

分类号 TP391.41 密级

UDC

西 安 交 通 大 学

硕 士 学 位 论 文

图像的视觉信息表象理论及其数学基础

王 郑 耀

指导教师姓名 程正兴 教 授 西安交通大学

申请学位级别 硕 士 专业名称 计算数学

论文提交日期 2006 年 2 月 论文答辩日期 2006 年 3 月

答辩委员会主席

评 阅 人

2006 年 3 月

**Visual Information Representation Theory of Image
and It's Mathematical Basis**

Dissertation Submitted to

Xi'an Jiaotong University

in partial fulfillment of the requirement

for the degree of

Master of Science

By

Wang Zhengyao

(Computation Mathematics)

Dissertation Supervisor: Prof. Cheng Zhengxing

March, 2006

图像的视觉信息表象理论及其数学基础

学科专业： 计算数学

学位申请人： 王郑耀

指导教师： 程正兴 教授

摘要

图像处理技术目前有了十分广泛的研究和应用。本文首先简单介绍了与图像处理相关的一些理论和技术：视觉计算理论、表象学、图像识别和分析技术、多媒体内容描述界面 / 接口 (MPEG-7) 和基于内容的图像搜索引擎技术。这些理论和技术存在一个共同的基本问题：如何描述图像中的视觉信息，也就是图像的视觉信息表象理论。

图像的视觉信息表象理论是研究如何把图像中的视觉信息表达清楚的形式化系统和说明这个系统如何获得、描述、重构图像中视觉信息的若干规则。视觉包含有三种感觉：形状、纹理和颜色。图像的视觉信息表象理论就是寻找恰当的这三种信息的形式化描述。本文给出了图像视觉信息表象理论的基本体系结构。

接着，本文在视觉感知理论的基础上，使用了一系列数学定理和数值试验，分析了多尺度零交叉点和小波变换系数的数学性质，说明：多尺度零交叉点是形状表象的合适选择；图像的小波变换系数是纹理表象的合适选择。

然后，本文说明了：零交叉点检测中的很多滤波器都可以看作是一个广义的小波函数。因而，零交叉点检测可以通过对小波变换后的数据做进一步的逻辑推断得到。因此，形状表象和纹理表象都可以在小波分析的框架下一致的得到，并给出了一个合适的模型。基于这样一个模型，可以将纹理看作是图像的复杂性和视觉系统能力有限性之间的平衡：如果图像区域中零交叉点充分多，视觉系统就使用统计方法进行区域描述，而不是仔细精确的形状描述。

最后，本文设计了图像视觉信息表象理论的实验平台：基于现有搜索引擎的特定图像检索软件 (SSIS)。利用这个平台，可以使用互联网上海量图像数据测试理论的可行性。

关键词： 计算视觉, 图像处理, 小波分析, 边缘检测, 形状表象, 纹理分析, 图像检索

论文类型： 应用基础

Visual Information Representation Theory of Image and It's Mathematical Basis

Major: Computation Mathematics

Candidate: Wang Zhengyaoyao

Supervisor: Prof.Cheng Zhengxing

Abstract

Image processing is important in many applications. This thesis firstly introduced some theories and techniques concerning image processing: Computing Vision, Representation Theory, Techniques of Image Recognition and Analysis, Multimedia Content Description Interface (MPEG-7) and Image Search Engine based on Image Content. All fields above have a common problem: how to describe visual information in image, i.e. visual information representation theory of image.

The visual information representation theory of image is a formal system to study the method of distinct description of visual information(shape, texture and color)in image and the rules of getting, describing and reconstructing visual information in image.

Based on visual perception, this thesis used mathematical theorems and numerical experiments to analyze the properties of zero-crossing points and wavelet transformation coefficients of the image and give a conclusion: multi-scale zero-crossing points is the appropriate shape representation and wavelet coefficients is the appropriate texture representation.

Because many filters in zero-crossing points detection algorithms can be considered as wavelet function, zero-crossing points can be detected by logical inference of wavelet coefficients. So, shape and texture representation can be detected using wavelet transformation, and this thesis gives a model to describe it. Based on this model,texture can be considered as the balance of the complexity of image and the capability of vision system.

In the final, this thesis designed an experimental platform, software of special image search based on current search engine (SSIS), to test the representation theory, algorithms and descriptions.

Key words: computing vision, image processing, wavelet analysis, edge detection, shape representation,texture analysis,image search

Thesis type: Application Foundation

目 录

第一章 绪论	1
§1.1 引言	1
§1.2 前期工作回顾	3
§1.3 研究方法	4
§1.4 本文的主要工作	8
第二章 图像视觉信息表象理论的背景及重要性	11
§2.1 视觉计算的简单模型	11
§2.1.1 视觉的生理学基础	11
§2.1.2 视觉心理学	14
§2.1.3 视觉计算的数学模型	20
§2.1.4 结论	27
§2.2 表象学	28
§2.2.1 视觉表象的形成和记忆	28
§2.2.2 视觉表象的内注意	30
§2.2.3 视觉表象的对照功能	30
§2.2.4 视觉表象的分拆和组合功能	30
§2.2.5 视觉表象的运用功能	30
§2.2.6 视觉表象的连接功能	31
§2.2.7 表象学研究 with 图像表象理论	31
§2.3 图像识别与图像分析	32
§2.3.1 机器视觉的一般技术	33
§2.3.2 识别技术	33
§2.3.3 图像表象在图像识别和分析中的重要作用	36
§2.4 多媒体管理系统与 MPEG7	36
§2.4.1 MPEG7 简介	37
§2.4.2 MPEG7 中描述视觉特征的描述符与图像表象	37
§2.5 基于内容的图像和视频搜索:WWW 图像检索的发展目标	39
§2.5.1 搜索引擎	39
§2.5.2 现有的图像搜索引擎技术	39
§2.5.3 图像搜索引擎发展方向	40

§2.5.4	图像搜索引擎与图像的视觉表象	42
第三章	图像的视觉信息表象理论	43
§3.1	图像表象理论的体系结构	43
§3.2	图像表象理论的研究现状	45
§3.3	二值表象	46
§3.3.1	二值区域的 Blob 描述	47
§3.3.2	二值区域的角映射系数描述	47
§3.3.3	二值区域的多尺度曲率空间描述	49
§3.3.4	二值区域的骨架描述	51
§3.3.5	图像二值表示的缺点	52
§3.4	图像表象理论的数学基础	52
第四章	图像的形状表象	53
§4.1	形状表象的基本要求	53
§4.2	零交叉点是重要的图像特征	53
§4.2.1	图像的复杂性	53
§4.2.2	零交叉点是图像中重要的特征	60
§4.2.3	小结	61
§4.3	零交叉点的完备性	61
§4.3.1	二阶方向微分和图像的拉普拉斯变换	63
§4.3.2	人眼的辨别力和图像的拉普拉斯变换	64
§4.3.3	Logan 定理	65
§4.3.4	Carlsson 的数值试验	66
§4.3.5	函数的指纹定理	68
§4.3.6	一类有限带宽的二维信号的零交叉点的完全性定理	69
§4.3.7	Hummel-Moniot 定理	69
§4.3.8	零交叉点的局限性	71
§4.3.9	Elder 数值试验	72
§4.3.10	结论: 多尺度零交叉点是合适的形状表象	72
§4.4	多尺度零交叉点检测算法	75
§4.4.1	边缘是指具有特定模式的灰度变化的点	75
§4.4.2	边缘模式	80
§4.4.3	边缘检测算法	89
§4.5	多尺度零交叉点对图像视觉信息的表象	92

§4.5.1	分层次的形状表象	92
§4.5.2	零交叉点的封闭性	92
§4.5.3	基于轮廓的颜色和纹理填充	93
§4.6	轮廓线的描述和识别	94
§4.6.1	多尺度轮廓线匹配	94
§4.6.2	自由曲线表示	95
§4.6.3	特殊的表示	96
§4.6.4	小结	96
第五章	图像的纹理表象	97
§5.1	纹理的统计描述	97
§5.1.1	视觉系统的纹理感知	97
§5.1.2	纹理表象与统计和随机	102
§5.1.3	统计分析的是什么?	103
§5.2	小波变换系数作为纹理表象的优越性	103
§5.2.1	函数的小波变换	104
§5.2.2	小波变换的频率分解能力	115
§5.2.3	小波变换在统计分析中的优点	115
§5.3	纹理的小波表象的描述	119
§5.3.1	子图内的统计量	119
§5.3.2	其它统计量	120
第六章	边缘检测和纹理特征提取的统一	123
§6.1	边缘检测的微分滤波器一般是小波函数	123
§6.2	形状和纹理的提取统一为小波分析	127
§6.2.1	图像区域的复杂性度量	128
§6.2.2	视觉系统的 Gabor 模型	133
第七章	图像表象的简单应用：基于内容的特定图像检索	135
§7.1	需求分析	135
§7.2	基于现有搜索引擎的特定内容图像检索方案设计	137
§7.2.1	基于现有搜索引擎的检索	137
§7.2.2	SSIS 中使用的搜索引擎	137
§7.2.3	总体方案设计图	138
§7.3	SSIS 中各模块实现的简单说明	139

§7.3.1	GoogleImage 关键字检索网页的获取	139
§7.3.2	网页分析	139
§7.3.3	服务器和图像链接的数据库管理	139
§7.3.4	图像下载	143
§7.3.5	图像读写库	143
§7.3.6	基本过滤	143
§7.3.7	图像表象的提取模块	144
§7.3.8	表象的 XML 语言描述	144
§7.3.9	图像中特定内容检测算法	144
§7.3.10	界面设计	146
第八章	结论与展望	149
	致谢	151
	参 考 文 献	153
附录 1	边缘检测 COM 组件: EdgeCom	163
附录 2	攻读硕士期间的主要成果	165

插图目录

1.1 一种典型研究方法	7
2.1 是花瓶还是侧面像	14
2.2 人类视觉系统具有非凡的能力	15
2.3 接近律	15
2.4 连续律	16
2.5 封闭律	16
2.6 Werner 实验	17
2.7 基于边缘的填充过程	18
2.8 轮廓在纹理感知中的作用	18
2.9 长度错觉	19
2.10 大小错觉	19
2.11 Kanizsa 三角形	19
2.12 信息处理过程	20
2.13 视觉信息处理过程	21
2.14 从图像到物体表象的过程	23
2.15 零交叉点检测	24
2.16 零交叉点检测的分组	24
2.17 启发式的连接	25
2.18 客观世界的约束和视觉器官的选择性	26
3.1 图像的视觉信息表象理论	44
3.2 图像的二值表象	46
3.3 曲线的尺度变换	50
3.4 二值图像的骨架	51
4.1 表面的反射与光源直射	55
4.2 光线引起的灰度变化	56
4.3 边缘模糊的原因	56
4.4 卷积效应	59
4.5 卷积效应	59
4.6 卷积效应	59
4.7 图像中边界的复杂性	60

4.8	零交叉点和导数极值点	61
4.9	边缘对图像信息的刻画	62
4.10	二阶方向微分和拉普拉斯变换	63
4.11	图像的压缩表示	67
4.12	作为函数指纹的零交叉点	69
4.13	证明示意图	70
4.14	多尺度零交叉点相同的两个函数	71
4.15	基于边缘的图像编码及其重构	73
4.16	基于边缘的图像编码及其重构	73
4.17	基于边缘的图像编码及其重构	74
4.18	灰度变化的模式	76
4.19	灰度变化的方向	77
4.20	灰度变化的尺度	78
4.21	边缘模式	79
4.22	边缘产生的物理原因	81
4.23	各种边缘	82
4.24	边缘模式形态: 理想的阶梯形边缘和斜坡型边缘	83
4.25	边缘模式形态: 线型边缘	83
4.26	边缘模式形态: 山脊型和屋顶型	84
4.27	各种类型的角点	84
4.28	阶跃边缘	85
4.29	斜坡型边缘模型	86
4.30	线型边缘的模型	86
4.31	L 型角点的模型的图像	87
4.32	Y 型角点的模型的图像	87
4.33	X 型角点的模型的图像	88
4.34	边缘闭合技术	93
4.35	封闭轮廓线	93
4.36	基于 Hausdroff 距离识别算法的演示	96
5.1	纹理图像的例子	98
5.2	人类主观对纹理的感受	98
5.3	纹理感知中的方向	99
5.4	纹理分割	100

5.5	纹理、零交叉点与尺度	101
5.6	个别的基元不称为纹理	102
5.7	纹理表象的基本框架	103
5.8	高斯函数及其离散化表示	106
5.9	连续图像及其离散化表示	107
5.10	连续卷积的离散化	107
5.11	图像的多尺度表示	108
5.12	图像的 Mallat 分解	112
5.13	图像的 Mallat 分解	113
5.14	图像的 Mallat 分解	114
5.15	小波的去相关能力	118
5.16	图像的离散小波变换得到的子图	119
6.1	常见的边缘检测滤波器	124
6.2	Roberts 模板在小波框架下的理解	125
6.3	Sobel 模板在小波框架下的理解	126
6.4	Prewitt 模板在小波框架下的理解	126
6.5	获取形状表象和纹理表象的统一	127
6.6	在一点处图像像素的结构	131
6.7	图像尺度空间中视觉信息度量	132
7.1	含有感兴趣内容的图像	136
7.2	AltaVista 搜索引擎界面	138
7.3	SSIS 的总体方案合计: 基于 Web 检索	140
7.4	SSIS 的总体方案合计: 基于 Image 检索	141
7.5	SSIS 中特定内容的图像检索模块	142
7.6	视觉表象的 XML 语言描述	145
7.7	圆环检测算法	145
7.8	圆精确定位	146
7.9	SSIS 的界面	147

第一章 绪论

§1.1 引言

伴随着计算机视觉和图像分析的发展，图像处理技术越来越受到大家的重视。图像处理目前有了十分广泛的研究和应用，例如：国防、工业、医疗等等。

总的来说，图像处理的这些研究的目的是为了服务以下三个方面：图像压缩编码、计算机视觉和图像分析。所有这三个方面的一个共同点是：它们都是从原始图像出发，经过一系列过程，获得对原始图像的一个新的描述。图像编码的目的是获得数据量少的一个全新的数据描述；图像分析的目的是对图像本身或者图像中的内容做出一个判断或者测量。

人类视觉的目的是从眼睛得到的图像中获得关于世界的符号化描述，例如“这里有一张桌子，前面坐着一个人，这个人李四”。在这个过程中，人的视觉系统必须要建立现实世界的模型，然后才能实现认知和推理。和生物视觉系统一样，计算机视觉系统的目的也是建立现实世界的模型，并且利用这些表示与现实世界进行交互。我们无法直接体验或者感知物理世界，感知系统必须在任何时候立刻创建世界或者环境的表示模型，所使用的信息是当前感知器官感知到的信息和以往的经验。事实上，正是大脑中这个世界和环境的模型使我们能够进行智能的各个部分，感知、认知、推理等的交互。换句话说，我们之所以能思考，是因为大脑中具有所思考物体的模型。这些物体的模型是通过视觉、触觉等感觉器官建立的。

本文讨论的是计算机视觉和图像分析，而不讨论图像压缩编码。在计算机视觉和图像分析中，实际上有三大类重要的描述或者模型。

第一类是物体的模型（本文只考虑与视觉有关的表象，物体的味道和温度是嗅觉和触觉获得的，也是物体表象的内容，但是这里不考虑）。即使用形式化模型来描述视觉感受到的物体。对物体的这种模型必须是丰富而又仔细的。这些模型应该使得我们能区分数以亿计的三维物体，而且这个模型应该能表示物体之间的细微差别和局部细节，另外也需要分层的和粗糙的模型。正是利用物体的模型，我们才可以在大脑中建立外部世界的三维模型，实现认知和推理，而这些都是智能的基础。

感知的一个目的是紧凑表示过去的经验，并且在需要的时候有效地恢复。为了能获得紧凑和自由的表示，物体的模型必须是结构化的。结构化的模型使用不同部分的参数表示。在恢复中使用索引格式。更重要的是，在每一个瞬间，系统建构的关于世界的知识是不完全的、不准确的和不确定的。模型必须满足这些不完全性和不确定性。

第二类是对知识和经验的描述。在建立世界或者环境模型的过程中，系统使用

了物体的知识，感知器的知识（近视眼）和系统应用领域的作用和位置知识。所有这些知识源的信息表示的本质和信息量的多少都是不同的。显然，信息的描述是视觉系统最重要的问题之一。生物具有的知识和经验一部分是由进化和遗传保留在个体基因中的先验知识，还有一部分是后天学习得到的。

在机器视觉中，系统也有很多先验的知识：系统的应用环境、相机标定信息和颜色校正信息等等。计算机的分析和推理依赖于这些先验知识。

第三类是图像的描述。物体的模型是通过视觉系统和大脑创建的，是视觉系统利用信息和经验对眼睛获得的图像进行深入加工分析的结果。因此，在获得物体的三维模型之前，视觉系统首先要建立二维图像的形式化描述。图像是三维物体的二维投影，对物体的表象是在图像的表象（多个图像表象）的基础上得到的，即利用一个或者多个图像表象和知识 / 经验建立物体的模型。

我们将这三类重要的描述或者模型可以统称为表象：物体的表象，知识和经验的表象以及图像的表象。表象，是指一种能把某种实体或某几类信息表达清楚的形式化系统，以及说明该系统如何行使其职能的若干规则。

显然，第三类表象是最基本的，是直接与视觉器官相关的。生理学和心理学实验表明，可以认为视觉系统分为两部分，第一部分也就是视觉前端建立了眼睛获得的外部图像的表象，这是一个无意识的自然的过程，是视觉前端自动完成的。在此基础上，视觉系统的第二部分根据知识 / 经验主动的加工处理，以创建外部世界的模型，也就是我们前面说的第一类表象，有了第一类表象后，才能获得更多的知识和经验。

本文研究的是第三类表象，也就是图像的表象。主要基于下面的几个原因：首先，图像的表象是视觉系统的基础，是建立物体的表象的先决条件。其次，图像的表象是视觉前端无意识自然的创建过程，所以受到人的心理活动的影响小，因此研究起来相对简单，可以看做是一个纯粹的数学建模过程。最后，第二章将会看到，图像表象是一系列应用的基础，具有广泛的应用前景。

视觉前端为了建立图像的表象，需要进行一系列操作。通过这些操作，逐渐从原始图像过渡到图像的表象述 / 模型，在这些过渡中有一些中间环节。原始图像、这些中间环节也可以称为图像的表象。原始彩色图像或者灰度图像本身就是一个表象。二值图像、边缘图像和小波变换后的数据等，都是从图像变换而来的，对图像中的某些实体或者信息做了形式化的表达，因此也都构成了原始图像的表象。

本文的目的就是说明什么是图像的视觉表象，它有什么重要性；寻找可以充当视觉表象的合适数据，并使用数学说明这些数据为什么可以作为图像的视觉表象；这些表象应该如何描述。

§1.2 前期工作回顾

与图像的视觉表象相关的领域非常多，它们都已经发展了很长时间，取得了很多成就。下面简要回顾一下这些领域已经取得的成就和存在的问题。

1. 视觉生理学、心理学和视觉计算理论。它们从三个不同的方面研究人类视觉，视觉生理学研究生物系统是如何实现视觉感知的；心理学研究视觉系统具有的普遍特性；视觉计算理论是在视觉生理学和视觉心理学基础上对视觉系统的建模，希望通过一系列计算过程还原视觉系统。视觉心理学很早就有人在研究，但是因为缺乏对人的大脑的了解，研究一直没有取得突破性的进展，引人注目的就是二十世纪初兴起的格式塔学派。视觉生理学是随着解剖学的发展而开始的，已经取得了很多实验结果，但是对这些实验结果还缺乏一定深度的理解和理论上的解释。视觉计算是十分复杂的问题，在计算机出现后，人们才有能力去模仿视觉计算过程。二十世纪六七十年代兴起了视觉计算理论的研究，这些研究集中体现在二十世纪八十年代初出现的一些理论和方法，例如视觉计算理论、多尺度理论等。但是，从那时候起直到现在，计算机视觉领域再也没有出现过激动人心的结果。计算机视觉在极其缓慢的发展着。

