减少到$n = 85$左右(实验结果详见第 2.6 节)。剩下的小于5%的像素所占的颜色被直方图中距离最近的颜色所代替。 图 2.3为典型的量化



(a) 像素颜色的显著性值　　(b) 对应的显著性图

图 2.4　颜色空间平滑前后像素颜色的显著性值(左) 以及对应的显著性图(右)。为了显示方便，所有显著性值被归一化到[0, 1]。

样例。请注意，基于时间性能的考虑，本章用简单的基于直方图量化的方法来代替为整张图像优化计算特定的调色板。

2.3.2　色彩空间平滑

虽然可以用颜色量化后的颜色直方图来高效地计算颜色对比度，但量化本身可能会产生瑕疵，因为一些相似的颜色可能被量化为不同的值。为了减少这种随机性给显著性值计算带来的噪声，本方法用平滑操作来改善每个颜色的显著性值。每个颜色的显著性值被替换为相似颜色(用$L^*a^*b^*$距离测量)显著性值的加权平均，这个过程实质上是颜色空间的一种平滑过程。作者选择$m = n/4$个最近邻颜色来改善颜色c的显著性值

$$S'(c) = \frac{1}{(m-1)T} \sum_{i=1}^{m} (T - D(c, c_i)) S(c_i) \tag{2-4}$$

其中 $T = \sum_{i=1}^{m} D(c, c_i)$ 为颜色 c和它的m 个最近邻c_i之间的距离，归一化因数由

$$\sum_{i=1}^{m} (T - D(c, c_i)) = (m-1)T. \tag{2-5}$$

得到。作者用了一个线性变化平滑权值($T - D(c, c_i)$)来为在颜色特征空间中距离 c 较近的颜色分配较大权值。在实验中，作者发现这种线性变化的权值比急剧下降的高斯权值的效果好。 图 2.4为颜色空间平滑后的效果，按显著性值降序排列。注意到相似的直方图区间在平滑过后会非常接近，表明相似的颜色非常可能分配到相似的显著性值，因此减少了量化的瑕疵(见图 2.12)。

2.3.3　具体实现

将颜色空间量化为12^3种不同的颜色的过程中，本方法将每个颜色通道均匀地划分为12个级别。虽然颜色量化在 RGB 颜色空间进行，但为了颜色距离计算与人类感知更加符合，作者在 $L^*a^*b^*$颜色空间来测量颜色间的距离。本方法并不直接在$L^*a^*b^*$颜色空间进行量化，因为并不是所有在区间$L^* \in [0, 100]$，$a^*, b^* \in [-127, 127]$中的颜色都对应真实的颜色。实验结果表明，直接在 $L^*a^*b^*$空间量化得到效果并不好。最佳结果是在 RGB 颜色空间量化，在$L^*a^*b^*$颜色空间测量距离得到的。

2.4　基于区域的全局对比度分析

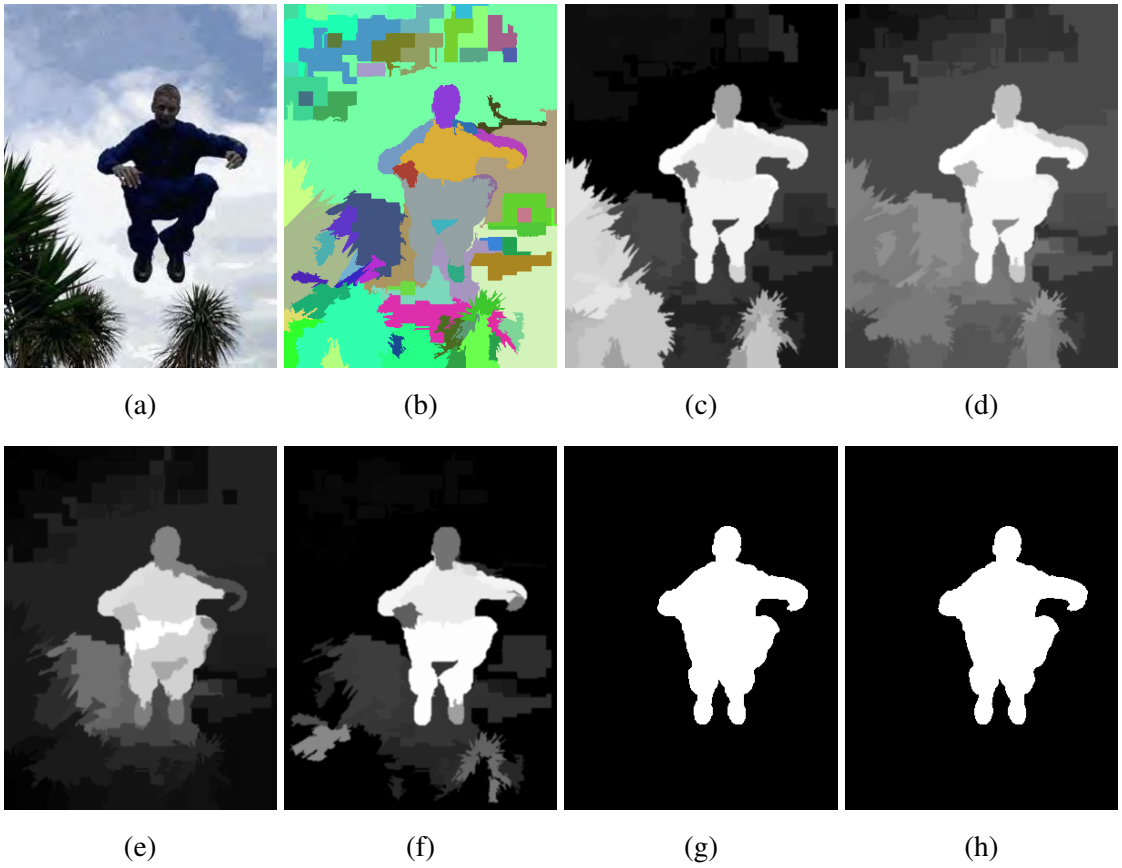
人们会更加注意到图像中和周围物体对比度较大的区域[108]。除了对比度之外，空间关系在人类注意力方面也起到重要的作用。相邻区域的高对比度比很远区域的高对比度更容易导致一个区域引起视觉注意。在计算像素级对比度时引进空间关系的计算代价会非常大，本章介绍一种对比度分析方法：区域对比度 (Region Contrast, RC)。以此来将空间关系和区域级对比度计算结合到一起。在RC方法中，作者首先将图像分割成若干区域，然后计算区域级颜色对比度，再用每个区域和其它区域对比度加权和来为此区域定义显著性值。权值由区域空间距离决定，较远的区域分配较小的权值。

2.4.1　基于稀疏直方图比较的区域对比度

首先，作者用基于图的图像分割方法将输入图像分割成若干区域[60]。然后为每个区域建立颜色直方图(见第 2.3节)。对每个区域r_k，通过测量它与图像其它区域的颜色对比度来计算它的显著性值

$$S(r_k) = \sum_{r_k \neq r_i} w(r_i) D_r(r_k, r_i), \tag{2-6}$$

其中$D_r(r_k, r_i)$为两个区域r_k和r_i的颜色距离度量， $w(r_i)$为区域r_i的权值。这里用区域 r_i里的像素数$w(r_i)$ 来强调与较大区域的颜色对比度。两个区域r_1和r_2的颜色距



(a)　(b)　(c)　(d)

(e)　(f)　(g)　(h)

图 2.5　基于区域的全局对比度分析示例：(a) 输入图像， (b) 由 Felzenszwalb 和 Huttenlocher 的分割方法[60] 得到的图像区域分割结果， (c) 不考虑距离加权和空间先验的区域对比度分析结果(式 (2-6))， (d) 进一步考虑距离加权的区域对比度分析结果， (e) 再进一步考虑空间先验的区域对比度分析结果(式 (2-8))， (f) 经过边界区域估计和色彩空间平滑后的区域对比度分析结果， (g) 感兴趣区域分割(第 2.5节)结果， (h) 人工标注的感兴趣物体区域。本方法得到的高质量的重要性分割结果可以和人工分割结果相媲美。

离为

$$D_r(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D(c_{1,i}, c_{2,j}) \tag{2-7}$$

其中$f(c_{k,i})$为第i个颜色$c_{k,i}$在第k个区域r_k的所有n_k种颜色中出现的概率，$k = \{1, 2\}$。注意，这里使用区域的概率密度函数(即归一化的颜色直方图)中颜色出现概率作为权值，以强调主要的颜色之间的区别。

因为每个区域只包含图像的直方图中很少数目的颜色，所以为每个区域存储和计算常规矩阵形式的直方图是低效的。本方法用稀疏直方图以使得存储和计算过程更加高效。

2.4.2　空间加权的区域对比度计算

更进一步，通过在式 (2-6)中引进空间权值，本方法将空间信息加入进来，来增加区域的空间位置关系对结果的影响。增大近邻区域的影响，减小较远区域的影响。特别地，对任意区域 r_k，基于空间加权区域对比度的显著性定义为

$$S(r_k) = w_s(r_k) \sum_{r_k \neq r_i} \exp\left(\frac{D_s(r_k, r_i)}{-\sigma_s^2}\right) w(r_i) D_r(r_k, r_i) \tag{2-8}$$

其中 $D_s(r_k, r_i)$ 为区域r_k 和 r_i的空间距离，$w_s(r_k)$ 是一个控制图像内部区域比边界区域更有可能吸引人类注意的控制项 (类似CB方法[96]), σ_s控制空间权值强度。本方法采用

$$w_s(r_k) = \exp(-9d_k^2), \tag{2-9}$$

其中 d_k是像素坐标归一化到$[0, 1]$后区域r_k 中的像素到图像中心的平均距离。 σ_s 越大，空间权值的影响越小，导致较远区域的对比度会对当前区域显著性值做出较大的贡献。两个区域的空间距离定义为两个区域重心的欧氏距离。在本章的实验中：$\sigma_s^2 = 0.4$，像素坐标归一化到$[0, 1]$区间。

2.4.3　区域对比度结果改进

本章进一步通过两个步骤来改进之前计算得到的RC方法显著性图。首先，利用空间先验来进一步估计非显著性(背景)区域，其次参考第 2.3.2节，利用颜色空间平滑来进一步改进结果。

本文把那些与图像边缘有大面积重合的图像区域称为边缘区域。作者观察到这部分区域通常对应于非显著性的背景区域，并利用这一观察作为第二个空间先验知识(第一个为式 (2-8)中的 $w_s(r_k)$)来帮助检测非显著性区域。在实验中，作者将区域中所含位于15个像素宽图像边缘的像素个数用该区域中的总像素个数归一化。对应该归一化数值大于某一阈值的区域即被当作边缘区域。实验中，这种硬约束不但改进了显著性图的精确度，而且通过改进初始值加速了显著性分割(第 2.5节)的收敛速度。进行边缘区域检测的动机是高准确率，而非高召回率。因此，作者采用一个严格的阈值来检测边缘区域以减少误报率。最终，该严格阈值选择为对应参考数据集中2%误报率的经验阈值。

为了均匀地突出图像中的整个显著性物体，作者计算了每种颜色对应的平均显著性值，并利用第 2.3.2节中描述的颜色空间平滑操作来改进RC方法的结果。平滑之后，有些边缘区域像素会获得非零的显著性值。作者将边缘区域像素的显著性值重置为零，并将每个区域中对应像素的平均显著性值作为该区域的显著性

值。由于RC方法中获得的初始显著性值比HC方法的结果更倾向于均匀地突出整个显著性区域，在这次平滑过程中，作者采用更小数目($m = n/10$)的邻近颜色对应的信息。 图 2.5(f)演示了这样的一个例子。在该例子中，与弹跳的人对应的区域相对于 图 2.5(e)被更加地均匀突出。

2.5 感兴趣区域的自动分割

在一个重要的早期工作中，GrabCut[63] 技术对 GraphCut框架做了一些关键的改进以使其能够更好地处理含有噪声的初始化。这种改动使得用户可以仅仅粗略地标出感兴趣的物体区域(例如：采用矩形空间进行选择)， GrabCut工具就能精确地提取用户期待的物体区域。利用计算出的显著性图，作者甚至消除了对这种简单的矩形区域选择交互的依赖。本节将介绍一种感兴趣区域自动分割方法，作者称其为*SaliencyCut*。 SaliencyCut利用自动计算的显著性图来辅助图像中感兴趣区域的自动分割。这就使得自动地分析大规模的网络图像数据立即成为可能。更进一步地说，作者对GrabCut[63]技术做了两大改进：“迭代更新” 和 “自适应拟合”。这些改进使得新算法可以处理明显具有更多噪音干扰的初始化。得益于新方法的鲁棒性，作者可以利用检测得到的显著性图自动地初始化感兴趣物体区域的分割过程。

2.5.1 基于显著性图的感兴趣区域初始化

与经典的GrabCut中利用用户选择的矩形区域进行初始化不同，本方法利用固定阈值二值化显著性图得到的分割结果自动地初始化感兴趣区域。类似于GrabCut，本方法同样采用不完全的Trimap来进行初始化。显著性值低于阈值的区域被初始化为背景区域，其余的区域对应Trimap中的未知部分。注意，此处本方法并未设定任何的固定前景区域。初始Trimap中的未知部分被用来训练前景颜色模型，并帮算法的后续步骤确定前景像素。

在GrabCut优化过程中，由于初始的背景区域会被保留，而其余区域将随着优化过程而改变，在Trimap的初始赋值时本方法倾向于将置信度非常高的非显著性区域初始化为背景区域。因此，本章采用对应于较高的潜在前景召回率的阈值来初始化GrabCut算法，并利用迭代过程改进其准确率。实验中，作者经验性地选择这个阈值为固定阈值分割实验中(第 2.6.2节) 对应95%召回率的阈值。在利用RC方法的显著性图进行初始化时，如果将所有显著性值归一化到区间$[0, 255]$范围内，这个阈值则是70。

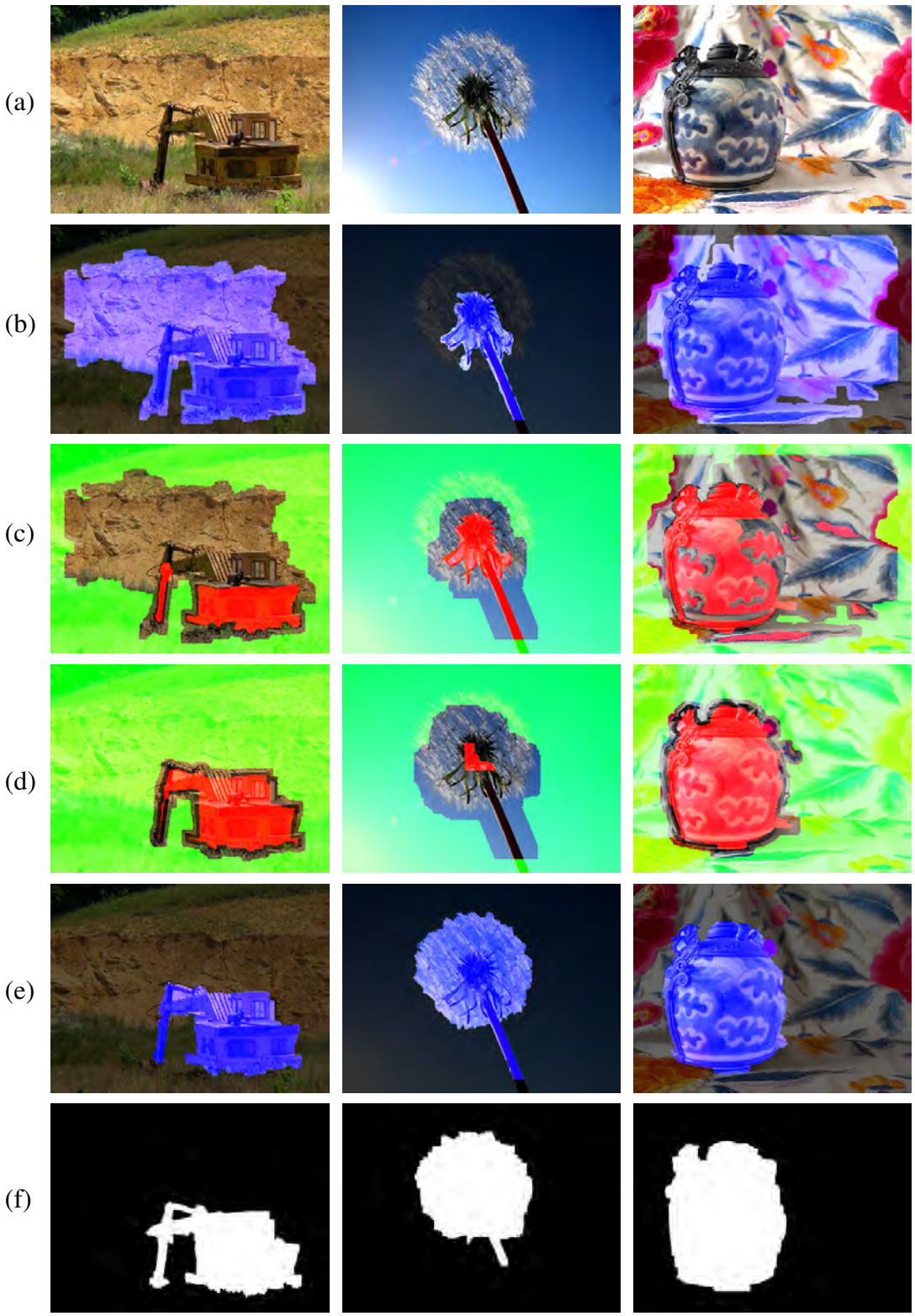
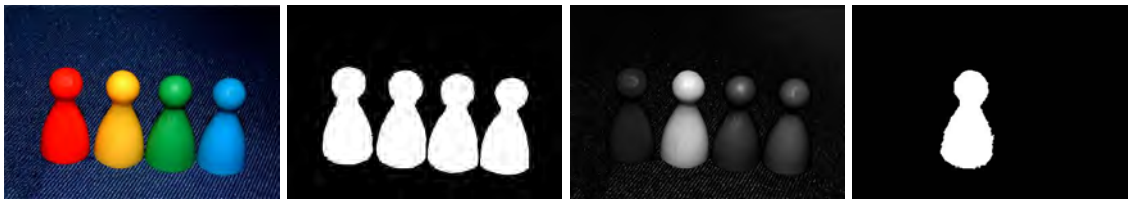


图 2.6　感兴趣区域的自动分割(SaliencyCut)示例：(a) 输入图像, (b) 基于固定阈值对显著性图的初始分割结果, (c) 第一次迭代后的Trimap, (d) 第二次迭代之后的Trimap, (e) 最终的分割结果, (f) 用户标注的基准数据。在分割结果中(b,e)，蓝色代表前景，灰色代表背景。在Trimap中(c,d)，红色代表前景，绿色代表背景，原图颜色区域为未知区域。

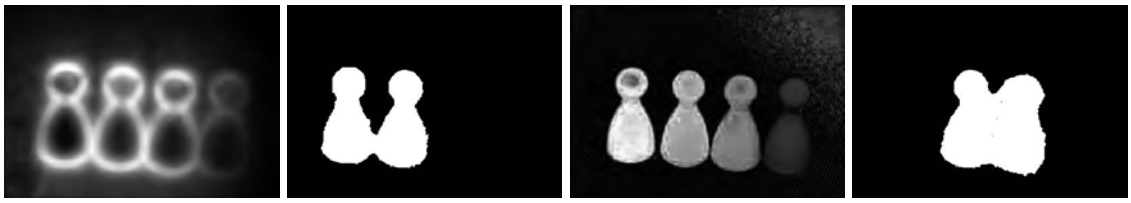
2.5.2　基于迭代的感兴趣区域分割

初始化之后，本方法迭代地运行GrabCut[63]来改进显著性分割结果 (在本章的实验中最多迭代 4 次)。在每一次迭代后，本方法用膨胀和腐蚀操作来得到新的Trimap以进行下一次迭代。如图 2.6(c, d)所示，膨胀后仍然落在外面的区域设置成背景，在腐蚀后区域内的设置成前景，其余的区域为Trimap中的未知。Grabcut本身是用高斯混合模型和Graph cut进行迭代。作者利用迭代的GrabCut来逐渐地修正显著性物体区域的分割。

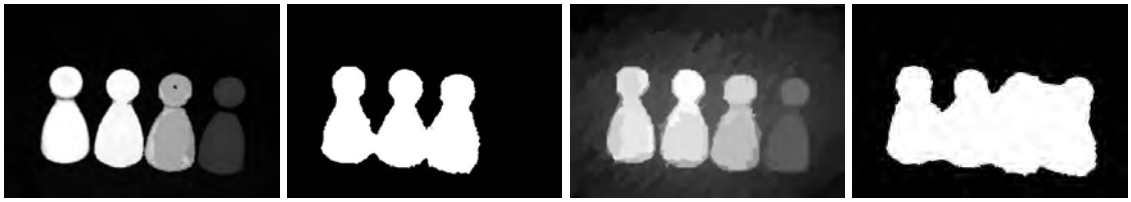
与单纯的一次使用GrabCut不同，新的SaliencyCut方法**迭代地更新**初始的显著性区域。这种迭代的设计对于处理粗糙的初始化非常重要。为本方法提供了免去人工标注，而利用显著性检测自动进行初始化的可能。在如图 2.6 (b)所示的“花”的例子中，初始的背景区域错误的包含了前景物体。虽然作者仍然可以通过GrabCut方法得到一个包含很多前景物体区域的分割结果，在初始背景区域部分的物体被强制性的标定(hard labeling)为背景，整个“花”的区域永远不可能被完整地利用GrabCut方法提取出来。在自动地处理大量图像的过程中，这种粗糙的初始化不可避免。一个好的感兴趣物体自动提取方法必须对这种粗糙的初始化具



(a) 输入　(b) 手工标定　(c) LC方法[98]对应的显著性图及分割



(d) CA方法[99]对应的显著性图及分割　(e) FT方法[8]对应的显著性图及分割



(f) HC方法对应的显著性图及分割　(g) RC方法对应的显著性图及分割

图 2.7　用不同显著性图初始化显著性分割的结果比较： (a) 输入图像，(b) 手工标定的分割结果， (c-g) 用各种不同方法计算得到的视觉显著性图(左)及以其初始化SaliencyCut 得到的分割结果(右)。

有鲁棒性。放宽GrabCut中的强制约束来解决这一问题也许是一种可能的选择。但是，本章的实验结果表明，这种约束放松并不合适。因为这会导致最终分割结果经常全部为前景或者背景。

本方法迭代地更新初始的分割结果并自适应地调整初始条件以适应新分割得到的显著性区域。这种**自适应的拟合**基于一个重要的观察结果：初始显著性物体区域附近的区域相对较远的区域更有可能是感兴趣物体中的一部分。因此，新的初始化方式使得GrabCut能够吸纳邻近的显著性区域，并根据颜色特征的不相似性剔除非显著性区域。在每一步GrabCut迭代之后，SaliencyCut 利用新得到的Trimap为约束，并根据之前的结果训练一个更好的表观模型(appearance model)。

图 2.6显示了三个例子。在第二行“花”的例子中，SaliencyCut 成功地对(由显著性图中直接得到的)初始的显著性区域进行扩展，使得算法最终收敛到精确的分割结果。在“起重机”和“茶壶”的例子中，多余的区域被成功地利用GrabCut迭代过程移除。这些中间过程显示了SaliencyCut 是如何成功地从这些非常有挑战性的例子中提取物体区域的。 第 2.6.3节描述了更加全面的定量评测。

2.6 实验验证

首先，作者在Achanta等人提供的公开测试集[8] 上测试了本章方法。据作者所知，此测试集是包含精确人工标注显著性区域的公开数据集中最大的一个。本章将基于全局对比度的方法与其它8个现有最先进的方法进行了比较。参照[8]，本章依据以下几个方面的因素来选择其它参考方法来进行对比：引用数多(IT方法[7] 和 SR方法[47])、方法较新(GB方法[100], SR 方法, AC方法[49], FT方法[8]和 CA方法[99])、多样性(IT 方法为生物驱动, MZ方法[65] 为纯计算, GB 方法为两者混合的， SR 方法在频域进行处理, AC 方法和FT 方法输出全分辨率显著性图)，和本章方法最接近(LC方法[98])。

为了进一步验证本章方法在更大数据集上的可推广性，作者创建了一个更大规模包含精确显著性区域标注的图像数据集THUS10000 (第 2.6.1.2节)。该数据集中包含的图像个数是具有精确区域标注的现有最大公开测试集[8]的10倍。作者通过几个重要的计算机视觉和计算机图形学应用对本章方法进行验证。这些应用包括：基于固定阈值分割的显著性区域提取、显著性区域自动分割、内容敏感的图像缩放以及非真实感绘制。在这个更大数据集上的验证中，作者同时扩大了参考方法的数目，一共与15种方法进行了比较。这些方法包括：SR方法[47]、 IT方法[7]、 IM方法[103]、 SUN方法[48]、 AC方法[49]、 S-

eR方法[101]、 AIM方法[50]、 GB方法[100]、 LC方法[98]、 CA方法[99]、 FT方法[8]、SWD方法[104]、 SEG方法[52]、 MSS方法[102]和 CB方法[96]。

本章的实验环境是一个具有Dual Core 2.6 GHz CPU、2GB内存的个人电脑。表 2.1比较了各种方法处理THUS10000数据集中的一个图片的平均用时。其中本章的HC方法和 RC方法是用C++语言实现的. 对于其它一些方法(IT, AIM, IM, MSS, SEG, SeR, SUN, GB, SR, AC, CA, FT 和 CB)，作者大多采用了这些方法的原作者给出的实现。由于找不到LC方法的原作者实现版本的代码，本文作者用C++语言自己进行了实现。对于典型的自然图像，本章的HC方法运行时间复杂度是$O(N)$。这样的效率足以满足实时应用的需求。由于用到了图像分割[60]，RC方法速度相对更慢一些。但RC方法方法的结果平均质量更高。 表 2.2 给出了本章的显著性区域分割方法和其它几个现有最先进的方法的平均用时(关于结果精度的详细比较请

表 2.1 用各种不同方法计算THUS10000数据集中图像的显著性图的平均用时。该数据集(参见作者主页)中大部分的图像分辨率为400×300。这里所示的所有方法的时间是在一个拥有Dual Core 2.6 GHz CPU，2GB内存的机器上测得的。对于所有Matlab程序，作者利用了Matlab的并行环境来快速的计算结果。

方法	时间(秒)	代码类型	代码共享地址
IT方法[7]	0.246	Matlab	http://www.saliencytoolbox.net/
AIM方法[50]	4.288	Matlab	http://www-sop.inria.fr/members/Neil.Bruce
IM方法[103]	0.991	Matlab	http://www.cat.uab.cat/
MSS方法[102]	0.106	C++	http://lcavwww.epfl.ch/~achanta/
SEG方法[52]	4.921	Matlab	http://www.cse.oulu.fi/CMV/Downloads
SeR方法[101]	1.019	Matlab	http://alumni.soe.ucsc.edu/~rokaf/
SUN方法[48]	1.116	Matlab	http://cseweb.ucsd.edu/~l6zhang/
SWD方法[104]	0.100	Matlab	
CB方法[96]	5.568	Matlab & C	https://sites.google.com/site/jianghz88/
GB方法[100]	1.614	Matlab	http://www.klab.caltech.edu/~harel/
SR方法[47]	0.064	Matlab	http://www.klab.caltech.edu/~xhou/
FT方法[8]	0.016	C++	http://lcavwww.epfl.ch/~achanta/
AC方法[49]	0.109	Matlab	http://lcavwww.epfl.ch/~achanta/
CA方法[99]	53.1	Matlab	http://webee.technion.ac.il/labs/cgm/
LC方法[98]	0.018	C++	http://cg.cs.tsinghua.edu.cn/people/~cmm/
HC方法	0.019	C++	
RC方法	0.254	C++	

参考第 2.6.3节)。

为了更好地验证本章方法进行显著性区域分割的正确率，作者采用了两种客观评价标准。在第一个实验中，作者直接用固定阈值对显著性图进行分割(参考Achanta等人的固定阈值分割实验[8])。在第二个实验中，作者对本章提出的SaliencyCut方法(第 2.5)进行了验证。