2. 图像处理、分析和机器视觉。在生产和生活中有很多问题，需要去客观的观察和记录下来，并希望利用这些观察结果迅速作出反馈。这些工作一开始是由人来完成的，随着生产线的运行速度的提高，工作量越来越大，迫切需要使用摄像头和计算机代替人进行处理和分析。例如，文字识别、条码技术、缺陷检测、指纹识别和人脸检测等。这些需求大大刺激了图像处理和研究的分析热情。对一些具体问题，已经取得了非常可喜的成就，例如条形码识别和计算机辅助医疗。但是，缺乏一般意义的表示和算法。

3. 多媒体应用技术和搜索引擎。最近十几年来，随着互联网络和数字成像技术的发展，图像、视频等多媒体资料越来越多，形成了海量的视觉信息。对这些信息需要有理想的压缩存储、管理和检索方法。图像压缩编码已经形成了一系列方案，具有代表性的是 MPEG 压缩标准。但是，令人满意的视觉信息检索方法还没有出现。

4. 数学。作为形式化的工具，数学可以描述、分析上述各个领域中所使用的算法。主要用到了数字信号处理（傅里叶变换）、函数分析、微分几何、统计学、随机过程、拓扑学、小波分析、微分方程理论、图论等等。其它领域中已经获得的数学成果、数学结论和数值计算方法，被有效地应用到图像处理和视觉计算过程中，大大促进了视觉计算领域的发展。但是，数学所能分析和描述的是研究者所提出的模型中限定的问题，在建模过程中基于特定哲学观念和试验结果的假设往往被

忽略了（这些哲学观念往往就是视觉心理学和生理学中领悟到的）。很多所谓的最优边缘检测算法，都只是在特定的模型下的最优。离开那些假设，它可能没有任何作用，所谓的最优也就失去了意义。图像的复杂性使得大部分简单的数学模型都有很大的局限性。新出现的一些算法，使用了一些非常深刻的数学理论，例如微分几何、尺度空间理论和小波分析。从目前看，使用这些理论的算法表现出了一些优势。但是，由于缺乏相应的数学基础，一般的工程技术人员很难使用这些算法。这就需要数学工作者具有相应的技术背景和程序开发经验，将这些技术应用到实际问题中。

本文就是在这些领域现有成果的基础上完成的。为了找到合理的视觉表象，需要结合视觉生理学和心理学的结果给出一些可能的候选。然后需要使用数学去分析它们是否满足实际问题的需要，例如是否满足一般性，是否具有完备性。最后，还需要在视觉计算理论、机器视觉、多媒体等领域的实践验证它们的可靠性，并提出可改进的意见。

本文可以看作是这些研究的在计算机视觉方面的系统化总结。这个系统化的总结对计算机视觉、图像分析和图像处理等相关的研究具有十分重要的意义。

§1.3 研究方法

本文所立足的领域是计算机视觉和图像分析，它们诞生于二十世纪四五十年代，到目前已经有五十多年的历史。尽管如此，计算机视觉和图像分析的研究目前还非常不成熟。原因很简单，计算机视觉的目的是使得机器能像人一样能“看”，所以计算机视觉参考的对象是人的视觉。而人的视觉是与人的智能有关的，我们之所以看到眼前的桌子，很大程度上是因为我们的大脑，而不仅仅是我们的眼睛。但是，众所周知，对人的智能和大脑的机理我们还不是非常清楚。

我们说计算机视觉和图像分析研究不成熟，是基于计算机视觉和图像分析研究中的一系列不正常的现象。1992年，Jain发表了一篇著名的文章：《计算机视觉系统中的无知、短视和天真》[1]。在这篇文章中，他们总结了计算机视觉研究中出现的一些问题。尽管十几年过去了，计算机视觉和图像分析出现了很多新的研究专家、新的研究方法和新的算法。但是研究中出现的问题依然和十几年前是类似的。下面我们就列举这篇文章中指出的三大基本问题：

1. 在计算机视觉研究的基本哲学上，患有“路灯下丢了钥匙的醉汉”和“皇帝的新装”综合症。

“我们常常是将大学里学过的一些好的工具应用到计算机视觉中，而不是认真研究计算机视觉中我们面对的问题然后发展合适的工具。这些工具包括数学方法、数据结构和其它方法。一般的说，我们想办法把我们从其他学科中学到的工具应用

到计算机视觉中，但是并没有认真的分析这样做是不是合适。我们的方法很多时候是工具驱动的。于是，如同一个喝醉的人一样，我们用已经发展起来的工具来寻找我们问题的答案。如果这个工具对这个问题和应用是合适的，就认为是好工具。为了寻找合适的工具或者改善它们，我们应该做实验，发展一些试验方法。如果我们不这样做，我们仍然在困惑的路灯下，寻找丢失在很远黑暗中丢失的钥匙。显然，大家都知道，要想找到钥匙，我们应该照亮我们丢失钥匙的地方，甚至在黑暗中寻找都要比在路灯下寻找更有价值。

每一个领域都有其时尚和流行的东西。Thomas Kuhn 提出，科学发展过程是一系列的相继的范式。在某一个时刻，某一个范式很流行，在这个时候，情况就类似于皇帝的新装。大部分人都相信这个范式，都不敢挑战它。即使有人感觉到有问题（其实大家都发现这个问题），他们也认为自己错了，所以他们仍然说他们看到的和大家（某一个著名的专家）是一样的。需要一个天真的儿童，纯真无邪的看到并且说：皇帝没有穿衣服（也就是专家希望大家相信的东西是错误的）。计算机视觉中这种情况很容易出现。计算机科学、心理学、神经科学、认知科学、电子工程和数学中都有研究者使用它们的工具来解决计算机视觉问题。如果一篇文章的作者是一个来自著名实验室的著名的研究者，并且使用了复杂的数学理论，计算机科学中的评论家，如果他没有听到这个方法，他要么相信这个方法能解决问题，要么干脆就默不作声。相似地，如果一个心理学家提出他的新的人类视觉的理论，计算机视觉中的研究者就会获得启发并且建立给予这个理论的系统，但是当这个系统实现的数据结构设计出来的时候，也许心理学中这个理论已经不流行了。Kuhn 指出，在大部分科学学科中，真正的进步发生在一个估计的范式被推翻。计算机视觉中推翻范式的进程非常慢，这就告诉我们这是死胡同。”

2. 在计算机视觉算法体系上，虽然研究了很多复杂的、高级的问题，但是很多基本问题没有解决，例如图像分割和边缘检测问题。另外，对很多基本的算法和滤波器的行为，我们还缺乏深刻的了解。

“计算机视觉系统中首先必须进行的操作中的一个便是将物体从背景中分离开来，这个问题一般称为分割问题。分割可以利用像素之间的相似性来进行，也可以利用图像中具有某些特性的不连续点来进行。显然，可能物体的模型可以帮助进行分割，但是在很多情况下我们不知道场景中的物体是什么。所以，一般不使用知识进行分割的，但是使用这些知识可以增强分割结果。

经过早期一段时间的活跃的研究之后，分割问题常常被忽略了。早期的大部分方法是基于简单的模型：物体表面的灰度特征。这些方法只有在物体比较简单并且图像灰度变化较小的情况下才能成功。基于知识的方法要么借助于物体的一般知识，要么是肤浅的图像知识。很少有人使用物体的三维模型和成像过程来提高图像

分割的效果。由于分割方法的局限性，取得的成就不大。但是，我们相信这是计算视觉中一个重要的问题。

过去十几年中研究者提出了很多问题，例如说 Shape From X 技术。这个问题的暗暗的假设要么图像分割问题已经解决，要么没有图像分割问题。但是，这两种假设都是错误的。在第一个假设下，几种不同的模块最终导致问题十分困难。第二个假设更是致命的，这个假设认为整个图像就是一个曲面。满足这一假设的图像是很稀少的并且即使有也是我们不感兴趣的。在这个假设的基础之上，发展起来的很多技术都与计算机视觉任务没有任何关系。

显然，物体的信息可能帮助分割。这个方向的努力基于早期过程中弱的、一般目的的操作。可惜的时，对这些操作的行为我们不理解。很难设计依赖于我们不是很理解的，性能不好的底层视觉处理的视觉系统。很多研究者强调了图像计算多分辨或者从图像计算多重内在性质的数据，然后将这些结果综合后得到信息。创建多重图像的思想是：每一个图像代表一个信息，也许是某一个尺度的信息；如果我们知道我们是如何合并这些信息的，那么这样会很有效。到目前为止，生成这些图像的标准似乎对人的视觉的作用要比从中获取信息要有用。”

3. 在计算机视觉的具体算法上，存在三个典型的问题：很少或者不合适的使用知识；空间和时间上看的不够远；缺乏试验传统。

3.1 很少或者不合适的使用知识

“感知系统依赖很多种知识。可惜的是，计算机视觉研究主要考虑的却是发展算子。知识的表示、推理、各种不同算子输出结果合并的和创建基于知识的算子性能特征化常常被忽略了。解释图像时，任何感知过程中的每一个时期都会使用知识和推理。这些知识包括图像中的实体和应用领域。一般说来，在早期过程中，模型时一般的，就像边缘模型（阶梯型灰度变化）。在后期的过程中，使用明确的物体模型。图像理解中需要的知识包括：环境（包括照明）；物体，他们的几何和其他性质；不同表示的关系；感知过程（成像过程）；不同处理任务操作的行为。

所有的系统都会或明或暗的使用上面的知识。显然，系统的效率和灵活性取决于使用的模型。使用的模型越一般，系统的灵活性越大。但是一般的模型实现的时候计算量就越大。视觉中一些流行的模型具有一定的局限性。“阶梯型边缘”是边缘的最常见的模型，但是其局限性是显而易见的。边界的变化往往是十分平滑的。那为什么他流行呢？因为他很容易模型化并容易计算。虽然我们是通过一个点附近的模型来定义边缘的，却老是抱怨边缘断裂。我们的模型是逐点的，但是我们评估的标准却是曲线。

3.2 空间和时间上看的不够远

视觉系统的另外一个问题就是：发展得到的技术在时间上和空间上都是近视

的。很多图像分析技术都假设结论对局部是正确的；他们没有考虑到一个事实：大部分图像包含很多曲面，所以当从一个曲面到另一个曲面的时候，很多局部的假设都违反了。这些操作只能在图像分割成很多曲面后使用。而且，因为感知中占主导的是推理而不是操作，错误的使用局部操作导致的结果就是丢失了问题的本质。

在动态视觉中，短视更明显。这个领域大部分的研究集中在从运动恢复结构问题中。这里，问题集中在使用最少数目的图像帧中的最少数目的点恢复结构。我们可以想象，在动态环境中，重点是恢复充分的信息，所利用的是一系列图像（可以等待到一个合适的时间获得足够的信息）。有趣的是，即使是很多光流场方法都只是使用很少的几帧图像。有些研究者鼓吹使用多分辨率的算法进行边缘检测，他们毫不犹豫的使用 51×51 的模板，但是却希望能使用很少的几帧图像恢复信息。

3.3 缺乏试验传统

第三个问题就是缺乏实验性的计算视觉。缺乏试验所导致的结果就是：研究者没有认真的思考过他提出的问题到底是什么。一般的研究方法参见图 1.1，如图中所示，当一个问题提出并且解决的时候，这个问题与计算机视觉的关系已经非常弱了。很多成熟的学科解决这个问题的方法是，提供了大量的标准化的实验。每一个成熟的学科都发展得到了对技术或者过程的适的评估方法，这是不奇怪的。如果没有这些评价方法，研究者就不会定义他们的问题。这也影响到理论。研究者进行理论研究没有明确的指定他们的假设，所以也就没有必要判断这些理论是否于计算机视觉有真正的联系。经常见到一些文章开始提出了计算机视觉中的问题，然后辩解说这些技术基于作者知道的心理学和神经科学。这种辩解让我们想起了 Martin A Gardner 所定义的伪科学。”

希望解决问题 X
 X 太复杂，简化为 X'
 X' 太复杂，简化为 X''
 X'' 太复杂，简化为 X'''
 $\dots$
 X'''''''''''''''''''' 解决了，宣布解决了 X
 因为解决了 X ，然后提出 Y 。

图 1.1: 一种典型研究方法

《计算机视觉系统中的无知、短视和天真》一文在批判计算机视觉研究中存在

的错误思路的同时，也给我们指出了研究计算机视觉的正确思路：正确的研究哲学应该首先弄清楚什么是视觉，去理解视觉中真正困难的问题所在，然后去研究最基础的问题。在研究中，应该从问题本身出发，通过实验（生理学实验，计算机编程实验）寻找和创造合适的工具来解决计算机视觉中的这些问题，即研究方法是问题驱动的。

§1.4 本文的主要工作

本文试图使用正确的方法去研究计算机视觉中的最基础、最重要的问题之一：图像的视觉信息表象，也可以简称为图像的视觉表象或者图像表象。视觉包括形状、颜色和纹理三种感觉，这三者基本上可以看作是相互独立的视觉感受。因此，图像的视觉表象也包括三大模块：颜色表象，形状表象和纹理表象。因为我个人对颜色没有研究，所以本文对颜色表象并没有深刻的讨论，主要讨论的是形状表象和纹理表象。对形状和颜色，所有人的理解都是一致的。但是，纹理却是很容易引起歧义的词汇。朗文词典中对纹理的解释是：由紧密交织的成分组成的事物；身体或物质的组成微粒构成的组织；某事物可见或可触摸的表面特征及外观。

本文中研究图像表象问题的方法是，首先从分析问题本身出发，分析现有的实验结果，然后提出问题的模型和解决问题的办法，接着使用数学方法说明这种模型的可行性和完备性，最后给出例子。

本文试图将最实际的应用问题和最理论的数学证明联系起来，联系它们的桥梁是数学模型。数学模型就是对于一个特定的对象为了一个特定目标，根据特有的内在规律，做出一些必要的简化和假设，运用适当的数学工具，得到的一个数学结构。数学结构可以是数学公式、算法、表格、图示等。数学模型的一边是最实际的应用问题，这是工程师解决的问题；数学模型的另一边是纯粹的数学问题，这是纯粹数学家研究的问题。数学模型联系这两个极端，是理论和应用联系的桥梁，是实际问题得以完整解决的关键。本文希望充当图像处理实践和数学理论联系的纽带作用：希望数学研究者明白他们的定理在图像处理和计算视觉领域的意义和价值，给他们启发以发现更好的定理；希望工程师们能从数学理论中明白他们算法和程序的局限性。

本文在一系列实验结论和数学分析的基础上说明：多尺度零交叉点是形状表象的合适选择；图像的小波变换系数是纹理表象的合适选择；并且可以在小波分析的框架下建立形状和纹理提取的统一模型。

本文设计了一个基于现有搜索引擎的特定图像搜索软件（SSIS）框架，利用它可以在互联网上搜索具有特定内容的图像。SSIS 可以作为视觉表象理论的验证平

台，测试图像集是互联网上日益增长的图像。

本文的结构如下：第二章通过对大量的实例分析，说明图像表象的重要理论价值和应用价值；第三章给出图像表象理论体系结构和一个简单的例子：二值表象；第四章说明零交叉点是合适的形状表象；第五章说明小波变换后的系数是合适的纹理表象，而对纹理的描述是通过对小波系数的统计描述实现的；第六章给出了小波框架下统一提取零交叉点和统计描述的模型；第七章设计了一个软件框架，通过它可以验证表象理论；第八章对全文做了总结，并对进一步的研究做了展望。

第二章 图像视觉信息表象理论的背景及重要性

本章简要介绍了有关图像表象的一些理论问题和应用系统视觉计算理论、表象学、图像分析识别技术、多媒体管理系统与 MPEG7 和基于内容的图像搜索引擎。提出了它们共有的、基础的环节：如何通过对图像本身的处理，提取出与图像内容有关的有意义的、完备的符号化表示，这就是图像的视觉信息表象理论。

§2.1 视觉计算的简单模型

本节把视觉计算问题当作一个数学建模问题，即给定一些视觉生理学和视觉心理学事实，如何用一个数学模型来描述视觉计算的过程。

§2.1.1 视觉的生理学基础

有关生物视觉的生理基础的资料参见 [3,4,5,6]，这里仅仅罗列一些本文需要的结果。这方面的研究方法主要有生物解剖学、神经科学和电生理物理学。

§2.1.1.1 视觉通路

光是由眼睛进入大脑的，眼球是眼睛的主要部分。视网膜是一种透明薄膜，是眼球的感光部分。眼球内包括晶体、房水及玻璃体，它们都是屈光介质。晶体为一扁球形的弹性透明体，位于玻璃体与虹膜之间，睫状体的收缩可改变晶体的屈光力，使外界的对象能在视网膜上形成清楚的影像。

视网膜可分为以视轴为中心直径约 6 毫米的中央区和周边区。中央区有一直径约 2 毫米 (折合 6 度视角) 呈黄色区域，称为黄斑。黄斑中央有一小凹，叫做中央凹 (fovea)，面积约 1 平方毫米。视网膜由视锥细胞 (cone cell) 和视杆细胞 (rod cell) 两种感光细胞组成，其中，锥体细胞约有 630 万到 680 万个，杆体细胞约有 1.1 亿到 1.25 亿个。在视网膜的中心凹处视锥细胞为每平方毫米约有 150000 个；中心凹处外几乎没有视杆。离开中央凹，锥体细胞急剧减少，而杆体细胞急剧增多，中心凹处外视杆细胞密度约为每平方毫米 23000 到 160000 个。在离开中央凹 20 度的地方，杆体细胞最多。中央凹的锥体细胞密度很高，是产生最清晰视觉的地方。感光细胞中，视杆细胞对光的敏感性较强，并且只对非常弱的光线起反应。视锥细胞是在光线强的时候工作，可以分为三类：分别对短波（蓝色）、中波（绿色）和长波（黄、橙和红色）起最佳反应。

在吸收光线之后，视杆细胞和视锥细胞通过突触将信息传递给其他神经元，如双极细胞，然后再经过突触和一些中介物质，由双极细胞将信息传递给神经节细胞。

视网膜上的光感受器 (杆体细胞和锥体细胞) 通过接受光并将它转换为输出神

经信号而来影响许多神经节细胞、外膝状体细胞以及视觉皮层中的神经细胞。反过来,大多数神经细胞的输出都依赖于视网膜上的许多光感受器。我们称直接或间接影响某一特定神经细胞的光感受器细胞的全体为该特定神经细胞的感受野 (receptive field)。视网膜上的感受野之间有一定的重叠,任意一个感受器都可同时将输入信息传递给不同的神经节细胞。

20 世纪 60 年代末,威塞尔 Wiesel 与 Hubel 一起发现了中枢视觉通路中细胞的感受野特性。在此之前,人们一直认为视网膜上的图像是以点对点的方式传递到大脑视觉中枢的。他们通过对视觉传入通路感受野特性的系统性研究,证明了视网膜上的图像信息在视觉系统中是被分极处理的,每一个神经细胞都具有一定的特征选择功能,只反应图像的某一特定的细节。他们还发现,功能相同的细胞在大脑皮层上聚集成特殊的柱形结构。这些成就被认为是 20 世纪视觉中枢研究的里程碑。为此,于 1981 年他们获诺贝尔生理学奖。

1953 年 Kuffler 首次阐明猫的视网膜神经节细胞 (GC) 的感受野在反应敏感性的空间分布是一个同心圆,即感受野一般是由中心的兴奋区域和周边的抑制区域构成的同心圆结构,称为 On 一型感受野,还有一类感受野是由中心抑制和周边兴奋区域的同心圆构成,称为 Off 一型感受野。

Rodieck 于 1965 年提出了关于同心圆拮抗式感受野的数学模型。它由一个兴奋作用强的中心机制和一个作用较弱但面积更大的抑制性周边机制所构成。这两个有相互拮抗作用的机制,具有高斯分布的性质,但中心机制有更高的峰值敏感度,而且彼此方向相反,即彼此抵消。Rodieck 使用 DOG (difference of two Gaussians, 两个高斯函数的差) 作为神经节细胞的感受野模型。这个模型与其中的 X 细胞的试验数据吻合很好,与 Y 细胞的数据不是很好。

从神经节细胞开始,视觉信息传递分成两部分:常常被称作主视觉系统 (primary visual system) 和次视觉系统 (secondary visual system)。主视觉系统是 (感光细胞→双极细胞→) 神经节细胞→侧膝体 (lateral geniculate nucleus, LGN) →视觉皮层。次视觉系统是 (感光细胞→双极细胞→) 神经节细胞→上丘脑→视觉皮层。因为次视觉系统在进化上早于主视觉系统。在低级动物中,这条通道可能是最重要的视觉系统,主视觉系统是后来发展演化的结果。

主、次视觉系统具有不同的功能。主视觉系统善于对视像进行细节区分,次视觉系统善于粗略的、准备性的分析。这使得上丘脑能够非常迅速地对那些强烈而又易于辨别的刺激作出反应。例如,捕捉或避免被捕捉都需要做出快速反应。另一个例子是上丘脑在眼动产生中有作用。次视觉系统对落在视网膜周边的刺激非常敏感。对灵长目动物而言,次视觉系统好像不能解决复杂图形的细节问题,必须由主视觉系统来解决。上丘脑的功能包括,当目标刺激视网膜周边时,移动眼睛,使注

视点朝向目标；上丘脑输出的信息到达控制颈部肌肉运动的神经元。特瑞瓦撒将次视觉系统调节的视觉称为边缘视觉，主视觉系统调节的视觉为中心视觉。也可以把这两种系统近似地看成是，次视觉系统负责物体的方向和定位，主视觉系统负责物体的辨认和细节分析。

在主视觉系统（从视网膜到视觉皮层的通路）中，也存在平行加工过程。证据表明，对应于分离的加工通道的起点有两种不同类型的神经节细胞，分别称作巨细胞 (mango cell) 和小细胞 (parvo cell)，巨细胞对刺激的快速运动以及明度差异非常敏感，但对颜色差异不敏感。小细胞对静止的视像和颜色敏感，也就是说，它们会被视像中红色与绿色的边界强烈地激活。

尽管这两种类型的细胞在视网膜上是混在一起的，但从侧膝体上，它们的功能就在解剖学方面分离开来，一直到大脑皮层，形成两个功能独立的系统，即巨系统 (mango system) 和小系统 (parvo system)。巨系统尤其适于知觉视像变化的调整。如果没有这个系统的帮助，本来可以看到的视野在几秒钟之内便会消失，因此，它负责运动物体的觉察。小系统专门专责分析细节，这是一个耗时间的过程，它也处理颜色差异。从进化的角度看，巨系统出现得相对较早，小系统是后来进化的结果。在主视觉皮层，信息存在着分割与整合，也就是说，存在看一些既有巨大输入的皮层神经元，也有小输入的皮层神经元。

在主视觉皮层区以外，其他皮层区分别对视像的其他特性进行分析，如形状、颜色、运动等。这就是扎基提出的视觉皮层功能特异性理论。由脑损伤造成视觉某一方面特性的偏盲可以作为功能特异性加工的证据。如果只是看不到颜色，这种偏盲称为色盲。

§2.1.1.2 结论

视觉是一个十分复杂的过程，视网膜发挥了十分重要的作用。视网膜含有三种功能神经元：第一种神经元为视锥细胞和视杆细胞，又称为感光细胞；第二种神经元是双极细胞，主要起连接作用，将来自第一种神经元的光感受信息传递给第三种神经元，也就是神经节细胞。神经节细胞的轴突形成视神经，经过侧膝状体到达视皮层，经过大脑的分析和综合形成视觉。这个过程称为视觉通路，即进入眼睛的光就是沿着这样的顺序逐步转变一些能被大脑理解的视觉信息。

尽管人们对生物视觉的了解还不是很深入，但是有一点是肯定的。人的视觉不是直接对视网膜上的图像进行分类、识别和记忆的，而是首先将视网膜上的图像进行多种层次的分析和特征提取，表示成“具有语义的描述”，然后根据这些“具有语义的描述”进行分类、识别和记忆。

在视觉通路上，首先在视网膜上形成图像，然后图像的成分和特征被一层层，一步步地“抽取”现来。然后到了视知觉的相对晚期，这些成分和特征被整合在一

起，从而形成整体感。

§2.1.2 视觉心理学

William James 说：“我们对世界的感知，部分依赖于对客观事物的感觉，另一部分，可能是更重要的一部分，来自于我们的思维。”

§2.1.2.1 格式塔心理学

格式塔心理学 [13] 是西方现代心理学的主要流派之一。1912 年在德国诞生，后来在美国得到进一步发展。这个学派主要考虑的是视觉心理学。格式塔学派认为视觉的主要问题是：图像是由背景和我们关心的物体组成的，如何将两者分割开来。这不是一个容易的问题，首先背景和物体的区分是与观察者的心理有关的。图 2.1 中，观察者既可以将这幅图看作是两个人的侧面像，也可以看作是一个花瓶，这两种情况同时只能出现一种。

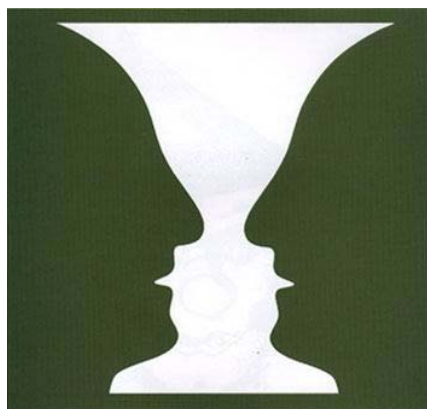


图 2.1: 是花瓶还是两个人的侧面像？源自 [14]

另外，人的视觉系统似乎有超凡的能力。例如图 2.2 中，人能将图像中的人看成是一个整体而不是四个不同的部分的。为了解决这个问题，格式塔学派提出了一系列理论。格式塔学派倾向于用共同性，差异性词汇描述整体的性质，并力图发现这些整体发生的规律和法则。主要体现在下面的四个定律：

(1) 接近律：我们倾向于将那些相互靠得很近且离其他相似物体较远的东西组合在一起。这在图 2.3 中就看得很明显。该图由许多规则的矩形阵列小黑点组成。你的大脑既可能将它们组织成水平线也可能组合成垂直线。但实际上，你把它们看成是垂直线。这是因为，一个点到其最近点的距离，在垂直方向要比水平方向短。其他实验显示，接近律通常指“空间上接近”，而非在视网膜上的接近性。

(2) 相似律（运动）：格式塔的相似律是说，我们将那些明显具有共同特性（如颜色。运动、方向等）的事物组合在一起。如果你看见一只正在跑的猫，你就会把



图 2.2: 人类视觉系统具有非凡的能力。一个完整的人还是四个不同的部分? [14]

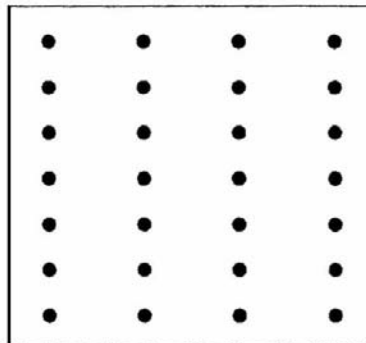


图 2.3: 接近律 [14]

它身体的各个部分组合在一起。因为一般来讲，当猫跑时，它的各个部分会在一个方向上运动。同样的原因，正在树丛中爬行的猫也会被识别出来。但是，如果它纹丝不动，我们就很难发现它。

(3) 连续律：良好的连续性定律可以由图 2.4 加以说明。该图的上部分显示两条相互交叉的曲线。我们的确把它看成是两条线，而不是像该图的下部分所显示的那种交汇于一点的四条线或是两个靠近的 V 型。我们同样倾向于把中断的线段看成是被某个物体遮挡一部分的连续直线。

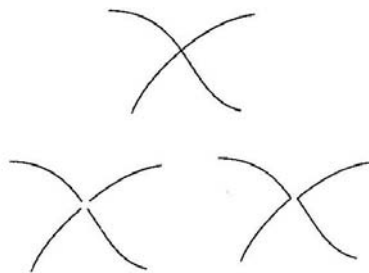


图 2.4: 连续律 [14]

(4) 封闭律：如图 2.5, 封闭性在线画图形中表现得最为明显。如果一条线形成了封闭的或几乎封闭的图形，那么我们就倾向于把它看成是被一条线包围起来的图形表面，而不仅仅是一条线。

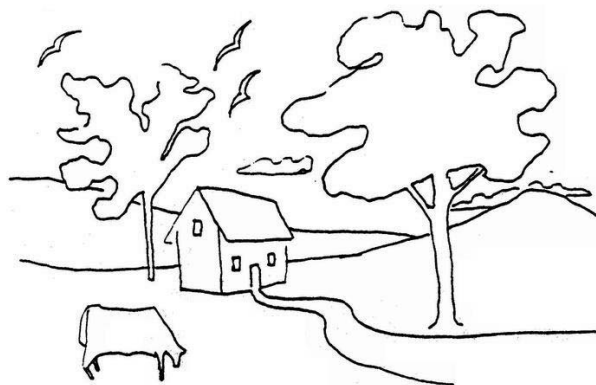


图 2.5: 封闭律, 源自 [14]

大脑在决定哪些视觉特征属于某个物体时，要依赖于大体上符合格式塔知觉定律的明显的视觉线索。因此，倘若一个物体较坚实（接近性）、具有明确的轮廓（封闭性）、朝一个方向运动（共命运），而且整个为红颜色（相似性），那么，我们就很可能认为这是一个运动的红球。

格式塔理论中封闭律启发我们，图像中基本轮廓形成的线图对视觉感知有十分重要的作用。

§2.1.2.2 Werner 实验

1935 年，Werner 做了一个试验，如图 2.6 所示，白色的背景上有一个黑色的圆盘，称为测试物，将这个画面持续 10 毫秒；然后持续放映一定时间的白色空白页面；最后是一个环状物，称为掩饰物，这个环状物恰好处于第一个画面的黑色圆盘周围。也就是说掩饰物和测试物在任何时间和空间上都不重叠，它们具有共同的轮廓（称为 metacontrast）。实验表明，观察者能看到环状物，但是无法看到圆盘。

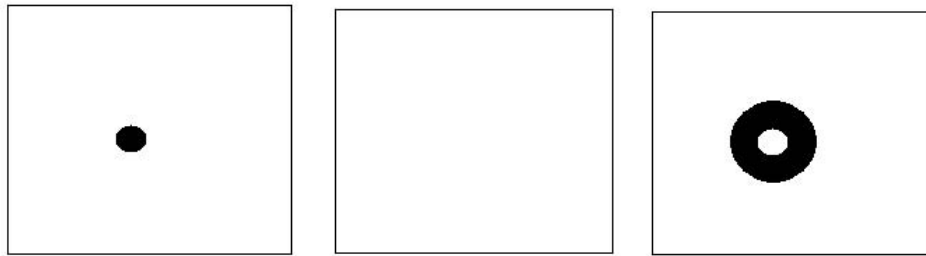


图 2.6: Werner 实验

于是，这就证明了，通过切断完整的边界轮廓，可以阻止观察者感知这个轮廓界定的物体的颜色或者灰度。后来的很多类似的实验支持了这一结论。这就产生了一个假设：在我们感知物体表面的灰度和颜色的时候存在着一个基于轮廓的填充 (filling-in) 过程。

这一过程可以通过画家绘画的过程加以说明。如图 2.7 所示画家在绘画的时候，首先绘制出基本的轮廓，决定出轮廓点上的颜色，然后将轮廓中间的颜色通过画笔平滑的连续起来。

实际上，对物体表面属性反射系数、颜色和纹理等性质的视觉感知过程中，界定物体的轮廓都有十分重要的作用。图 2.8 反映了轮廓对于纹理感知的影响，图 2.8(a) 中环状结构是均匀的纹理组成的，但是我们看起来左边半个圆环要比右边半个圆环的点要密；如图 2.8(b) 通过添加一条直线，就可以将圆环分成两个半圆环；如图 2.8(c)，通过添加两条轮廓线，就可以把圆环从背景中分离开来。

§2.1.2.3 视错觉

虽然说“眼见为实”，但是并不是所有眼睛看到的都一定是真实的。视觉系统可能会产生错误，这就是视错觉 (Visual illusions)。

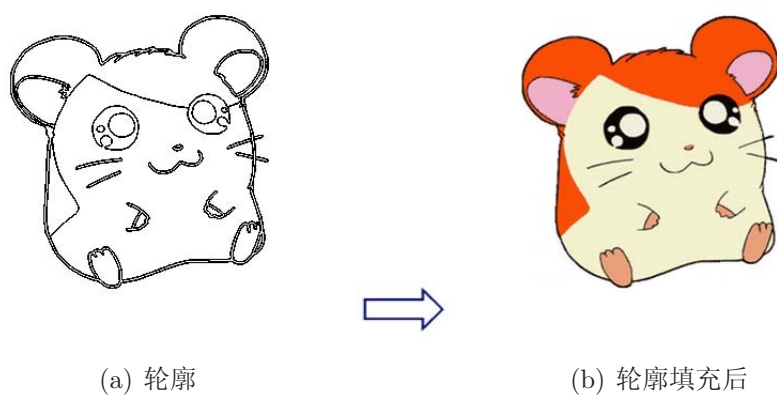


图 2.7: 基于边缘的填充过程

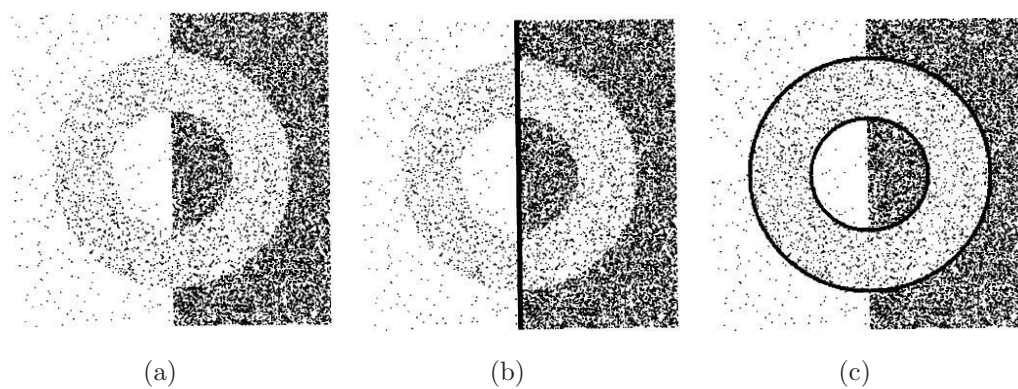


图 2.8: 轮廓在纹理感知中的作用, [15]

图 2.9 是长度错觉的例子。图中两条相互垂直的线段的长度是相等的，但看起来似乎竖直的线比水平的线更长一些。

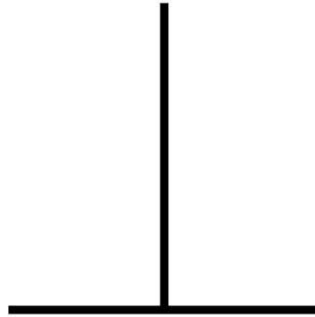


图 2.9: 长度错觉

图 2.10 是大小错觉的例子。左右两个黑色的圆点是同样大小的，但看起来似乎右边的圆点比左边的圆点大。

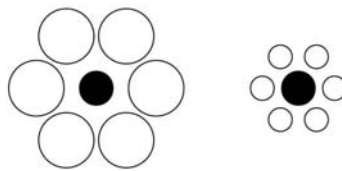


图 2.10: 大小错觉

图 2.11 是著名的 Kanizsa 三角形。我们可以看到，三个缺口的圆盘形成一个很大的白色三角形。

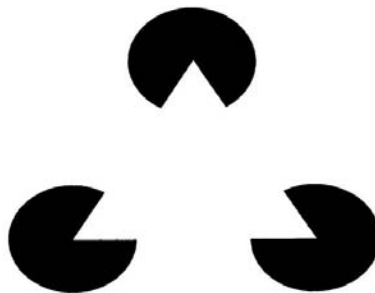


图 2.11: Kanizsa 三角形

§2.1.2.4 心理物理学

心理物理学主要研究人脑是如何表达物体的，也就是视觉的表象问题。研究者设计了一些简单的物体（积木），在三维空间中进行转动或者再加一个镜面反射，让试验者判断两个物体是不是仅仅差一个三维旋转或者再加一个镜面反射。试验结果表明，判断所需要的时间与旋转角成线性关系。

对这个实验结果的一个解释就是：大脑中确实有某种对应的结构表达视觉中观察到的物体，在识别物体的过程中，大脑中的这一结构也在进行旋转。这种“对应的结构”称为表象。

§2.1.3 视觉计算的数学模型

本节参考了 [8,9]。

关于视觉的生理基础的研究和视觉心理学的研究，给我们很多知识和启示：我们大致知道了视觉通路上的基本流程，我们可以知道某些地方需要存储器，我们也知道了视觉本身所具有的一般规律。但是，仅仅到这里显然是不够的。我们仍然不知道视觉感知过程中，需要计算什么（假设计算是必要的话）；也不知道视颜料中包含了那些信息，它们是如何组织的；纵然你知道某个系统中需要一个存储器，但是你不知道到底要存储什么。

我们需要一个视觉理论体系，就是在已有的生物视觉和视觉心理学基本事实的基础上，建立一个数学模型去理解和解释这些事实。

§2.1.3.1 视觉是一个主动的信息处理过程

为了生存和繁衍后代，我们需要快速的获取外界的信息。视觉就是人通过视觉观察去感知有什么东西在什么地方的过程。换句话说，视觉是一种处理过程，经过这个过程从图象中发现外部世界中有什么东西和它们在什么地方。因此，视觉首先是一个信息处理过程。图 2.12 为信息处理过程的一般结构。

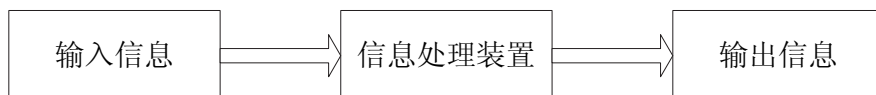


图 2.12: 信息处理过程

如图 2.13 所示，对于视觉信息处理过程，它的输入是视网膜上的反射光子形成的像；而输出信息应该是视觉表象。视觉感知系统完成的就输入信息与输出信息之间的转换。

对这样一个过程，并且应该从三个不同层次来研究和理解：

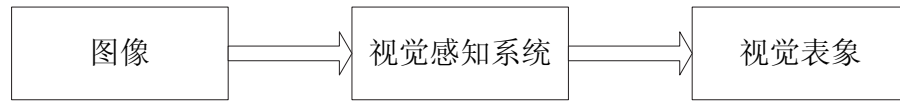


图 2.13: 视觉信息处理过程

第一个层次是，要计算的视觉表象到底是什么？是如何表示的？为什么要进行这些计算？这些计算是不是必要的？这个层次就是视觉信息处理的计算理论。第二个层次是如何才能完成这种输入和输出的转换，也就是如何设计算法去还原视觉感知系统。第三个层次是实现算法的机制或硬件，即如何用具体的硬件去实现这种算法。例如，根据 Fourier 分析理论，任意连续函数可用它的 Fourier 频谱来表示，因此 Fourier 变换是属于第一层的理论。而计算 Fourier 变换的算法，如各种快速 Fourier 算法是属于第二个层次的。至于实现快速 Fourier 算法的硬件实现当然就属于硬件的层次。