2.6.1　实验数据集

2.6.1.1　Achanta数据集

随着视觉显著性研究的日益深入，对显著性检测结果的客观评价变得尤为重要。 Liu等人在他们CVPR 2007会议上发表的论文[59]中公开了一个包含 20000 多张图片的测试数据集。在这个数据集中，每一个图像都包含一个没有歧义的显著性物体。整个数据集涵盖了不同种类、颜色、形状、大小的显著性物体。换句话说，这些物体除了在所在图像中最具显著性以外，没有任何别的先验知识或者约束。每个图片都对应3-9个用户手工标注的矩形显著性物体区域。虽然这个数据集是非常有价值的显著性检测验证的资源。但正如被Wang和Li[109]，以及Achanta等人[8] 指出的那样，仅仅只有包围矩形的标注太过粗糙，并不适合精细的评测。

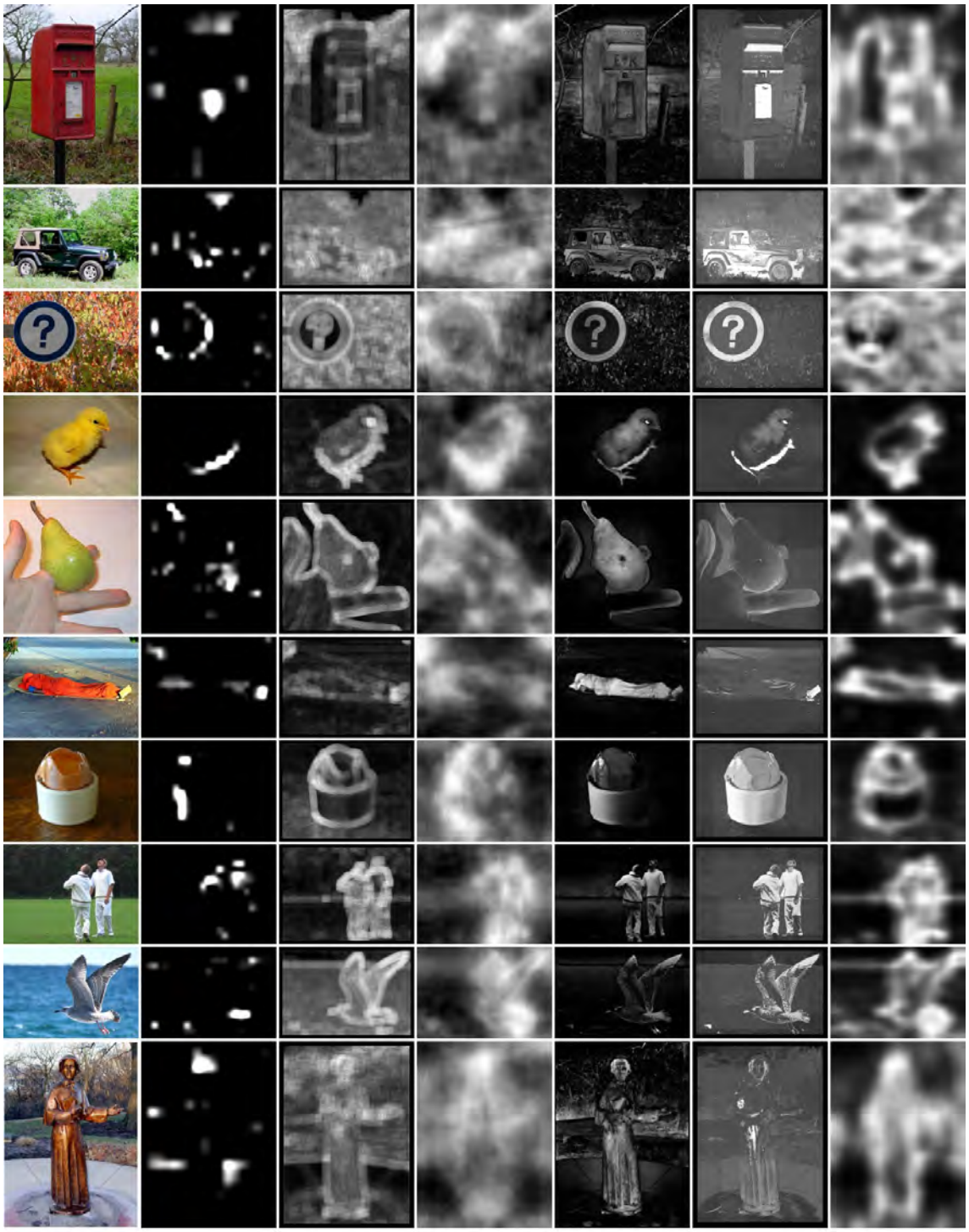
为了客观、精确地验证感兴趣物体分割方法的可用性，Achanta等人从Liu等人的数据集中选出了1000张图像，并为每个图像标出具有像素精度的显著性区域范围。该数据集随着他们CVPR 2009的论文被一起公开，是具有精确显著性区域标注的现有最大公开测试集。因而，该数据集被广泛地用于显著性区域检测和分割算法的验证中。

2.6.1.2　THUS10000数据集

为了验证本章算法在更大数据集上的鲁棒性，作者以Liu等人[59]的数据集为基础，从中随机选出了10000张在该数据集中包围矩形标注一致性较好的图像。这里一致性的度量和Liu等人论文中构建图像数据集B的方法一样。如图 2.11所示，作者精确地标注了显著性区域内部的像素。由于该数据集中包含10000张

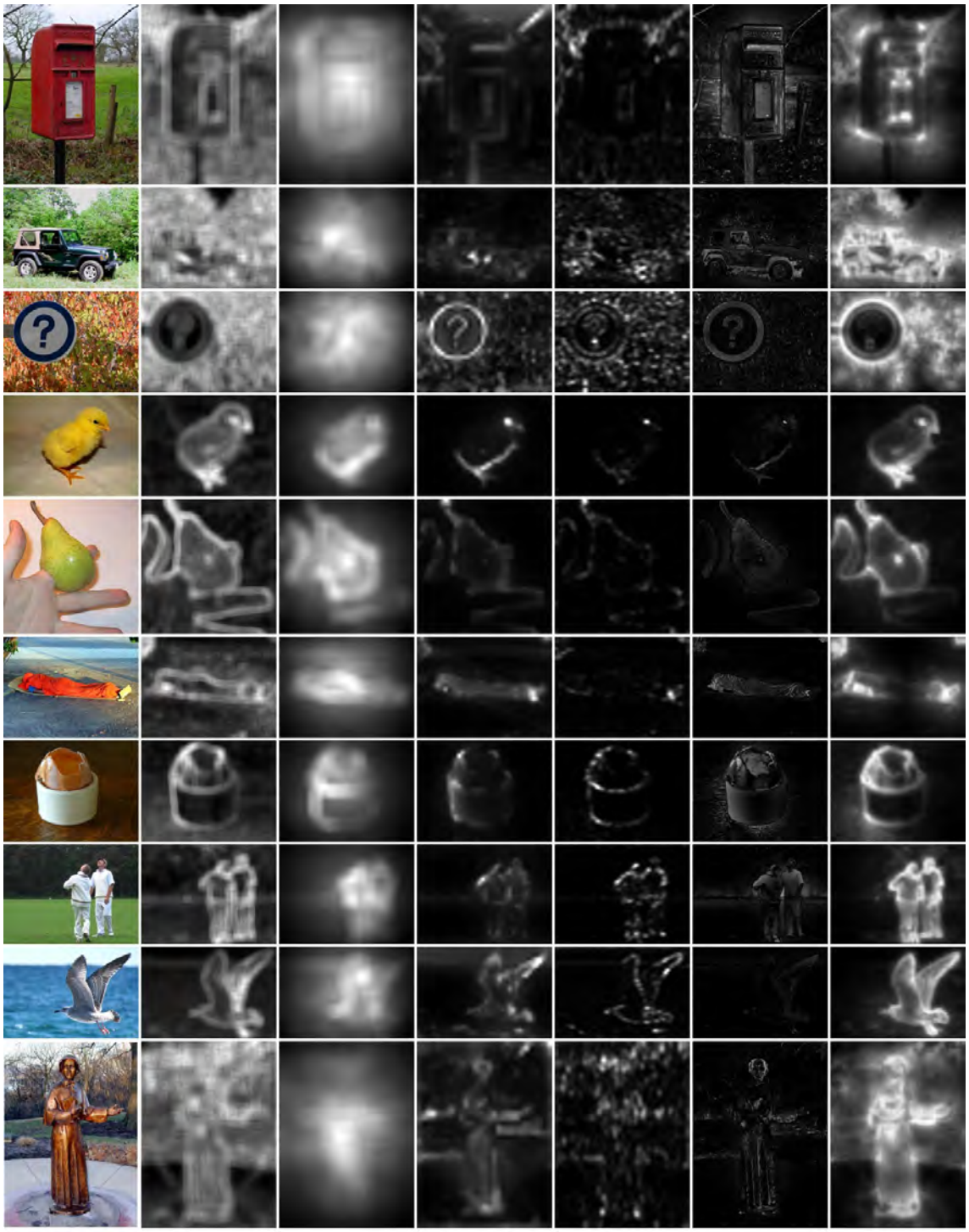
表 2.2　不同显著性区域分割方法的平均用时。THUS10000数据集上用不同显著性区域方法得到的结果可以从该项目的主页上找到。

方法	FT方法[8]	SEG方法[52]	CB方法[96]	本章的方法
时间 (s)	0.247	7.48	36.5	0.621
代码类型	Matlab	Matlab & C++	Matlab & C++	C++



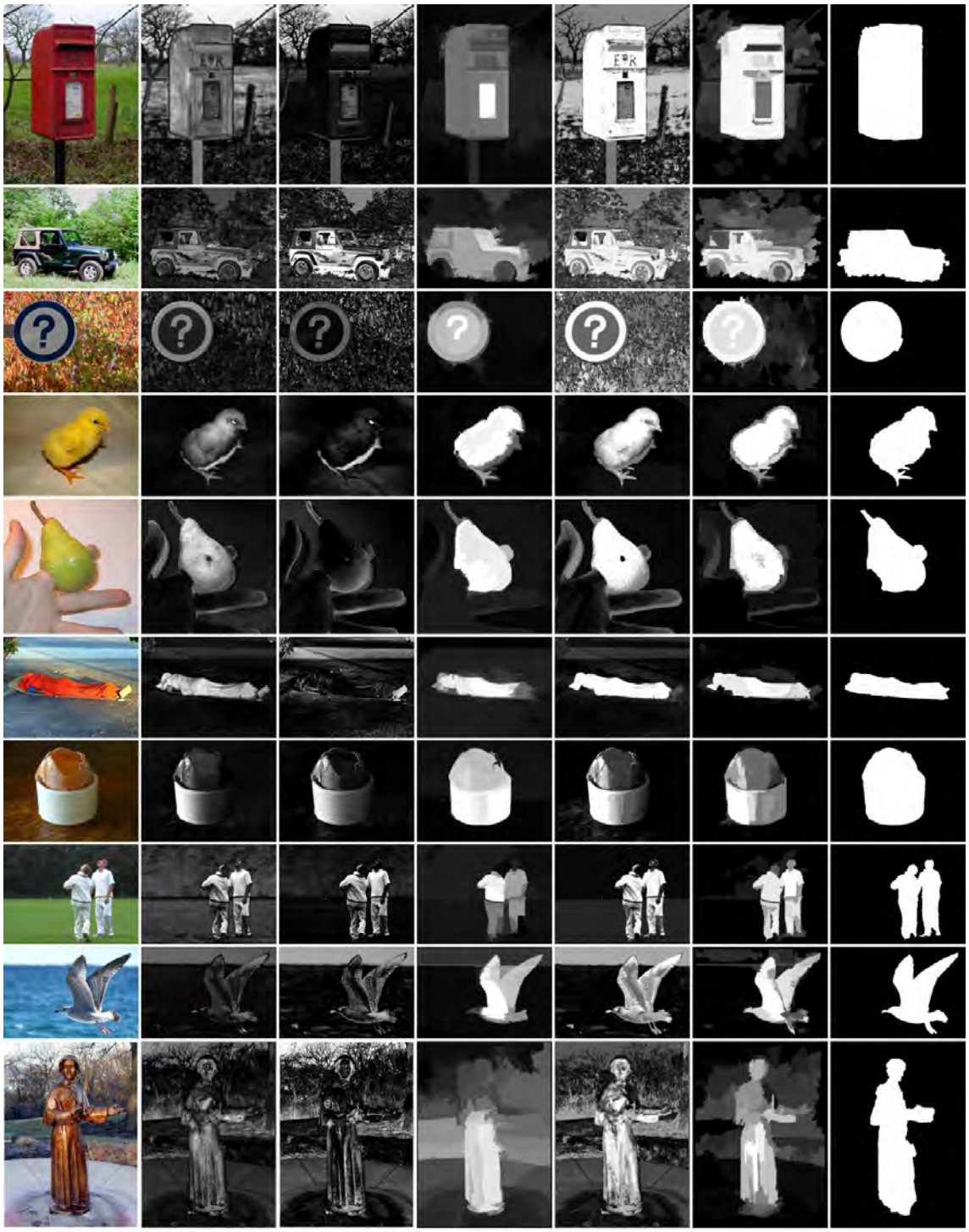
(a) 原图　(b) IT　(c) AIM　(d) IM　(e) MSS　(f) SEG　(g) SeR

图 2.8　显著性图的视觉效果对比。(a)原图，由以下方法产生的显著性图: (b) Itti 等[7], (c) Bruce 和 Tsotsos[50], (d) Murray 等[103], (e) Achanta 和 Süsstrunk[102] (f) Rahtu 等[52]，和 (g) Seo 和 Milanfar[101]. 同一组图与更多方法的比较见图 2.9和图 2.10。整个数据集上的所有检测结果可以从作者的项目主页上获取。



(a) 原图　(b) SUN　(c) SWD　(d) GB　(e) SR　(f) AC　(g) CA

图 2.9　显著性图的视觉效果对比。(a)原图，由以下方法产生的显著性图: (b) Zhang 等[48], (c) Duan 等[104], (d) Harel 等[100], (e) Hou 和 Zhang[47], (f) Achanta 等[49], 和 (g) Goferman 等[99]. 同一组图与更多方法的比较见图 2.8和图 2.10。整个数据集上的所有检测结果可以从作者的项目主页上获取。



(a) 原图　(b) FT　(c) LC　(d) CB　(e) HC　(f) RC　(g) RCC

图 2.10　显著性图的视觉效果对比。(a)原图，由以下方法产生的显著性图: (b) Achanta 等[8], (c) Zhai 和 Shah[98], (d) Jiang 等[96], (e) 本章的HC和(f)RC方法，和(g)基于RC图的显著性区域分割结果. 同一组图与更多方法的比较见图 2.8和图 2.9。整个数据集上的所有检测结果可以从作者的项目主页上获取。

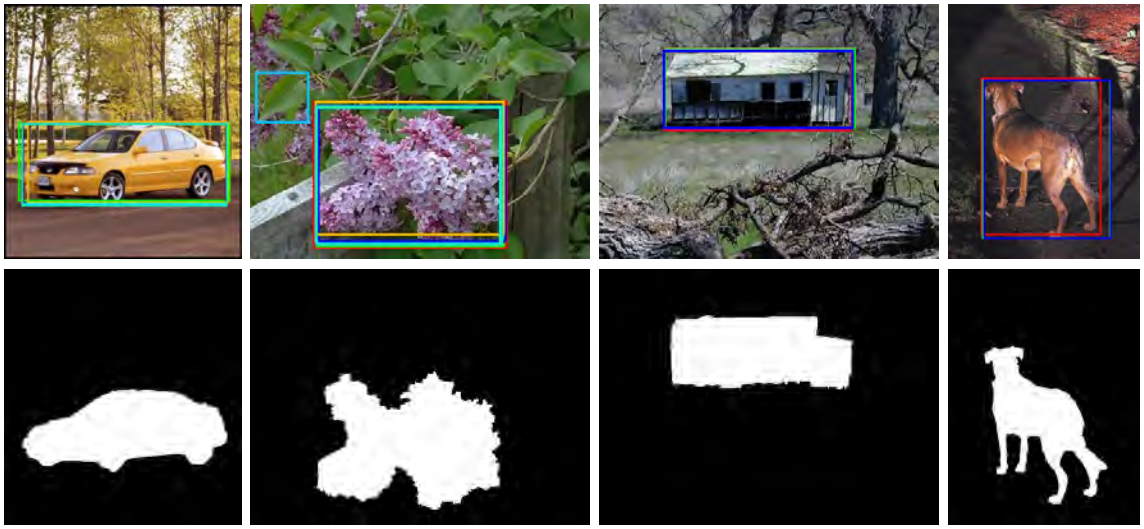


图 2.11　THUS10000测试集中的基准数据示例：(第一行) 原始图像及相应的(来自[105]的)显著性矩形区域标注； (第二行)，本章的THUS10000数据集中包含显著性区域像素级精度精确标注的基准数据。

精确显著性区域标注的图像，作者将其称为THUS10000 (Tsing Hua University Saliency detection dataset with 10000 images) 该数据集比之前同类公开测试集中最大的一个[8]还要大10倍，之后将在作者的项目主页中公开。

2.6.2　固定阈值分割

得到显著性物体二值分割的最简单方法就是设定一个$T_f \in [0, 255]$的阈值。为了可靠地比较多种显著性检测方法检测显著性物体的效果，作者将阈值T_f设定为在0到255之间变化。在Achanta等人[8]的数据集以及THUS10000数据集上得到的精确度召回率曲线结果分别为图 2.12和图 2.13。在图 2.12中，作者还显示了加入色彩空间平滑以及空间权值后的效果以及与其它方法的比较。这些方法得到的显著性图的视觉直观比较在图 2.2 和图 2.8-2.10中。

正确率、召回率曲线清楚地展示出本章的RC方法优于其它的方法。曲线的极限很有趣：在$T_f = 0$处召回率最大，所有的像素被认为是前景。所以，所有的方法得到相同的正确率和召回率；这点的正确率为0.2，召回率为1.0。这表明，在标注数据中，平均20%的图像像素属于显著性区域。曲线另一端，本章方法的最小召回率比其它方法高，因为本方法的显著性图更加平滑，且包含更多显著性值为255的像素。本章的HC方法得到的结果也比其它具有相似计算消耗的快速检测方法(SR方法[47], FT方法[8], 和 LC方法[98])得到的结果要更好。

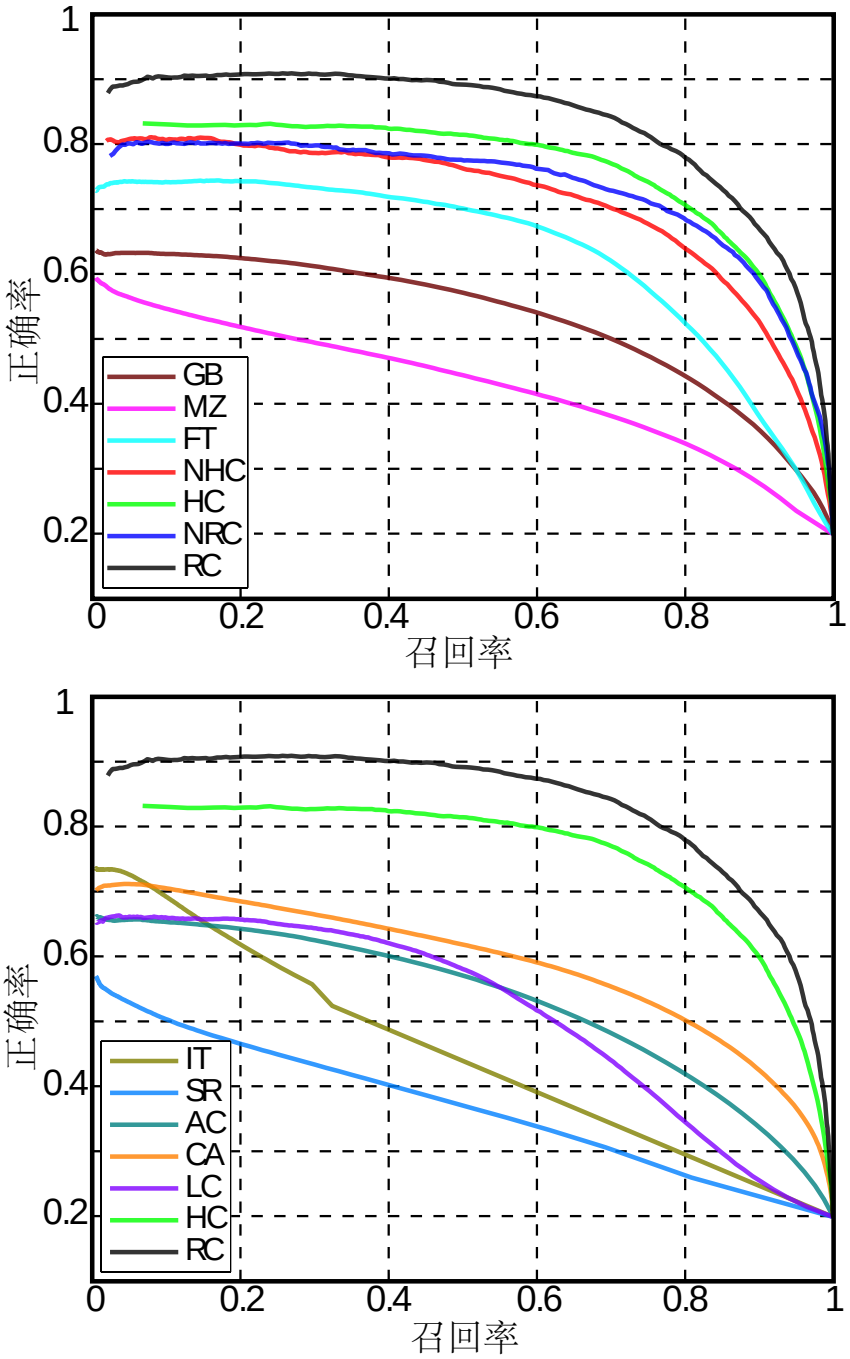


图 2.12　在Achanta等人[8]提供的包含1000副图像的公开测试集上各种方法检测得到的显著性图经过简单阈值分割得到结果的正确率召回率曲线。上图和下图是本章方法不同配置与GB方法[100], MZ方法[65], FT方法[8], IT方法[7], SR方法[47], AC方法[49], CA方法[99],和LC方法[98]等方法的比较。 NHC代表本章HC方法不带颜色空间平滑，NRC代表本章的RC方法不带空间加权。本章的RC方法在1000副图像数据集上的结果具高的精度和召回率。 (相关的结果图像从作者的项目主页上可以得到。)

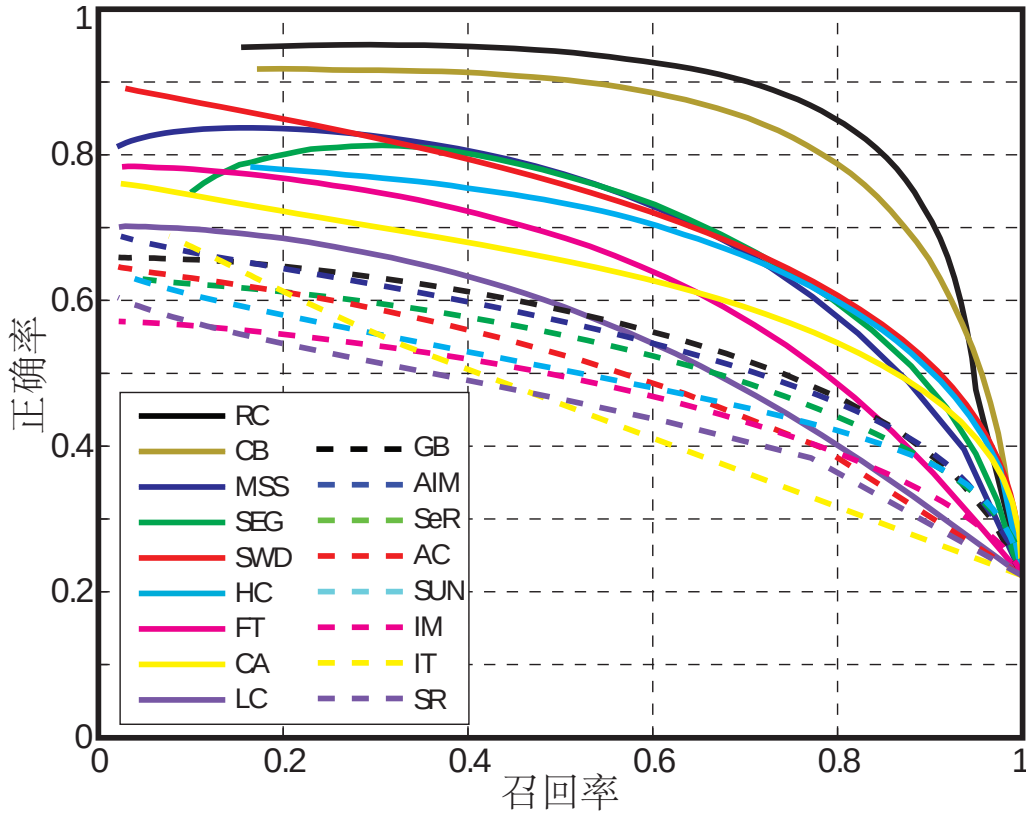


图 2.13　在含有10000张图像精确标注的THUS10000测试集上，各种现有最先进的方法检测得到的显著性图经过简单的固定阈值分割得到结果的正确率召回率曲线。本章的HC方法和 RC方法与15种现有最先进的方法进行了比较。这些方法包括：SR方法[47], IT方法[7], IM方法[103], SUN方法[48], AC方法[49], SeR方法[101], AIM方法[50], GB方法[100], LC方法[98], CA方法[99], FT方法[8], SWD方法[104], SEG方法[52], MSS方法[102], 和 CB方法[96]. (整个数据集上的相关方法对应的所有结果可以从作者的项目主页上得到。)

2.6.3　显著性区域自动提取

为了客观地评价本章的感兴趣物体自动提取方法(以RC方法的结果进行自动初始化)，作者在THUS10000数据集上将本方法的结果和现有最先进的其它显著性区域自动分割方法的结果进行了比较。这些方法包括FT方法[8], SEG方法[52], 和CB方法[96]。具体的客观评价准则包括：平均正确率、平均召回率以及平均F_β度量。其中

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}. \tag{2-10}$$

和Achanta等人[8]一样，作者用$\beta^2 = 0.3$来使正确率的权重高于召回率。正如Liu等人[105]指出的那样，在显著性物体检测中，正确率比召回率更重要。因为可以简单地将所有区域标为显著性区域而达到100%的召回率，而这毫无意义。 图 3.9显示了用不同的显著性图初始化SaliencyCut的结果比较。可以看出基

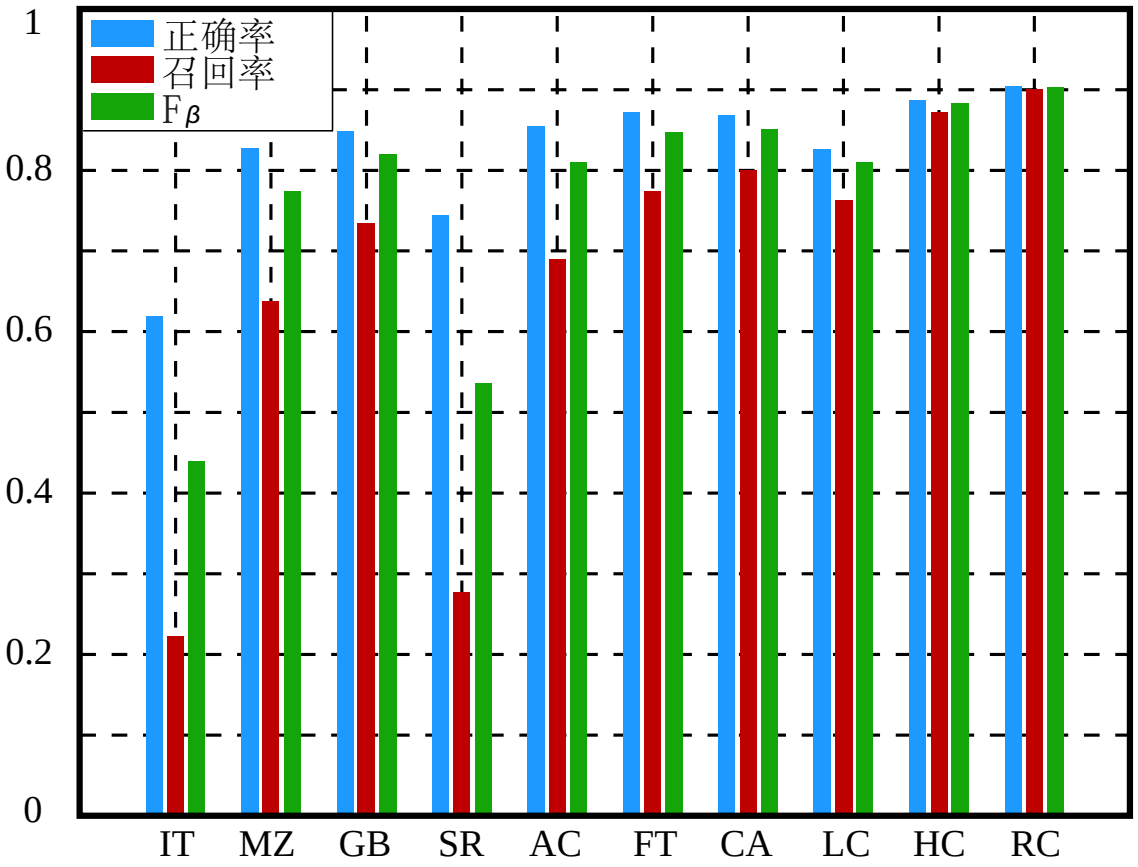


图 2.14　用不同的显著性图初始化SaliencyCut的结果比较 (在Achanta等人[8]的数据集上的结果)。

于RC方法显著性图的结果明显优于其它方法。与FT方法[8], SEG方法[52], CB方法[96]相比，本章的算法将F_β度量上的错误率分别降低了 55%, 49% 和 20%。在这个含有10000张图像的大数据集上的验证结果表明，本章方法的结果明显优于CB方法[96]，同时计算效率也明显更高(参见表 2.2)。 (作者的演示程序和源代码可以从项目主页中得到。)