对视觉系统的建模问题就是要解决第一个层次的问题，即我们要从视网膜上的图像中计算什么和为什么要计算它们。这是一个根本性的问题。只有合理的解决了这个问题，才有可能正确的解决第二、第三层次的问题。

视觉是一个信息处理过程，而且它并不是被动的接受和记录进入眼睛中的视觉信息，而且在进行主动的建构。这可以从视错觉的例子中得到证实，Kanizsa 三角形正是这种主动建构的结果（图 2.11）。我们知道，真实的三角形是不存在的，它是我们的视觉系统自主创造出来的。

§2.1.3.2 视觉最重要的是获取形状信息

视觉最重要的是获取形状信息。支撑这个观点的依据有两个：一个就是前面提到的 Werner 实验，另一个是临床神经病学方面的。

1973 年，临床神经病学家 Warrington 发现，左颅壁损伤的病人虽然一般没有语言能力，无法说出物体的名称，也不知道被观测物体的用途等；但是，他能感知到物体的形状。

根据这样一个现象，我们有下面的结论：首先，关于物体形状的表象和用途的表象是存储在脑中不同的地方的。按照我们前面的说法，他们是通过表象之间的连接相互作用的。其次，“即使病人在理解物体的用法、用途这个传统的意义上不能认出被观察的是什么物体但仅靠视觉就能传递该物体形状的一个内部描述。”也就是说，视觉感知过程中，即使没有大脑的有意识的活动（左颅壁损伤的病人，他们回忆物体的用途和用法的表象的能力都没有了），我们仍然能获取形状信息。因此，获取形状、空间、空间位型等信息是视觉的主要目的。

颜色和形状是视觉表象的两个重要方面, Werner 实验告诉我们: 视觉感知中, 对颜色和纹理的感知是依赖于对轮廓的感知的。也就是说先有对形状的感知, 而后才在形状感知的基础之上进行颜色和纹理的感知。

因此, 对形状的感知比对颜色的感知更加重要。

§2.1.3.3 视觉系统的表象

如前所述, 视觉器官一个重要功能就是获取视觉表象。因为如果我们能够认识外部世界有什么东西在什么地方, 那么我们的大脑必须以某种方式表象 (representating) 这种信息——以尽可能丰富的色彩、形状、美感、运动和细节等多方面来表象这种信息。所以, 视觉研究不仅必须探讨我们是如何从图象中提取外部世界中对我们有用的各个方面, 而且还必须探讨内部的表象。

所谓表象是指一种能把某种实体或某几类信息表达清楚的形式化系统, 以及说明该系统如何行使其职能的若干规则。使用某一种表象描述一个给定的实体所得到的结果称为该实体在这种表象下的一个描述。

那么, 视觉计算过程中的表象是什么呢? 视网膜上的光经过多个层次的十分复杂的变换得到了视觉表象, 因此可以认为, 存在着多个层次的图像表象。这些表象开始于直接从图象中得出的描述, 而每一个表象都将有助于后面进行的处理工作, 经过一步一步的处理, 逐步恢复物体形状的越来越客观的物理性质。

这些表象应该是越来越高层次的符号化描述。例如, 我看到桌子上有一本书。这是视觉系统主动获取的视觉信息。可以想象, 视觉表象是以一个整体来描述视野中的“书”这个符号, 它不可能只是记录反射光子的物理性质。

真实的物体是三维的, 但是一幅图像却是平面的。从三维到二维可见表面的过程是降维的过程, 降维过程中必然丢失一些信息。三维物体的信息全部蕴含在像素灰度的变化中。从可见表面获取视觉表象的过程就是从像素变化恢复外界三维信息的过程, 需要利用客观世界的约束。通过对反射率函数的普遍性质进行了考察, 可以提出了基本的物理假设 [8]:

1. 表面存在: 可以认为可见世界是由一些平滑的表面组成的, 这种表面的反射率函数的空间结构可能是非常复杂的。
2. 分层次的组织: 一个表面的反射率函数的空间组织, 往往是由一些不同的成像过程产生的; 每一个过程是在不同的尺度上起作用的。
3. 相似性: 对任一给定表面, 通过以某一给定尺度产生反射率变换的过程所得到的各个项目, 就其大小、局部反差、颜色、空间组织而言, 它们之间的相似性比它们与该表面上其它项目之间的相似性总是要大的多。
4. 空间的连续性: 由单一过程在任一表面上所产生的标记往往是有空间组织

的，它们排成一条直线或者曲线，也许还能形成更复杂的模式。

5. 不连续中的连续性: 深度或表面朝向不连续的轨迹几乎处处平滑。

6. 运动的连续性: 如果运动的方向在多于一个点上（例如沿一条直线）不连续，那么就意味着出现一个物体的边界。

这些约束对我们的视觉系统而言是先验的，视觉系统在长期的进化过程中已经具有了这些知识。

结合前面的结论，我们可以做一个大胆的假设：视觉系统在先验的约束的基础上主动地从图像中提取一系列不同层次的符号化的图像表示，并最终得到原始图像的符号化表象。这个符号化的视觉表象中包含：闭合的轮廓线（与具体的物体的形状相关）；闭合轮廓线中的颜色和纹理（物体表面的颜色和纹理）；轮廓线之间的空间位置关系（物体之间的空间关系），并且这些表象是以物体为中心的坐标系描述的。

§2.1.3.4 视觉表象的主动建构

现在的问题是，大脑是如何主动建构这些表象的。可以认为，从图像推得视觉表象的过程如下：

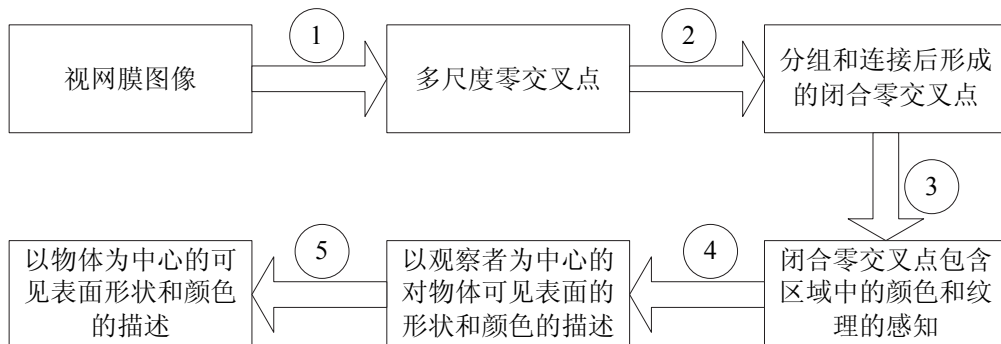


图 2.14: 从图像到物体表象的过程

step 1: 多尺度零交叉点检测

一维的零交叉点对应着一元函数梯度的极值点，也就包含了梯度极大值点。而梯度极大值点对应着图像上亮度变化剧烈的位置。推广到二维，二维的零交叉点对应的是图像中梯度方向上的二阶导数为零的点。这些点中包含了绝大多数的图像的形状信息。

梯度方向上的二阶零交叉点对应着物体的边界，是界定物体的根据。零交叉点对应图像像素亮度发生剧烈变化的位置。在早期的计算机视觉中，利用物体的边界

可以进行物体识别和匹配，这些成功说明了零交叉点作用。

另一方面，由于视觉具有多尺度特征，在零交叉点检测过程中采用多尺度的零交叉点检测。

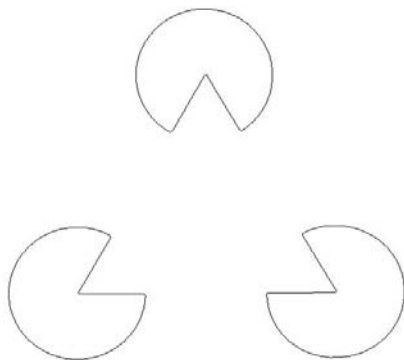


图 2.15: 零交叉点检测

step 2: 零交叉点的分组与连接

如图 2.15 所示，虽然我们看到了白色的三角形，但是零交叉点检测中对这个白色三角形而言，零交叉点是断裂的、不完整的。所以，视觉系统必须通过主动的神经活动将这些断裂的零交叉点进行连接起来得到封闭的边界轮廓线。这一步可以通过下面的步骤完成：分组和启发式的连接。

首先是分组，靠近的、具有相似特征（梯度方向接近等等）的零交叉点被认为是一组。按照这个思路，图 2.15 中的边缘被划分为下面 9 组。第一组和第二组能被分开，是因为在第一组和第二组之间有一个尖点，其他类似。

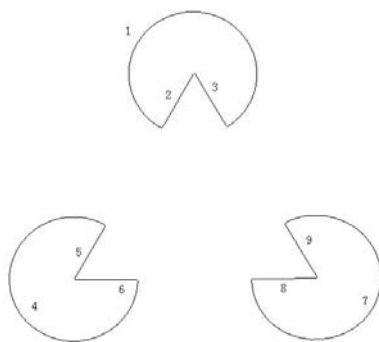


图 2.16: 零交叉点检测的分组

然后，在分组的基础上进行启发式的连接。

启发式的连接有很多种结果：

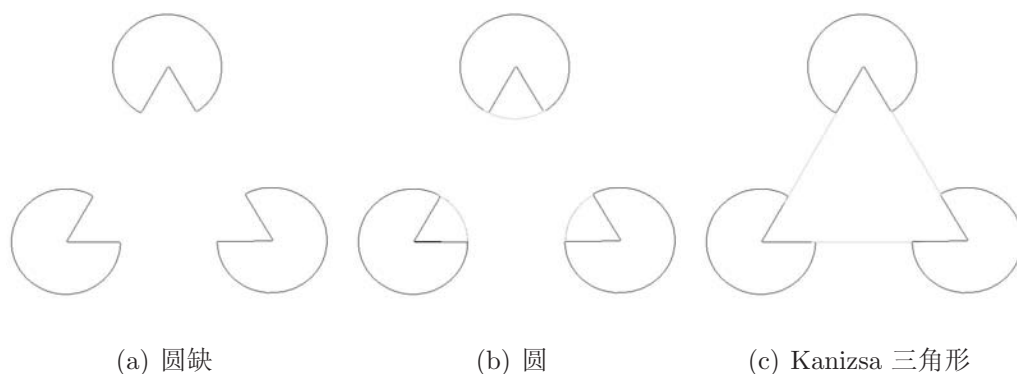


图 2.17: 启发式的连接

将第 1,2,3 组, 第 4,5,6 组和 7, 8, 9 组零交叉点分别连接, 我们得到三个圆缺, 见图 2.17(a)。

将 1 组, 第 4 组和第 7 组零交叉点分别自己启发式的首尾连接起来, 出现三个圆形。我们将原图中三个图形称为圆缺, 说明我们在感知到这个形状时, 确实感知到了强烈的圆形, 见图 2.17(b)。

将 2, 3, 5, 6, 8 和 9 启发式的连接起来, 就出现了白色三角形, 见图 2.17(c)。

也就是说, 对同样的零交叉点, 启发式的连接可以有很多种可能。不管怎样, 每种可能的连接都使得感觉器官获得合适的形状表象。

可以把这个过程理解为多种构图的竞争过程。分组后的图像实际上形成了多条零交叉点形成的线。连接过程就是感觉器官试图将这些线条连接得到闭合的轮廓的过程。这种连接显然有多种可能, 感觉器官试图获得“最优”连接, 在寻求最优连接的过程中, 若干种“较优”的连接也可以形成表象。

可能的零交叉点的启发式连接有多种可能, 从某种意义上说, 任何的连接都可以解释。但, 实际上并不是所有的连接都可能被感知到, 图 2.18 中这种连接就不会产生。简单的说, 这种情况可能是下面两种原因。

首先, 客观世界的约束。视觉器官不能做无限的启发, 只能在已知图像的基础上进行合理的启发式连接。并且, 连接的结果要么有客观世界(图像种的亮度变化)的强有力的支持, 启发式的连接线不能太长; 要么符合客观世界的常规模式, 三角形, 圆就是常规模式, 光滑得曲线也是常规模式。

其次, 视觉感知器有一定的选择性, 这些选择性可以用格式塔心理学的规律去解释: 只有靠的比較近、具有类似结构的的线才有可能被连接。

step 3: 基于填充的颜色和纹理感知

如果封闭轮廓中颜色是接近的, 就感知颜色; 否则进一步感知闭合曲线中的颜

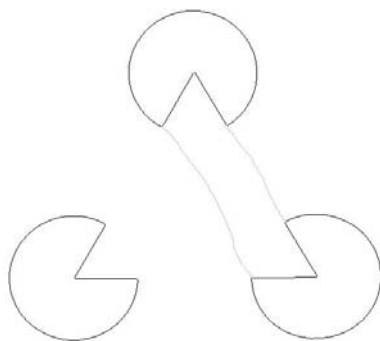


图 2.18: 客观世界的约束和视觉器官的选择性

色变化。如果颜色变化具有某种统计模式，这就是纹理。

通过上面这些过程，我们获得了封闭的曲线和封闭曲线中填充的曲面的颜色和纹理的符号化描述表象。这些描述都是建立在图像上的，也就是图像的视觉表象。图像的视觉表象可以看作是视觉前端自动完成的。

step 4: 以观察者为中心的对物体可见表面几何特征的描述

在第三步结果的基础上，需要进一步的计算以在大脑中恢复物体的三维信息：双目匹配成像获得立体感觉，利用运动估计物体形状等等，推导出一个能反映可见表面几何特征的表象，其中包括表面朝向，观察者的距离，以及朝向和距离的不连续性，表面的反射情况，以及对主要照明情况的某种粗略的描述。这些过程是基于多个图像的表象的。

step 5: 三维表象

最后获的物体的三维表象：被观察形状的三维结构组织在以物体为中心的坐标系中表象，以及在这种坐标系下对物体形状和表面性质的描述。

§2.1.3.5 对现象的解释力

理论的建立是为了解释现象、进行预测或者利用理论构造相应的系统。其中，解释现象是最基本的。下面，我们在这个理论的基础上解释一些生理学和心理学上的现象。

(1) 视觉前端神经节细胞的感受野恰恰具有类似 LoG 的结构。对这个解释就是：LoG 滤波器可以作为零交叉点检测滤波器，神经节细胞上的 LoG 状结构就是进行了零交叉点检测。

(2) 根据 Hubel 和 Wiesel [10, 11] 关于大脑视皮层细胞感受野的论述，以及 Campbell 和 Robson [12] 关于视觉系统空间频率通道的论述，视觉感知是一个特征检测过程或者或者是某种空间 Fourier 分析过程，但是这两种观点都只能部分的解释试

验现象。

多尺度零交叉点检测综合了特征检测和频率分析两种思想，大脑通过多个通道相互独立的特征（零交叉点）检测来提取各种不同尺度下的特征。这就说明了 Hubel 和 Wiesel 关于大脑视皮层细胞感受野的论述和 Campbell 和 Robson 关于视觉系统空间频率通道的论述都是正确的。视觉感知既是一个特征检测过程或者，同时视觉感知过程又是涉及频率的多尺度的分析过程。所以，这两种理论中的任何都只能部分解释试验现象，只有综合运用它们才能合理解释视觉现象。

(3) 格式塔心理学提示我们仅仅通过线图（也就是说图像中亮度剧变点），就能够给我们丰富的视觉信息。Werner 实验提示我们，感知颜色首先要感知轮廓，然后使用填充过程来感知颜色。这两个现象在这样的理论下是直接的。

§2.1.3.6 Marr 视觉理论

本节前面的内容很多地方的根据 Marr 视觉理论 [8,9]，但是本文并没有完全照搬 Marr 理论，而是按照自己的理解，重新搭建了视觉计算的框架。

Marr 和 Hildreth 最早开始寻找更符号化但是仍然是完全的图像表示，取得了一定的成果，并在此基础上建立了视觉的计算理论。Marr 计算视觉理论对计算机视觉的影响非常大，1982 年 Marr 的遗著 [8] 出版以后，受到十分广泛的关注。随后几年涌现出了一大批提取描述形状的边缘的算法和有关的论文。特别值得一提的是，1986 年 IEEE 上 PAMI 中集中出现了一批特别值得阅读的论文，包括著名的 Canny 算子及边缘检测技术。1988 年，科学出版社出版了由姚国正、刘磊和汪云九翻译的这本书的中译版 [8]。

§2.1.4 结论

视觉包含三种独立的感受：形状，颜色和纹理。物体视觉表象也应该包含这三个方面的内容，这样才能区分多姿多彩的各种物体。建立物体视觉表象的前提是视觉前端能获得视网膜上图像的表象。视觉前端自适应的获得了图像表象，然后大脑根据先验知识和经验计算物体和客观世界的视觉表象。在计算物体和客观世界的表象的过程中，大脑是主动的启发式的建构过程。因为是启发式的过程，所以建构的结果不唯一。

图像的表象是建立物体和世界表象的基础，视觉生理学和视觉心理学试验表明：对形状的感知是最重要的，是产生颜色感知和纹理感知的基础。颜色感知纹理感知是通过对形状轮廓的填充过程来实现的，多尺度零交叉点是描述形状的重要工

具。

§2.2 表象学

表象学 [16] 是从人对事物认识的最基础原理开始研究人的思维和认识过程的科学。简单的说,表象学是专门研究人类思维中物体表象的科学:物体表象的形成以及人是如何使用这些表象进行思维和推理的。表象学中所说的表象包括视觉表象、听觉表象、嗅觉表象、味觉表象和肤觉表象五种。其中,视觉表象占 80% 以上,是最重要的表象。下面我们从表象学的角度简单介绍一下视觉表象。

视觉表象是大脑借助于视觉感官反映外部物体和空间形成的表象,是我们思考外部物体的形状、颜色和纹理的工具。在表象学中,大脑具有重要的功能:表象形成功能,内注意表象功能,对照表象功能,分拆、组合表象功能,运用表象功能和表象连接功能。

§2.2.1 视觉表象的形成和记忆

在表象学中,视觉通路上的视觉信息转换可以用下面的术语描述:

(1) 光作为视刺激作用于人的视觉感官,是视觉表象形成的起点。

(2) 视网膜将光转换为视递质,视觉通路对视递质进行处理得到视颜料(神经性材料)。

(3) 视颜料是直接构成视觉表象的物质材料。

视觉表象与真实的物体是不同的。首先,构成物体的材料是各种物质分子,但是构成视觉表象的材料是神经性的物质成分。另外,存在各种大小的物体,他们的表象都不大于人的大脑。因此,视觉表象是一种压缩性的存储。

另一方面,视觉表象和物体之间具有一致性。一致性表现在,一定的物体只能反射一定的光(这是由物理规律决定的),一定的光只能刺激一定的视递质,一定的视递质只能加工得到一定的视颜料。一致性还表现在,大小是比例一致的。因此,大脑中的视觉表象和真实的物体具有一致性,我们可以通过视觉表象准确地思考客观世界。这是人能通过视觉表象获取外界信息的原因。

视觉表象的普遍特征是形状和颜色。颜色表象和形状表象都是视觉表象的一部分,它们都是可分离的。可分离性在于,我们能独立思考颜色或者形状,也就是颜色感觉或者形状感觉能被单独注意到。一个典型的例子是,我们能抽象出只具有形状意义的三角形或者平行四边形,它们与颜色没有任何关系。

物质材料具有吸收和反射确定波长光子的属性,称为反射属性。颜色表象仅仅与反射光子有关。颜色表象的产生过程为

色刺激 → 色递质 → 色感受器 → 颜色表象

颜色表象有重要的意义。颜色表象是我们思考外物颜色的工具。颜色表象与语言文字、语音建立神经连接，转化为颜色类表象，从而使其对外物的各种颜色具有广泛的识别和区分能力。例如，这是红色，那不是绿色等等。颜色表象具有判断作用，判断一个物体是或者不是某物。

形状表象是视觉表象的另一个重要的特征。形状表象的产生过程为：

形状刺激 → 视网膜象 → 形状递质传输模式
→ 形状递质分布模式 → 形状颜料分布模式

形状刺激是反射光子空间分布模式，也称反射光谱模式。光不具有特定的形状，光的形状主要是由一定的物体造成的空间形状决定的，物体表面的形状直接决定了反射光谱模式，所以我们能通过物体的反射光获得其形状信息。但是，物体的形状是由物质成分在空间上的分布形成的；而反射光谱模式只是由反射光子的分布形成。这种转换是通过视通路实现的，它使得外在的物体的形状能被人感觉到。

确定的反射光谱模式（形状刺激）是由入射光和反光面共同决定的，并且形状刺激的存在是一种持续的存在，无论我们是否看见，物体的光谱模式一直存在着，并且具有特定的角度和空间范围。在这个范围以外，我们就无法看到物体。反射光谱具有多向性，能够从多个角度看到各种有差异的形状。需要注意得是，其它感觉也可以获取形状表象，例如触觉。

这种反射光谱模式的也具有很大的局限性：首先，它依赖于光照；其次，依赖于角度和尺度：我们的眼睛只有与物体保持适当的角度和距离时，才能接受反射光谱模式刺激。离物体越远，看到的物体越模糊，这就是尺度效应。当距离很远时，反射光谱的光子分散开来，变的很稀薄，人眼单位面积上接受到的光子越少，无法获得清晰的表象，我们就无法看清物体。只有在观察角度和观察距离固定时，才能得到确定的反射光谱模式，从而获得确定的形状表象。第三，细微和遥远的物体产生的反射光谱模式，不能刺激视感官形成表象。最后，我们只能看到物体的表面。

获取视觉表象后，需要记忆表象。表象的记忆发生在大脑皮质上，是大脑的功能之一。记忆是保持在大脑中的对过去事物的印象，而不是事物本身。换句话说，我们记忆的就是感觉器官获取的物体的表象。

记忆的本质就是表象与大脑皮质的粘合关系，合成的神经（视觉）颜料与皮质的结合过程。这个过程可以类比为拿颜料在白纸上绘画的过程，这也就是我们将表象的物质材料成为视颜料（形状颜料）的原因，大脑中的这张“白纸”就是胞柱膜。视觉表象（当然其它表象也是如此，但是我们这里强调视觉表象）被获取后，将被永久的涂抹在特定的胞柱膜上。不同时间获取的相同的表象，将重复的涂抹在同一个地方。但是这是分层的重复涂抹，不同时间获取的表象涂抹在不同的层

中。这样，一方面节约了胞柱膜，增加了大脑的记忆容量（大脑具有无限记忆的能力），另一方面使得大脑具有了“类”的观念，即涂抹在同一位置的所有层中的表象，是同一类的，可以称为表象类。分类是人工智能先决条件和重要问题，这是抽象思维的基础。

§2.2.2 视觉表象的内注意

视觉表象被涂抹到胞柱膜，这是永久的涂抹。但是，并不是所有被涂抹的表象都能被回忆。回忆就是注意涂抹在胞柱膜上的表象；这种注意有别于我们的感官对外在物体的注意。一般称前者为内注意，称后者为外注意。只有当一个表象能被内注意，他才能被回忆起来。如果一个表象不能被内注意，那么这个表象就是被遗忘了。

大脑具有能使物体的视觉表象在大脑中再现的能力，这就是大脑的内注意功能。例如，我们能闭着眼睛回忆一幅美丽的图片。

§2.2.3 视觉表象的对照功能

视觉表象的对照功能是建立在表象被内注意的基础之上的，当两个表象同时被内注意时，它们便呈现一定的异同关系。对照功能有横向对照和纵向对照两种，横向对照指的是不同物体的表象的对照，横向对照使我们知道不同物体的差异；纵向对照指对同一个物体，大脑中记忆的表象和当前获得的表象之间的对比，这种对照使得我们具有对物体的识别能力。

由于任何两个表象都可能被同时内注意而发生对照，所以大脑能比较事物之间的异同：是同一个物体或者不是同一个物体。判断两个物体是否相同，是逻辑思维的基础。

§2.2.4 视觉表象的分拆和组合功能

胞柱膜上涂抹了很多表象，大脑能内注意这些表象，而且能自由调整内注意的对象和范围。于是，大脑就能只内注意物体表象的一部分。这就是大脑的表象分拆功能。同时，大脑也能由若干个表象的若干部分组合形成一个新的表象，这就是大脑的表象组合功能。例如，任何人没有见过龙，但是很多人绘出了龙的图画。

大脑能进行(视觉)表象的分拆和组合，这是人类创造性思维的基础。

§2.2.5 视觉表象的运用功能

大脑将已有的表象解释未有或者未知的物体的功能，分为：未有运用，未知运用和未来运用三种。

未有运用是指，本来不存在某物，但是大脑运用已有物体的表象，认为存在某物。例如，很多人认为上帝是存在的，或者本来盒子里没有东西，但是我说盒子里面有一个杯子。

未知运行是指，大脑利用已有的表象解释没有在大脑中形成完全表象的物体。在一定的时期，并不是所有的物体或者物体所有的部分和属性都能刺激人的感觉器官，而在大脑中形成表象。例如，在没有显微镜以前，人类无法看到细菌和病毒以及它们的活动，但是人还是能在很多现象中感觉到它们的存在，而想象它们的形态，这就是假说。提出假说是自然科学的基本任务，假说可以引导进行科学探索。

未来运用是指，将表象运用与说明未来事物或者未来事物的状态。例如，我们经常设想自己的将来，我想象自己已经拿到了硕士毕业证书。

§2.2.6 视觉表象的连接功能

在神经连接基础上，大脑中的各种表象和表象类形成了各种复杂的连接，这种连接的形成是以多种表象被同时内注意到而形成的。因为，我们经常将乌鸦和黑色这两个表象同时内注意，所以，它们之间就会建立连接。

表象之间的连接和同一表象的分层涂抹使得归纳和推理成为可能。同一位置的所有层中的表象都和某一表象 P 有连接，那么这一类表象就和 P 有一定的关系。我们看到了 100 只乌鸦是黑色的，在大脑胞柱膜的某一位置上 100 层的乌鸦表象都和黑色建立了连接，我们就可以归纳出：所有的乌鸦都是黑色的。

§2.2.7 表象学研究与图像表象理论

在《表象学》中，作者从事物的外部表象入手，对思维的形成过程进行了深入的研究与探讨，分析了思维的一般模式与正确途径，给人耳目一新的感觉。书中阐述了（视觉）表象与类、分类、比较、假设、归纳法、假说和创造性思维有密切的关系，这些是智能、逻辑和科学的基础。

但是，表象学中的许多结论缺乏视觉理论的支持或者不深入。例如，表象学中指出，（视觉）表象中的视颜料存储的是物体的表象，它是思维的基础。但是，表象学并没有告诉我们视颜料中具体存储了什么数据，这些数据行使自己表象视觉信息的任务的机制以及视颜料在大脑胞柱膜上可重复涂抹的机制。表象学对视觉表象产生的机理、过程和结果并没有深究，而只是一般性地研究了在这个过程中的一些属性。实际上，这个过程正是视觉通路所完成的任务，是视觉计算理论所要解决的问题。表象学应该是建立在视觉计算理论的基础之上的，但如前所述，视觉计算理论本身还很不完善，所以（视觉）表象学的基础还不完善。

表象学中所说的从视递质到视颜料的过程（也就是视觉通路）是一个十分复杂的过程，图像的表象是处于视递质和视颜料中间的一个环节，是获得视颜料的重要一步。表象学本身目前对这个过程了解不多。需要我们对生物视觉、视觉计算理论和人类思维推理过程做进一步的研究和分析。

显然，图像的表象理论对表象学也是十分重要的。

§2.3 图像识别与图像分析

视觉的计算理论是对生物视觉的理论解释。如果我们能还原生物视觉的表示和过程，就能想办法使用计算机模拟视觉过程，进而制造“人工智能”：机器人。有了人的智能、钢铁的身躯和闪电般的反应速度，机器人可以代替人从事危险的、快速或者重复的工作。

不过，目前的计算机视觉研究还很成熟，还无法还原人的视觉过程，所以还无法制造具有足够智能的机器人。但是，我们还是可以利用图像分析和处理技术，帮助人更好的生产和生活。例如，在工业自动化中，可以使用图像分析和处理技术检查产品是否具有视觉缺陷，确定图像中物体的位置；在医学中，使用图像分析和处理技术帮助医生检查和治疗；在日常生活中，使用图像分析和处理技术帮助我们更好的生活：字符识别技术，条形码和二维码技术，生物识别技术（人脸识别，指纹识别，虹膜识别等等）。

需要强调的是，这些技术可能和生物视觉没有任何关系。也就是说，虽然这些应用中的表示和技术在某些方面取得了成功，但是它们并不就是生物视觉中使用的表示和技术，即使具有类似的功能。这些技术和视觉计算过程都可以归结为二维或者三维的信号处理技术。人的视觉是一种特别的二维或者三维信号处理，是研究图像处理和分析技术的样本。之所以这样，是因为人的视觉使用的信息处理机制是生物物种经过漫长的历史进化而产生的，理论上具有一定的优势，而且人类对它已经产生了适应性。但是，不能排下面的可能：存在与人的视觉机制不同但效果相同的信号处理机制的可能；也不能排除存在比人的视觉机制更好的信号处理机制。

本节主要以工业自动化为背景，介绍像识别和图像分析技术。在现代化的大生产之中，视觉检测往往是不可缺少的环节。比如，药品包装的正误，纺织品是否存在缺陷，产品是否存在杂质等等，都需要众多的检测工人通过肉眼或结合显微镜进行观测检验。大量的人工检测不仅影响工厂效率，而且带来不可靠的因素，直接影响产品质量与成本。在有些环境下，例如食品生产，我们要尽量减少人的介入。除此以外，许多检测不仅需要判断，而且需要准确获取检测数据，比如零件的宽度等。有时候，还要求这些数据必须满足一定的可重复性，即多次检测同一数据结果差异不大。这些都是人工检测很难完成的，需要借助于机器视觉技术。

工业自动化中的图像识别和图像分析一般称为机器视觉。机器视觉是计算机视觉目前最产业化和市场化的部分，据估计 2000 年的市场规模是 11 亿美元，而到 2009 年的机器视觉市场规模预计将达 26.2 亿美元。简单的说，机器视觉是指为了完成特定的目的，用摄像机和计算机代替人眼对目标进行识别、跟踪和测量等。与计

算视觉理论相比，它所完成的任务单一，除了特定的场合以外，它没有任何用途。

§2.3.1 机器视觉的一般技术

机器视觉系统一般采用 CCD 摄相机获取数字图像，再采用先进的计算机硬件与软件技术对图像数字信号进行处理，从而得到所需要的各种目标图像特征值，并由此实现模式识别、数据计算等多种功能，然后再根据识别和计算结果显示图像、输出数据、发出指令等，以完成好坏筛选，数据统计等自动化流程。与人工视觉相比较，机器视觉的最大优点是精确、快速、可靠、可重复性以及数字化。

一个典型的机器视觉系统由图像获取系统、图像处理、识别系统以及反馈控制系统等四部分组成。

1. 图像获取系统：图像的获取实际上是将被测物体的可视化图像和内在特征转换成能被计算机处理的一系列数据，它主要由三部分组成：照明、镜头、图像采集设备（图像采集卡、CCD 摄像头）。照明是通过添加人工光源使得所采集到的图像具有最佳的效果（针对特定的目标）。镜头是为了获取效果最佳的图像。图像采集设备将光信号转化为数字信号。

2. 图像处理：包括图像编码、传输，包括图像增强、图像分割、特征提取等等。

3. 分析和识别过程：分析提取到的特征，然后识别、理解和分类等等。

4. 反馈控制系统：将识别过程、分类的结果反馈，利用机械控制装置进行控制。

机器视觉的内容非常广泛：字符 / 物体识别、缺陷检测、图像测量技术和三维形状恢复。但是最基本的任务是图像识别。图像测量技术和三维形状恢复是建立在识别技术基础之上的。缺陷有形状缺陷、颜色缺陷（彩色印刷品）和纹理缺陷（纺织物表面纹理）。

§2.3.2 识别技术

图像识别（image recognition）是指辨识出经验过的某一图像的过程。在图像识别中，既要有当前获得的视觉信息，也要有记忆中存储的信息。只有通过存储的视觉信息与当前的视觉信息进行比较，才能实现对图像的识别。物体识别应用十分广泛。例如，在工业检查任务中将物体可能存在的图像与物体的模型图像进行匹配，得到的是物体的个数（有没有物体）以及物体的姿态，从而可以进行挑选、配置或者质量控制。纺织物生产质量控制中，需要识别当前生产的纺织物颜色和纹理是不是设计的颜色和纹理。

在图像识别中，已经存储的视觉信息称为（物体的）模型。一般情况下，模型来自物体的图像。模型中包括形状、纹理和颜色信息。几乎所有的图像识别算法都

分为两步：第一步：生成待识别物的视觉描述；第二步：对待识别的图像进行处理得到图像的视觉描述；第三步：使用待识别物的描述在图像的视觉描述中匹配、寻找待识别物体。

下面分别介绍基于三种视觉信息的图像识别技术。本文主要强调这些算法所使用的图像表象，其它与本文主旨无关的内容不做介绍。

§2.3.2.1 基于形状的认识技术

模型中物体的摆放可能与物体中摆放姿态不同，从数学上来说他们之间差一个特定类型的变换（transformation），例如：刚体变换 (rigid transformation)、相似变换 (similarity transformation) 或者更一般的说是 2 维仿射变换 (affine transformation)。

变换类记做 T ，可能的变换类包括：平移、欧式变换、相似变换、仿射变换或者任意的投影变换。变换类 T 指定了物体的自由度，也就是待搜索的图像上物体可能的姿态变换。这些变换类被离散化为大量的参数固定的变换 $\{T_i\}$ 。然后，模型 M 被变换为 $\{M_i^t = T_i M\}$ ，对每一个变换后的模型 M_i^t ，计算它与待检测图像的表示对应部分的相似性度量。从数学的观点看，这些变换类构成了变换群。因此，识别可以看作是寻找在变换群下不变的性质，然后通过这些不变性质判定物体到底是什么。

文献 [17] 列举了七种典型的基于形状的认识算法，包括对商业软件 PaxMax® 的介绍。它们都是相对比较成熟的识别技术，已经获得广泛应用。在这些算法中，有的使用用了灰度值本身描述形状，有的使用边缘点和角点等描述形状，这些描述可以看作是图像中形状信息的表象。比较基于灰度值的表象和基于边缘点、角点的表象，后者的速度快并且具有很好的鲁棒性，因为灰度值本身数据量太大，冗余信息太多。如此相对应的是，人类的视觉系统也并没有直接将视网膜上的图像直接作为视觉表象，而是提取了特征作为表象。

§2.3.2.2 颜色描述与基于颜色的识别技术

颜色是物体的基本视觉感受，是区分不同物体的重要条件，因此也是图像分析中识别物体的重要因素。

与基于形状的认识一样。基于颜色认识的第一步就是要对参考图像和待识别图像进行处理，以获得图像的颜色表示。困难的是，数字相机获得图像，其颜色和真实的颜色是不同的，需要进行颜色校正。另外，存在非常多的颜色表示系统，需要从中挑选合适的使用。

颜色描述的主要方法是：指定区域的主演色，使用直方图描述颜色等。由于颜

色不是本文研究的重点，这里不再详细介绍，具体参见文献 [24] 的第 4 章和 §2.4.2。

§2.3.2.3 纹理描述与基于纹理的识别技术

纹理在图像中广泛存在的，但是在计算机视觉和图像处理中，纹理是一个容易引起误解的词语。我们能感觉到什么是纹理，但是我们无法定义什么是纹理。尽管纹理是十分重要的概念，并且非常多的算法来描述和分析纹理，但是到目前为止还没有唯一和明确的定义。每一种纹理分析方法都是根据自己从图像中提取出来的特征来刻画纹理的。因此，纹理分析不仅取决于所研究的图像，还取决于应用目的和从图像中提取出来的特征。尽管没有一致的定义，所有的研究者都同意下面的几个说法：第一，在纹理图像中，邻近像素的灰度存在剧烈的变化；也就是说，在分辨率的限制下，灰度是变化的；第二，在大于图像分辨率的空间尺度上，纹理区域具有相似的性质。

在识别物体的时候，纹理是重要的曲面特征。一幅图像的纹理就是特征化图像灰度曲面的东西。灰度是对一个像素的度量，而纹理只有当图像区域足够大的时候才能感知到，与像素相比，纹理是一种区域性质。下面是纹理的一些定义 [21]。

Haralick [18] 的定义：我们所考虑的图像纹理是非呈现的（non-figurative）和细胞化的。图像的纹理使用基元的数量和类型以及基元在空间上的组织和布局来描述。

Faugeras 和 Pratt [19] 的定义：纹理样本的基本模式和重复频率可以被无形的感知，尽管量化是存在的...。确定而简洁的纹理是基本模式在一定区域上周期或者非周期的重复。

Jain 和 Karu [20] 的定义为：纹理不仅仅是给定像素的灰度值的特征，而且也是围绕着这个像素的相邻像素的灰度值的变化模式。

图像纹理可以认为是颜色（灰度）在空间上的一定形式的灰度变化而产生的感知模式，是真实图像区域固有的特征。图像的纹理是物体的纹理的视觉表现形式，任何物体的表面都会，如果一直放大下去观察的话，都会观察到纹理。由于纹理是物体表面的重要特征，所以基于纹理的识别就有重要的作用。而且在很多时候，图像中的物体并没有明确形状和颜色，能对物体进行区分的只有纹理（纺织物表面）。

基于纹理的物体识别一般过程是：第一步，对物体的表面图像（参考图像）进行处理，抽取一些特征量，用这些特征来描述参考图像的纹理，这就是参考图像的纹理表示；第二步，对待识别的图像，按照同样的方法抽取描述纹理的特征量，也就是使用纹理描述子进行图像表示；第三步，用参考图像的纹理表示分析待识别图像的纹理表示。在这个过程中，最重要的就是图像的纹理表示。为了系统的一般性，我们需要一个一般的纹理表示系统，它能够区分任意人眼能区分的两个物体的

表面。

现有的纹理表示方法可参考 [22] 和 §2.4.2。

§2.3.3 图像表象在图像识别和分析中的重要作用

我们可以看到，在图像识别技术中，图像表示起着核心的作用。图像表示的一般性和精度就决定了系统的一般性和精度。

视觉系统的目的可以看作是为了完成特定的任务，例如判断生产线上零件是否合格，而松散地整合在一起的一些算法。从这个角度来说，算法与表示的设计依赖于所完成任务的特定约束和已知知识。早期的视觉系统中都是根据特定的任务制定特定的表象，目前的很多系统也是如此。

另一种思路就是寻找一般的算法和表示方法使得适应大量的视觉系统，例如 Marr 提出的初始简图和要素图就是希望满足所有的视觉计算需要的表示。这种方法受到了生物视觉的启发，很多迹象表明人的大脑皮层中的视觉模块需要视网膜等器官上具有一般的计算和表示。

一般目的视觉系统必须具有多种表示，并且能够在各种计算中进行数据交换。我们这里考虑的是最初的、直接从图像中得到的一般意义上的表示，这种表示能够代替原始图像为大量的高层计算视觉模型提供数据。

§2.4 多媒体管理系统与 MPEG7

多媒体技术是综合处理声、文、图、音频、视频等多种媒体信息载体的技术。多媒体技术广泛的应用到电子政务、数字图书馆、教育培训、电子出版、通信、文化娱乐等领域。随着多媒体技术的发展，多媒体资源（包括影片、图片、动画、声音等等）越来越庞大。随之带来了两个问题：第一个问题是多媒体信息的压缩存储和传输；第二个问题是多媒体资料的管理和检索。