2.6.4　内容敏感的图像缩放

在内容敏感的图像缩放中，显著性图像经常用来指定图像的相对重要区域 (见[110])。作者用提取出的显著性图像进行了图像缩放实验。实验中采用了Zhang等人提出的[14]内容敏感图像缩放方法 ①。该方法通过变形能量将变形分配到相对非显著性区域，同时保持全局和局部的图像特征。 图 2.17比较了用RC方法和CA方法[99]显著性图像得到的图像缩放结果。本章方法得到的显著性物体区域的显著性值是成片光滑的，这点对于基于能量的缩放非常重要，因此本

① 采用原作者公开的代码

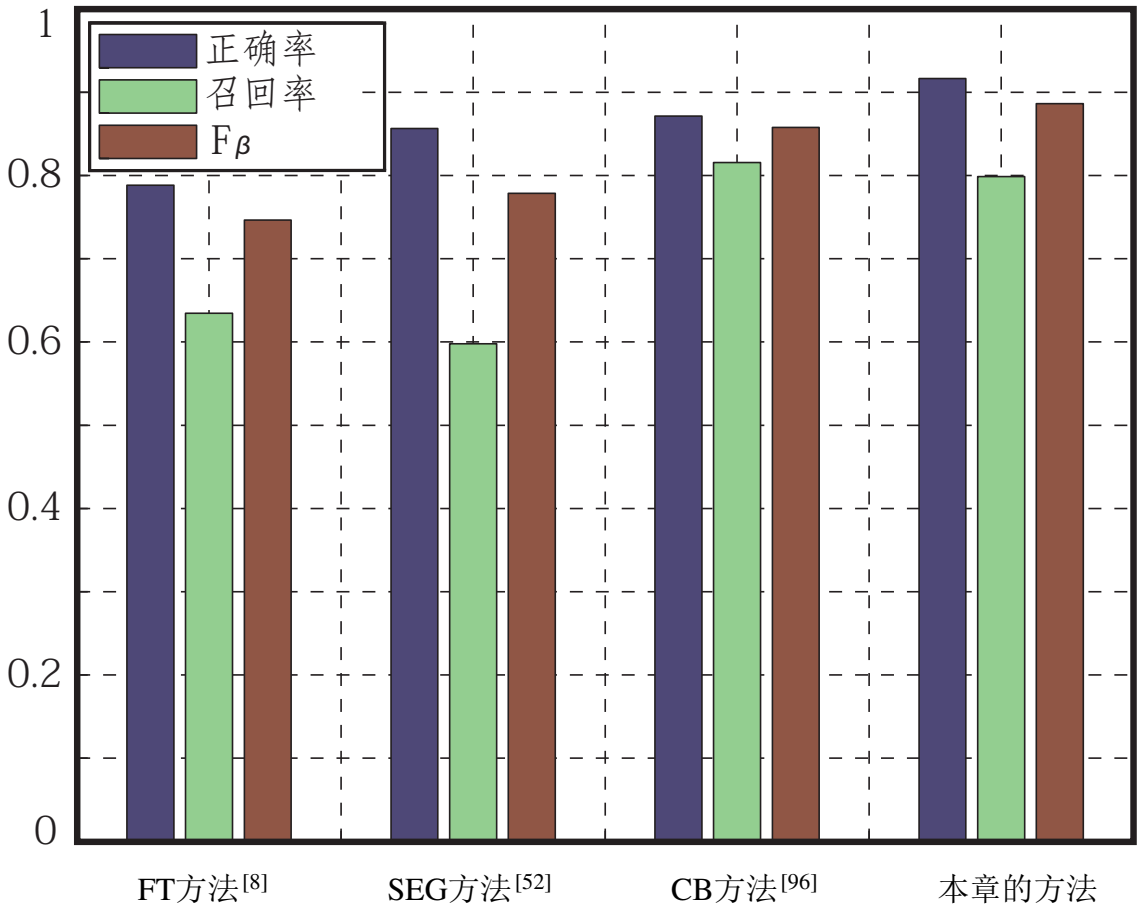


图 2.15　在THUS10000数据集上，各种方法的正确率、召回率以及F_β度量的比较。(相关结果可从项目主页上获取。)

章的 RC 显著性图产生更好的缩放结果。 CA显著性图像在物体的边界有更高的显著性值，但这并不适于图像缩放等应用，这些应用要求整个显著性物体一致地被突出。

2.6.5　非真实感渲染

艺术家们经常对图像进行抽象并突出有意义的部分，同时淡化非重要区域[112]，从而使得图像中的重点更为突出。受此现象启发，一系列用显著性值来进行非真实感渲染的方法应运而生，并产生了有趣的结果[113]。作者将本章提出的显著性检测方法和Achanta等人[8]的显著性检测方法用在非真实感渲染技术[111]中，并对结果进行了比较(见图 2.18)。本章的RC方法提供更好的重要物体区域信息，这种信息可以帮助非真实感渲染方法更好地保留重要图像部分以及区域边界的细节，同时平滑其它部分。

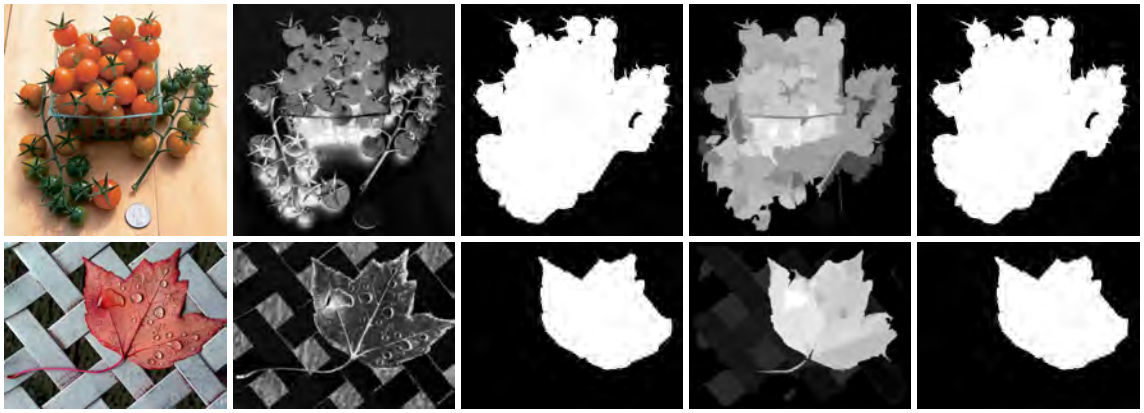
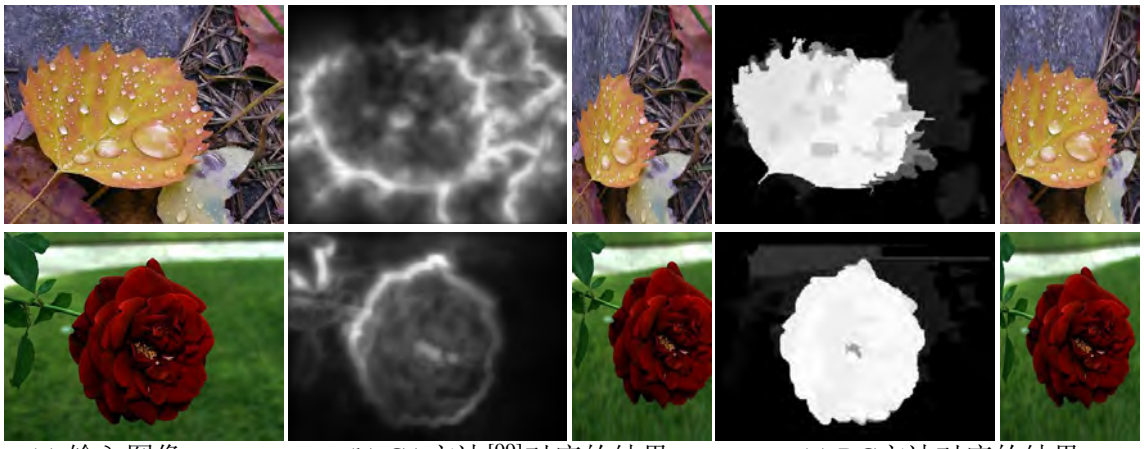


图 2.16　对于本章提出的基于直方图的方法来说一个有挑战性的例子：(上图)显著性区域和非显著性区域具有相似的颜色，(下图)图像具有复杂的纹理背景。从左到右依次是输入图像、HC显著性图、HC显著性区域分割、RC显著性图、 RC 显著性区域分割.



(a) 输入图像　　(b) CA方法[99]对应的结果　　(c) RC方法对应的结果

图 2.17　用CA方法[99]和本章的RC显著性图进行内容敏感的图像缩放[14]的结果比较。

2.7　小结与讨论

本章提出了基于全局对比度的显著性计算方法，即基于直方图对比度 (HC) 和基于空间信息增强的区域对比度 (RC)方法。 HC方法速度快，并且产生具有精细细节的结果，RC方法可以产生空间增强的高质量显著性图像，但与此同时具有相对较低的计算效率。作者在国际上现有最大的公开数据集上测试了本章的方法，并与十多种现有最先进的方法进行了比较。实验结果表明，本章提出的方法在正确率和召回率上都明显优于其它方法，并且简单而高效。

在未来的工作中，作者计划研究包含空间关系且保留详细细节的显著性图像的高效计算算法，并且希望研究能够处理具有复杂纹理的背景图像的显著性检测算法，以克服本章算法在处理这类情况中存在的缺陷。最后，作者还希望在显著性图像的检测过程中进一步考虑人脸、对称性等高级因素。作者相信本章算法得到的显著性图像可以应用于高效物体检测[114–116]，可靠图像分类，鲁棒的图像景

(a) 输入图像　　(b) FT方法[8]对应的结果　　(c) RC方法对应的结果

图 2.18　FT方法[8]和RC方法的显著性图分别被用于非真实感绘制(又称风格化绘制)[111]。本章的方法生成了更好的显著性图，使得风格化绘制保持了重要部分的细节。例如马头和栏杆部分细节更好地被保留了。

物分析[85,117]等应用中，并提高图像检索效果[118–120]。

第 3 章 基于相似性分析的群组图像显著性区域提取与检索

本章主要研究基于相似性分析的群组图像显著性区域提取与检索。与第 2章中单张图像的视觉显著性区域检测与分割相比，本章的讨论对象是一组相关、低质、不可靠的网络图像(例如：一组网络图像)，将重点研究如何利用一组图像中公共显著性物体区域的相关信息提高显著性区域提取的正确率，并利用显著性区域提取结果进行基于草图的图像检索。 第 3.1介绍背景知识、研究动机和解决方法概要； 第 3.2节给出与本章方法紧密联系的相关工作概述； 第 3.3节和第 3.4节分别介绍本章算法的两个核心部分：单张图像的无监督分割和群组图像的视觉显著性区域提取。 第 3.5给出实验验证和结果分析。最后，第 3.6节对全章进行总结。

3.1 引言

3.1.1 背景知识

3.1.1.1 图像分割与感兴趣区域提取

对图像进行分割，进而提取出有意义的分块或者感兴趣的物体区域，是许多图像处理技术的基础，为图像识别、检索、跟踪、分割、合成、编辑等重要应用提供支撑。经典的图像自动分割方法，通常根据像素的颜色或纹理等表观特征的一致性将图像分割为几个大块的区域。这类方法主要包含两大类：谱分割方法和图分割方法。其中最为典型的工作分别是： Jianbo Shi 和 Jitendra Malik于2000年提出的 Normalized Cut[32] 方法与Pedro F. Felzenszwalb和Daniel P. Huttenlocher于2004年提出的基于图的快速分割[60] 方法。虽然这些方法能够有效地避免细碎的纹理区域对分割结果的影响，但分割结果的每一个区域依然是一个纹理颜色等可视特征较为统一的区域。但现实生活中的很多物体是由颜色纹理特征差异巨大的许多部分组成的 (例如穿不同色彩风格外套和裤子的一个人)，因而这种可视特征较为统一的区域并不能直接反应用户所理解的场景物体。

基于交互的图像分割可以通过简单的用户交互来避免自动分割所面临的困境，其中最为典型的工作分别是：Leo Grady 提出的基于随机游走模型的图像分割方法[61,62] 与 Yuri Boykov 和Gareth Funka-Lea 提出的基于图分割的图像分割方法[33]。这些方法以简单的几笔用户输入来指示目标结果区域的典型部分，并以此为依据得到高质量的分割结果。虽然通过引入颜色以外的其它指导信息 (在以上

两种方法中这种信息是用户标定的样例区域)，能够有效地提高用户感兴趣区域获取的质量，然而以上方法都需要较大的用户交互工作量。

通过更少的用户交互获取可能包含不同颜色纹理特征的整个感兴趣物体区域，是计算机视觉和图形学研究的一个重要趋势。这种较小的用户交互给用户提供了更加智能的编辑体验，同时也给分割算法提出了更高的鲁棒性要求，特别是这些算法对低质量初始化的鲁棒性要求。 2004年，Carsten Rother 等人[63] 提出了GrabCut 分割方法。用户只需要输入简单的矩形框来指明前景区域，就可以得到非常高质量的分割结果。 2008年，Shai Bagon等人[64]提出了一种统一的分割框架来处理各种图像区域，包括：相似颜色区域、相似纹理区域、对称性区域等。该系统仅需用户提供一个感兴趣点的信息，该系统就能计算出包含该点的“最好”图像区域。虽然单个兴趣点的输入对用户进行单张图像的处理来说已经非常方便，但是当用户所要面对的是成千上万的图像数据时，即使每个图像仅需一个点的交互对用户来说也是非常繁重的交互。

单张图像感兴趣区域的无监督分割： 利用视觉显著性区域进行指导，从图像中自动地获取感兴趣物体区域成为一个重要的研究课题。这种无监督的物体分割 (unsupervised object segmentation) 提供了对从海量数据进行分析的可能，为图像的有效组织、检索和分析提供了有力的支持。 2003年，微软亚洲研究院的Yu-Fei Ma和Hong-Jiang Zhang[65] 利用模糊区域增长的方法从显著性图中得到矩形的显著性区域。这种模糊增长过程通过模拟人类视觉自底向上的搜索过程对感兴趣区域进行分割。 2006年，香港中文大学的Junwei Han等人[66] 在马尔可夫随机场(Markov random field) 理论框架下对颜色、纹理以及边缘特征进行建模，并利用区域增长策略从显著性图中的种子点得到视觉显著性区域。 2010年，芬兰赫尔辛基大学的Esa Rahtu等人[52]在条件随机场算法框架下利用显著性图去初始化一个基于能量最小化的分割方法，以便得到完整的显著性物体。该方法进一步利用积分图 (integral histogram) 方法和基于图分割的求解策略，从而实现快速计算。以上方法，在单张图像感兴趣区域无监督分割领域做出了许多重要的尝试，使这种无监督分割的正确率持续改善。但由于没有任何用户输入，这种无监督分割方法的自动初始化，远远不如基于交互的分割算法中人工指定的初始信息精确。对于低质量输入的鲁棒性是解决这种无监督分割问题的一个关键问题。

多张图像感兴趣区域的无监督分割： 大量相关图像的感兴趣区域自动分割，在图像检索[15]、编辑[16,67]、合成[18,68]、知识学习[17]等领域有着重要应用。相似图像的公共部分含有很多关于感兴趣区域的重要信息，如果能够有效地利用这些

信息，就很有可能改善单张图像的感兴趣区域分割结果。然而，利用这种相似信息并不容易。一方面，许多相关图像集合 (特别是用同一关键字得到的网络图像) 本身可能包含很大比例的无关图像，从这种低质量的图像集合中获取公共信息本身存在着巨大的挑战。另一方面，针对大量图像公共信息的获取和利用通常需要大量的计算消耗。如果不能有效地降低这种计算量，任何算法都无法真正用于实际应用中。鉴于多张图像感兴趣区域无监督分割本身所固有的这些难题，研究者们首先尝试了一些相对简单的情况，并取得了一定的成功。2010年，国立清华大学的Hwann-Tzong Chen[69] 利用两幅图像中的联合信息来抑制仅在一幅图像中出现的图像区域显著性。该方法能够在不使用匹配算法的情况下，对视觉注意机制进行模拟，并从一对图像中找到公共的显著性物体。虽然仅仅是处理两幅图像，但该工作对多张图像的公共显著性区域检测也可算是一次大胆尝试。2011年，成都电子科技大学的 Hongliang Li 和香港中文大学的 King N. Ngan 共同提出了一种检测两张图像中公共显著性区域的方法。该方法是由单张图像显著性图和公共显著性图两部分的线性组合构成。其中，线性组合的第一部分通过三种已有显著性检测技术获得，第二部分通过在两幅图像尺度金字塔结构表达上定义的图结构来计算。同年，台湾中央研究院经济研究所的Kai-Yueh Chang等人提出了一个多张图像的公共显著性区域检测和分割算法[70]。该方法在马尔可夫随机场算法框架下，利用公共显著性先验知识，作为可能的前景区域的线索，并通过一个优化过程求解公共显著性。以上模型都要求在几乎所有输入图像中都至少包含公共显著性物体的一部分，这种严格的要求对于组成复杂的网络图像集合来说明显是难以满足的。并且，这些方法的计算量会随着图像数目的增加而快速增长，进而限制了其处理更大规模数据的能力。在这些工作的实验中，所演示的最多同时处理的图像数目也只有30多个[70]，而本章工作的目标是从数以千计更为复杂的网络图像中检测并分割出公共显著性物体区域。

3.1.1.2　基于草图的图像检索

形状是人类识别和检测[71,72]物体最重要的视觉依据[21]，人类可以仅靠形状信息就能对物体进行有效地识别[22]。相对于颜色或者纹理等其它视觉特征，形状特征不仅表达能力更强大, 而且更方便用户输入。因而，通过简单的草图形状对图像进行检索具有非常大的研究价值。特别是在智能手机等触摸式终端普及的今天，这种输入方式对用户更有吸引力。

早在1992年，Kyoji Hirata 和 Toshikazu Kato[73] 就提出了一种利用用户草图输入检索具有相似边缘形状图形的算法。然而，该方法对用户输入草图的要求过高，现实生活中用户很难进行如此精确的输入，因而该系统难以用于实际的搜索

中。

为了改进算法对具有一定差异的用户输入草图形状的鲁棒性，意大利佛罗伦萨大学的 Alberto Del Bimbo 和 Pietro Pala[74] 于1997年提出了一种采用弹性匹配机制对用户输入草图和图像边缘进行匹配的方法。但是，由于这种非线性匹配耗时过大，该系统很难用于实时交互搜索应用之中。

1999年，加州大学伯克利分校的Chad Carson等人[75]提出了一种名为 Blobworld 的方法。该方法通过期望最大化对图像像素依据其颜色、位置、纹理等因素进行分割，以得到底层视觉特征较为一致的几个大尺度的图像区域，然后为每个区域赋予颜色和纹理信息。然而，由于很多物体本身并不是由底层视觉信息一致的单个区域组成，以这种方法得到的形状信息很多时候不足以用于基于形状的检索。虽然该方法也提供了基于草图的界面，但只是利用草图来绘制颜色区域，而没有利用形状信息。

考虑到基于草图的自然图像检索本身的巨大困难。也有很多研究工作针对某些特殊的图像类别进行检索，如：简单的图案[76,77]或者剪贴画[78]等。在用户草图本身的分类方面，Eitz等人的研究也取得了重大进展[79]。这种分类技术有望给用户输入的草图关联一些类别信息，从而利用基于关键字技术的已有成果。

近些年，随着智能手机等触摸式输入设备的普及和相关图像分析技术的发展，基于草图的图像检索受到了越来越多的重视和研究。

英国萨里大学的Rui Hu等人[80,81]提出了一种多分辨率的区域表达来防止背景区域的影响，并采用一种基于码本的检索机制实现快速的用户相应。

微软亚洲研究院的Changhu Wang等人建立了MindFinder[82] 系统。该系统首先将输入草图和数据库中的图像均匀地分割为10 × 10个区块，并对每个块中的特征建立索引。虽然这种索引机制使得检索过程非常高效，但是该方法不具有平移、缩放等不变性。这种期望用户输入与目标图中物体在相似位置和相似角度的要求对普通用户来说具有一定的难度。

柏林工业大学的Eitz等人[24] 采用局部特征进行基于形状的检索并达到了现有最好的检索性能。该方法的成功之处主要在于局部特征对平移不变性的支持，同时他们采用较大的(实验中他们用的大小是图像对角线距离的20% – 25%)局部特征来保留大尺度的特性。但是，由于如此大的窗口为草图平移留下的空间并不太多，这种大窗口限制了平移不变性。

和所有以上方法不同，作者利用显著性分割 (SaliencyCut)来自动地检测和提取目标区域，并通过形状比较对分割结果进行排序。通过迭代地使用从排序结果靠前的图像分割中学习到的表观一致性信息来提高目标提取的正确率，进而改善检索准确率。因此，本章的方法对复杂背景更具鲁棒性，并天然支持那些需要精

确物体区域信息的应用[15–18,67,68]。

3.1.2　研究动机

伴随着数字照相机和智能手机的普及，人们迎来了一个图像信息大爆炸的时代。这些存放在个人电脑或者网络相册里的图像构成了人们日常生活中分享信息和记录事件的重要媒介。然而，对这些海量图像数据的有效组织和高效检索仍然是个巨大的挑战，同时也是一个非常有意义的研究课题。通常情况下，这些相册集合具有数量巨大、内容多样、干扰信息多等众多对传统图像查找和检索技术不利的因素。而现有的商业搜索引擎主要依赖图像的元数据(例如：关键字、GPS数据等)。但是，这些元数据通常并不太可靠，也不全面。

为了克服元数据方法存在的缺陷，研究者们尝试建立了许多基于内容的图像检索系统。这些系统通常利用图像的颜色、纹理、形状等视觉特征来查找用户期待的结果[20]. 在这些众多的因素当中，形状信息包含了重要的区域信息，是最强大的视觉线索[21]。认知心理学研究表明，即使只有形状信息，人类也可以进行有效地物体识别[22]。基于草图的图像检索(Sketch Based Image Retrieval, SBIR)就是一种允许用户绘制轮廓线条，并以此为搜索依据的图像检索方式。然而，要求用户绘制和图像轮廓精确相似的草图并不现实。为了适应这种输入的不精确性，许多基于草图的图像检索系统利用全局的形状特征进行匹配。这些全局特征描述在仿射变换下的匹配结果通常并不鲁棒。因此，Eitz等人[24]提出了一种基于局部描述子的方法，并达到了现有最先进的基于形状的图像检索水平。

与基于草图的图像检索领域遇到瓶颈不同，在没有背景干扰的纯形状比较领域，相关的研究则相对成熟。即使对非常有挑战性的MPEG-7数据集，现有最先进的方法也能达到91.6%的检索率[23]。 Shape Context[71] 和 Chamfer Matching[121] 等经典的形状匹配算法对于没有背景干扰的形状比较非常鲁棒。没有无关边缘干扰的纯物体轮廓可以很自然地改进现有的基于草图的检索系统。基于这一观察，Bai 等人[122] 提出了一种形状带模型(Shape Band Model) 来选择候选的边缘片段，并用Shape Context距离来提取最优的匹配。然而，形状带是和用户输入草图有关的。由于其不能被预计算，因而不适合交互式的查询。

一组相关图像集合中公共显著性区域的表观信息包含了这组图像集合中感兴趣区域的重要信息。这些信息为改善感兴趣物体区域的自动提取提供了重要依据。所得到的感兴趣物体区域排除了背景区域的干扰，使得纯形状比较领域的相关成果可以直接应用于基于草图的图像检索应用中，改善检索性能。

3.1.3　解决方法概要

本章将介绍一种显著性形状(*SalientShape*)算法框架。该技术利用含有相似物体的一组图像(例如，同一关键字对应的网络图像)之间的颜色与形状的相似性来进行快速有效地图像检索。首先，作者利用关键字检索一组候选图像。由于检索词本身的二义性和标签之间的差异性，以上候选图像包含大量噪声数据(例如不包含关键字对应目标的图像)。该方法利用视觉显著性区域检测和分割算法[107]为每一个候选图像提取显著性区域，并同时估计出前景和背景的表观模型。作者观察到了一个有趣的现象：即使对于这种含有非常多噪音信息的图像集合，它们的显著性区域在表观和形状性质上也有某种一致性。

本方法首先滤除那些含有复杂场景的或者显著性分割质量不高的候选图像，然后利用当前类查询结果中形状排序靠前(其显著性区域的形状与用户输入的形状较一致) 的一批图像来建立一个全局的前背景表观模型。这种全局表观模型被用来改进显著性区域的分割过程。最终，作者利用显著性分割结果和用户输入形状的一致性对图像进行排序，利用这种学习全局表观模型改进显著性区域检测的过程经过几次(实验中作者采用2 次)迭代后，便得到了更好的显著性区域分割和图像检索结果。为了更好地对这种群组图像的显著性区域分割结果进行验证。作者构建了一个含有15000张图像的参考数据集，并起名为THUR15000。该数据集是由

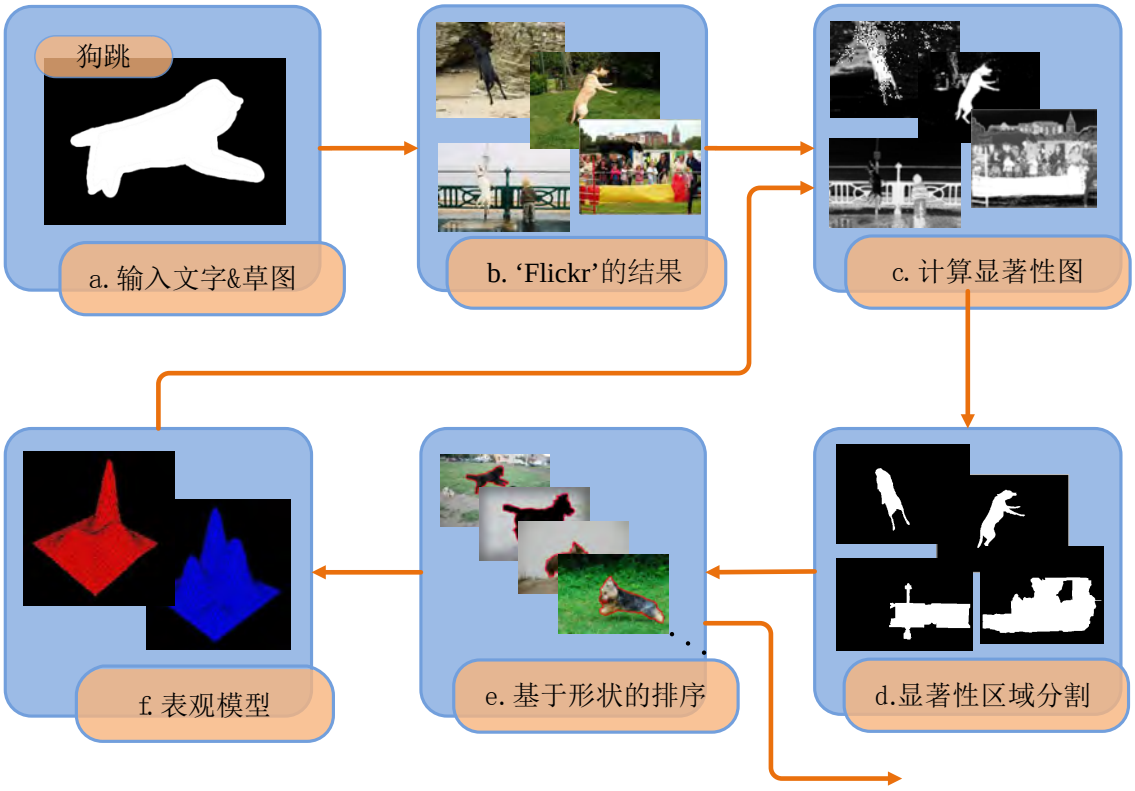


图 3.1　系统流程图。

从Flickr上下载的图像和这些图像对应的像素精度显著性区域标注(如果存在这样的区域则标注)共同组成的。作者在该数据集上对本章提出的算法进行了广泛的测试。实验结果表明，这种群组图像显著性区域检测和分割算法明显优于已有单张图像的显著性区域分析算法。这些分割结果直接与输入形状进行比较，就能取得优于现有最先进的形状检索算法的结果。