针对这两个问题，建立了一系列标准，这就是 MPEG 标准。MPEG 表示 Moving Picture Experts Group（运动图像专家组），是一系列视频 / 音频压缩编码标准的名字。MPEG 优于其他影像压缩方案的地方是：具有很高压缩比和很好的兼容性。到目前为止，MPEG 已经提出了 MPEG-1, MPEG-2, MPEG-4, MPEG-7 四个标准并已经启动了 MPEG-21 的研发。其中，MPEG-1, MPEG-2, MPEG-4 主要解决视频 / 音频的压缩编码和传输问题，这个方面不是本文的核心，本文不再深入介绍，详情请查看 MPEG 的主页。

多媒体管理系统应该提供基于内容的视觉信息检索功能，例如我希望能立刻找到神舟 6 号发射升空的画面，又例如绘制一个草图能找到一个乒乓球拍等等。基于内容的视觉信息检索是介于用户和数据库之间的信息服务系统。用户提交查询要求后，基于内容的视觉信息检索系统首先对查询数据进行处理，主要是提取与视觉内

容相关的一些特征；然后使用这些特征与多媒体数据库中的视觉信息提取出的相关特征进行匹配，以找到所需要的资源。如果视觉媒体自身能保存关于自己内容的描述，那么基于视觉信息的检索就会相对简单。MPEG-7 就具有这样的功能。

§2.4.1 MPEG7 简介

MPEG-7 [23, 24, 25, 26] 是国际化标准组织制定的国际标准的简称，其正式名称是多媒体内容描述界面 / 接口 (multimedia content description interface)。MPEG-7 并不是一种压缩编码方法，而是一个多媒体信息内容表达描述的标准，它试图规范对不同种类多媒体信息内容的描述而不受表达形式的限制。MPEG-7 的目标是建立对不同多媒体信息（视频 / 音频）内容的标准描述（指定描述符和描述方案），这些描述符与信息的内容密切相关，以便快速有效的查询和访问多媒体信息。MPEG-7 只是规定了多媒体内容使用哪些描述符（特征）进行描述，但是并没有规定如何提取这些特征。

MPEG-7 所描述的对象是各种多媒体资源，包括静止图像、序列 / 运动图像、计算机图形、3-D 模型、动画、语言、声音等等。

MPEG7 中使用了很多概念，这里只介绍下面需要的几个基本概念。数据 (data) 是指用 MPEG-7 描述的视频 / 音频信息，例如一幅彩色图像或者一首歌曲。特征 (feature) 指描述数据本身特色的特性，比如图像的颜色。描述符 (descriptor) 指定义特征表示句法和语法的表达与结构，例如颜色的直方图。描述符值 (descriptor value) 指描述符对特定数据的取值，例如直方图各元素的值。描述方案 (descriptor scheme) 表示指定其元素（描述符和描述方案）关系的结构和语义，例如一幅彩色图像使用颜色直方图表示其颜色种类和数量。描述 (description) 指用描述方案和描述符描述数据，例如，用颜色直方图描述图像。编码的描述 (coded description) 指为满足一定需要而已经编码的描述，例如直方图的距。描述定义语言 (description definition language) 是指用以产生新的描述方案和描述符的语言，也允许修改和扩展已有的描述方案和描述符。

§2.4.2 MPEG7 中描述视觉特征的描述符与图像表象

MPEG-7 中考虑了五类基本的视觉特征，对应的使用了五类描述符：颜色描述符、纹理描述符、形状描述符、运动描述符和位置描述符。其中，前三类描述符都是针对单幅静止图像的描述，是图像表象的三个重要方面。由于 MPEG-7 的目的是建立对图像信息完备而又一致的描述，因此，MPEG-7 中提出的颜色描述符、纹理描述符和形状描述符等单类描述符必须尽可能的完备而又一致：使用这些描述符能尽可能完备地表达图像中的视觉信息，并且更重要的是这些描述符适用于所有图像。而这两者则正是图像表象所努力追求的目标：我们就是希望找到一种表达体

系，能完备而又一致的描述所有的图像中的所有视觉信息。本文的目的正是从理论上分析什么是满足这两个要求的视觉表象。

MPEG-7 中提出的这些描述符是全世界相关研究机构和应用部门共同讨论的结果，具有理论基础和实验基础，是本文后面分析“什么是正确的图像表象”的重要参考。因此，下面简要介绍 MPEG-7 中这三个方面的描述符。

MPEG-7 中的形状描述符有三种，第一种是区域形状描述符 (Region Shape Descriptor)，使用一组角放射变换系数来描述单个连通区域和多个不连通的区域。Angular Radial Transform (§3.3.2) 定义了一组二维的复值正交基函数函数，将二维区域投射到这些基函数上，得到的系数归一化后就可以描述区域的形状。第二种是轮廓形状描述符 (Contour Shape Descriptor)，利用轮廓的多尺度曲率空间 (Curvature Scale Space, CSS) 来描述封闭的轮廓 (§3.3.3)。具体的说，对封闭轮廓线进行多尺度变换，得到多个尺度的轮廓线，随着尺度的变换，轮廓线越来越平滑。多尺度变换可以使用高斯函数对曲线上的点进行平滑来完成。然后，在每一个尺度计算曲线的曲率，得到曲率变化的零交叉点，使用这些点就可以描述轮廓的形状。第三种是 3D 形状描述。

MPEG-7 中的颜色描述符有很多。颜色空间描述符 (Color space Descriptor)，描述了 MPEG-7 颜色描述符的颜色空间，包括 RGB 和 HSV 等以及各种颜色系统与 RGB 的线性变换矩阵。颜色量化描述符 (Color Quantization Descriptor)，描述了颜色空间的均匀量化（需要和主颜色等配合使用）。主颜色描述符 (Dominant Color Descriptor)，对任意形状的区域，指定一组主颜色。可伸缩颜色描述符 (Scalable Color Descriptor)，定义 HSV 空间的颜色直方图，然后用 Harr 变换编码，它的二进制表达在 Bin 的数量和 Bit 表达精度上都是可伸缩的。颜色布局描述符 (Color Layout Descriptor)，描述了整幅图像或者图像的部分区域的颜色空间分布状况。颜色结构描述符 (Color-Structure Descriptor)，既描述颜色内容本身，也借助结构元素描述颜色内容的结构。

MPEG-7 中的纹理描述符有三种。同质纹理描述符 (Homogenous Texture Descriptors)，同质纹理描述符使用 62 个数字（每一个都量化为 8bits）提供了纹理信息的量化表示。这 62 个数字的计算方法是，将被图像用一组带有方向和尺度参数的 Gabor 滤波器进行滤波，一共是六个方向和五个尺度。一共得到 30 个滤波结果。每一个滤波结果频率域的一阶矩和二阶矩就是两个描述数字。纹理浏览描述符 (Texture Browsing Descriptors)，描述纹理的感知特点，例如规则性、方向性和粗糙性。计算方法和同质纹理描述符类似，首先使用一组带有方向和尺度参数的 Gabor 滤波器进行滤波，然后通过分析滤波结果，找到纹理主要的方向。接着分析滤波后的图像沿着这两个（第二个主方向是可选的）主方向投影，来确定纹理的规则性和

稀疏性。同质纹理描述符和纹理浏览描述符提供了表示相似纹理区域 (homogeneous texture regions) 的多尺度方法。边缘直方图描述符 (Edge Histogram Descriptors), 描述了 5 种边缘 (4 种方向边缘和一种非方向边缘) 的空间分布统计。

§2.5 基于内容的图像和视频搜索:WWW 图像检索的发展目标

互联网改变了人与人之间的连接方式, 空间上的距离已经不再是人与人之间交流的障碍。对很多人来说, 互联网已经成为他们生活中不可缺少的一部分。

§2.5.1 搜索引擎

随着互联网的发展, 互联网的用户越来越多, 互联网上的信息也越来越多。人们一般称为互联网上的信息量是海量。但是, 互联网上的信息是分散的、杂乱的。我们知道互联网上存在我们需要的信息, 但是我们不知道在那儿能找到我们需要的信息, 大量的时间花在信息的寻找和整理。

搜索引擎 [27] 就是针对这个问题出现的。搜索引擎是为用户提供信息“检索”服务的网站, 它使用某些程序把互联网上的所有信息归类, 以帮助人们在茫茫网海中搜寻到所需要的信息。

早期的搜索引擎是把互联网中的资源服务器的地址收集起来, 由其提供的资源的类型不同而分成不同的目录, 再一层层地进行分类。人们要找自己想要的信息可按他们的分类一层层进入, 就能最后到达目的地, 找到自己想要的信息。这其实是最原始的方式, 只适用于互联网信息并不多多的时候。随着互联网信息按几何式增长, 出现了真正意义上的搜索引擎, 这些搜索引擎知道网站上每一页的开始, 随后搜索互联网上的所有超级链接, 把代表超级链接的所有词汇放入一个数据库。这就是现在搜索引擎的原型。

互联网上的信息有多种形式, 文本信息 (HTML 网页, WORD 文档, PDF 文档等)、图形图像信息 (Logo, Bar, 照片、设计图等等)、视频信息和音频信息。对这些形式, 我们都需要设计相应的搜索引擎。但是, 只有文本搜索引擎是成熟的技术。其他类型的信息的搜索引擎在技术上还需要深入的研究。

§2.5.2 现有的图像搜索引擎技术

图形图像是互联网上非常常见的信息形式, 几乎每一个网页上都有图形图像, 已经成为表示网页内容不可缺少的组成部分。可能是网站的 Logo 图像, 也可能是网站设置的广告 (Bar), 还可以是人的照片、风景图像或者设备的图像。图像比文字描述更直观, 但是与视频信息相比数据量少, 浏览方便, 因此是一种重要的信息来源。

实现对网页中文本信息提取的同时, 如何再为用户抽取所需的图片资料, 是搜

索引引擎技术中一个重要的方面。于是,各种基于 Web 的图像检索系统应运而生,极大地方便了用户对网上图像进行检索。

AllTheWeb [28], AltaVista [29], Cobion [30], DINO online [31], Ditto [32], Google [33], Picsearch [34], Tiscali [35] 是比较著名的几个图像搜索引擎。目前,这些图像搜索引擎主要通过以下两种方法识别图像:

(1) 自动查找图像文件。通过两个 HTML 标签,即 `imgsrc` 和 `href` 来检测是否存在可显示的图像文件, `imgsrc` 表示“显示下面的图像文件”,链接的是嵌入式图像;而 `href` 则表示“下面是一个链接”,导向的是被链接的图像。搜索引擎通过检查文件扩展名来判断其导向的是否为图像文件,如果文件扩展名是“.gif”,“.jpg”或者“.png”等等,即是一个可显示的图像。

(2) 人工干预找出图像。进行分类,由人工对网上的图像及站点进行选择。这种方法可以产生准确的查询体系,但劳动强度太大,因此要限制处理图像的数量。

由于图像不同于文本,需要人们按照各自的理解来说明其蕴含的意义,因此图像检索比文本的查询和匹配要困难得多。目前,在多媒体网络中,主要有两类图像搜索引擎技术正在研究和应用之中,一类采用的是传统的基于关键字的图像检索技术,另一类采用的是基于内容的图像检索技术。

传统的图像检索技术是基于关键字的精确匹配检索,系统内的图像用关键字标识,检索线索是与标识相一致的关键字,即输入是关键字,输出是图像。它又包括两种途径:

(1) 基于图像外部信息进行检索。即根据图像的文件名或目录名、路径名、链路、ALT 标签以及图像周围的文本信息等外部信息进行检索,这是目前图像搜索引擎采用最多的方法。在找出图像文件后,图像搜索引擎通过查看文件名或路径名确定文件内容,也可以通过查看图像的标题来匹配检索词。

(2) 基于手工标注的检索。手工对图像的内容(如物体、背景、颜色等)进行描述并分类,将其标注为一系列关键字,并建立索引。检索时,将主要在这些描述词中搜索用户输入的关键字。这种查询方式是比较准确的,一般可以获得较好的查准率,但需人工参与,劳动强度大,因而限制了可处理的图像数量。另外,由于图像所包含的信息量庞大,不同用户对于同一张图像的看法又不尽相同,导致了对图像的标注缺乏统一标准。

§2.5.3 图像搜索引擎发展方向

利用文本信息或者外部信息的图像搜索引擎具有明显的弊端。例如,在 Google 图像搜索引擎中输入“moon”检索,搜出的图像中,只有一部分图像中有月亮。它们出现在搜索结果中,是因为图像所在的网页中有“moon”这个关键字或者链接中、注释中有这个关键字。我们需要基于图像内容的搜索引擎。

基于图像内容的检索主要是由图像分析软件自动抽取图像的颜色、形状、纹理等特征，建立特征索引库，其输入为用户要查找的图像的大致特征描述或示例，通过一定相似性匹配规则，输出为与之具有相近特征的图像，按相似程度排列，供用户选择，从而把在传统图像检索技术中一般用户难以完成的图像特征描述、提取与识别等难题，交由系统去解决。

这是一种基于图像本身特征层次的检索，特别适用于检索目标明确的查询要求（例如对商标的检索），但目前这种较成熟的检索技术主要应用于图像数据库的检索。在图像搜索引擎中应用这种检索技术还有一定困难，但已有部分图像搜索引擎尝试了这种检索方法 [39, 40]。

MIT 的 Rob Fergus [36] 在牛津大学学习时提出了一种图像检索算法 [37, 38]。这种算法具有一定智能，并且可操作性很强。在这个算法中，描述图像内容的特征有两种：曲线段 (Curve Segments) 和像素片 (Pixel Patches)。

首先使用边缘检测算法 (Canny 算法) 获得边缘轮廓，然后利用曲线上曲率大的点 (角点) 将边缘分成相对平滑的若干段，这就是曲线段。每一个曲线段做一个仿射变换，使得每一个曲线的起点为 (0, 0)，终点为 (1, 0)。在变换后的曲线上，等间距取 13 个点，用这 13 个点表示曲线段的形状特征，并用另外两个数据表示曲线的方向。每个曲线段用这个 15 维的向量来描述。像素片 (Pixel Patches) 是通过显著点 (Salient points) 检测算法 [41] 得到，显著点包括几何特征点 (角点)，稀有点和复杂点。视觉系统对这些点更敏感。在这个点处，根据特征点的尺度，提取出 11×11 的灰度片段，称为像素片。对 11×11 的灰度片段进行主成分分析，使用前 15 个量来描述像素片。另外，曲线段和像素片都有其位置和尺度 (特征所占空间的大小)。总体来说，使用位置 $\mathbf{X}$ ，尺度 $\mathbf{S}$ 和描述子 $\mathbf{D}$ 来描述图像中存在的曲线段或者像素片。

图像中所有曲线段或者像素片的描述，构成了图像内容描述的特征集。

基于这些特征集和 Bayesian 推断，在大批量数据的基础上 (Google 图像搜索引擎)，估计图像中存在某类物体 *Object* 的存在的概率 $p(\text{存在}Object|Image)$ 和不存在 *Object* 的概率 $p(\text{不存在}Object|Image)$ ，如果

$$R = \frac{p(\text{存在}Object|Image)}{p(\text{不存在}Object|Image)} \quad (2.1)$$

大于某一阈值 T ，就可以认为图像中存在 *Object*。

Fergus 算法是给予内容的搜索引擎算法 [40] 中的优秀代表。首先，使用一组特征来描述图像中的视觉信息，然后使用统计学习理论学习，以获得物体类的视觉描述。然后用学习获得知识去指导进行下一步的检索。Fergus 算法最大的特点是：它

是建立在现有搜索引擎的基础之上的，统计学习的数据库就是现有 Google 图像检索的结果。于是，很容易实现下面的任务：

1. 通过输入关键词“自行车”，查询 Google 图像检索引擎（传统图像搜索引擎技术）获得大批图像，其中很多并不含有“自行车”的；
2. 利用 Google 返回的结果，进行特征提取和统计学习，学习“自行车”的视觉模型；
3. 利用“自行车”的视觉模型，提升图像检索的结果。

§2.5.4 图像搜索引擎与图像的视觉表象

图像搜索引擎问题分为三个子问题：

1. 图像视觉信息的描述（图像的视觉信息表象），即把图像用具有语义的线条和点、颜色和纹理等描述，得到图像的一个语义描述，形成图像的特征集合。这样的特征集合应该尽可能完全的表示图像中的信息。
 2. 抽象概念理解，也就是语义库的建设：一个抽象的概念与哪些特征（线，点，颜色，纹理）相关。这个过程可以通过机器学习来完成，例如 Rob Fergus 等的研究；语义库建设完全建立在图像表象的基础上。
 3. 图像检索，利用语义库和图像的特征集合来实现图像检索。
- 显然，图像搜索引擎算法的优劣直接取决于所使用的图像的视觉表象。

第三章 图像的视觉信息表象理论

上一章我们简要介绍了视觉计算理论、表象学、图像分析技术、媒体内容描述界面 / 接口 (MPEG-7) 和 WWW 图像检索技术。通过简单的分析后我们发现, 这些领域中存在一个共同的基本问题: 如何描述图像中的视觉信息, 也就是图像的视觉信息表象理论。本章提出了该理论的基本框架。

§3.1 图像表象理论的体系结构

所谓表象是指一种能把某种实体或某几类信息表达清楚的形式化系统, 以及说明该系统如何行使其职能的若干规则。使用某一种表象描述一个给定的实体所得到的结果称为该实体在这种表象下的一个描述。例如, 阿拉伯数制, 二进制数制, 它们都是表象“数”的形式化系统。阿拉伯数字表象由数集 $(0, 1, 2, \dots, 8, 9)$ 中所抽取的一串符号组成; 构造特定整数 n 的描述规则是: 先把 n 分解为数符与位值的乘积之和, 位值用 10 的幂表示, 再按幂的大小把数符自左向右排成一行。因此三十七等于 $3 \times 10 + 7 \times 1$ 变成 37, 即该数在阿拉伯制下的描述。这个描述明确表达了数按 10 的整数幂所作的分解。

图像的视觉信息表象理论 (Representation Theory of Visual Information in Image) 是研究如何把图像中的形状、纹理和颜色等视觉信息表达清楚的形式化系统和说明这个系统如何获得、描述、重构图像中视觉信息的若干规则。以下简称图像的表象理论 (Representation Theory of Image)。图像形状信息、纹理信息和颜色信息的这种描述成称为图像的表象: 图像的形状表象 (Shape Representation of Image)、图像的纹理表象 (Texture Representation of Image) 和图像的颜色表象 (Color Representation of Image)。表象也可以称之为表示。

图像表象理论包括以下几方面的问题:

1. 图像的表象应该满足哪些基本要求?