作者利用显著性估计以及学习到的表观模型共同来处理一组相关的图像，并在质量参差不齐的网络图像集上展示了该方法快速鲁棒的性能。本章的系统同时利用网络元数据(Meta-Data)①、视觉显著性以及形状相似性，从图像中显式地提取显著性区域，并使得形状检索过程不受背景噪声的干扰。改进后的显著性区域提取使得本方法能够自然地利用现有形状比较算法[21,123,124]来实现尺度、旋转、平移等不变性。本章的主要贡献包括四个方面：

(1) 提出了一种群组显著性(*group saliency*)的方法来自动从一组相关但组成多样的网络图像中提取感兴趣物体区域。
(2) 提出了一组无监督图像分割结果可靠性的度量方法。
(3) 利用显著性区域的提取和自动提取结果的度量方法，建立了一个简单、快速高效的基于草图的图像检索系统。
(4) 建立了一个用于评测一组相关图像集合的无监督分割和草图检索结果的基准数据集。该数据集含有15000张图像，是现有最大单张图像显著性区域精确标注数据集的15倍。

3.2　相关工作

3.2.1　视觉显著性区域提取

本章中将要介绍的工作属于视觉显著性区域提取领域的研究范围。Ko和Nam[106]用图像分割特征数据上训练的支持向量机(Support Vector Machine, SVM) 来选择显著性区域。紧接着对这些显著性区域进行聚类以提取显著性物体。 Han 等人[66] 用马尔可夫随机场(Markov Random Field, MRF) 框架对颜色、纹理和边缘特征进行建模，并从显著性图中的种子点处进行区域增长来得到显著性物体区域。在一个很有影响力的工作中，Achanta等人[8]对均值漂移分割结果中每一个区域内的显著性值进行平均，并选择平均显著性值高于整个图像平均显著性值一定倍数的区域作为显著性物体区域。 第 2章中介绍的显著性区域分割方

① 元数据是图像检索的现有工业标准。它被广泛地用于各大商业引擎，如： Google图像(http://images.google.com/)、 Baidu图像 http://image.baidu.com/、搜搜图像 http://image.soso.com/、 Flickr http://www.flickr.com/、等等。

法[107]在国际上现有最大的公开测试集上取得了现有最好的效果。近期，关于两幅图像[69,125]或多幅图像[70] 中公共显著性物体区域提取的研究已经开始展开。然而，以上模型都要求几乎所有输入图像的视觉显著性区域都至少包含前景物体的一部分。对于组成复杂的网络图像，这种严格的要求明显是难以满足的。而且，这些方法的计算量会随着图像数目的增加而快速增长，这一点限制了其处理更大规模数据的能力。在这些工作的实验中，最多同时处理的图像数目只有30多个，而本章的目标是处理数以千计的更为复杂的网络图像。

3.2.2　网络图像重排

本工作和网络图像重排领域的研究也有着紧密的联系。 Fergus 等人[126] 利用网络图像搜索引擎反馈的排名靠前的结果训练分类器，并用训练得到的分类器对搜索结果进行过滤。 Ben-Haim 等人[127] 自动地把图像分割为多个区域，并为每个区域建立颜色直方图。他们随后对这些区域直方图进行聚类。最终依据图像区域与主要聚类中心之间的特征距离对图像进行排序。 Cui 等人[128] 首先将待查询图像分为事先定义好的若干类中的某一类。然后利用该类对应的一组特定的相似性度量对图像特征进行组合并依据与查询图像的距离对结果进行排序。Popescu等人[129]根据查询结果的视觉一致性对图像进行排序。他们同时采用了一种控制多样化的函数来避免接近重复的查询结果，并确保查询的不同侧面都能呈献给用户。然而，以上方法都没有利用到视觉注意机制和目标物体的形状信息。与以上工作不同，本方法利用视觉注意机制和目标物体的形状来采集潜在目标物体可能具有的较大范围表观特征。这些特征被进一步用来改善目标物体提取和基于形状的图像检索的结果。

3.3　单张图像的无监督分割

作者首先为给定的每个关键词 (例如： ‘jumping dogs’, ‘plane’, 等)，从Flickr上检索得到一组候选图像(见图 3.1)，其中每组大约包含3000张图像。按照本章中将要描述的方法，作者对每一个图像都进行无监督的分割，以获取这些图像的显著性物体区域。然后，作者利用相似图像之间的相关性估计群组显著性(Group Saliency)，并改进单张图像显著性区域提取的结果(见第 3.4节)，最终提高基于显著性区域形状比较的检索过程的正确率。

3.3.1　基于视觉显著性的图像分割

对一张由RGB色彩空间中的像素I_i组成的彩色图像$I = \{I_i\}$进行分割相当于为

(a) 输入图像　(b) 分割结果[60]　(c) 显著性图[107]

图 3.2　复杂或者混乱场景的图像对应的分割结果经常是由许多零碎的区域组成的。这些图像的显著性检测质量通常不高。

每一点求解对应的透明度值$\alpha = \{\alpha_i\}$ (其中，$\alpha_i \in [0,1]$)。对于硬分割，$\alpha_i \in \{0,1\}$。其中0表示背景，1表示前景。在进行无监督的分割过程中，作者首先为前背景颜色分布建立一个高斯混合模型G；然后直接用这个高斯混合模型获取二值化的分割结果，从而避免人工指定阈值。

作者通过优化吉布斯能量方程(*Gibbs* Energy Function) E 来实现图像分割过程

$$\min_{\alpha} E(\alpha, G, I) = \min_{\alpha}(U(\alpha, G, I) + V(\alpha, I)) \tag{3-1}$$

其中，$U(\alpha, G, I)$ 度量了在给定颜色模型G的情况下，透明度分布α与图像数据I的一致性； $V(\alpha, I)$ 度量了 α 的平滑性 (具体细节请参考GrabCut方法[63])。本方法迭代地优化这个能量方程来自动地提取显著性区域。

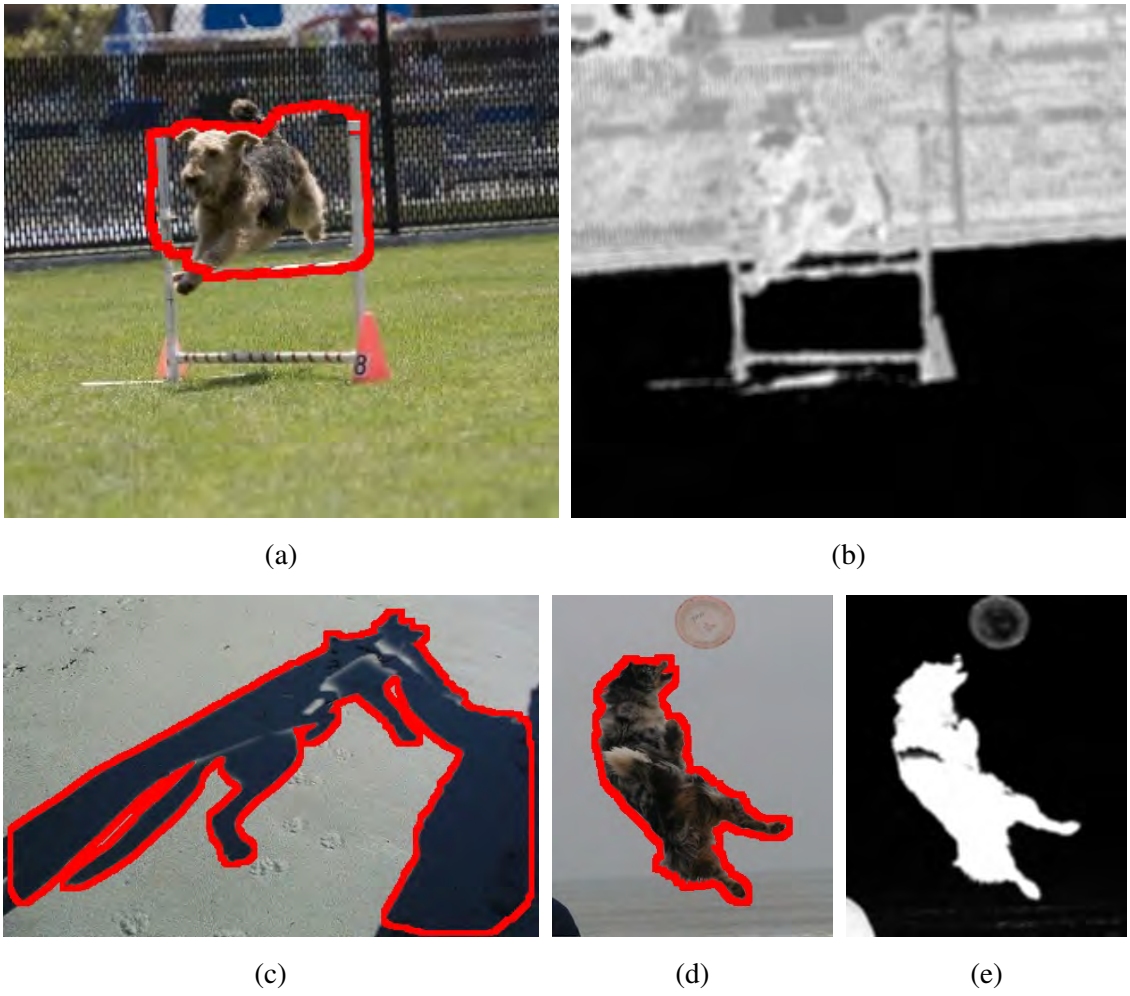
作为初始化，本章首先利用连续的显著性图[107]来学习颜色的初始高斯混合模型G。然后，作者利用迭代的GrabCut交替地提取二值的透明度值和改善高斯混合模型的估计，进而得到显著性区域的二值化分割结果。由于本章采用连续的显著性图，因而避免了类似[107]中选取二值化阈值的问题。实验中，作者发现这种软赋值在改善分割结果的同时并未增加太多计算量(< 10%)。

3.3.2　显著性分割的可靠性度量

基于关键字的图像检索结果经常含有大量的无关图像(见第 3.5.3节)。在本章的图像检索应用中，用户更关心排名靠前的检索结果(例如前50张图)的正确率，而非所有结果(通常是成千上万个)中的召回率。因此，本方法可以大量的滤除可能的异常图像。具体步骤将在本节中详细介绍。

3.3.2.1　场景复杂度

如图 3.2所示，对于场景复杂或者混乱的图像，其显著性图的检测结果通常质量较低。作者利用图像自动分割方法[60]所得到的区域数作为场景复杂性的度



(a)　(b)

(c)　(d)　(e)

图 3.3　一些典型的显著性区域自动分割示例 (其中，分割结果用红色轮廓表示)：(a) 是一个不精确的分割；(c) 由于受到图像边界的影响，其感兴趣区域不完整；(d) 是一个较高质量的分割。这些不精确的或者不完整的分割可以通过度量相应的前景概率图(b, e)来实现。

量。直观上说，含有较少分割区域的图像通常对应着较简单的场景。本方法按照区域数递增的顺序对图像进行排序，并仅为后续处理步骤保留排序前T_R 的图像。在本章的所有实验中，作者采用$T_R = 70\%$。也就是说，本方法通常在这步中丢弃大约1000张图像。

3.3.2.2　分割质量

即使对相对简单的场景，显著性分割也经常会产生不完整或者不精确的分割边界。作者通过两种方法来分别检测这两种情况。

为了判断分割结果是否精确，作者首先利用显著性区域分割的结果来训练前景和背景的高斯混合模型。然后依据前背景的高斯混合模型估计图像像素的

前景概率。作者计算显著性区域分割结果周围一个(30个像素宽的)带状范围内的像素前景概率的平均值。这个平均值越高，分割结果不精确的可能性就越大(如图 3.3(a))。本方法利用图像边界20个像素宽度以内区域所含显著性区域分割前景像素的数目作为该区域不完整性的度量(见图 3.3(c))。这个值越高，其对应的显著性分割结果就越有可能不完整。

本方法按照以上两种度量的升序对剩余的图像进行排序，并分别保留前T_P和T_B的图像用于后续步骤。实验中，作者采用$T_P = 80\%$ 和 $T_B = 80\%$。剩余的图像将在后续过程中被用于对图像集的一致性进行分析。

3.4 群组图像视觉显著性区域提取

由同一个关键词从互联网上搜索得到的一组图像之间存在着紧密的联系，也同样会由于姿态、表观等因素的不同产生个体差异性 (相关例子请参考图 3.10的第一列)。作者用粗略的笔画来指定用户感兴趣的物体姿态(例如，当用户想获取某个特定姿态的“dog jump”时)，并用检索结果图像之间的一致性来提取相关物体合理的表观模型 (例如，“dog”的颜色，以及其经常出现的地方的背景颜色)。作者称这种图像的显著性为群组显著性(*group saliency*)。某个图像区域的群组显著性较高则意味着该区域与其它图像区域之间具有较高的相似性(*similarities*) 同时在图像内部又具有特异性(*distinctness*)。更确切地说，作者首先用一种高效的瀑布模型对单张图像的无监督分割结果按照其与用户输入草图形状的一致性进行排序。然后用排序靠前的若干结果去训练感兴趣物体的全局表观统计模型，并用这个模型来改进群组图像显著性检测与分割的结果。

3.4.1 基于瀑布模型的形状检索

基于初始图像的显著性分割的物体轮廓，作者利用现有的形状比较算法[21,124]来检索与用户输入草图相一致的图像。本方法使用了以下这些在背景干净的图像区域上容易计算的简单度量：

(1) 圆度(Circularity)[124]，
(2) 实心率(Solidity)[130]，
(3) 傅里叶描述子(Fourier Descriptor)[131]，
(4) 形状上下文(Shape Context)[71]。

其中圆度定义为周长P与区域面积A之间的比值P/A，实心率定义为区域面积A与物体凸包面积H的比值A/H。对于由离散像素组成的图像，其周长和面积可由其轮廓点的坐标直接计算。设区域形状由轮廓点所组成的N多边形的$N+1$个顶点逆

图 3.4　从“dog jump”图像组的显著性物体区域中所学到的全局颜色先验 (以样本色彩的直方图方式呈现)。左边部分显示了典型的前景颜色。而右边的部分则显示了典型的背景颜色。在这个直方图中，概率值$\{p\}$按照降序排列，直方图的高度为$|p-0.5|$。为了显示更加清晰，本方法略去了概率值接近0.5的颜色样本。直方图中间的插图显示了一个典型的样本。

时针方向定义，即 $(x_0,y_0),(x_1,y_1),\cdots,(x_N,y_N)$，且 $(x_N,y_N)=(x_0,y_0)$。则有

$$P=\sum_{n=0}^{N-1}\sqrt{(x_{n+1}-x_n)^2+(y_{n+1}-y_n)^2} \tag{3-2}$$

$$A=\sum_{n=0}^{N-1}\sqrt{x_{n+1}y_n-x_ny_{n+1}} \tag{3-3}$$

本方法依照瀑布模型的样式使用这些形状度量。对于每一个度量，作者以该度量下提取的显著性区域与用户草图形状的相似性对显著性区域进行降序排列，并仅保留排序靠前的候选项。为了更高效地在早期排除极不相似的候选项，作者按照从简单到复杂的顺序使用这些形状相似性度量。在实验中，作者按照圆度、实心率和傅里叶描述子，依次仅保留了前 $T_C=80\%$、$T_S=80\%$和 $T_F=70\%$的图像。这些描述子的长度依次为1、1和15。在形状比较中，作者简单地使用这些描述子与用户输入草图的相关特征的欧式距离。最终，作者使用以上描述子中最复杂却最有效的Shape Context描述子对剩余结果进行排序。虽然用户也可以选择使用其它更为复杂的描述子(例如[24])，作者发现以上组合已经能够有效地移除绝大部分异常的形状。请注意，经过这个阶段后，作者剩余的图像数目为初始数目的 $T_R\cdot T_P\cdot T_B\cdot T_C\cdot T_S\cdot T_F\approx 20.0\%$。这些形状被进一步用于表观一致性的学习。

3.4.2　全局先验的统计模型

经过瀑布模型过滤之后，排序靠前的图像通常具有很好的准确率。本方法用前50张图像作为一个高质量的训练集来学习全局表观先验，并用以指导后续的群组显著性检测与分割。本方法选择采用高斯混合模型来刻画这种先验$\bar{G}$主要是基于两个方面的考虑： (i) 高斯混合模型在较小数据集上的训练结果比直方图模型

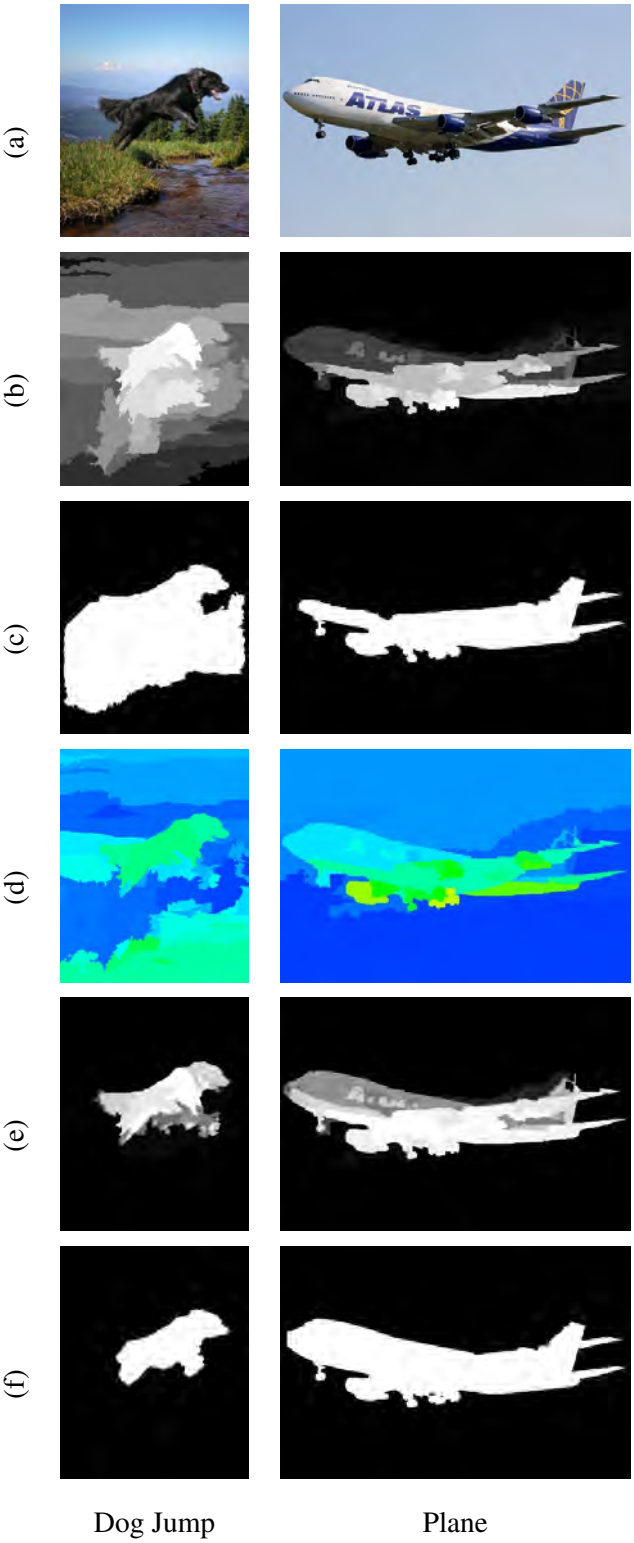


图 3.5　利用先验统计信息改进无监督分割结果的例子： (a) 输入图像， (b) 单张图像的显著性图， (c) 显著性区域分割(用 式 (3-1))， (d) 全局颜色先验 (采用伪彩色显示，暖色调代表更有可能是前景，冷色调代表更有可能是背景) , (e) 群组显著性图 (f) 群组显著性分割(用 式 (3-4))。

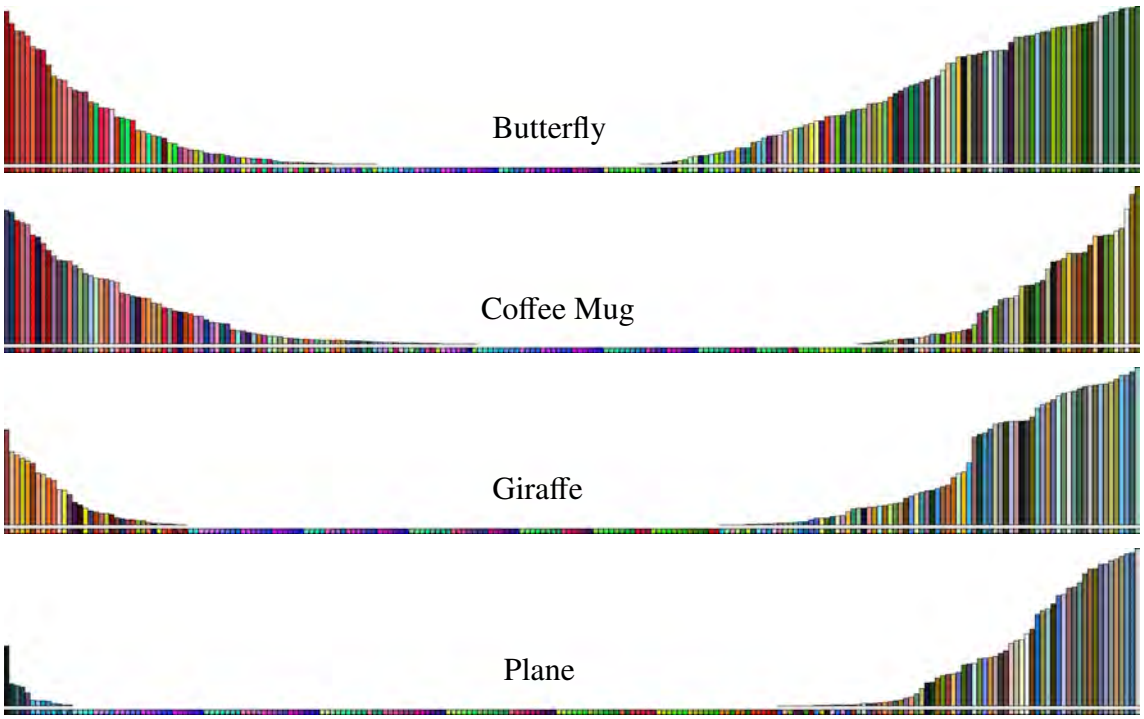


图 3.6　与图 3.3相似的另外四组图像的全局表观先验。这里的先验信息和人们对这些类别的直观理解相一致。例如，Giraffe的通常前景是土黄色的(统计结果的左侧)；其经常出现在蓝色的天空(从近处拍长颈鹿通常只能仰视，因而拍到天空) 或者绿色的树木(从远处拍到长颈鹿吃树叶的通常情景)背景之下。

的结果可推广性更好； (ii) 高斯混合模型可以很容易地集成到作者的无监督分割框架中(见第 3.3节)。例如，图 3.4展示了“Dog Jump”例子中的前背景高斯混合模型。该结果表明，狗通常是黄色或者暗色，而它们喜欢在绿色的草地上玩耍。虽然也可以得到纹理或者其它视觉特征方面的表观先验，为了简单起见，本方法目前只使用了颜色信息。 图 3.6展示了更多类别的例子。

由于形状特征和表观属性通常无关，利用形状的检索结果很好地保持了图像集中样本的多样性，所以更适合被用来学习代表性表观模型。通常情况下这些图像中只有一部分(在作者的测试中通常是15% – 57%)包含目标物体。这些物体可能具有不同的颜色和纹理。即使是某个单独的物体，也有可能是由许多表观特征差异很大的区域组成(例如蝴蝶)。仅仅考虑使用最大的表观聚类[127]或者是排序靠前的网络图像[126]作为初始集合是不合适的。在另外一个相关的工作中，Chang 等人[70] 使用图像之间的重复性作为多个图像的全局先验。该方法假设绝大部分图像都包含至少一部分前景物体。这个假设对于网络图像来说并不现实。而且，对于每个图像在计算全局项的过程中都需要考虑其它所有图像。当扩展到大规模图像检索应用中的时候，这种计算开销显然是无法承受的 (例如，他们的例子中最大的规模是30，而本方法要处理数千张图像)。

3.4.3　基于全局先验的视觉显著性区域检测与分割

最终，本章采用学习到的全局表观统计信息来改进每幅图的显著性区域检测与分割。由于估算得到的全局颜色先验$\bar{G}$是以高斯混合模型形式表达的，作者可以直接在单张图像无监督分割能量方程(式 (3-1))中加入一个全局先验约束。新的能量方程为

$$E(\alpha, G, \bar{G}, I) = \lambda U(\alpha, \bar{G}, I) + (1-\lambda)U(\alpha, G, I) + V(\alpha, I), \tag{3-4}$$

其中新增加的项$U(\alpha, \bar{G}, I)$度量了透明度分布α和全局颜色先验$\bar{G}$的匹配程度，而权重λ (在作者的实验中为0.3)用于控制全局颜色先验和单张图像颜色分布的比例。这里的全局颜色先验$\bar{G}$反映了目标物体之间的相似性(*similarities*)。每张图像的颜色分布G是由单个图像的显著性图分割统计得到的。G度量了在图像内部的特异性(*distinctness*)。

作者按照与优化式 (3-1)相同的方式优化式 (3-4)以获取群组显著性分割的结果。所得到的感兴趣区域分割结果和图像中的其余区域被分别用于训练感兴趣区域和背景区域的高斯混合模型。按照这个高斯混合模型计算出来的每个像素颜色属于感兴趣区域的概率值被作为其显著性值。

图 3.5展示了利用全局颜色先验改进显著性区域分割的典型例子。在“Dog Jump”的例子中(图 3.5d)，图像中的冷色调区域表示依照颜色先验，这部分更有可能属于背景区域。类似的，在“Plane”的例子中，缺失的物体区域在全局统计信息的帮助下被正确的恢复出来。

3.5　实验结果与讨论

作者用C++语言实现了本章中介绍的算法框架，并在一个具有两个Quad Core i7 920 CPU 和 6G 内存的个人电脑上进行了实验。如图 3.1所示，本方法利用基于

表 3.1　处理各组图像所需的运行时间。

	预处理时间	在线检索时间
Butterfly	31.5 min	0.98 s
Coffee mug	31.9 min	0.90 s
Dog jump	29.5 min	0.83 s
Giraffe	27.1 min	0.83 s
Plane	32.2 min	0.89 s

表 3.2　THUR15000数据集的一些统计信息，其中包括各个类别的图像数目、含有目标物体的图像数目及比例。

	图像数目	含目标物体的数目	含目标物体的比例
Butterfly	3457	1242	35.9%
Coffee Mug	3282	1884	57.4%
Dog Jump	2893	1614	55.8%
Giraffe	2933	461	15.7%
Plane	2970	1032	34.7%

群组显著性的检索结果重新训练新的表观特征统计模型，并迭代地改善显著性分割。实验中，作者发现两轮的迭代通常就已经足够了。

本方法首先对数据库中的每一个图像进行基于单张图像的无监督图像分割[107] (大概每分钟能处理100张图像左右)。对于数据库中的每一类(某个关键词对应的一组图像)，作者预先用代表性的草图形状初始化一个好的表观学习模型，并基于上述迭代方式为每个图像计算群组显著性分割结果。在执行用户检索的时候，本方法仅需对用户输入草图和数据库中的显著性物体形状进行基于瀑布模型的形状检索即可(见第 3.4.1节)。处理THUR15000数据集(见第 3.5.1节）中各组图像集合所需的预处理和检索时间情况见表 3.1。虽然提取群组显著性区域的过程比较耗时，但是这部分可以离线进行。这种机制从一个包含3000张图像的初始集合中检索一次大约需要1秒钟时间，因而在线检索部分的时间消耗能够满足交互式检索的需求。对于更大的数据集，作者建议采用更加快速的形状检索机制[123]。

作者用第 3.5.1节中将要介绍的基准数据集对本章中所提的方法进行了验证。该验证包含以下几个应用下本算法的表现：

(1) 群组显著性图的固定阈值分割；

(2) 感兴趣区域的自动分割；

(3) 基于草图的图像检索(Sketch Based Image Retrieval, SBIR)。

对于前两个应用，作者仅考虑包含目标物体的图像的平均分割性能 (哪些图像包含目标物体可从基准数据集的标注中获知)。

3.5.1　测试数据集THUR15000

由于本方法是第一个试图对大量复杂的网络图像提取群组显著性区域的工作，目前没有合适的公开测试集。为了全面地测试本章算法在处理大量复杂的网络图像时的性能，作者利用关键字从Flickr上下载并收集了大约15000张图像

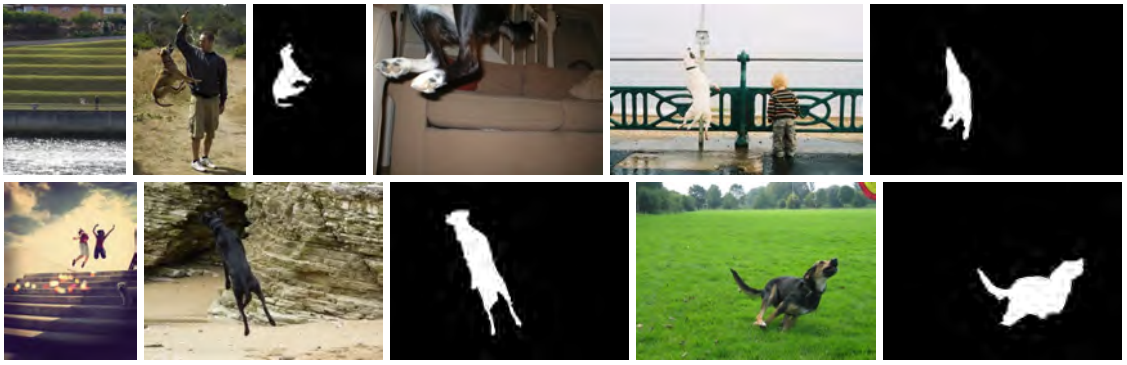


图 3.7　基准数据集中关键词“Dog Jump”对应的图像示例。对于含有感兴趣区域的图像(这里的7张图中有4张有)，其像素精度的标定数据在相应图像的右侧展示。

①，并参考THUS10000数据集(见第 2.6.1.2节)，为所有含关键字对应目标物体的图像进行像素精度的显著性物体区域标注。作者将这个用于验证群组显著性区域检测与图像检索应用性能的基准数据集命名为THUR15000 (Tsing Hua University Retrieval dataset with 15000 images)。在数据收集过程中，作者为以下关键字分别下载了大约3000张图像：“Butterfly”，“Coffee Mug”，“Dog Jump”，“Giraffe”，和“Plane”。由于网络图像的质量参差不齐，数据集中的有些图像并不包含对应关键字匹配的显著性物体。对于这类图像，作者不进行显著性区域标注。此外，由于被部分遮挡的物体在形状匹配时并不鲁棒，作者仅对绝大部分未被遮挡的物体进行标注。 图 3.7展示了THUR15000数据集中图像及对应的标注的部分示例。在这5类图像中，每个图像集合仅仅有一部分(15% – 57%)包含目标物体。各个类别包含原图、目标物体的数目及比例的统计信息见表 3.2。

Eitz 等人 [24] 曾经建立了一个用于验证基于草图的形状检索系统的基准数据集。该数据集中包含了给定的草图与图像对(sketch image pair)之间的匹配程度。在作者的基准数据集中不仅包含图像中是否有目标物体的信息，而且包含了关于目标物体的像素精度的区域信息。因而，作者的基准数据集能够验证相应的区域分割算法。

3.5.2　群组显著性图的固定阈值分割

参考Achanta等人[8]对显著性图的验证方法，作者用阈值$T \in [0, 255]$对群组显著性图进行分割，并用分割结果和用户标注的基准结果进行比较，以验证该显著性检测算法结果的正确性。 图 3.8中的正确率召回率曲线显示，本章中的算法在两次迭代之后就基本收敛, 而且群组显著性检测结果明显优于现有最先进的单个图像显著性检测算法[107]。

① 同样具有像素精度显著性区域标注的现有最大的公开测试集包含1000张图像。

表 3.3 前50和前100个检索结果中的Flickr结果、本文方法的结果和SHoG方法[24]的结果中包含的正确检索结果的数目。与其余方法仅包含是否有物体的信息不同，本文的方法同时知道物体的具体区域。在这个比较中，本文方法的结果分为两种评价标准：(i) 包含目标物体则算正确； (ii) 包含目标物体且其具体区域基本被正确的找到，即$F_\beta > 0.7$。

		Flickr	本文的结果		SHoG方法[24]的结果
			$F_\beta > 0.7$	包含目标物体	
Butterfly	前50	14	19	27	18
	前100	28	36	50	40
Coffee mug	前50	29	46	47	41
	前100	51	88	90	78
Dog jump	前50	28	42	46	37
	前100	55	78	85	73
Giraffe	前50	15	29	32	9
	前100	25	48	51	18
Plane	前50	22	41	47	45
	前100	48	70	93	91

3.5.3 群组图像的感兴趣物体分割

另外，作者也验证了本章算法从质量参差不齐的网络图像中提取目标物体的正确性。对于包含目标物体的图像，作者将群组显著性分割结果与THUR15000中像素级别的区域标注进行准确率、召回率、F_β度量等方面的比较。其中F_β度量定义为：

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}. \tag{3-5}$$

在显著性区域分割应用中，由于可以简单地通过将所有图像区域判定为显著性区域就能轻松地达到 100%的召回率，所以正确率比召回率要更为重要[8,107]。在网络图像检索应用中，由于几个错误的检测结果比数千个结果中漏掉几个正确的结果要糟糕的多，准确率相对召回率的重要性要更为明显。为了公平起见，作者在与现有最先进的单张图像显著性区域提取算法[107] 进行比较的过程中，也采用$\beta^2 = 0.3$，让正确率的重要性比召回率更高[8,107]。如图 3.9所示，群组显著性区域分割结果的F_β明显高于单张图像的显著性区域提取结果。

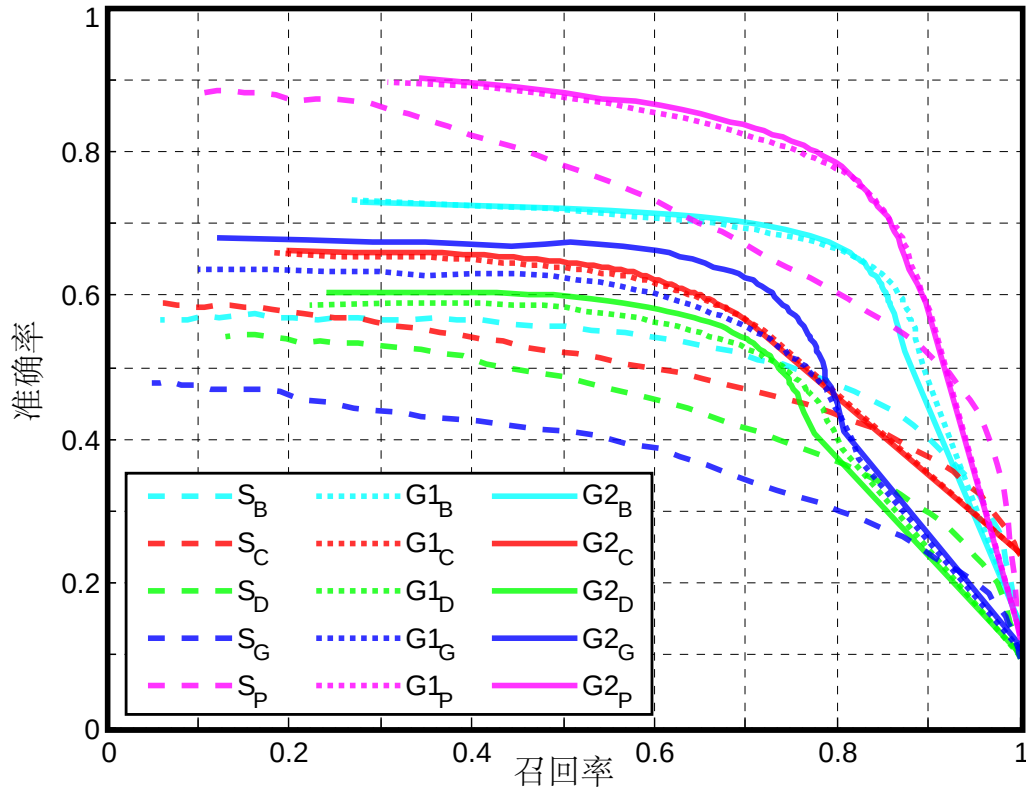


图 3.8　对显著性图进行固定阈值分割得到的正确率-召回率曲线。 S、 G1 和 G2 分别表示单张图像的显著性图、经过第一次和第二次迭代之后的群组显著性图。后缀B、 C、 D、 G 和 P 分别表示“Butterfly”，“Coffee Mug”，“Dog Jump”，“Giraffe”和“Plane”。

3.5.4　基于草图的图像检索

作者将本章的检索方法与Eitz等人[24]提出的现有最先进的基于草图的检索方法进行了比较 (比较中用了该作者自己的算法实现)。由于本章提出的算法显式地从图像中提取感兴趣物体区域，因而可以有效地利用现有形状比较技术。对于质量参差不齐的网络图像，群组显著性分割和形状比较的结合有效地选出了包含目标物体的图像。为了验证算法在更多类别图像上进行基于草图的图像检索的性能，作者利用另外的25组关键字从 Flickr上自动下载了更多的图像，并将这部分图像和THUR15000数据集中的5组图像一起用于草图检索应用中算法的验证。表 3.3和表 3.4展示了定量的比较结果。该结果显示，本章方法检索结果的正确率优于Eitz等人[24]的结果。 图 3.10-3.12给出这30组图像的检索结果与Flickr结果以及SHoG方法[24]结果的比较示例。本方法的结果除了具有是否含有目标物体的判定以外，也包含了目标物体的具体区域。这种区域信息为很多需要利用目标物体的相关像素信息的应用[15–18,67,68] 提供了有力的支持。

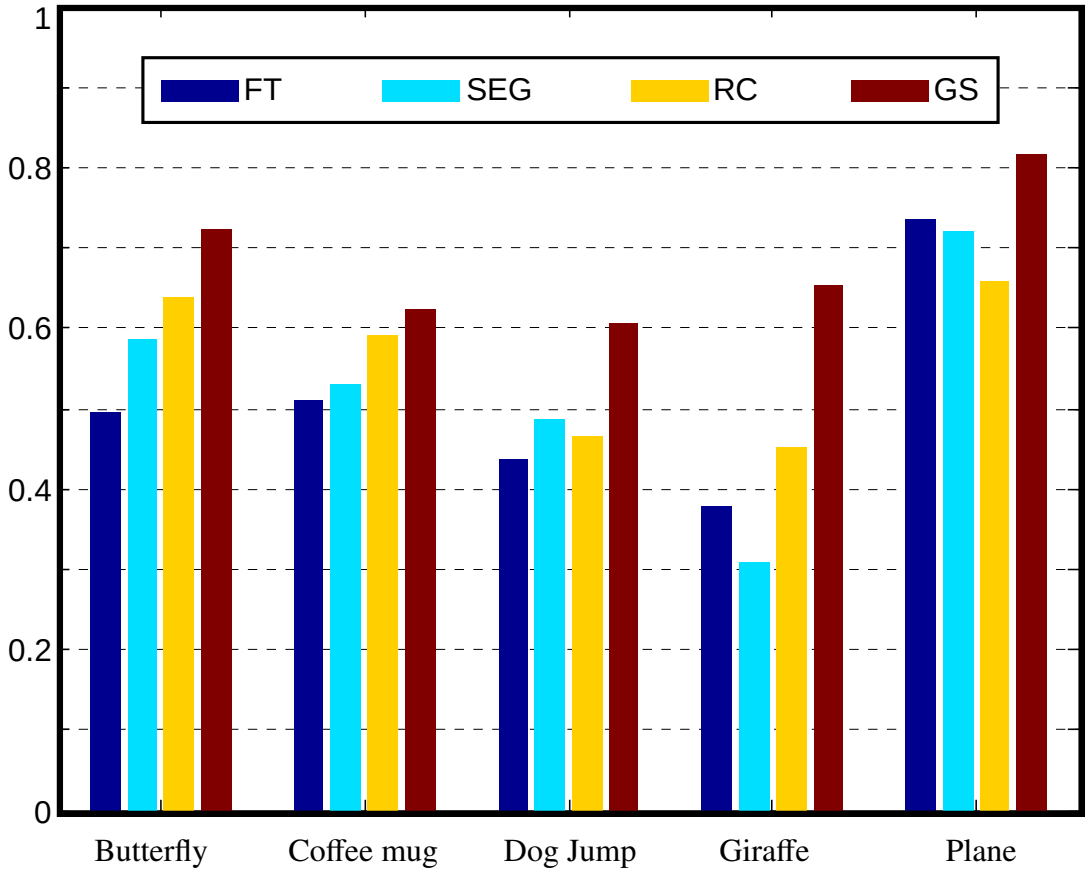


图 3.9　在F_β度量下几种现有最先进的单张图像显著性区域检测算法 (FT[8], SEG[52], RC[107]) 与群组显著性算法的结果的比较。F_β 越高表示性能越好。

3.6　本章小结

本章提出了一种利用相关、低质的网络集合(类别)中图像之间的公共显著性区域信息来改进图像显著性物体区域检测与分割结果的方法。利用简单的初始形状，本算法自动地挑选出高质量的前背景区域样本并学习相应类别的表观模型。这种表观模型继而被用于改进显著性区域的自动分割。作者用5组关键字自动地从 Flickr (http://www.flickr.com/) 上搜索并下载的图像及这些图像相应的像素精度显著性区域标注建立了一个基准数据集。在该数据集上的实验结果显示，本章算法能够对复杂的网络图像产生高质量的显著性区域检测与分割结果。这种分割结果使得背景区域的干扰被排除，因而能够支持基于草图的图像检索应用。实验结果表明，基于该分割结果的形状检索结果也优于现有最先进的基于草图的检索系统，同时提供额外的物体区域信息以支持更广泛的应用需求。

在将来的工作中，学习形状[132]和纹理特征有望进一步改进本章方法的自动分割结果。如何利用更加高效的形状比较和索引策略[133,134] 也是一个重要的研究方向。

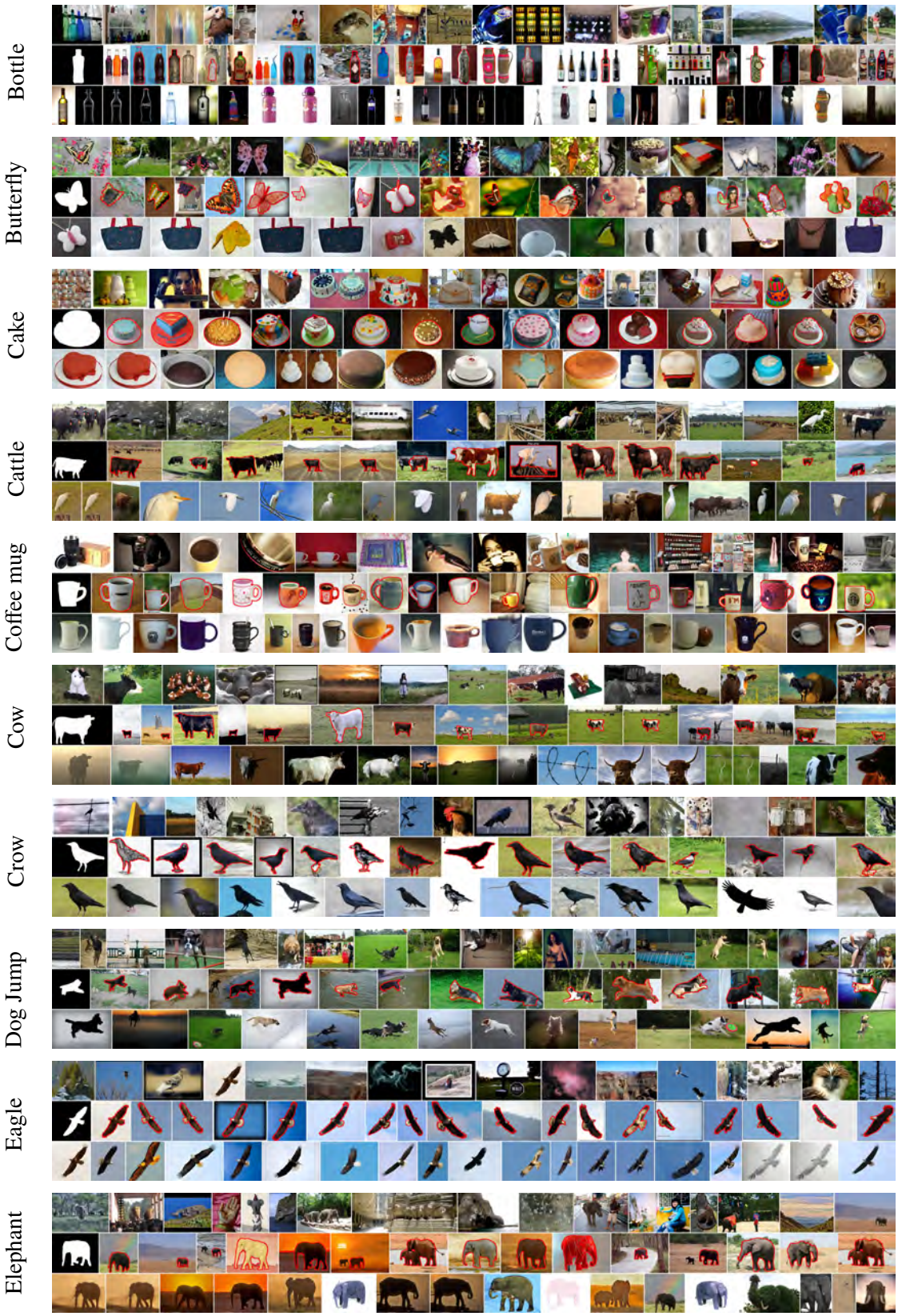


图 3.10　基于草图的图像检索结果比较示例。每组结果中，从上到下的三行分别表示：Flickr结果、本文的结果和SHoG方法[24]的结果。

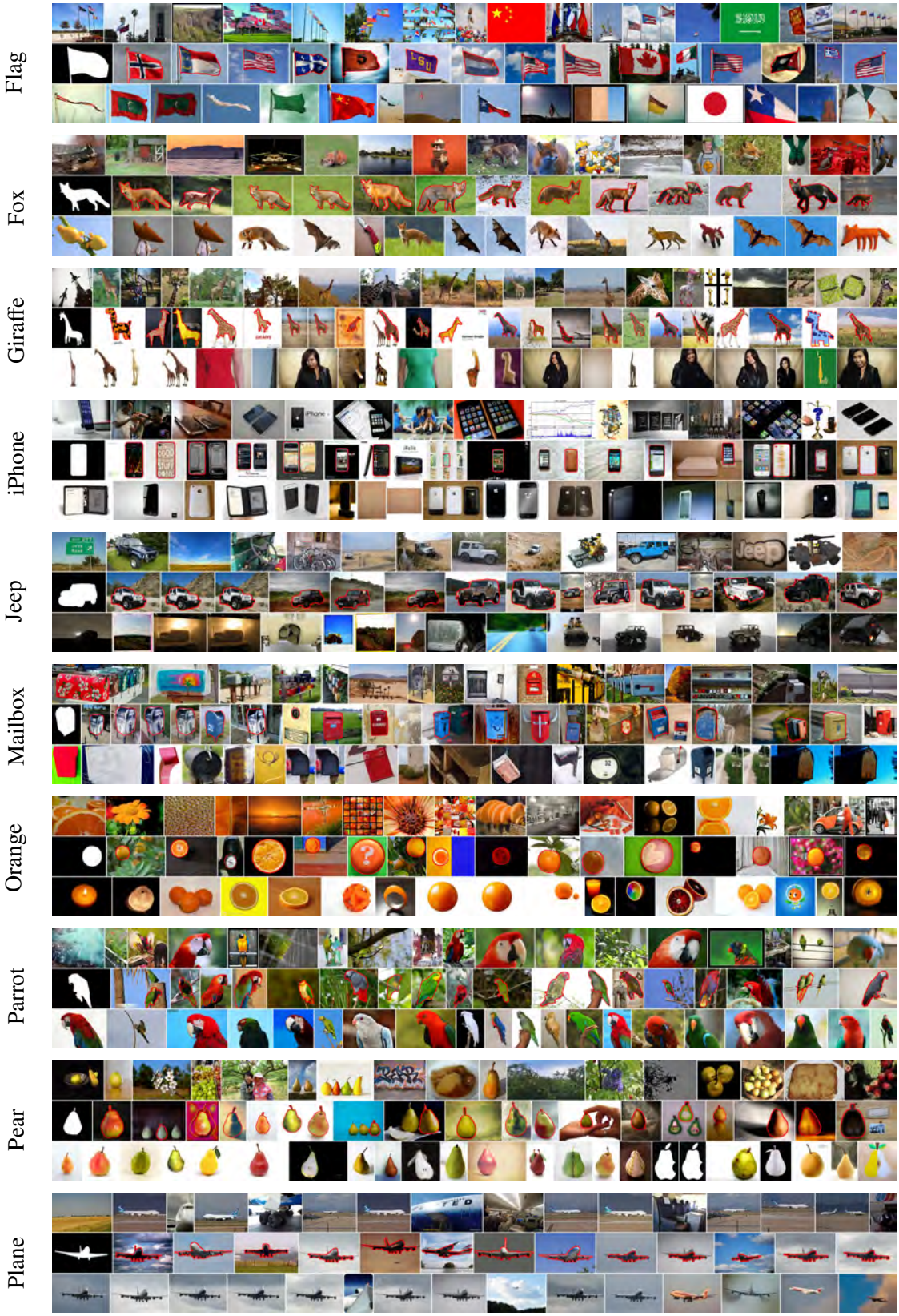


图 3.11　基于草图的图像检索结果比较示例。每组结果中，从上到下的三行分别表示：Flickr结果、本文的结果和SHoG方法[24]的结果。

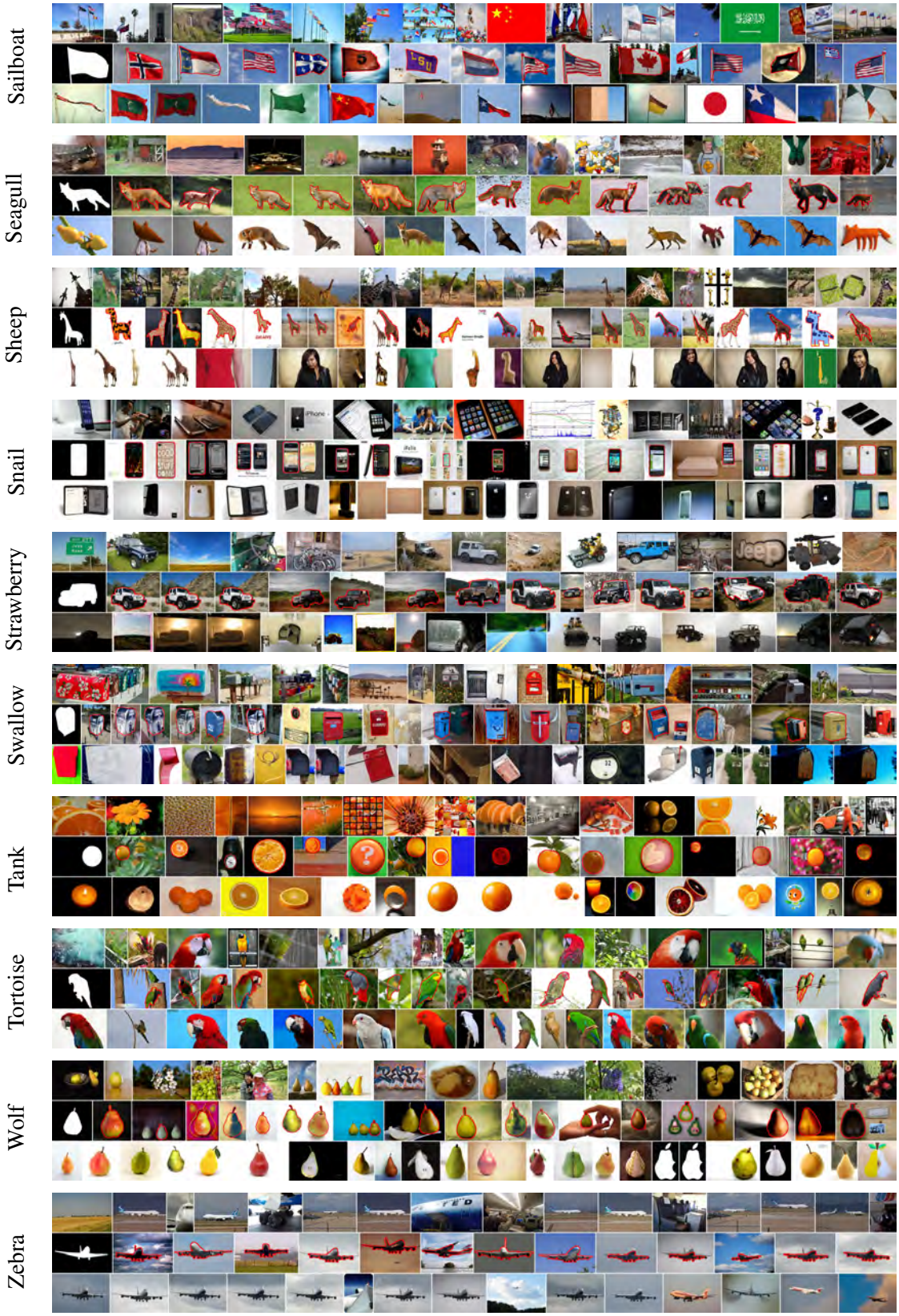


图 3.12　基于草图的图像检索结果比较示例。每组结果中，从上到下的三行分别表示：Flickr结果、本文的结果和SHoG方法[24]的结果。

表 3.4　更多类别的前50和前100个检索结果中，Flickr结果、本章方法结果和SHoG方法[24]的结果中包含的正确检索结果的数目。

	前50个			前100个		
	Flickr	本文的结果	SHoG[24]	Flickr	本文的结果	SHoG[24]
Bottle	29	49	42	60	94	82
Cake	28	42	39	55	81	79
Cattle	24	47	7	37	81	12
Cow	26	47	27	54	92	55
Crow	18	48	38	38	86	61
Eagle	9	47	41	23	95	80
Elephant	20	48	38	38	87	68
Flag	28	48	27	52	91	57
Fox	16	48	19	37	80	42
iPhone	24	48	34	43	90	54
Jeep	27	47	19	59	93	43
Mailbox	26	43	34	52	87	60
Orange	12	35	35	19	62	52
Parrot	29	45	30	56	79	57
Pear	32	42	41	63	82	81
Sailboat	36	43	28	67	81	54
Seagull	28	46	44	59	90	86
Sheep	27	39	22	50	81	40
Snail	31	47	30	63	87	61
Strawberry	25	35	38	53	69	67
Swallow	12	38	37	31	66	68
Tank	24	44	28	44	84	45
Tortoise	24	31	23	48	62	45
Wolf	12	25	15	25	55	29
Zebra	15	33	14	25	56	31
平均准确率	46.6%	85.2%	60.0%	46.0%	80.4%	56.3%

第 4 章　基于相似结构分析的图像编辑

本章主要研究基于相似结构分析的图像编辑方法。与前两章中仅仅分析感兴趣物体区域信息不同，本章主要讨论在交互式图像编辑应用中，如何通过简单的用户交互指导，获取和利用相似物体的区域、层次、遮挡、对应等信息，对含有相似物体的图像进行场景物体级别的图像编辑。 第 4.1节给出本章的研究动机和交互式相似物体检测与分析算法框架； 第 4.2节给出与本章方法紧密联系的相关工作概述； 第 4.3节和第 4.4节分别介绍基于轮廓带图的相似物体高效检测算法和图像中相似物体结构的分析算法。 第 4.5节展示了基于相似物体检测分析结果的多种场景物体级别图像编辑应用。 第 4.6节对全章进行总结。