首先, 图像表象应该具有一般性, 即自然图像中发现的一些结构能够被表示, 人眼能区分的视觉信息应该都能够被区分; 图像表象应该具有简洁性, 即这种表象能够减少原始图像中存在的对视觉感知无意义的冗余; 图像的表象应该具有明确性, 即这种表象能够将原始图像中存在的重要结构明确地表示出来; 图像表象应该具有完全性, 即能够满足大量的高层视觉领域的需要, 能够提供所有生物视觉系统能提供的视觉信息; 视觉表象应该是分层次的组织, 即在需要时能迅速恢复不同程度的细节。

2. 图像的形状表象、纹理表象和颜色表象到底是什么, 也就是说应该用什么 (特征) 数据来描述它们? 为什么它们能表达图像中的形状、纹理和颜色信息? 图

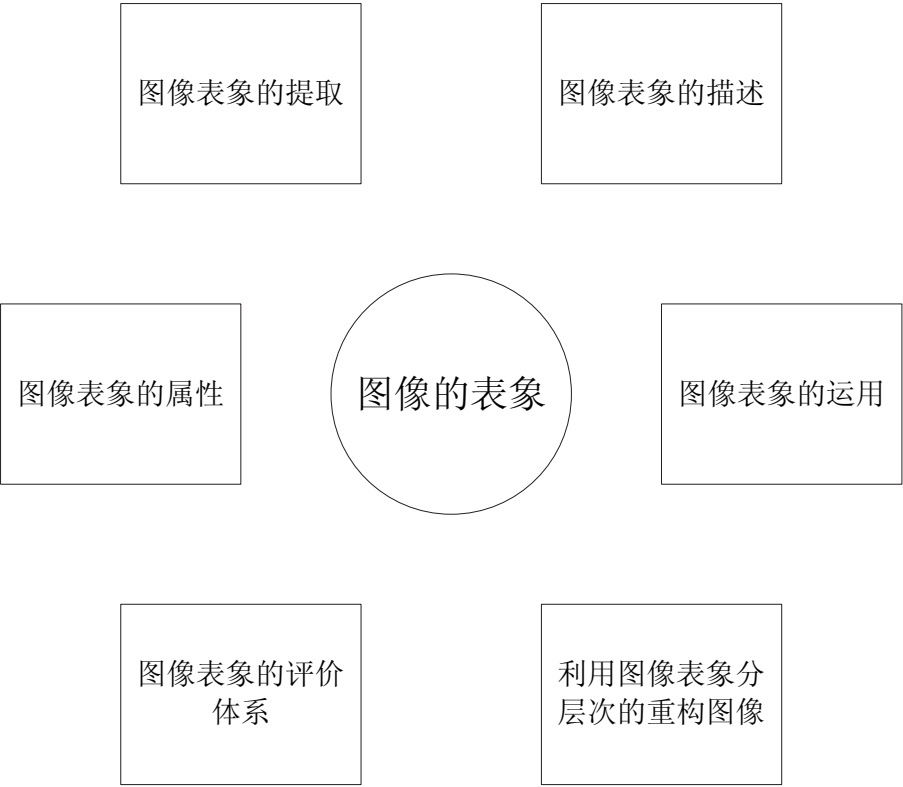


图 3.1: 图像的视觉信息表象理论

像表象具有什么性质？具有什么局限性？

这是图像表象理论的核心问题。

3. 如何从图像中提取这些表象？

需要设计提取这些表象的算法，这些算法必须足够快。同一种表象的提取可能有很多种方法，需要知道这些算法各有什么特点？它们的适合什么应用。

4. 如何对图像的表象进行描述？

提取得到图像表象后，需要用合适的、紧凑的数据将图像表象存储下来，这种存储是压缩性的存储。更重要的是，这些表象系统在计算机中应该如何描述。

5. 如何使用图像的表象恢复各种不同层次的图像？

为了适应不同应用的需要，表象系统应该能提供各种不同细节等级的视觉信息。这种不同层次的细节恢复必须足够方便和快捷。

6. 如何评价图像的表象体系？

评价图像表象体系优劣的参照物是人的视觉感知能力！也就是说，人的视觉能发现的形状、能区分的纹理和能感知到的颜色差异都必须能在图像表象中体现出来。应该建立客观的评价平台评价现存的表象体系。

7. 如何运用视觉表象？

例如，如何使用这些表象进行图像识别？如何利用图像的表象恢复物体的 3D 视觉表象等。

§3.2 图像表象理论的研究现状

1966 年，Gibson [8] 试图解决与图像表象问题很接近的问题：人是如何在不断变化着的感觉的基础上获得恒定的知觉呢？它的结论是：感知系统的作用是检测不变量。于是，按照图像表象理论的解释，Gibson 认为图像的表象是其表面的不变量。

Marr 和 Hildreth [8] 最早开始寻找更符号化但是仍然是完全的图像表示。1977 年, Bell 实验室的 Logan 定理提出了一个定理：“带通的一类一维信号的零交叉点构成了原始信号的一个完全的表示。” Marr 猜想这个定理可以推广到二维任意连续函数，于是 LoG 算子检测出的图像的零交叉点就可以完全决定图像曲面，于是边缘的完全性就可以保证。

图像的各种变换，图像的离散余弦变换、图像的傅里叶变换和图像的小波变换等都可以看作是图像的表象。使用这些表象可以进行识别任务。

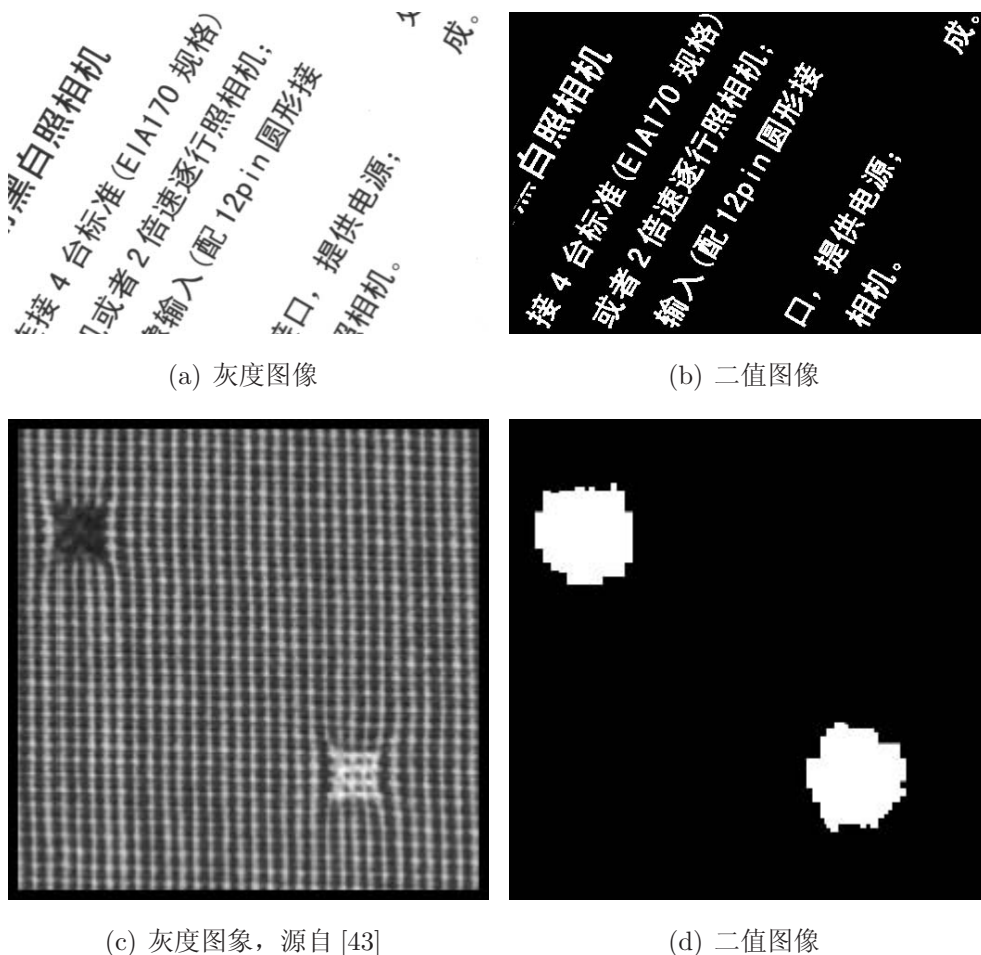


图 3.2: 图像的二值表象

在实践中，已经提出了大量的表示，目前最常用的是图像的二值图像。

§3.3 二值表象

二值视觉是最简单的视觉，也是目前机器视觉领域中最常用的技术。所谓的二值视觉就是在特定环境下，场景中的物体相对简单以至于可以使用简单的二值图像就能几乎完全的表达我们所需要的所有信息。

二值视觉的应用有很多。如图 3.2 所示，文档图像处理中字符区域都可以用二值图像来表示，字符所在的位置为 0，背景所在区域为 1。在工业域中，为了进行产品质量控制，常常要进行产品的缺陷检测。为了实现自动化，有时候可以先对产品关心的区域成像，然后通过图像处理的方法进行缺陷检测。缺陷检测中常常将图像转化为二值图像，有缺陷的地方为 1，没有缺陷的地方为 0。以下均假设值为 1 的部

分表示感兴趣的区域，值为 0 的部分表示不感兴趣的区域。

显然，二值视觉或者基于二值的图像分析使用的是二值逻辑。处理的问题相对简单，在机器视觉中，为了能获取简单并且高质量的二值图像，往往使用额外的辅助光源。获得二值表象过程称为图像的二值化，图像的二值化算法参考 [42]。

二值化获得是图像上的一块区域，这一块区域表示一个物体或者一个抽象区域。对于这个区域，需要进一步使用量化的特征刻画这个区域，下面是几种常用的技术。

§3.3.1 二值区域的 Blob 描述

Blob 是指具有相似图像特征（如颜色、纹理等）而且在空间上是连通的像素组成的块或者斑点。二值化的结果是获得一个个 Blob。Blob 分析的主要作用是获得 Blob 的几何特征，如尺寸、参数和密度等，经常用于物件的检索和定位。Blob 分析实际上就是根据图像的灰度值对图像进行分割，然后对分割出的图像进行各种统计分析。通过 Blob 分析可以完成计数统计、位置、面积、紧凑度、惯性主轴、周长、方向等计算，还可以提供相关斑点间的拓扑结构，是一种十分有用的分析手段。

Blob 分析主要适用于以下机器视觉应用：二维目标图像、高对比度图像、存在/缺席检测、数量范围和旋转不变性需求。另一方面，Blob 分析并不适用于以下机器视觉应用、低对比度视觉、不能够用 2 个灰度表示的特性、图形检测需求。

Blob 描述，为了描述而知区域的形状，很多公司都开发了 Blob 计测函数库 [44]。

§3.3.2 二值区域的角映射系数描述

MPEG-7 推荐使用的基于区域的二值图像描述符是角径向变换 (Angular Radial Transform, ART) [45, 46, 47] 系数。

角径向变换首先定义了一组极坐标系下二维复值正交基函数函数 $V_{n,m}(\rho, \theta)$ （以下称为角径向变换的基函数）， n 和 m 是非负整数，称为 $V_{n,m}(\rho, \theta)$ 的阶。

$V_{n,m}(\rho, \theta)$ 定义如下，

$$V_{n,m}(\rho, \theta) = R_n(\rho)A_m(\theta) \quad (3.1)$$

其中， $R_n(\rho)$ 是用余弦函数定义的实值径向函数

$$R_n(\rho) = \begin{cases} 1, & \text{if } n = 0 \\ 2 \cos(\pi n \rho), & \text{其它} \end{cases} \quad (3.2)$$

$A_m(\theta)$ 是角基函数。为了具有旋转不变性，角基函数使用指数函数定义

$$A_m(\theta) = \frac{1}{2\pi} \exp(jm\theta) \quad (3.3)$$

对任意 2 维函数 $f(\rho, \theta)$ ，阶数为 m 和 n 的角径向变换定义为

$$F_{n,m} = \int_0^{2\pi} \int_0^1 V_{n,m}(\rho, \theta) f(\rho, \theta) d\rho d\theta \quad (3.4)$$

可以证明，角径向变换的基函数 $V_{n,m}(\rho, \theta)$ 在单位圆上是正交的。即，只要 $n_2 \neq n_1$ 或者 $m_2 \neq m_1$ ，就有

$$\int_0^{2\pi} \int_0^1 V_{n_1,m_1}(\rho, \theta) V_{n_2,m_2}(\rho, \theta) d\rho d\theta = 0 \quad (3.5)$$

这就保证了角径向变换的各个系数之间是相互独立的，即系数中不含有信息冗余。

$F_{n,m}$ 是复数，一般使用 $\{M_{n,m}\}$ 作为区域形状的描述子，其中 $M_{n,m} = \|F_{n,m}\|$ 。可以证明， $M_{n,m}$ 在图像 $f(x, y)$ 的旋转变换下是不变的 [47]。

随着 n, m 的增加， $V_{n,m}(\rho_j, \theta_j)$ 中高频信息越多，视觉感兴趣的信息越少。所以，在应用中，只对若干个较小的 n, m 计算 $V_{n,m}(\rho_j, \theta_j)$, $0 \leq n \leq n_0, 0 \leq m \leq m_0$ 。给定一个二值图像， $M_{n,m}$ 的计算步骤如下：

1. 对 θ 和 ρ 进行离散化，分别得到离散点列 $\{\theta_i, i = 0, \dots, P\}$ 和 $\{\rho_j, j = 0, \dots, Q\}$ 。根据公式 3.1 计算 $V_{n,m}(\rho_j, \theta_j, i = 0, \dots, P, j = 0, \dots, Q)$ 在这些点上的值；
2. 计算值为 0 的区域的重心位置 (m_x, m_y) ；
3. 计算感兴趣区域中与重心位置距离最大的点到重心 (m_x, m_y) 的距离 D ，称为区域的半径；
4. 利用 D 将感兴趣的区域变换到单位圆中；
5. 利用公式 3.4 和 $V_{n,m}(\rho_j, \theta_j, i = 0, \dots, P, j = 0, \dots, Q)$ 计算 $F_{n,m}$ 和 $M_{n,m}$ ，在计算中，积分变为求和即可；注意，二值区域 $f(\rho, \theta)$ 表示为

$$f(\rho, \theta) = \begin{cases} 1, & (\rho, \theta) \in \text{Shape} \\ 0, & \text{其它} \end{cases} \quad (3.6)$$

$\{M_{n,m}\}$ 可以描述单个连通区域和多个不连通的区域的形状。其中 $\{M_{0,0}\}$ 表示感兴趣区域的面积，由于我们感兴趣的只是区域的形状，所以常常用 $\{M_{0,0}\}$ 将 $\{M_{m,n}\}$ 归一化，得到 $\tilde{M}_{m,n}$ ，

$$\tilde{M}_{m,n} = \frac{M_{m,n}}{M_{0,0}} \quad (3.7)$$

对区域形状的表达用 $\{\tilde{M}_{n,m}\}$ ，而两个区域形状 $S1, S2$ 相似性 $\text{SimMeasure}(S1, S2)$ 的可以使用

$$\text{SimMeasure}(S1, S2) = \sum_{n=0}^{n=n_0} \sum_{m=0}^{m=m_0} |\tilde{M}_{m,n}^{S1} - \tilde{M}_{m,n}^{S2}| \quad (3.8)$$

一般选取 $m_0 = 11, n_0 = 2$ 就可实现对二值区域的形状描述。在很多应用中, 为了存储的有效性, 还要对 $M_{n,m}$ 进行量化。例如, 在 MPEG7 中, 每个归一化后的系数 $\tilde{M}_{m,n}$ 被量化为 16 个等级 (4 比特)。这样, 就可以在保持形状识别能力的前提下, 实现高压缩率的存储。

§3.3.3 二值区域的多尺度曲率空间描述

利用区域轮廓的尺度曲率空间 (Curvature Scale Space, CSS) [48] 也是 MPEG-7 推荐使用的一种描述区域形状的方法。由于区域和其轮廓是一一对应的, 所以对区域轮廓的描述就等价于对二值区域的描述。

假设区域的轮廓线 (平面曲线), 用向量值参数方程 $\mathbf{r}(u) = (x(u), y(u))$ 表示。 $\mathbf{r}(u)$ 的导数 $\dot{\mathbf{r}}(u)$ 定义为,

$$\dot{\mathbf{r}}(u) = d\mathbf{r}/du = (\dot{x}(u), \dot{y}(u)) \quad (3.9)$$

于是, 对任意参数 u , 有

$$\|\dot{\mathbf{r}}(u)\| = (\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}} \quad (3.10)$$

曲线的切向量 $\mathbf{t}(u)$ 和法向量 $\mathbf{n}(u)$ 定义为

$$\mathbf{t}(u) = \frac{\dot{\mathbf{r}}}{\|\dot{\mathbf{r}}\|} = \left(\frac{\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}}, \frac{\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}} \right) \quad (3.11)$$

$$\mathbf{n}(u) = \left(\frac{-\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}}, \frac{\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}} \right) \quad (3.12)$$

曲率是几何体不平坦程度的一种衡量。平坦对不同的几何体有不同的意义 (对曲线就是直线, 对曲面就是平面)。对于一个以参数化形式给出的平面曲线 $\mathbf{r}(u) = (x(u), y(u))$, 参数 u 处的曲率 $\kappa(u)$ 定义 [48] 为

$$\kappa(u) = \frac{\dot{x}(u)\ddot{y}(u) - \dot{y}(u)\ddot{x}(u)}{(\dot{x}(u)^2 + \dot{y}(u)^2)^{3/2}} \quad (3.13)$$

如果 u 为归一化的弧长参数 $0 \leq s \leq 1$, 则

$$\kappa(s) = \dot{x}(s)\ddot{y}(s) - \dot{y}(s)\ddot{x}(s) \quad (3.14)$$

$\kappa(u)$ 的零交叉点, 即 $\kappa(u) = 0$, 可以作为曲线的特征点。

假设 $\Gamma = \{(x(s), y(s)) | s \in [0, 1]\}$ 表示归一化的弧长为 s 为参数的平面参数曲线。则尺度为 $\sigma > 0$ 的曲线 Γ_σ 定义为 $\Gamma_\sigma = \{(X_\sigma(s), Y_\sigma(s)) | s \in [0, 1]\}$, 其中

$$X_\sigma(s) = x(s) \otimes G(s; \sigma) = \int_{-\infty}^{+\infty} x(t)G(s-t; \sigma)dt \quad (3.15)$$

$$Y_\sigma(s) = y(s) \otimes G(s; \sigma) = \int_{-\infty}^{+\infty} y(t)G(s-t; \sigma)dt \quad (3.16)$$

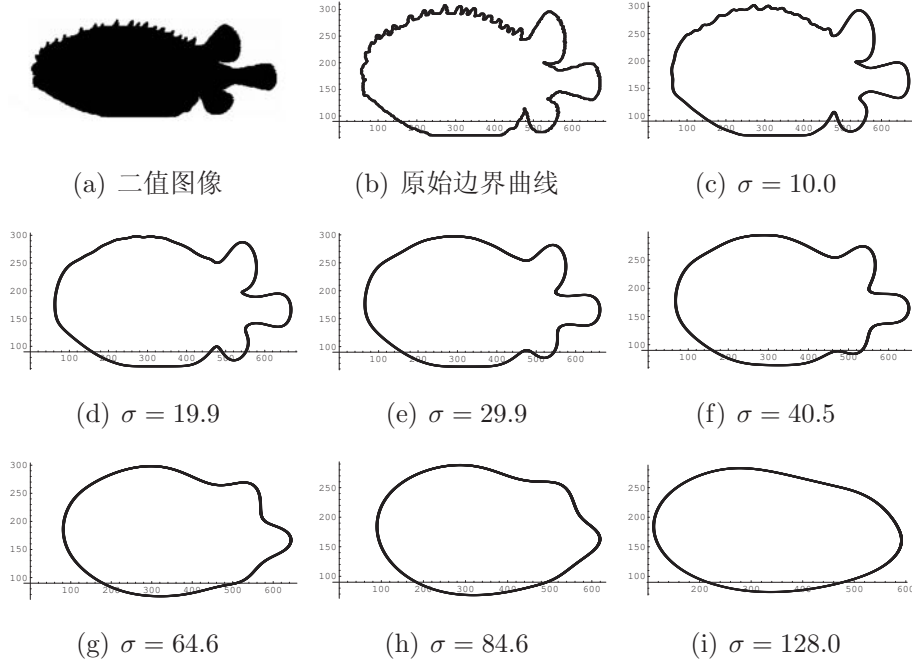


图 3.3: 曲线的尺度变换

这里, $G(s; \sigma), \sigma > 0$ 表示高斯函数,

$$G(s; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{s^2}{2\sigma^2}} \quad (3.17)$$

令 $\Gamma_\sigma = \Gamma$, 则通过多尺度变换, 将 Γ 嵌入到一族曲线 $\{\Gamma_\sigma, \sigma > 0\}$ 中, 见图 3.3。而这一族曲线中的每一条曲线都可以计算其曲率 $\kappa(s, \sigma)$, 进而求得其曲率的零交叉点 $\{(s, \sigma) | \kappa(s, \sigma) = 0\}$ 。

$\{(s, \sigma) | \kappa(s, \sigma) = 0\}$ 就称为曲线 Γ 的曲率尺度空间 (Curvature Scale Space, CSS) 图像。

可以证明曲线和其曲率尺度空间是一一对应的 [48], 所以曲率尺度空间可以用来描述曲线, 所以也称为曲线的指纹。关于曲率尺度空间理论和应用请参考专著 [49] 和相关网址 [50, 51]。下面给出提取曲率尺度空间的大致步骤:

1. 从二值图像上追踪得到边界, 得到边界点 $\{(x_n, y_n), n = 0, \dots, N\}$ (整数对序列), 算法可以参考 [52];
2. 对 $\{x_n, n = 0, \dots, N\}$ 和 $\{y_n, n = 0, \dots, N\}$ 进行高斯滤波可以得到各种尺度的曲线 $\{(\tilde{x}_n, \tilde{y}_n), n = 0, \dots, N\}$, 参见图。在每一个尺度, 重复 3, 4, 5 提取曲率零交叉点;
3. 将 $\{(\tilde{x}_n, \tilde{y}_n), n = 0, \dots, N\}$ 重新采样 $\{(\bar{x}_m, \bar{y}_m), m = 0, \dots, M\}$, 使得相邻两点间的距离 (近似等于弧长参数) 相等;

4. 使用中心差分计算每个点的各阶导数；使用式 3.14 计算各点的曲率 $\kappa(m), m = 0, \dots, M$;

5. 计算曲率 $\kappa(m), m = 0, \dots, M$ 的零交叉点.