4.1　引言

4.1.1　研究动机

图像是现实生活中的客观场景的一种数字化描述。传统图像编辑手段大多仅从像素和区块级别对图像进行编辑，忽略了图像的潜在场景语义。基于底层元素的操作方式导致在图像的编辑过程中，用户难以将自己思维理解中的编辑意图和对底层元素的操作对应起来。如何自动的对图形中的场景进行分析，提取出符合人类感知和理解规律的物体级别场景信息，是一个有意义的研究课题。

为了能够让用户可以像操纵现实世界中物体一样对图像进行直观、方便的编辑，首先需要以下几种重要信息：有意义的物体区域、这些物体的层次、对应、遮挡等相互关系。相对于从任意的图像中提取这些信息通常所具有的困难性，含有相似结构物体的图像通常含有关于场景物体潜在位置、区域、层次、对应等关系的更多信息。同时，相似结构在自然和人造场景中普遍存在。由于遮挡、部分缺失、物体间形变、光照变化等因素的影响，要想编辑这样的图像并保持其中的相似结构之间的关系不被破坏非常困难。通过手工的方式保留这种相互关系非常费时，而且容易出错。本章的主要研究目的是以含有相似结构的图像为入手点，研究基于场景物体级别高层语义信息的图像编辑方法。

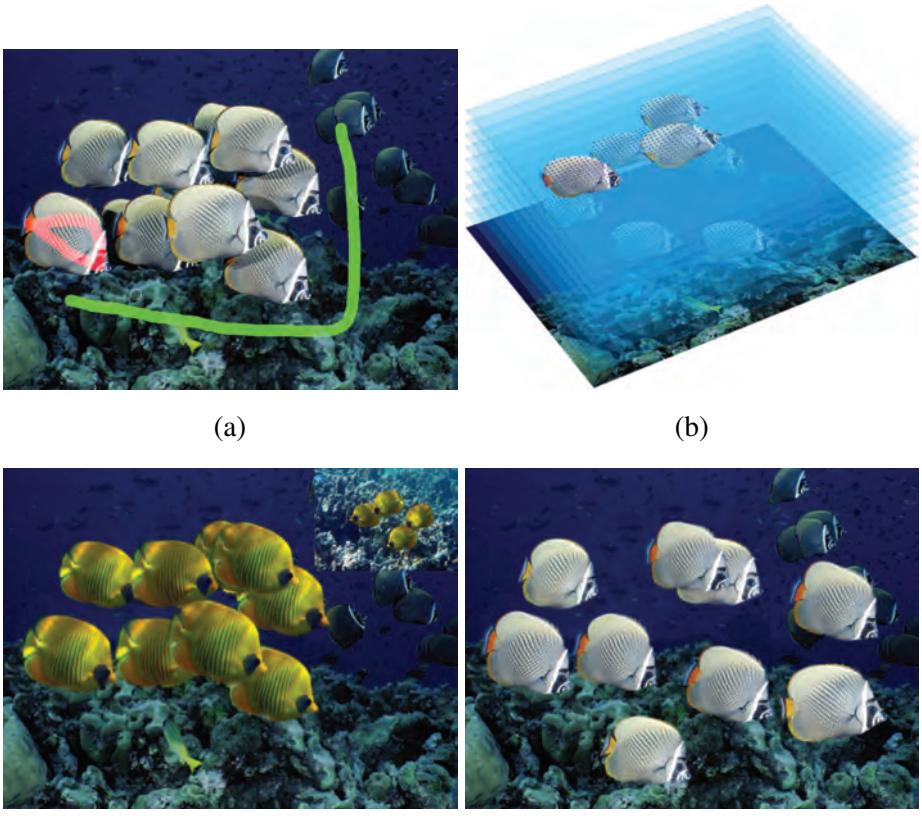
4.1.2　交互式相似物体检测与分析算法框架

本章提出了一种可以有效地提取图像中相似结构的算法框架。通过这种分析，本系统可以方便地对这些相似结构进行同步的编辑。 图 4.2展示了本章系统

的一个流程图。即使相似结构在形状上有一定的变形和遮挡，本系统仍然能够利用用户输入笔刷标定的指导，有效地检测出这些相似元素。该系统的核心是一种轮廓带图(Boundary Band Map, BBM)匹配算法(见图 4.3 和第 4.3节)。作者进而利用主动轮廓模型方法(Active Contour Method)和形状先验来改进检测结果与真实图像的一致性(第 4.4.1节)。然后，本章依据相似结构之间两两的深度关系，并用拓扑排序方法来诱导并建立所有相似结构之间的深度顺序 (见第 4.4.2节)。最终，本系统为相似结构进行稠密对应关系估计和遮挡部分补全。

采用基于轮廓的方法来匹配相似结构，而非特征点方法，是本章系统设计中一个重要的设计决策。作者发现，由于图像中的相似结构通常包含很多相近的特征点，所以基于SIFT点的匹配中有太多的二义性。这种二义性很难在后续的过程中得以解决。

本章的新系统可以自动地检测图像的指定区域中相似物体之间的相互关系，并建立场景物体级别的图像结构。这种场景结构从一个侧面部分地刻画了潜在的场景语义，使得场景物体级别图像编辑操作成为可能。以此为基础的编辑操作，



(a)　(b)　(c)　(d)

图 4.1　相似结构的检测与编辑示例： (a) 原始图像以及用户粗略标注的模板(红色)和背景(绿色)区域； (b) 检测出并对遮挡部分进行补全后的相似结构； (c) 原图中的鱼被用参考图(右上角的内嵌图)中的另外一种鱼替换； (d) 重排之后的鱼。

考虑了图像组成部分之间相互的几何和层次关系。这种强大直观的编辑过程通过现有其它途径实现的时候非常费时费力 (见第 4.5节)。

4.2　相关工作

4.2.1　近似单元提取

相似物体在自然和人工场景中广泛存在，对这种相似结构的分析在图像场景分析[83,84]、智能编辑[85,86]、和内容敏感图像缩放[87] 等领域都有着重要应用。如何检测并分析这些相似的单元及其相互关系一直是计算机视觉、几何处理和计算对称性分析等领域的重要问题。

加州大学伯克利分校的 Thomas Leung 和 Jitendra Malik [88]提出了一种基于图结构的方法。这种图结构以图像单元作为节点，并以它们之间的仿射变换为边。该方法利用这个图结构对节点单元进行增长和组合来提取显著的相似单元，从而显著地改进了重复结构的检测。

卡内基梅隆大学的Yanxi Liu等人[89]提出了一种基于晶体组数学理论的周期性模式感知计算模型。该理论认为，N维欧式空间中有限数量的对称组可以刻画周

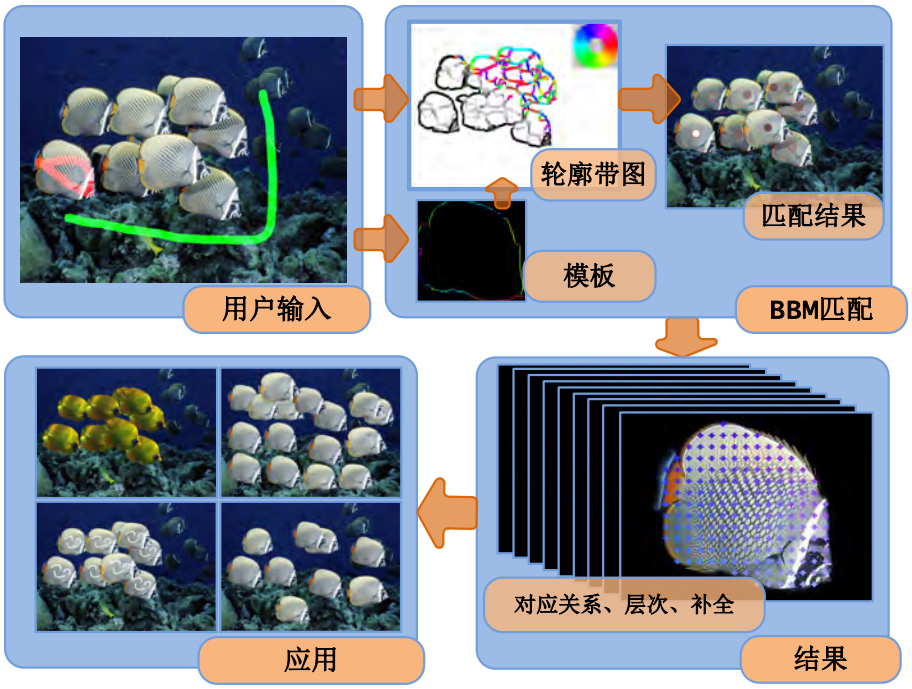


图 4.2　图像中相似结构分析与编辑流程图. 用户粗略的标出一个物体模板(红色部分)和背景区域(绿色部分)。本章的轮廓带图算法可以检测到其它近似的物体。通过进一步建立这些相似物体之间的稠密对应关系，本方法对相似部分进行补全，并从一定程度上建立这些物体之间的深度顺序。这些提取出来的图像结构使得一系列直观、高效的图像编辑应用成为可能。

期性模式无限的多样性。在二维空间中，存在描述一维方向上单色模式的7种带状组和描述两个线性独立方向上的17 种壁纸组模式。该方法通过一系列的计算过程来捕获这种模式，并自动地理解给定的周期性图像输入。

图像中经常包含不规则分布的相似单元，对于具有一定遮挡关系的不规则分布相似单元，以上方法均不能有效地处理。

伊利诺大学香槟分校的Narendra Ahuja 和 Sinisa Todorovic[90] 共同提出了一种纹理单元提取算法。该方法首先建立了一种表示整个图像分割结构的树形表达，然后通过在这个表达中寻找相似子树的方式提取自然的纹理元素。虽然该方法可以自动地从具有一定遮挡关系的复杂场景中检测到近似的单元，但即使是对中等大小的图片，该方法也需要数十秒的计算时间，不能满足交互式的图像编辑应用需求。

特征点经常被用于建立相似物体之间的对应关系，并被进一步用于物体检测、识别等应用。加拿大英属哥伦比亚大学的 David G. Lowe[91]提出的SIFT特征点和苏黎世理工学院的Herbert Bay等人[92]提出的SURF特征点是应用最为广泛的特征点匹配技术。这些特征点能够为同一物体在不同视角和光照条件下所得到的图像间建立联系，因而被广泛地用于物体检测任务中。加州大学伯克利分校的Alexander C. Berg 等人[93] 提出了一种基于局部特征点匹配和全局几何形变建模的形状匹配算法。该算法在整数规划(integer programming)框架下对两个形状之间的稠密对应关系进行有效求解。然而，这些描述子仍然只能为局部图像区域间建立联系，而不能刻画高层的场景结构。

由于遮挡关系、光照变化、物体间形状差异等因素的影响，从自然图像中自动、高效地提取相似物体单元仍然存在一定困难。为了满足对这类图像的高效编辑需求，作者通过简单的用户输入来简化这一问题，并提出了一种同时包含全局几何信息、局部几何信息和轮廓点置信度的数据结构来高效地提取相似物体单元，并进一步用于多种物体级别图像实时编辑应用。

4.2.2　图像编辑

出色的图像编辑工具需要具有快速、可交互、直观等特性。理想情况下，用户应该能够利用少量不精确的笔刷来无缝地进行期望的编辑和操作。这种工具开发过程中一个基本的挑战在于很难从图像中获取场景的语义或者潜在的物体间关系。即使是关于图像部分之间基本的分割和链接关系就可以被用来进行直观的形状敏感几何编辑 (例如：图像部分的变形[135–137])。作为语义表达的另一个选择，Lalonde等人[138]提出了Photo Clip Art技术。该技术利用数据驱动的方法为图像中插入元素，通过这种从其它图像中拷贝相关的部分的方式以取得以假乱真的

效果。然而，利用该方法很难对目标物体进行精细的控制，也不可能编辑得到图像库中未曾出现的新异内容。

An 和 Pellacini 提出了一种 AppProp [139] 方法。该方法通过能量最小化的方式将用户对某个区域的编辑扩散到其它外观相似的区域。之后，Xu等人[140]通过KD Tree对AppProp方法进行了加速。然而，这些方法并不考虑潜在的几何和重复关系。 Hoiem等人[141]从2D的图像中推断深度信息，并用该深度信息生成有趣的图像内容立体化效果。该方法针对室外场景，并要求用户对室外场景的每个区域标注地面、水平线或者天空。近期出现的 PatchMatch 方法[30] 利用随机通讯方法近似计算最近邻场。这种最近邻场的快速计算使得一系列重要的图像编辑应用可以实时进行。这些应用包括图像缩放、补全以及重排。 Landes 和 Soler[142] 通过分析纹理中的层次结构来改进合成过程。然而，以上这些工作都不能在图像场景中近似的物体单元之间进行编辑传播。

4.2.3　物体检测

基于窗口关联分析(window based correlation analysis)的图像分割利用颜色或者纹理信息对图像进行分割。在含有遮挡、部分缺失或者光照变化的情况下，这类方法可能会变得不稳定。很多基于轮廓的物体检测方法已经被提出来，其中包括Belongie等人的[71]形状上下文(Shape Context)方法。一个形状上任何一点处的形状上下文可以通过其邻近点相对极坐标的直方图来计算，并通过最小化每一组匹配代价的总和以求解出全局最优的关联。然而，这种离散直方图的创建过程并不适合杂乱的图像场景。在受到严重数据缺失的影响下，这种方法的表现可能还不如一些更简单的方法。 Thayananthan等人[144]提出了一种利用Chamfer匹配(chamfer matching) 来处理杂乱图像场景中的形状匹配问题。

4.3　基于轮廓带图的相似物体高效检测算法

受到相互遮挡、部分缺失、摄像投影、光照变化、物体间形变等多种因素的影响，图像中相似的场景物体通常在表观(appearance)上有所不同。任何一种试图提取这种近似物体单元的方法都需要排除表观差异的因素，分割出图像中的近似单元并建立它们之间的关联。这种分割以及关联的建立一直是计算机视觉中的一个很具挑战性的问题[90]。即使是现有最先进的方法也不能满足交互式图像编辑的速度需求，并在检测重复单元的过程中经常失败，在含有较大表观差异及遮挡的情况下这种不足更为明显。

本章提出了一种基于**轮廓带图**(*Boundary Band Map, BBM*)的方法来利用简单

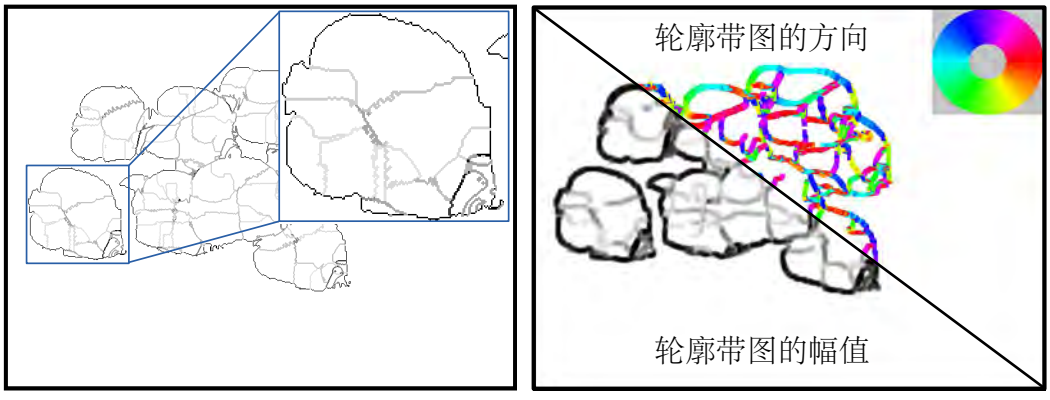
的用户笔刷标注，检测并分割出相似物体单元。在本节中，作者首先将对这种包含候选物体全局和局部几何信息的轮廓带图表达进行详细介绍，接着将介绍如何将物体模板与轮廓带图进行匹配。这种匹配过程可以通过快速傅里叶变换高效地计算。

4.3.1　轮廓带图的构建

轮廓带图$M = \{m_p\}_{H\times W}$是一个和大小为$H \times W$的图像对应的二维向量场。在轮廓带图中，每一个像素p都和一个包含该像素点附近局部几何信息的二维向量m_p相对应。更具体的说，这个二维向量就是边缘图中的局部方向和强度信息。作为一个例子， 图 4.3b 展示了图 4.1对应的轮廓带图。

轮廓带图是依据轮廓图计算得到的。为了得到轮廓图，本系统要求用户在原始图像上绘制少量的简单笔刷来表示一个物体模板和背景区域(如图 4.1)。根据这些用户笔刷，本系统采用基于Graph-Cut的方法[33,63] 来获取物体模板并分割出整个前景区域。继而，利用层次化Mean-Shift分割技术[34] 以获取前景区域的轮廓图(见 图 4.3a)。采用这种层次化分割的主要原因是为了避免检测结果过多地受到纹理边缘的影响。同时，层次化分割结果包含了检测出的轮廓的置信度(或者说强度)信息(见 图 4.3a-b).

为了适应模板和其它相似物体间可能的形状变化，本算法将初始的轮廓向外扩展一个宽度b。为了产生轮廓带图$M = \{m_p\}_{H\times W}$，本算法将每个轮廓点p对应的m_p赋值为轮廓图中像素p处的切向量。这个向量的幅值和方向分别代表p点处的轮廓强度和局部轮廓方向。轮廓宽度为b范围内的其它点处的向量根据算法 1 进



(a) 轮廓图像　　(b) 轮廓带图

图 4.3　基于轮廓带图 (Boundary Band Map, BBM) 算法的相似场景物体检测： (a) 通过前景区域层次分割得到的边缘图像, (b) 由边缘图像导出的轮廓带图。该二维向量场的方向用彩色表示。其中颜色与方向的对应见子图b的右上角

行赋值：该点处的轮廓带图向量即为其相邻的已赋值点处的向量的平均值。在计算向量和以及平均值的过程中，相反方向的向量事实上意味着同样的潜在局部几何信息。因此，本算法并不是简单地对相邻向量计算总和，而是考虑每个向量的两个版本 — 这个向量本身及其反向量。在计算局部和的过程中，本算法使用这两个版本中与其邻域更一致的那个。对于含有多种不同方向向量的局部邻域，其对应的总和和均值的幅值会比较小。由于这种区域处计算得到的轮廓方向本身具有奇异性并且不够可靠，这种抵消现象降低了这些区域在随后的匹配过程中的作用，因而对匹配过程是有利的。

算法 1 利用轮廓图像建立轮廓带图

初始化边界像素集合$B \leftarrow \{p\}$;
初始化队列$Q \leftarrow \emptyset$;
for 每一个边缘像素$p \in B$ **do**
    $m_p \leftarrow$ 边缘图在p点处的切向量;
    如果p的所有邻居点q满足$q \notin B$，则将p的所有邻居点q加入队列Q;
end for
初始化 $i \leftarrow |B| * b$, 其中$|B|$是集合B的大小, b是一个宽度参数;
while $i \neq 0$ 且 $Q \neq \emptyset$ **do**
    从队列中 Q 弹出 p;
    $m_p \leftarrow 0, num \leftarrow 0$;
    for 对于p的每一个已经初始化的邻居点q **do**
        if $|m_p + m_q| > |m_p - m_q|$ **then**
            $m_p \leftarrow m_p + m_q$;
        else
            $m_p \leftarrow m_p - m_q$;
        end if
        $num \leftarrow num + 1$;
    end for
    $m_p \leftarrow m_p/num$;
    将p的每一个尚未被初始化的邻居点加入队列Q;
    $i \leftarrow i - 1$;
end while
对于其余未被初始化的像素p, 令 $m_p \leftarrow 0$;

4.3.2　基于轮廓带图的快速匹配

为了将一个 $h \times w$ 的物体模板与 $H \times W$ 的图像进行匹配，本算法首先为模板建立一个向量场$T = \{t_p\}_{h \times w}$。模板轮廓点处的幅值被设置为1。其余点处的幅值被设置为0。非零点处的向量方向被设为该点处模板轮廓的局部方向。在原始图像上每一个可能的像素位置，本算法都计算模板的向量场T和该点处被模板覆盖的图像轮廓带图区域的向量场之间的相关系数。因此，图像中点(u, v)处模板T 和轮廓带图 M的匹配值为：

$$D_{(u,v)}(T, M) = \sum_{j=1}^{h} \sum_{i=1}^{w} (t_{(i,j)} \cdot m_{(i+u-1, j+v-1)})^2. \tag{4-1}$$

在每一点处，点积的平方和$(t_{(i,j)} \cdot m_{(i+u-1, j+v-1)})^2$ (简记为 $(t \cdot m)^2$)度量了模板的局部几何和相应的轮廓图的匹配程度。本方法用点积的平方和来解决向量与其反向量之间的二义性。

轮廓带图中包含了前景物体的全局几何信息。不难想象，如果一个物体和模板的形状很不相同，那么在匹配过程中，模板的很多部分将落在待检测物体的轮廓带以外。这种情况下，这些区域的匹配值为零，因而总体的匹配值也会降低。更进一步，本章提出的轮廓带图方法允许模板在轮廓图的一定宽度范围内的非刚体形状变化。

本算法受到了计算机视觉领域最新提出的 Shape Band[122] 方法的启发。Shape Band 表示了一定宽度范围内的可变形模板，模板与图像中候选相似物体之间的匹配仅在这个宽度范围内考虑。本方法选择扩展轮廓图而非模板形状主要是基于两点考虑。首先，如果对模板进行扩展，模板一定宽度范围内不相关的轮廓会导致匹配过程中不正确的响应。不相关的边缘在自然场景中非常常见，特别是在含有大量相互重叠和遮挡的相似物体的情况下，这种影响更为明显。作者发现，扩展模板比扩展边缘更加鲁棒。其次，在杂乱的图像场景中估计的边缘信息通常并不是非常可靠。利用采用轮廓带图，通过局部求和与均值估算出的轮廓带中的二维向量在杂乱场景处的幅值较小，使得匹配值更低，从而降低了这些不可靠区域对匹配过程的影响。

本算法通过每一个像素点处的轮廓带图匹配值来选择候选的物体区域，这些匹配值被归一化到区间$[0, 1]$之间，位于局部最大值点处的模板位置被选定为候选物体位置。实现中，本系统忽略匹配值低于阈值t_m的局部最大值点。如果两个在阈值d距离范围以内存在多个局部最大匹配值，则仅保留匹配值最大的那个。在实验中，令阈值$t_m = 0.3$，并将阈值d设置为模板大小的一半。 图 4.4展示了图 4.1的轮廓带图匹配结果。

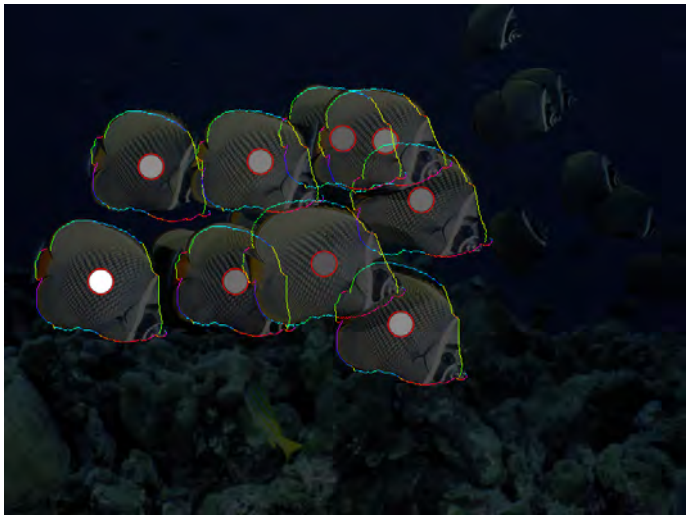


图 4.4　图 4.1的轮廓带图匹配结果。图中，物体中间圆圈区域的亮度表示模板和轮廓带图中检测出的候选物体之间的匹配值。检测出的所有相似物体的边缘覆盖在原图的上方。图中的轮廓颜色表示了局部轮廓方向(颜色与方向的对应见图 4.3b右上角)。

为了能够更好地适应更多的相似物体形状差异，本系统对模板进行旋转和缩放。在所有的实验中，作者采用了三个方向的旋转$\{-10°, 0°, 10°\}$ 和三个尺度的缩放$\{0.8, 1.0, 1.2\}$。如果有必要，用户可以采用更多的旋转量和尺度值。轮廓带的宽度b控制了轮廓带图所允许的形状变化范围。虽然更大的b可以使得轮廓带图处理更大范围的形状差异，但是b值过大有可能导致轮廓带扩展时的自交现象。在本章的所有实验中，b的值都被设置为15。

4.3.2.1　基于快速傅里叶变换的加速

如果通过朴素的方法直接计算式 (4-1)中的点积平方和将会非常耗时。这个计算过程可以利用快速傅里叶变换 (Fast Fourier Transforms, FFT)来加速。 式 (4-1)可以重写为：

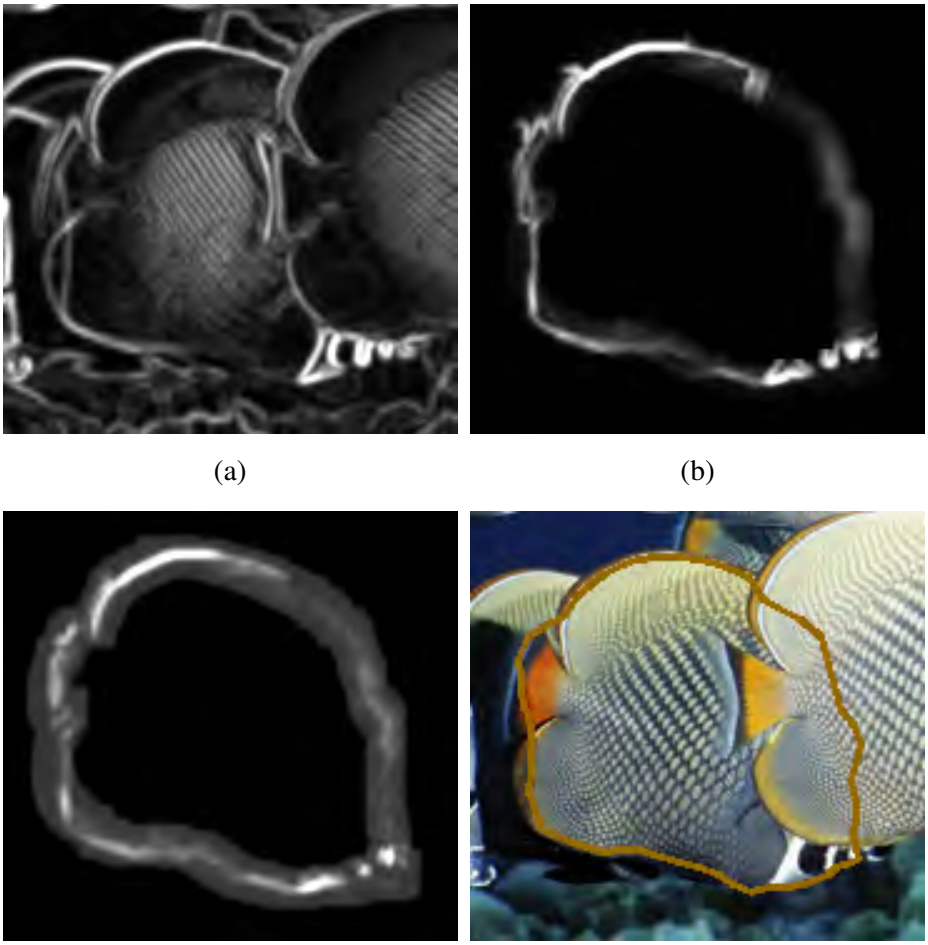
$$D_{(u,v)}(T, M) = \sum_{j=1}^{h}\sum_{i=1}^{w}(t_x^2 m_x^2 + t_y^2 m_y^2 + 2t_x t_y m_x m_y). \tag{4-2}$$

其中的t_x^2, t_y^2, $t_x t_y$, m_x^2, m_y^2 和 $m_x m_y$ 是相互独立的项。每一项都可以通过快速傅里叶变换求解乘积和的方式进行加速计算[145]。

4.4　图像中相似结构分析

4.4.1　基于形状先验的检测结果改进

本系统将那些与前景区域交集所含的像素小于该候选区域总像素数30%的候



(a)　(b)

(c)　(d)

图 4.5　利用主动轮廓模型进行检测结果改进的过程示例：(a) 边强度，(b) 加入形状先验信息后的能量图，(c) 进一步考虑局部方向一致性后的能量图，(d) 利用主动轮廓模型得到的物体边缘。

选区域舍弃掉。然后，每一个候选物体区域都被更新为它们与前景区域蒙板的交集。本章采用一种基于形状先验的主动轮廓模型来进一步改进物体轮廓的精度。为了获取形状先验，本方法首先利用Ho等人[146]提出的形状注册方法来估计模板与当前候选物体区域的最优仿射变换。模板形状经过该仿射变换之后的结果将被作为该候选物体区域的形状先验。这个注册过程使得本方法能够获得模板形状与目标物体最匹配的姿态，而不是在之前的候选物体检测阶段(见第 4.3节)所采用的有限的几个 (例如3个)旋转和缩放。然后，本方法采用主动轮廓模型[147]来进一步细化物体轮廓。与传统的主动轮廓模型中直接利用局部边缘强度不同，本方法同时考虑了物体轮廓与形状先验的全局和局部几何匹配信息。全局的形状相似性通过候选物体轮廓上的点集和仿射变换后的模板轮廓点集之间的欧氏距离来度量。局部的形状相似性通过归一化后的模板与候选区域方向向量的点积来度量。以

下三项中的每一项都被归一化到区间[0.1, 1]之间：边强度、点的距离、和方向差异。本方法用归一化之后的三项的乘积作为主动轮廓模型中的总能量。这种设计使得结果区域的轮廓与形状先验(即仿射变换后的模板)相一致。这种一致不仅体现在图像梯度上，也体现在全局和局部几何形状信息上(参见 图 4.5)。本章的方法利用了形状先验信息，能够更好地对付干扰边的影响。

4.4.2　层次关系估计

在实际的三维线索缺失的情况下，相对的深度顺序或者层次信息对图像编辑非常重要。本方法通过分析两个物体区域之间的重叠部分(记为R_I)来计算它们之间的层次关系。首先，本系统为重叠区域R_I建立描述其颜色的高斯混合模型(Gaussian mixture model, GMM)。其次，对于每个相似物体，寻找其匹配模板上与R_I对应的区域。记一个物体对应的这种区域为R_{T_1}，另一个物体对应的这种区域为R_{T_2}，然后分别为区域R_{T_1}和R_{T_2}建立表示其颜色的高斯混合模型。哪个物体对应的模板区域与图像区域R_I的颜色模型最符合，这个物体就被认为在更前面。例如图 4.5d中所示的左右两个鱼的遮挡部分，分别对应模板中的鱼头和鱼尾位置。由公共区域与模板的鱼头及鱼尾的颜色相似度不难得出，右边的鱼在左边鱼的上方(即公共区域更像鱼尾)。由于这个简单的二值关系的判断中有重叠区域的大量信息作为参考，因而非常鲁棒。最后利用拓扑排序算法，从图像所包含的这种相互重叠的两两物体对的层次关系中，推导出与局部次序相兼容的全局层次顺序。

4.4.3　透明度抠图

虽然恢复了物体的轮廓信息，由于两大因素的影响，这些信息还不足以被直接用于图像编辑。首先，主动轮廓模型会倾向于产生过分平滑的物体轮廓，因而缺乏足够的几何细节 (见图 4.5d)。其次，自然图像I可以被认为是前景物体F和背景区域B的线性组合，记为：$I = \alpha F + (1 - \alpha)B$，其中$\alpha$是前景的透明度。这种透明度经常被称为透明度抠图 (*Alpha Matte*)，其中的透明度Alpha值在物体轮廓附近平滑的变化。如果把图像中的物体直接挪到另一个区域而不考虑这种透明度因素就会产生明显的瑕疵。

因此，本系统进一步利用已经获取的层次信息来改进对物体区域的估计，并计算相关的透明度参数。本方法采用Levin等人的透明度抠图技术[38]来获取精确的物体区域及其相关的透明度值。为了获取一个物体的透明度矩阵，本方法首先需要一个Trimap。所谓Trimap，就是将图像中的像素分为三类：确定的前景像素、确定的背景像素、和未知区域。本系统希望能够处理含有相似的物体单元的

(a)　　(b)　　(c)

图 4.6　透明度抠图示例。图 4.5d 中鱼的例子相应的透明度抠图结果： (a) Trimap， (b) 透明度值， (c) 前景颜色。

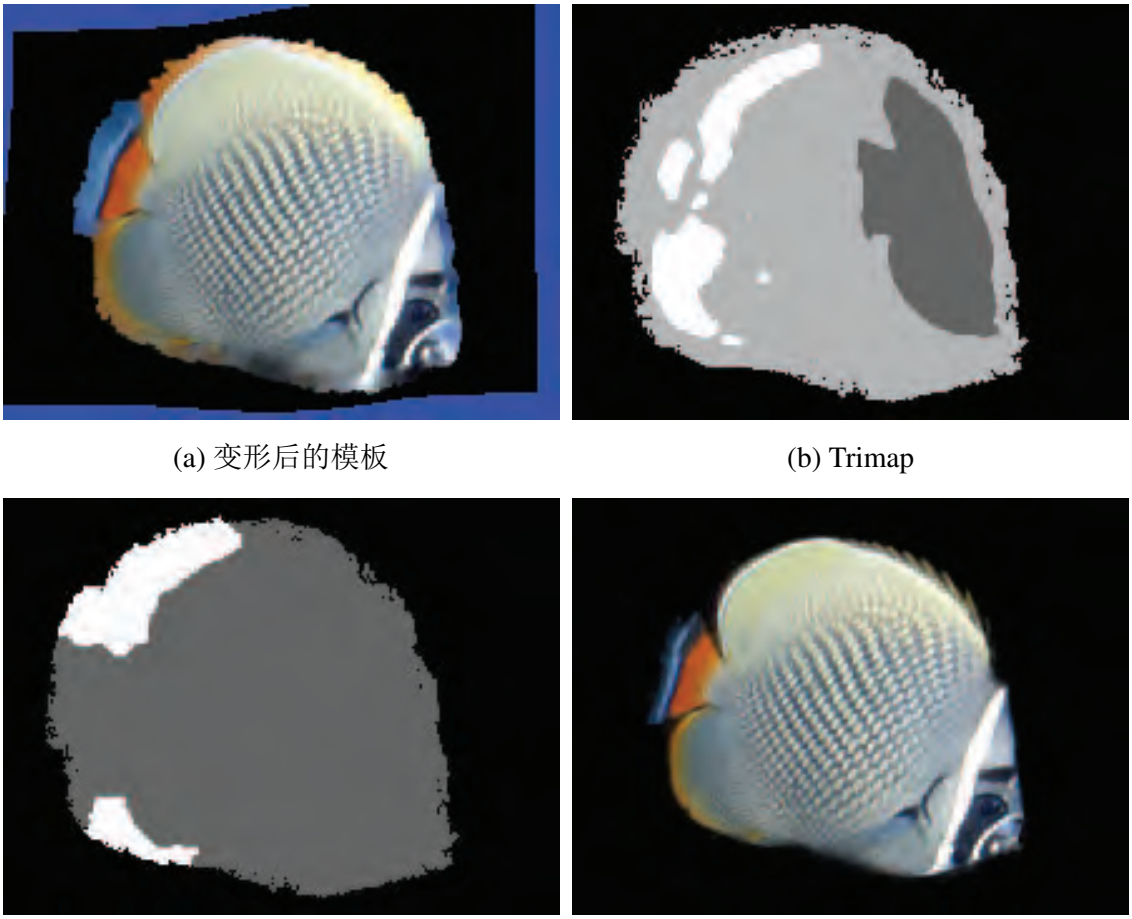
图像。由于这种相似物体单元之间经常具有非常相似的前景和背景颜色，因此必须仔细地设定Trimap以防止出现较差的结果。本方法将可见的物体轮廓向内外两侧分别扩展20个像素宽，并在这个带状区域内部计算透明度值。由于物体轮廓被遮挡的部分不含有可区分前背景颜色的信息，在构建Trimap的时候这部分不应该被扩展。 图 4.6展示了图 4.5d对应的透明度抠图结果。

4.4.4　稠密对应关系估计

给定一张图像中检测到的两个相似物体，本方法希望计算一个平面变换 $T : \mathbb{R}^2 \to \mathbb{R}^2$ 将一个物体上任意一点映射到另一个物体上的对应点。为此，本方法首先利用Thayananthan等人提出的Shape Context算法[144] 计算两个物体的轮廓点之间的对应关系。这种轮廓点通常是图像中较为可靠的特征点，因此它们之间的对应关系估计比其它点更可靠。本方法进一步利用构建弹性坐标变换中常用的薄板样条插值[148] (Thin Plate Splines, TPS) 方法来计算相似物体对之间的稠密像素对应关系。

4.4.5　遮挡部分补全

在含有相似物体的图像中，物体之间经常存在遮挡现象。为了便于进行后续的图像编辑操作，本系统需要对这些缺失区域进行合理的补全。本算法利用相似物体之间信息表达的冗余性来解决这一问题。为了补全被部分遮挡的物体，作者首先将没有被遮挡的参考物体映射到这个物体上。 图 4.7a 展示了这样一个映射的例子。参考图像映射后落在遮挡区域的像素在图 4.7b中用深灰色表示。如果未被遮挡的像素和映射后的参考图像的相应像素颜色区别较小，就采用未被遮挡的原始像素(见图 4.7b中的白色区域)。对于其余像素 (图 4.7b中的深灰色)，采用GraphCut技术[33]来获取一条补全路径。这条路径将使得变形后的参考图像区域



(a) 变形后的模板　(b) Trimap

(c) Graph-cut 结果　(d) 遮挡部分补全的结果

图 4.7　基于补全蒙板的物体补全。 (a) 将一个未被遮挡的参考模板变形并映射到一个部分遮挡的物体(图 4.5d)区域。 (b) 初始的补全蒙板。其中的白色部分代表采用原始图像的像素，深灰色代表采用变形后的模板的像素，在其余的浅灰色区域，本方法利用GraphCut来寻找最优的补全路径。 (c) 利用GraphCut计算的补全结果蒙板。 (d) 最终的补全结果。为了更好的显示，该图将前景颜色与透明度值做了乘积。

和原始物体区域相接处颜色变化最小，以达到无缝补全的目的。通常情况下，本方法采用三个未被遮挡的物体分别进行这一补全操作，并最终选择接缝代价最小(两边颜色差异最小)的那个作为最终的补全结果。

4.5　图像编辑应用

本章介绍的系统是一种交互式的场景物体级别图像编辑工具。本章的实验环境是一个含有2个Xeon (R) 2.27GHz CPU和12GB内存的个人计算机。经过快速傅里叶变换加速后的轮廓带图检测算法在考虑三个缩放尺度和三种旋转量的情况下处理一张 800×600 的图像通常需要0.4秒。自动补全和稠密对应关系计算的速度取决于相似物体的数目和大小，这些步骤通常需要0.4 – 0.8秒。

(a) 输入图像　　(b) 重排后的结果

图 4.8　基于物体重排的场景物体重复效果。

在少量用户笔刷的指引下，本章的系统能够自动地完成图像中相似物体单元的检测、遮挡部分的补全以及层次关系的估计。所有这些计算的速度都能满足交互式图像编辑的性能要求。同时，本系统为相似物体间建立稠密的对应关系。在图 4.1的例子中，本方法除了检测形似的鱼以外，还分析出了它们之间的稠密对应关系和层次关系。这些信息使得一系列图像编辑应用得以实现。本节将详细介绍这些应用。

4.5.1　物体重排

在现实世界中，人们经常移动物体的位置和改变它们的顺序关系。图像是现实世界的一个镜像，用户也希望能够在图像中方便地进行类似的物体位置和顺序的调整。然而，图像本身所固有的像素表达形式使得这种简单直观的编辑操作不可能实现。通过提取的图像结构，本系统允许用户在一定程度范围内对图像场景进行重排，同时对新暴露出的背景区域进行合理的补全。 图 4.1展示了重排一组鱼的例子。作者将图像上部中间区域被遮挡的鱼挪到了最前面的一层，并改变了其相对于其它鱼的空间位置。由于所有的鱼都被检测出来，并且遮挡区域也被补全了，这个过程非常简单。图中的背景区域通过Barnes等人提出的PatchMatch[30]

技术进行了补全。本章的重排应用类似于已有的Reshuffling的工作[28,149]。与这些方法的主要区别包括：本方法不仅可以改变图像场景中的物体对象的相互位置关系，而且可以编辑他们的层次关系；不同于已有方法编辑无语义的图像区块，本章中的算法直接在场景物体级别对图像进行编辑，因而更加符合人类的直观理解。

在图 4.13a 中，作者重新调整了柿子的位置以使其排列成一条直线。同时改变了这些柿子之间的层次关系，以使其符合透视现象。更加灵活的层次编辑可以进一步考虑集成局部层次(local layering)方法[150]来实现。

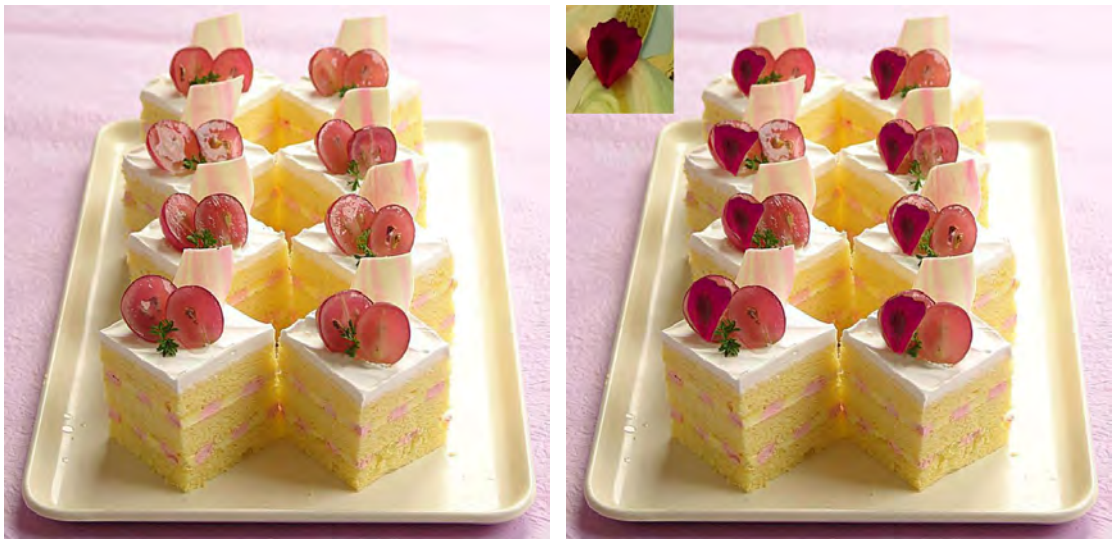
在纹理合成领域，对纹理图进行重复以生成一个更大的纹理图已经是一个研究的比较透彻的问题。可是，当图像中包含具有明显结构的相似物体的时候，许多合成算法就会碰到问题。 图 4.8展示了一个利用本章算法给图像中插入更多相似物体并改变其位置，以实现物体重复效果的例子。这个重复“花”的例子很好地体现了本章算法进行场景物体级别图像编辑的一个优势。由于原图中不含有任何一个同时包含红色果实和黄色花的区块，该结果很难通过以往基于区块的方法做到。基于本章的算法，用户具有更大的自由度，可以按照自己的意愿随意地控制最终结果，从而达到自己想要的结果。 图 4.13e展示了另外一个通过物体重排进行内容重复的例子。在这个例子中，新产生的茶杯并不是现有茶杯的一个简单的拷贝。由于本章的算法利用GraphCut自动地寻找一条最优的补全路径对来自不同相似物体的图像区域进行无缝的拼接，因此产生的新物体和其余的相似物体并不相同。例如，左上角的新杯子同时含有多个其余杯子的特征。

4.5.2　编辑传播

在用户希望对图像中一组相似物体进行编辑的时候，保持相似对象间编辑的一致性对结果的质量和视觉效果非常重要。然而，仅靠手工来保持这种一致性对用户来说太过困难。利用本算法计算出的稠密对应关系，用户很容易就能在相似物体之间传播用户编辑。在图 4.13d所展示的例子中，用户抓取了一个门的区域，并将其贴到了一个“蒙古包”上。这种编辑操作被传播到其它的“蒙古包”上，这样，所有的蒙古包都被适当地进行了编辑。同样，在图 4.9b中，从另外一个图中得到的玫瑰花瓣被传播到了所有的蛋糕上。

An等人的Appprop[139] 方法同样可以将用户编辑传播到邻近的相似区域。与这种像素或者区块级别的编辑不同，本章方法的设计目标是在较高的语义层级上对编辑进行扩散。例如，对物体的一些特定区域的编辑进行扩散，以使得其它相似物体的对应区域得到相似的编辑。这种物体级别的编辑方式更加符合用户的直觉，并接近人类对潜在场景中这些物体关系的理解。

(a)



(b)

图 4.9　编辑传播示例： (a) 对一个蛋糕的编辑传播到其它蛋糕上，提高了编辑的一致性和效率。 (b) 在相似物体之间传播用户笔画绘制结果。

在图 4.9a所示的例子中，用户在一个蛋糕上画的图案被传播到其余蛋糕上的对应位置。由于基于用户笔刷的图像编辑是一种非常常见的用户输入方式，被广泛地应用于很多应用中。这些应用包括：抠图[38], 编辑扩散[139], 和图像补全[37]等。因此，传播用户笔刷具有很强的普遍意义。相似的用户编辑同样可以被用来改进检测和分割结果。在图 4.1鱼的例子中，初始分割中鱼尾部分有少量缺失。本方法通过一个简单的笔刷去更正了其中的一个，这种编辑被传播到了其它所有的鱼上。所有的鱼尾部分都在这一指导下被正确的补全了。

(a) 输入图像　　(b) 同步变形后的结果

图 4.10　在其中一个人身上选择6个控制点并控制这个人进行运动。本章的系统根据相似物体间的稠密对应关系，自动地将控制点及其移动情况传递到其它相似物体上，并对遮挡区域进行补全。通过这些控制点的传播，其他的人也进行了相似的变形。

4.5.3　变形传播

利用本章的系统，用户同样可以对一组相似的物体进行变形。用户对一组相似物体中的一个进行变形，该物体变形前后的控制点被传递到其它所有相似物体上。因此用户仅需对一组相似物体中的一个进行变形，所有相似物体都会同时被改变。作者实现了Igarashi等人的变形方法[135]。其余的一些变形方法[136,137] 同样也可以用在本章的系统中。变形传播的例子可以在图 4.10 和 图 4.13b中找到。

4.5.4　物体替换

本方法可以进一步地将一幅图像中的相似物体全部替换为另一幅图中的不同物体，同时保持原始物体的位置、大小、姿势和层次关系。这种操作可以通过在两幅图像中分别检测相似物体，并把其中一个图像中的相似物体替换为另一个图像中的相似物体来实现。在替换过程中，新的物体被按照原始物体之间的相互几何和空间关系进行了正确的缩放、旋转和次序调整。

图 4.11展示了把一个图片中的灯替换为另一种不同样式的灯的例子。通过在两个图片上分别画上简单的笔刷，两种灯都分别被检测和提取出来。然后用户通过拖拽另一种灯来替换原始图像中的一批灯。由于在替换过程中考虑了原始物体的几何和空间关系，新结果中的灯很好地融入了原始的图像场景之中。图 4.13c展示了一组将鸭子替换成鹅的例子。在该例子中，鹅的方向、尺度和层次关系都被调整以符合原图中鸭子之间的排列。由于存在很强的形状不一致性，该例子中有少量的鹅形状恢复并不是特别理想。 图 4.1展示了另外一组例子。

(a) 输入图像　　(b) 物体替换后的结果

图 4.11　通过物体替换，将一个图中的灯换成别的图中不同样式的灯。

4.5.5　算法的局限性

本章提出的相似物体检测算法对于相似物体间的光照和表观变化比较鲁棒，对遮挡和形变也有一定的鲁棒性。但是，和其它所有检测算法一样，本章的算法在一些特定的情况下也会失败。特别是当图像中有严重遮挡的相似物体，或者相似物体之间存在着较大的形状差异的时候，本章的算法有可能会因此漏检一些物体。 图 4.12展示了这样的两个失败的例子。但是，由于本章的编辑框架是交互式的，用户可以通过制定额外的标定来轻松地克服这一问题。

图 4.12　本系统不能很好处理的一些典型情况：左图中，相似物体之间的形状差异过大；右图中，物体之间的遮挡过于严重。这两种因素会影响到本章基于形状的检测算法，使其不能完全的检测出所有相似物体。

图 4.13　基于本算法框架的各种场景物体级别图像编辑应用。

4.6　小结与讨论

本章中介绍的算法可以利用非常少量的用户交互有效地检测出图像中的相似物体。通过自动地对这些相似物体进行遮挡部分补全、深度关系估计以及透明度抠图，本章的系统能够支持一系列直观方便的图像编辑应用。这些应用包括：图像重排、编辑传播、同步变形和物体替换。 图 4.13展示了本章系统支持的应用的更多结果。如果没有相似物体之间的层次、位置以及关联信息，这些编辑操作很难实现。

本章的系统框架可以在多个方向上进行扩展，这里仅对一小部分激动人心的未来工作方向进行讨论。首先，本系统中现有的稠密对应关系估计仅考虑形状因素，因此，对应点之间并不能保证具有相似的表观。在已有工作的基础上进一步考虑表观相似性约束，使得相关点之间的局部表观更加一致，是一个有趣的方向。其次，现有的框架要求图像中至少有一个完整可见的物体存在，并以此物体的形状作为模板寻找其它相似物体。处理那些所有物体都被部分遮挡的情况将是一个值得研究的问题。最后，虽然已经展示了多种基于本章算法框架的图像编辑应用，作者仍然计划实现更多类似的图像编辑应用，如图像分解和基于单幅图的动画。

本章提出了一种从图像场景中有效地检测和提取相似物体的算法框架，并同时确定它们之间的关系。这些关系包括：深度顺序、相对位置、朝向等。与此同时，本方法为相似物体之间建立了像素级别的稠密对应关系。这些信息使得一系列直观的图像编辑应用和操作得以实现。与以往在像素或者区块级别对图像进行编辑的方法不同，本方法允许用户在场景物体级别对图像进行编辑，因而提供了更贴近自然感受的用户编辑体验。场景物体级别图像编辑是对现实生活中用户体验进行模拟的一种有意义的尝试。在这个方向上，本章的方法仅仅走出了一小步。

第 5 章　总结与展望

本文研究工作的核心内容是图像场景内容分析及其应用，研究方法是显著性检测与相似性分析，主要的研究成果包括图像视觉显著性区域检测与分割方法、群组图像显著性区域提取与检索方法、以及图像中相似单元检测与结构分析方法。作者将本文的研究成果用于内容敏感的图像缩放、非真实感绘制、图像检索、图像重排、变形传播、编辑扩散、物体替换等图像处理应用中，得到了很好的结果。

5.1　工作总结

图像是记述信息、传递思想和表达情感的重要媒介。随着数码相机和智能手机等硬件设备的普及，特别是社交网络、微博、网络相册等大规模共享平台的快速兴起，数字图像已经成为日常生活中最受欢迎的一种知识传播和信息共享的媒介。对数字图像的处理也已经成为信息科学、工程学、生物学、医学、心理学甚至社会科学等领域的重要研究对象。随着计算机处理能力的持续提高以及相关领域分析手段的发展，改变传统基于像素、区块、信号等底层数据信息级别的图像处理方式，让计算机像人一样理解图像中的高层语义对象，从而提供自然、方便、直观、符合人类认知规律的图像处理工具，已经成为数字图像处理领域的重要发展趋势，值得深入研究。

本文首先介绍了数字图像处理技术产生和发展的背景，并对新时期数字图像处理所面临的主要问题和发展机遇进行了简单的分析。然后结合本文的研究内容，对数字图像处理中的几个重要研究方向进行了介绍，并回顾了相关的研究工作。这些研究方向包括：内容敏感图像处理、基于内容的图像检索技术、图像场景建模与智能编辑。

第 2章提出了一种基于全局对比度分析的图像视觉显著性区域快速鲁棒的检测方法。该方法通过对图像区域间的全局对比度和空间相关性进行建模，能够快速有效地检测并分割出图像中的视觉显著性区域。作者定性的展示了本算法的结果在内容敏感图像缩放、非真实感绘制、以及感兴趣物体区域自动提取等应用中相对传统方法的优势。为了定量的验证本算法结果的有效性，作者在国际上现有含有显著性区域精确标注的最大 (含1000张图像) 公开测试集上，对本文的方法和其它已有方法进行了比较。定量的实验结果表明，该方法的检测结果明显优于已有方法，其显著性区域分割结果的正确率与召回率 (正确率=90%，召回率=90%)

相对之前的最好结果 (正确率=75%，召回率=83%) 有大幅提升。为了进一步验证该算法在更大数据集上的可扩展性，作者建立了一个同样含有显著性区域精确标注的基准数据集THUS10000 (其数据量是之前最大数据集的10倍。作者将公开这一数据集以方便该领域的进一步研究)。在此数据集上将本文算法与国际上现有的15种主流方法进行了进一步的定量比较，实验结果再次验证了该算法的鲁棒性。

在第 3章，本文从同一个关键词所对应的一组相关网络图像集合中获取该类图像的表观特征统计模型, 并利用该统计模型来改善单张图像显著性区域检测和分割的结果。基于关键字所获得的初始图像集合通常不可避免的含有很多噪声。例如，通过关键字从Flickr上搜索图像，通常仅有一小半图像和输入的关键词相关。为了能够从这些低质、不可靠的网络数据中学习到能够尽量完整地反映该类别物体表观特征的高质量统计信息，本方法从单张图像的显著性区域提取结果中选择形状与用户输入草图较一致的一部分作为表观模型的学习样例。由于形状特征和颜色、纹理等表观特征相关性较小，这种基于形状的过滤方式在获得高质量初始训练集合的过程中不会对样本表观特征的多样性产生多少影响。本文进一步利用这种学习到的表观特征，增强那些与显著性物体表观特征一致的图像区域的显著性值，并抑制背景区域经常出现的表观特征对应的图像区域的显著性值。为了能够定量的分析这种表观统计模型对相关、低质的网络图像显著性区域提取的影响，我利用5类关键词从Flickr上自动下载了15000张图像并对其进行像素精度的标注。在该数据集上对显著性区域检测结果准确率和召回率的分析表明，本文提出的群组显著性区域提取方法明显优于现有最好的基于单张图像的提取方法。该方法进一步利形状匹配算法对这种显著性区域提取结果和用户输入草图做比较，以实现基于草图的图像检索。基于这种自动分割结果的草图图像检索系统 (SBIR) 不但在正确率方面优于之前最好的SBIR系统，而且提供了额外的目标物体区域信息以支持更广泛的应用需求。

第 4章提出了一种基于简单交互的相似物体快速检测与分析方法。本文利用简单的用户交互提取相似物体的一个模板和前景区域，对前景区域进行层次化的分割以获取待检测相似物体的潜在轮廓信息，并基于这种潜在轮廓构建轮廓带图表达。所提出的轮廓带图表达同时含有潜在物体的全局几何信息、局部几何信息、轮廓置信度，并能够处理一定范围内的物体间形状差异带来的影响。本文利用快速傅里叶变换对基于轮廓带图表达的配进过程行加速，实现了快速物体检测；通过拓扑排序来建立相互遮挡的物体之间的部分深度顺序；并利用未被遮挡物体的信息对遮挡物体进行补全。这些场景物体及其相互关系的信息使得一系列场景物体级别图像编辑应用得以实现，包括以下几大类：图像重排、编辑扩散、

变形传播以及物体替换。

5.2　未来工作的展望

从自然图像中获取场景中感兴趣物体的区域、层次、几何、相互关系等场景物体级别图像信息是使得计算机能够像人一样从语义级别对图像进行理解的重要步骤。这些信息的获取将为快速智能的图像处理技术、自然方便的图像检索方法、和直观高效的图像编辑体验提供重要基础。在这个大方向上，本文的工作仅仅只是个开始，还有这很多重要题目等待着更深入的研究。这里仅列出其中的几个大的方面：

(1) 在第 2章，本文利用图像区域颜色的全局对比度和空间相关性信息得到了高质量的视觉显著性区域检测算法，并展示了所得到的显著性图在一系列相关的图像编辑中的应用。对于具有复杂纹理的图像和具有明显运动信息的视频，仅仅分析颜色信息有时候并不足以获得完美的检测结果。在本文算法的基础框架之上，进一步分析纹理特征和运动特征是一个值得研究的方向。

(2) 第 3章中的群组显著性区域检测算法利用了一组相关网络图像中的公共颜色信息成功地改进了单张图像显著性区域的检测与提取，并在基于草图的图像检索应用中显示了巨大的潜力。在今后的工作中，进一步考虑这些相关图像之间公共的纹理和形状信息，对更好地提取图像中重要的物体区域信息具有重要意义，值得深入研究。目前基于显著性形状的图像检索方法虽然在正确率方面有优势，但形状比较速度仍然不能满足海量图像检索的需求。进一步研究形状索引技术对于更加有效地利用显著性区域提取结果具有重要意义。

(3) 第 4章中的相似物体检测与分析算法从含有相似物体的图像中提取出了物体区域、层次关系、对应关系等重要的场景物体级别信息。这些场景物体级别信息从一定程度上反映了人类理解这些图像场景的方式，因而在一系列直观、方便的图像编辑应用中取得了成功。如何将处理对象从含有相似物体的图像，推广到更广泛的自然图像中，是值得进一步研究的重要方向。另外，本文的方法虽然能够得到物体间的层次关系并从一定程度上反映了场景的几何信息，但是这些几何信息还远远不够精确。使用其它人性化的交互方式从图像中获得更加细致的几何信息和物体间关系也将为更加丰富多样的图像编辑体验提供支持。

参考文献

[1] Gonzalez R C, Woods R E. Digital image processing (3rd edition). Prentice Hall, 2007.

[2] Koffka K. Principles of gestalt psychology. Lund Humphries, 1935.

[3] Teuber H. Physiological psychology. Annual Review of Psychology, 1955, 6(1):267–296.

[4] Wolfe J M, Horowitz T S. What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience, 2004, 5:1–7.

[5] Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 1995, 18(1):193–222.

[6] Mannan S K, Kennard C, Husain M. The role of visual salience in directing eye movements in visual object agnosia. Current Biology, 2009, 19(6):247–248.

[7] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 1998, 20(11):1254–1259.

[8] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009. 1597–1604.

[9] Rutishauser U, Walther D, Koch C, et al. Is bottom-up attention useful for object recognition? Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004. 37–44.

[10] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 still image coding system: an overview. IEEE Transactions on Consumer Electronics, 2002, 46(4):1103–1127.

[11] Avidan S, Shamir A. Seam carving for content-aware image resizing. ACM Transactions on Graphics (SIGGRAPH), 2007, 26.

[12] Wang Y S, Tai C L, Sorkine O, et al. Optimized scale-and-stretch for image resizing. ACM Transactions on Graphics (SIGGRAPH Asia), 2008, 27(5):118:1–8.

[13] Rubinstein M, Shamir A, Avidan S. Multi-operator media retargeting. ACM Transactions on Graphics (SIGGRAPH), 2009, 28(3):1–11.

[14] Zhang G X, Cheng M M, Hu S M, et al. A shape-preserving approach to image resizing. Computer Graphics Forum, 2009, 28(7):1897–1906.

[15] Chen T, Cheng M M, Tan P, et al. Sketch2Photo: internet image montage. ACM Transactions on Graphics (SIGGRAPH Asia), 2009, 28(5):124:1–10.

[16] Chia Y S, Zhuo S, Gupta R K, et al. Semantic colorization with internet images. ACM Transactions on Graphics (SIGGRAPH Asia), 2011, 30(6).

[17] Liu H, Zhang L, Huang H. Web-image driven best views of 3D shapes. The Visual Computer, 2012, 28:279–287.

[18] Huang H, Zhang L, Zhang H C. Arcimboldo-like collage using internet images. ACM Transactions on Graphics (SIGGRAPH Asia), 2011, 30(6):155:1–155:8.

[19] Smeulders A, Worring M, Santini S, et al. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2000, 22(12):1349–1380.

[20] Datta R, Joshi D, Li J, et al. Image retrieval: ideas, influences, and trends of the new age. ACM Computing Surveys, 2008, 40(2):1–60.

[21] Bouet M, Khenchaf A, Briand H. Shape representation for image retrieval. Proceedings of ACM Multimedia, 1999. 1–4.

[22] Hoffman D, Singh M. Salience of visual parts. Cognition, 1997, 63(1):29–78.

[23] Bai X, Yang X, Latecki L, et al. Learning context-sensitive shape similarity by graph transduction. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2010. 861–874.

[24] Eitz M, Hildebrand K, Boubekeur T, et al. Sketch-based image retrieval: benchmark and bag-of-features descriptors. IEEE Transactions on Visualization and Computer Graphics (TVCG), 2011. Preprint.

[25] Criminisi A, Perez P, Toyama K. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing (TIP), 2004, 13(9):1200–1212.

[26] Eisemann E, Durand F. Flash photography enhancement via intrinsic relighting. ACM Transactions on Graphics (SIGGRAPH), 2004, 23(3):673–678.

[27] Adams A, Gelfand N, Dolson J, et al. Gaussian KD-trees for fast high-dimensional filtering. ACM Transactions on Graphics (SIGGRAPH), 2009, 28(3):21:1–12.

[28] Simakov D, Caspi Y, Shechtman E, et al. Summarizing visual data using bidirectional similarity. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008. 1–8.

[29] Bai X, Wang J, Simons D, et al. Video SnapCut: robust video object cutout using localized classifiers. ACM Transactions on Graphics (SIGGRAPH), 2009, 28:70:1–70:11.

[30] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics (SIGGRAPH), 2009, 28(3):24:1–11.

[31] Shamir A, Avidan S. Seam carving for media retargeting. Communications of the ACM, 2009, 52(1):77–85.

[32] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2000, 22(8):888–905.

[33] Boykov Y Y, Lea G F. Graph cuts and efficient N-D image segmentation. International Journal on Computer Vision (IJCV), 2006, 70(2):109–131.

[34] Paris S, Durand F. A topological approach to hierarchical segmentation using mean shift. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007. 1–8.

[35] Lempitsky V, Kohli P, Rother C, et al. Image segmentation with a bounding box prior. Proceedings of IEEE International Conference on Computer Vision (ICCV), 2009. 1–8.

[36] Jia Y, Hu S, Martin R. Video completion using tracking and fragment merging. The Visual Computer, 2005, 21(8):601–610.

[37] Sun J, Yuan L, Jia J, et al. Image completion with structure propagation. ACM Transactions on Graphics (SIGGRAPH), 2005, 24(3):861–868.

[38] Levin A, Lischinski D, Weiss Y. A closed-form solution to natural image matting. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2008, 30(2):228–242.

[39] Brox T, Kleinschmidt O, Cremers D. Efficient nonlocal means for denoising of textural patterns. IEEE Transactions on Image Processing (TIP), 2008, 17(7):1083–1092.

[40] Zheng Q, Sharf A, Wan G, et al. Non-local scan consolidation for 3D urban scene. ACM Transactions on Graphics (SIGGRAPH), 2010, 29(3):to appear.

[41] Triesman A M, Gelade G. A feature-integration theory of attention. Cognitive Psychology, 1980, 12(1):97–136.

[42] Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurbiology, 1985, 4:219–227.

[43] Vaquero D, Turk M, Pulli K, et al. A survey of image retargeting techniques. SPIE Applications of Digital Image Processing XXXIII, 2010, 7798(1).

[44] Treisman A, Gelade G. A feature-integration theory of attention. Cognitive psychology, 1980, 12(1):97–136.

[45] Gottlieb J, Kusunoki M, Goldberg M, et al. The representation of visual salience in monkey parietal cortex. Nature, 1998, 391(6666):481–484.

[46] Robinson D, Petersen S. The pulvinar and visual salience. Trends in Neurosciences, 1992, 15(4):127–132.

[47] Hou X, Zhang L. Saliency detection: a spectral residual approach. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007. 1–8.

[48] Zhang L, Tong M, Marks T, et al. SUN: a Bayesian framework for saliency using natural statistics. Journal of Vision, 2008, 8(7):32:1–20.

[49] Achanta R, Estrada F, Wils P, et al. Salient region detection and segmentation. Proceedings of IEEE International Conference on Computer Vision Systems (ICVS), 2008. 66–75.

[50] Bruce N, Tsotsos J. Saliency, attention, and visual search: an information theoretic approach. Journal of Vision, 2009, 9(3):5:1–24.

[51] Seo H, Milanfar P. Nonparametric bottom-up saliency detection by self-resemblance. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR). IEEE, 2009. 45–52.

[52] Rahtu E, Kannala J, Salo M, et al. Segmenting salient objects from images and videos. European Conference on Computer Vision (ECCV), 2010. 366–379.

[53] 张鹏, 王润生. 静态图像中的感兴趣区域检测技术. 中国图象图形学报: A 辑, 2005, 10(2):142–148.

[54] Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 1986, 8(6):679–698.

[55] Burns J, Hanson A, Riseman E. Extracting straight lines. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 1986, 8(4):425–455.

[56] Mokhtarian F, Suomela R. Robust image corner detection through curvature scale space. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 1998, 20(12):1376–1381.

[57] Viola P, Jones M. Robust real-time face detection. International Journal on Computer Vision (IJCV), 2004, 57(2):137–154.

[58] Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, 2005. 886–893.

[59] Liu T, Sun J, Zheng N N, et al. Learning to detect a salient object. Proceedings of Proc. of CVPR, 2007.

[60] Felzenszwalb P, Huttenlocher D. Efficient graph-based image segmentation. International Journal on Computer Vision (IJCV), 2004, 59(2):167–181.

[61] Grady L. Random walks for image segmentation. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2006, 28(11):1768–1783.

[62] Cheng M M, Zhang G X. Connectedness of random walk segmentation. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2011, 33(1):200 –202.

[63] Rother C, Kolmogorov V, Blake A. "GrabCut"– Interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics (SIGGRAPH), 2004, 23(3):309–314.

[64] Bagon S, Boiman O, Irani M. What is a good image segment? a unified approach to segment extraction. European Conference on Computer Vision (ECCV), 2008. 30–44.

[65] Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing. Proceedings of ACM Multimedia, 2003. 374–381.

[66] Han J, Ngan K, Li M, et al. Unsupervised extraction of visual attention objects in color images. IEEE Transactions on Circuits and Systems for Video Technology, 2006, 16(1):141–145.

[67] Goldberg C, Chen T, Zhang F L, et al. Data-driven object manipulation in images. Computer Graphics Forum (European Graphics), 2012, 31(2):to appear.

[68] Chen T, Tan P, Ma L Q, et al. PoseShop: a human image database and personalized content synthesis. Research Report TR-100906, Tsinghua University, 2010. http://cg.cs.tsinghua.edu.cn/papers/poseshop.pdf.

[69] Chen H. Preattentive co-saliency detection. Proceedings of IEEE International Conference on Image Processing (ICIP), 2010. 1117–1120.

[70] Chang K, Liu T, Lai S. From co-saliency to co-segmentation: an efficient and fully unsupervised energy minimization model. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. 2129–2136.

[71] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2002, 24(4):509–522.

[72] Shotton J, Blake A, Cipolla R. Multiscale categorical object recognition using contour fragments. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2008, 30(7):1270–1281.

[73] Hirata K, Kato T. Query by visual example-content based image retrieval. Proceedings of Advances in Database Technology-EDBT, 1992. 56–71.

[74] Del Bimbo A, Pala P. Visual image retrieval by elastic matching of user sketches. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 1997, 19(2):121–132.

[75] Carson C, Thomas M, Belongie S, et al. Blobworld: a system for region-based image indexing and retrieval. Proceedings of Visual Information and Information Systems. Springer, 1999. 660–660.

[76] Leung W, Chen T. Retrieval of sketches based on spatial relation between strokes. Proceedings of IEEE International Conference on Image Processing (ICIP), volume 1, 2002. I–908.

[77] Liang S, Sun Z. Sketch retrieval and relevance feedback with biased SVM classification. Pattern Recognition Letters, 2008, 29(12):1733–1741.

[78] Fonseca M, Ferreira A, Jorge J. Sketch-based retrieval of complex drawings using hierarchical topology and geometry. Computer-Aided Design, 2009, 41(12):1067–1081.

[79] Eitz M, Hays J. Learning to classify human object sketches. Proceedings of SIGGRAPH 2011: Talks, 2011.

[80] Hu R, Barnard M, Collomosse J. Gradient field descriptor for sketch based retrieval and localization. Proceedings of IEEE International Conference on Image Processing (ICIP), 2010. 1025–1028.

[81] Hu R, Wang T, Collomosse J. A bag-of-regions approach to sketch-based image retrieval. Proceedings of IEEE International Conference on Image Processing (ICIP), 2011. 3661–3664.

[82] Wang C, Li Z, Zhang L. MindFinder: image search by interactive sketching and tagging. Proceedings of ACM International World Wide Web Conference (WWW), 2010. 1309–1312.

[83] Huang H, Zhang L, Zhang H. RepSnapping: efficient image cutout for repeated scene elements. Computer Graphics Forum, 2011, 30(7):2059–2066.

[84] Fan B, Wu F, Hu Z. Towards reliable matching of images containing repetitive patterns. Pattern Recognition Letters, 2011, 32(14):1851–1859.

[85] Cheng M M, Zhang F L, Mitra N J, et al. RepFinder: finding approximately repeated scene elements for image editing. ACM Transactions on Graphics (SIGGRAPH), 2010, 29(4):83:1–8.

[86] Zhang F L, Cheng M M, Jia J, et al. ImageAdmixture: putting together dissimilar objects from groups. IEEE Transactions on Visualization and Computer Graphics (TVCG). to appear.

[87] Wu H, Wang Y, Feng K, et al. Resizing by symmetry-summarization. ACM Transactions on Graphics (SIGGRAPH Asia), 2010, 29(6):159.

[88] Leung T, Malik J. Detecting, localizing and grouping repeated scene elements from an image. Proceedings of European Conference on Computer Vision (ECCV), 1996. I:546–555.

[89] Liu Y, Collins R T, Tsin Y. A computational model for periodic pattern perception based on frieze and wallpaper groups. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2003, 26(3):354–371.

[90] Ahuja N, Todorovic S. Extracting texels in 2.1D natural textures. Proceedings of IEEE International Conference on Computer Vision (ICCV), 2007. 1–8.

[91] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal on Computer Vision (IJCV), 2004, 60(2):91–110.

[92] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF). Computer Vision and Image Understanding, 2008, 110(3):346–359.

[93] Berg A C, Berg T L, Malik J. Shape matching and object recognition using low distortion correspondences. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005. I: 26–33.

[94] Reynolds J, Desimone R. Interacting roles of attention and visual salience in V4. Neuron, 2003, 37(5):853–863.

[95] Chu H K, Hsu W H, Mitra N J, et al. Camouflage images. ACM Transactions on Graphics (SIGGRAPH), 2010, 29:51:1–51:8.

[96] Jiang H, Wang J, Yuan Z, et al. Automatic salient object segmentation based on context and shape prior. Proceedings of British Machine Vision Conference, 2011. 1–12.

[97] Harel J, Koch C, Perona P. Graph-based visual saliency. Proceedings of Neural Information Processing Systems (NIPS), 2006. 545–552.

[98] Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues. Proceedings of ACM Multimedia, 2006. 815–824.

[99] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010. 2376–2383.

[100] Harel J, Koch C, Perona P. Graph-based visual saliency. Proceedings of Neural Information Processing Systems (NIPS), 2007. 545–552.

[101] Seo H, Milanfar P. Static and space-time visual saliency detection by self-resemblance. Journal of vision, 2009, 9(12):15:1–27.

[102] Achanta R, Süsstrunk S. Saliency detection using maximum symmetric surround. Proceedings of IEEE International Conference on Image Processing (ICIP), 2010. 2653–2656.

[103] Murray N, Vanrell M, Otazu X, et al. Saliency estimation using a non-parametric low-level vision model. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. 433–440.

[104] Duan L, Wu C, Miao J, et al. Visual saliency detection by spatially weighted dissimilarity. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. 473–480.

[105] Liu T, Yuan Z, Sun J, et al. Learning to detect a salient object. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 2011, 33(2):353 –367.

[106] Ko B, Nam J. Object-of-interest image segmentation based on human attention and semantic region clustering. Journal of the Optical Society of America, 2006, 23(10):2462–2470.

[107] Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. 409–416.

[108] Eihhauser W, Konig P. Does luminance-constrast contribute to a saliency map for overt visual attention? European Journal of Neuroscience, 2003, 17:1089–1097.

[109] Wang Z, Li B. A two-stage approach to saliency detection in images. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2008. 965–968.

[110] Achanta R, Süsstrunk S. Saliency detection for content-aware image resizing. Proceedings of IEEE International Conference on Image Processing (ICIP), 2009.

[111] Huang H, Zhang L, Fu T N. Video Painting via Motion Layer Manipulation. Computer Graphics Forum, 2010, 29(7):2055–2064.

[112] Zeki S, Nash J. Inner vision: an exploration of art and the brain. Nature, 2000, 404(6774):123–123.

[113] DeCarlo D, Santella A. Stylization and abstraction of photographs. ACM Transactions on Graphics (TOG), 2002, 21(3):769–776.

[114] Loeff N, Arora H, Sorokin A, et al. Efficient unsupervised learning for localization and detection in object categories. Neural Information Processing Systems (NIPS), 2006, 18:811.

[115] Tuytelaars T, Lampert C, Blaschko M, et al. Unsupervised object discovery: a comparison. International Journal on Computer Vision (IJCV), 2010, 88(2):284–302.

[116] Rosenfeld A, Weinshall D. Extracting foreground masks towards object recognition. Proceedings of IEEE International Conference on Computer Vision (ICCV), 2011. 1–8.

[117] Isola P, Xiao J, Torralba A, et al. What makes an image memorable? Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. 145–152.

[118] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models. Proceedings of ACM Special Interest Group on Information Retrieval (SIGIR), 2003. 119–126.

[119] Torresani L, Szummer M, Fitzgibbon A. Learning query-dependent prefilters for scalable image retrieval. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009. 2615–2622.

[120] Lampert C. Detecting objects in large image collections and videos by efficient subimage retrieval. Proceedings of IEEE International Conference on Computer Vision (ICCV), 2009. 987–994.

[121] Thayananthan A, Stenger B, Torr P, et al. Shape context and chamfer matching in cluttered scenes. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, 2003. 127–133.

[122] Bai X, Li Q N, Latecki L J, et al. Shape band: a deformable object detection approach. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009. 1335–1342.

[123] Mori G, Belongie S, Malik J. Shape contexts enable efficient retrieval of similar shapes. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2001. 723–730.

[124] Zhang D, Lu G. Review of shape representation and description techniques. Pattern Recognition, 2004, 37(1):1–19.

[125] Li H, Ngan K N. A co-saliency model of image pairs. IEEE Transactions on Image Processing (TIP), 2011, 20(12):3365 –3375.

[126] Fergus R, Perona P, Zisserman A. A visual category filter for google images. Proceedings of European Conference on Computer Vision (ECCV), 2004. 242–256.

[127] Ben-Haim N, Babenko B, Belongie S. Improving web-based image search via content based clustering. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), 2006. 106–106.

[128] Cui J, Wen F, Tang X. Real time google and live image search re-ranking. Proceedings of ACM Multimedia, 2008. 729–732.

[129] Popescu A, Moëllic P, Kanellos I, et al. Lightweight web image reranking. Proceedings of ACM Multimedia, 2009. 657–660.

[130] Flickner M, Sawhney H, Niblack W, et al. Query by image and video content: the QBIC system. Computer, 1995, 28(9):23–32.

[131] Zhang D S, Lu G J. Shape-based image retrieval using generic Fourier descriptor. Signal Processing: Image Communication, 2002, 17(10):825–848.

[132] Charpiat G, Faugeras O, Keriven R. Shape statistics for image segmentation with prior. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007. 1–6.

[133] Peter A, Rangarajan A, Ho J. Shape l'ane rouge: sliding wavelets for indexing and retrieval. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008. 1–8.

[134] Soma Biswas G A, Chellappa R. An efficient and robust algorithm for shape indexing and retrieval. IEEE Transactions on Multimedia, 2010, 12:371–385.

[135] Igarashi T, Moscovich T, Hughes J F. As-rigid-as-possible shape manipulation. ACM Transactions on Graphics (SIGGRAPH), 2005, 24(3):1134–1141.

[136] Schaefer S, McPhail T, Warren J. Image deformation using moving least squares. ACM Transactions on Graphics (SIGGRAPH), 2006, 25(3):533–540.

[137] Karni Z, Freedman D, Gotsman C. Energy-based image deformation. Computer Graphics Forum, 2009, 28(5):1257–1268.

[138] Lalonde J F, Hoiem D, Efros A A, et al. Photo clip art. ACM Transactions on Graphics (SIGGRAPH), 2007, 26(3):3:1–10.

[139] An X, Pellacini F. AppProp: all-pairs appearance-space edit propagation. ACM Transactions on Graphics (SIGGRAPH), 2008, 27(3):40: 1–9.

[140] Xu K, Li Y, Ju T, et al. Efficient affinity-based edit propagation using K-D tree. ACM Transactions on Graphics (SIGGRAPH Asia), 2009, 28(5):118:1–118:6.

[141] Hoiem D, Efros A A, Hebert M. Automatic photo pop-up. ACM Transactions on Graphics (SIGGRAPH), 2005, 24(3):577–584.

[142] Landes P E, Soler C. Content-aware texture synthesis. Research Report RR-6959, INRIA, 2009. http://hal.inria.fr/inria-00394262.

[143] Pauly M, Mitra N J, Wallner J, et al. Discovering structural regularity in 3D geometry. ACM Transactions on Graphics (SIGGRAPH), 2008, 27(3):43:1–11.

[144] Thayananthan A, Stenger B, Torr P H S, et al. Shape context and chamfer matching in cluttered scenes. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2003. I: 127–133.

[145] Kilthau S L, Drew M S, Moller T. Full search content independent block matching based on the fast Fourier transform. Proceedings of IEEE International Conference on Image Processing (ICIP), 2002. I: 669–672.

[146] Ho J, Peter A, Rangarajan A, et al. An algebraic approach to affine registration of point sets. Proceedings of IEEE International Conference on Computer Vision (ICCV), 2009. 1–8.

[147] Sapiro G, Kimmel R, Caselles V. Geodesic active contours. Proceedings of IEEE International Conference on Computer Vision (ICCV), 1995. 694–699.

[148] Bookstein F. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Recognition and Machine Intelligence (PAMI), 1989, 11(6):567–585.

[149] Cho T S, Butman M, Avidan S, et al. The patch transform and its applications to image editing. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008. 1–8.

[150] McCann J, Pollard N S. Local layering. ACM Transactions on Graphics (SIGGRAPH), 2009, 28(3):84:1–7.

致　谢

衷心感谢我的导师胡事民教授对我在科研上精心指导；在职业选择上细心规划；更在为人处世上为我树立榜样。胡老师对我方方面面的指导，让我受益良多。同时感谢我的联合导师，腾讯研究院院长郑全战博士对我的指导和帮助。与工业界的接触和沟通，让我能更好的把握社会和国家的重大需求，做有用的研究。

感谢实验室的张松海、徐昆、杨永亮、丁濛、卢少平、李勇、杜松沛、李先颖、沈超慧等同学。尤其感谢一起合作项目的陈韬、张国鑫、张方略，以及伦敦大学学院的Niloy J. Mitra 博士、英国卡迪夫大学的Ralph R. Martin教授、Lehigh大学的XiaoLei Huang博士。

本文的研究工作受到核高基项目、国家973计划项目、国家863计划项目和国家自然科学基金项目的资助，特此致谢。

感谢我的家人和所有帮助过我的老师和同学！

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____ 日　期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1985 年 9 月 5 日出生于陕西省乾县。

2003 年 9 月考入西安电子科技大学计算机系计算机科学与技术专业, 2007 年 7 月本科毕业并获得工学士学位。

2007 年 9 月免试进入清华大学计算机系攻读博士学位至今。

发表的学术论文

[1] **Ming-Ming Cheng** and Guo-Xin Zhang, Connectedness of Random Walk Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2011, 33(1):200-202. (SCI源刊, 检索号: 681AC, 影响因子5.308.)

[2] **Ming-Ming Cheng**, Fang-Lue Zhang, Niloy J. Mitra, Xiaolei Huang and Shi-Min Hu. RepFinder: Finding Approximately Repeated Scene Elements for Image Editing. ACM Transactions on Graphics (SIGGRAPH), 2010, 29(4):83:1-8. (SCI源刊, 检索号: 624GZ, 影响因子3.632.)

[3] **Ming-Ming Cheng**, Guo-Xin Zhang, Niloy J. Mitra, Xiaolei Huang and Shi-Min Hu. Global Contrast based Salient Region Detection. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, 409-416. (EI源刊, 检索号: 20113814351657.)

[4] Tao Chen, **Ming-Ming Cheng**, Ping Tan, Ariel Shamir, Shi-Min Hu. Sketch2Photo: Internet Image Montage. ACM Transactions on Graphics (SIGGRAPH Asia), 2009, 28(5):124:1-10. (SCI源刊, 检索号:540EY, 影响因子3.619.)

[5] Youyi Zheng, Xiang Chen, **Ming-Ming Cheng**, Kun Zhou, and Shi-Min Hu, Niloy J. Mitra. Interactive Images: Proxy-based Scene Understanding for Smart Manipulation. ACM Transactions on Graphics (SIGGRAPH), 2012, 31(4):1-10. (SCI源刊, 检索号: 624GZ, 影响因子3.632.)

[6] Guo-Xin Zhang, **Ming-Ming Cheng**, Shi-Min Hu and Ralph R. Martin. A Shape-Preserving Approach to Image Resizing. Computer Graphics Forum, 2009, 28(7):1897-1906. (SCI源刊, 检索号:506ME, 影响因子1.681.)

[7] Fang-Lue Zhang, **Ming-Ming Cheng**, Jiaya Jia and Shi-Min Hu. ImageAdmixture: Putting Together Dissimilar Objects from Groups. IEEE Transactions on Visualization and Computer Graphics (TVCG), 2012, 18, to appear. (SCI源刊, 检索号:175HZ, 影响因子1.922.)

已授权国家发明专利

[1] 胡事民, 程明明, 张国鑫, 基于色彩直方图和全局对比度的图像视觉显著性计算方法, 专利号：ZL201110062520.1.

[2] 胡事民, 程明明, 张方略. 一种基于轮廓带图的相似单元检测方法, 专利号: ZL201010159931.8, (**PCT国际专利申请号: PCT/CN2011/000691**).

[3] 胡事民, 程明明, 张国鑫, 一种基于共形能量的内容敏感图像缩放方法, 专利号: ZL200910092756.2.

[4] 胡事民, 程明明, 陈韬, 张松海, 一种基于草图的网络图元自动提取方法和系统, 专利号: ZL200910081069.0.

[5] 胡事民, 程明明, 陈韬, 张松海, 基于卡通片的高质量线结构提取方法, 专利号: ZL200810106664.0.

[6] 胡事民, 陈韬, 程明明, 张松海, 基于图像库的图像合成质量自动评测方法, 专利号: ZL200910086937.4.

[7] 胡事民, 陈韬, 程明明, 张松海, 基于混合梯度场和混合边界条件的图像合成方法和装置, 专利号: ZL200910084769.5.

[8] 胡事民, 张一飞, 程明明, 视频像素可伸缩性的计算方法, 专利号: ZL200810114466.9.

参与的科研项目

[1] 国家科技重大专项, 新一代搜索引擎关键技术研究, 2011-01-01~2011-12-31.

[2] 国家973计划, 网络可视媒体的交互与合成, 2011-01-01~2015-12-31.

[3] 国家973计划, 可视媒体的交互与融合处理, 2006-09-01~2011-08-31.

[4] 国家自然科学基金重点项目, 基于认知模型的图像不变性特征理论及其应用研究, 2011-01-01~2014-12-31.

[5] 国家自然科学基金面上项目, 基于结构分析的视频卡通风格绘制技术研究, 2010-1-1~2012-12-31.

攻读博士学位期间的获奖情况

[1] Google PhD Fellowship, 颁奖部门: Google公司, 2010 年. (2010年全球共31名博士生获奖)

[2] IBM PhD Fellowship, 颁奖部门: IBM公司, 2011年.

[3] 博士研究生学术新人奖, 颁奖部门: 教育部, 2010年.

[4] 清华大学学术新秀提名奖, 颁奖部门: 清华大学, 2011 年.

[5] 清华大学计算机系学术新秀, 颁奖部门: 清华大学, 2010 年.

[6] 计算机图形学教学软件, 清华大学优秀教学软件一等奖, 颁奖部门: 清华大学, 2008 年, 第四获奖人.