§3.3.4 二值区域的骨架描述

1967 年 Blum 通过烧草模型 (Grassfire Model) 首先给出了骨架的定义 [53], 他假设图形边界点同时着火, 火源向图形内部各个方向等速的燃烧直至熄灭, 所有熄灭点的点集构成了该图形的骨架。Blum 还给出了骨架的另一种等效定义, 即骨架是所有最大圆 (Maximal Disks) 的圆心集合, 最大圆是完全包含在图形内部的圆, 并且不被任何其它包含在图形内部的圆所包含。骨架的定义还有很多种, 如脊点投影骨架, 形态学骨架等, 可以证明这些定义都是等效的。

骨架是原始二值图像的一种压缩表示, 与原始图形保持了相同的拓扑结构, 并且存在于图形的对称轴上, 能够同时反映图形的拓扑与形状信息。

对于骨架的求法, 当前的研究主要可以分为以下两类, 这两类研究是根据前面的两种定义发展而来的:

1. 细化的方法。这种方法就是在模拟烧草的模型, 即逐层均匀的剥掉图形的边界, 最后剩下最里层的已经不能再剥掉 (否则会影响连通性) 的部分就构成了图形的骨架。

2. 基于距离变换的方法。根据最大圆的定义可以推导出最大圆必是图形的内切圆, 因此考察以某点为圆心作的内切圆, 如果该圆不被其它的任何内切圆包含, 则此点作的内切圆是最大圆, 此点是骨架点。距离变换记录了图形内所有点到边界的最短距离, 利用距离变换可以生成任意点的内切圆及比较两个内切圆的包含关系。

基本算法是:

该算法的步骤如下:

- (1) 计算二值图像的距离变换;
- (2) 寻找骨架种子点。该步的目标是找到一个合适的起始骨架点;
- (3) 骨架生长。这一步考虑如何从一个已知的种子点经过反复迭代的过程找出其它所有的骨架点。

图 3.4 给出了二值化图像及其骨架的例子。



图 3.4: 二值图像的骨架

§3.3.5 图像二值表示的缺点

基于二值的图像表示是目前使用最广泛的表象。但是，它存在以下几个缺点。

首先，基于二值的图像表示缺乏人类视觉的支持。二值化视觉存在下面的假设，对物体区域的感知是先于对轮廓的感知的，这和生物视觉是矛盾的。

第二，二值化算法困难。尽管二值化算法非常多，但是对很多问题实际而言，可用的算法很难找到。因此，常常需要使用昂贵的光源。

§3.4 图像表象理论的数学基础

本文提出了图像表象理论的整体体系结构，显然这是一个很大的课题，本文只能尝试性的分析了问题几个方面的：形状表象的基本数学理论、形状表象的提取算法、纹理表象的基本数学理论、纹理表象的提取算法和描述方法、在小波框架下形状表象和纹理表象的提取，最后给出了一个实验平台的框架。所有这些讨论基本上都只局限在其中的数学问题，所以本文后面主要研究图像视觉信息表象的数学基础。

由于缺乏颜色处理的经验，本文没有分析图像的颜色表象的数学基础。因此，本文后面的讨论都是基于灰度图像进行的，在没有明确说明的情况下，所说的图像都是指灰度图像。

另外，本文后面的所有讨论都是参照人的视觉感知能力。图像处理和分析技术可能会用于与视觉没有任何关系的领域；虽然这些领域中也有“图像”，但它们都不是通过成像过程获得的，只是二维信号的图像化表示。在这些领域中所用到的信息并不是本文所说的“视觉信息”，所以不属于本文讨论的范围。

第四章 图像的形状表象

本章从分析成像过程和图像入手，证明零交叉点对表示图像信息的重要性。首先，在合理数学模型的基础上，证明了物体与物体之间、物体与背景之间的边界对应着零交叉点，而物体的边界既刻画了物体的形状，又给出了物体的位置和姿态，也就是说零交叉点是图像中的重要特征。然后，在一系列理论和实验的基础上证明，零交叉点几乎包含了图像中的全部信息。

§4.1 形状表象的基本要求

§2.1.2.1、§2.1.2.2 和 §2.1.3.2 中都指出了形状表象在视觉感知中的重要性。诞生于几千年前的初等几何学，从千变万化的世界中抽象出来没有颜色和纹理的点、线和面，作为认知世界的工具，这足以说明形状的重要性。

利用形状人类可以区分数以亿计的物体，而且这种才能是天生的。这就说明，人类视觉系统对形状信息描述和记忆细致且细节丰富，而且视觉系统对形状特征的提取和描述是非常可靠的而且一般的。而且，形状与物体的位置有密切关系，为了快速准确地获得位置和距离信息，对形状的感知必须有一定的精度。

总结视觉系统在形状感知中的特点，形状表象必须满足下面的一些基本的要求：

1. 一般性，即自然图像中发现的一些结构能够被表示，并且能检测到；
2. 可靠性，即对应着场景中特征的图像特征能够可靠地被表示（较低的负错误率）；同时，处理过程中产生的虚假结构不要被表示（较低的正错误率）；
3. 精确性，即图像中特征的定位要准确；
4. 简洁性，即这种表示能够减少原始图像中存在的无意义的冗余；
5. 明确性，即这种表示能够将原始图像中存在的重要结构明确地表示出来；
6. 完全性，即能够满足大量的高层视觉领域的需要，能够提供所有的视觉信息。

§4.2 零交叉点是重要的图像特征

§4.2.1 图像的复杂性

首先讨论视觉系统中的成像问题 [2]。

§4.2.1.1 分块均匀的物理世界

图像是对光的度量。图像是物体的各个发光点发射出来的光进入我们的眼睛后给我们的刺激从而产生的一种称之为颜色的感觉。颜色是由于存在或者缺失一种光

而在人眼中产生的一种感觉，颜色在视觉平面内的分布及其变化在一定的知识系统下带给观察主体信息。光是一种自然现象，可见光只是一种特殊的电磁波；但是颜色却不仅仅与眼睛接受到的光有关，而且与感觉主体有关。

为了理解颜色，我们首先得理解光的本质。物理学家已经确定无疑地建立了光的“波一粒二象性”理论：光是由光子组成的，运动着的光是由一种波确定的，波在某一点的密度表示在这个点发现光子的概率。运动着的光子碰到视网膜后就会产生电脉冲，当这些电脉冲到达人的大脑后就会让人有颜色的感觉。人眼接收到的不同波长的光带来不同的颜色的感觉。人能感觉到的光的范围是有限的，这个范围各个人都不是完全相同的，大约是 380 纳米 – 780 纳米。牛顿用实验证明了白光是由各种颜色的光组成的并且简单地用棱镜就可以将白光分解为各种成分。

一般来说，除了很少的例外，物体上的某个点发出的光以及人接收到的物体上的点发出的光都不是由同一种波长的光子组成的，它是由各种不同波长的光子组成的。不同波长的光子的数量形成光谱分布函数 $C(\lambda)$

$$C(x, y; \lambda) = l(x, y; \lambda) \cdot r(x, y; \lambda) \quad (4.1)$$

其中， $l(x, y; \lambda)$ 表示点 (x, y) 处波长为 λ 的入射光的光强， $r(x, y; \lambda)$ 表示点 (x, y) 处对波长为 λ 的光的反射系数。颜色就是各种波长的光子按照特定的比例混合得到的光谱分布函数 $C(\lambda)$ ，被人眼接收后在人脑中形成的感觉。

所以，我们说的一种颜色其实就是一个光谱分布函数 $C(\lambda)$ 。而纯色对应的光谱分布函数就是一个单脉冲函数。从这个意义上讲，颜色空间是一个无穷维空间。

场景中的物质可以是丰富的，但是我们总可以假设：组成场景中物体的物质在局部上是均匀的。对我们而言，其反射系数 $r(x, y; \lambda)$ 在局部是相同的，是由物质的物理性质决定的。这种假设广泛存在于各个领域，例如，在研究岩石力学的时候，虽然岩石的组成很复杂，但是至少在局部可以假设是均匀的。于是，理想的图像是分块常数函数。同样的假设出现在图形图像处理中的很多文献中。图像处理中的“Retinex theory”是研究如何补偿图像中的由于光线不均而引起灰度渐变效应。其中的一些算法，例如基于 Poisson 方程的算法是建立在 Land 的关于物质世界的 Mondrian 模型的基础上。这个模型的基本假设就是：自然界中物体的反射率是分片常数 [54]。文献 [55] 中分析了各向异性扩散微分方程在图像处理中应用和性能，在性能分析中，就是以这种理想的阶梯形边缘为理想模型进行分析的。

在这种理想情况下，忽略光线的影响， $C(x, y; \lambda)$ 是分块常数函数。在 $C(x, y; \lambda)$ 上对应着物体边界位置的是阶跃函数。

§4.2.1.2 光的影响

光是视觉的媒介，图像是对光的度量。光将物体的物理性质 $r(x, y; \lambda)$ 传递给眼

睛，大脑就是通过解读光中承载的 $r(x, y; \lambda)$ 而感觉世界的。理想的光照条件是均匀光照，这时光线的影响最小。但是，场景中所有的物体上的每一个点都会收到的别的点的光（要么来自于某个光源直射、要么来自于别的物体反射的光）；然后它将这光按照特定的方向反射出去，见图 4.1，这个过程中满足的是一些物理定律。

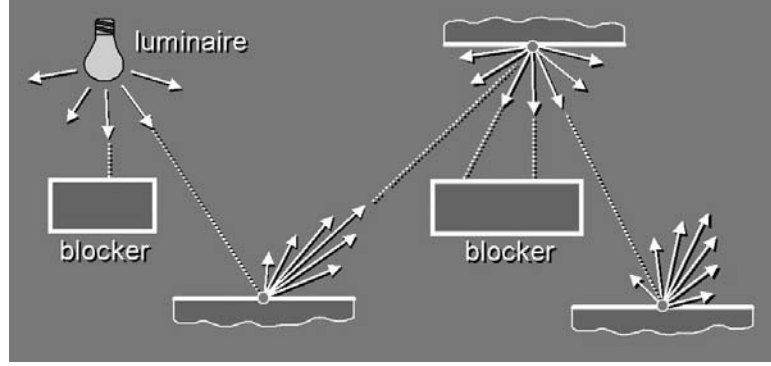


图 4.1: 表面的反射与光源直射。

这个过程用数学可以用式 4.2 描述

$$f = s + KGf \quad (4.2)$$

其中， M 表示场景中的所有物体的表面，这是一个二维流形。 f 是表面 M 上的亮度， s 为表面 M 上的可能的光源， G 是全局的算子，表示表面 M 的几何关系，即那些点可以接收到哪些点发射来的光等， K 表示局部的反射率系数。给定场景（即决定了表面 M 的几何 G 和物理性质 K ）和光源 s 之后，表面 M 上每一个点的光强实际上是由上述流形 M 上的方程决定的。这是一个十分复杂的方程，一般使用 Monte Carlo 方法近似求解，它在图形学的图像合成中有十分重要的价值。

这个公式说明，每个点的光的来源是十分复杂的。即使光源是可以控制的，但是由于其它物体的遮挡反光等等，还是会出现十分复杂的情况。下面是两种最常见的由于光（而不是物体本身的反射系数）引起的图像灰度值的变化，参见图 4.2。

§4.2.1.3 卷积效应

成像过程中具模糊和平滑的作用。Elder [56] 分析了三种产生成像模糊情形，如图 4.3 所示。

第一种是聚焦引起的模糊，图 4.3 中左图所示， R 为镜头中透镜的半径， i 是透镜和成像平面的距离， δ 是成像平面和传感器的距离， r 是模糊的大小。显然 $r = R \frac{\delta}{i}$ ，于是

$$I(x, y) = g(x, y) * u(-x) \quad (4.3)$$



图 4.2: 光线引起的灰度变化。左图为光的缓变引起的灰度变化; 右图为遮挡造成的灰度变化

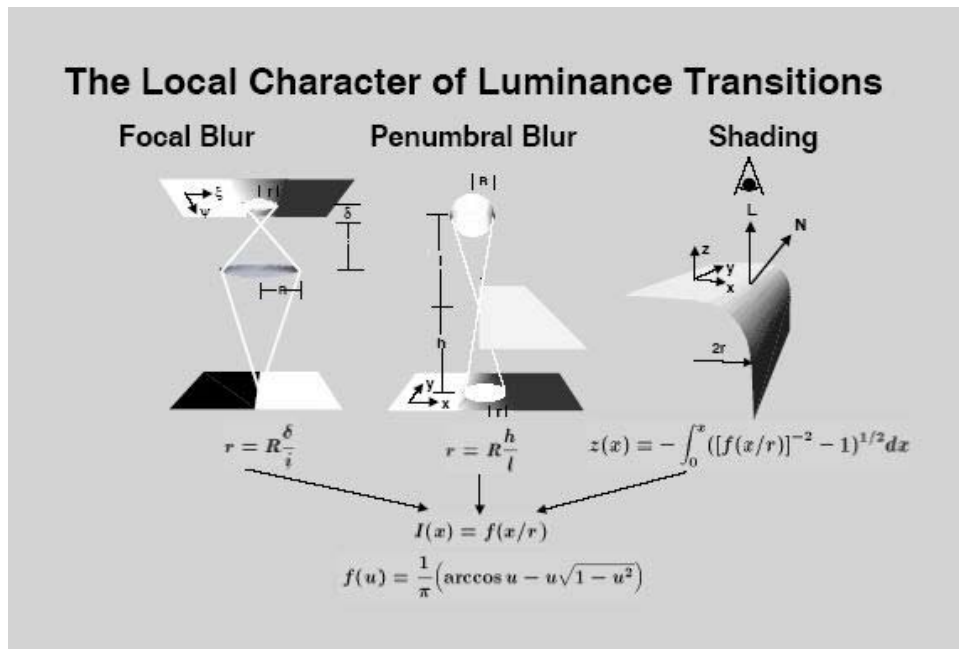


图 4.3: 边缘模糊的原因, 源自 [56]

$$g(x, y) = \begin{cases} \frac{1}{\pi r^2} & , \quad x^2 + y^2 < r^2 \\ 0 & , \quad \text{其它} \end{cases} \quad (4.4)$$

记 $\tilde{u}(x) = u(-x)$, 计算得

$$I(x, y) = g(x, y) * \tilde{u}(x) \quad (4.5)$$

$$= \int_{-\infty}^{\infty} g(x', y) \tilde{u}(x - x') dx' \quad (4.6)$$

$$= \int_{-\infty}^{\infty} g(x - x', y) \tilde{u}(x') dx' \quad (4.7)$$

$$= \int_{-\infty}^{\infty} g(x', y) u(x' - x) dx' \quad (4.8)$$

$$= \int_{-\infty}^{\infty} g(x', y) u(x' - x) d(x' - x) \quad (4.9)$$

$$= \int_{-\infty}^{\infty} g(t + x, y) u(t) d(t) \quad (4.10)$$

$$= \int_0^{\infty} g(t + x, y) d(t) \quad (4.11)$$

$$(4.12)$$

即,

$$I(x) = f(x/r) \quad (4.13)$$

$$f(u) = \frac{1}{\pi} (\arccos u - u \sqrt{1 - u^2}) \quad (4.14)$$

第二种是浓淡模糊, 图 4.3 中中间一幅图像所示。考虑最简单的情况: 球状光源直接投射到遮挡物的边沿上。并且与地平面坐标轴平行。对于点光源, 遮挡形成一个阶梯型边缘。假设不是点光源, 则在地平面上每一点, 照明决定于从这个点看光源时看到的光源点的比率。同聚焦过程一样, 这个过程可以用卷积表示为

$$I(x, y) = g(x, y) * u(-x) \quad (4.15)$$

$$g(x, y) = \begin{cases} \frac{1}{\pi r^2} & , \quad x^2 + y^2 < r^2 \\ 0 & , \quad \text{其它} \end{cases} \quad (4.16)$$

并且,

$$r = R \frac{l}{h} \quad (4.17)$$

其中, R 是点光源的半径, h 是遮挡的高度, l 是遮挡物与光源的距离。于是, 仍然有

$$I(x) = f(x/r) \quad (4.18)$$

$$f(u) = \frac{1}{\pi}(\arccos u - u\sqrt{1-u^2}) \quad (4.19)$$

光源的几何形状不同, 在图像中形成的明暗过渡也就不同。球状光源是最简单的, 并且常常可以作为光源的表的模型, 在这种情况下, 上述两种模糊具有相同的数学结构, 所以是无法区分的。

第三种是物体的平滑边沿引起的模糊, 图 4.3 中右图像所示。简单的, 假设物体是由两个相互垂直的平面平滑连接得到的。变沿平行于 y 轴并且保持不变。假设 lambertian 曲面和光源是从上而下的, 于是曲面的亮度由下式给出

$$I(x) = n(x) \cdot l \quad (4.20)$$

其中, $n(x)$ 是曲面的法向, 而 $l = (0, 0, 1)$ 是光源方向上的单位向量, 用 $z = h(x)$ 表示曲面, 则

$$n(x) = \frac{(-h'(x), 0, 1)}{\sqrt{1 + (h'(x))^2}} \quad (4.21)$$

所以, 有

$$I(x) = n(x) \cdot l = \frac{1}{\sqrt{1 + (h'(x))^2}} \quad (4.22)$$

即, 物体的几何性状决定了物体上各点的亮度。

另一方面, 由于

$$h'(x) = \left(\frac{1}{(I(x))^2} - 1 \right)^{\frac{1}{2}} \quad (4.23)$$

$$z(x) = h(x) = \int_0^x \left(\frac{1}{(I(x))^2} - 1 \right)^{\frac{1}{2}} dx \quad (4.24)$$

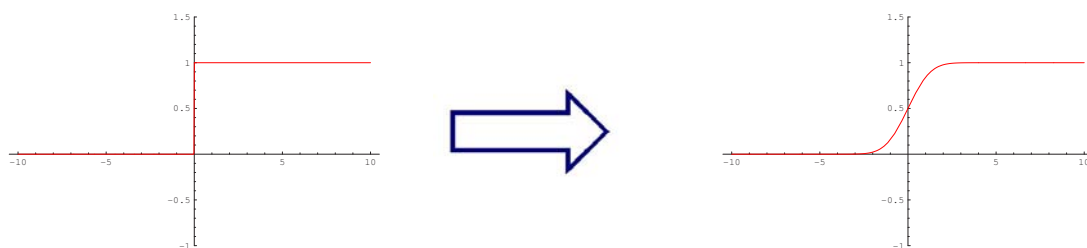
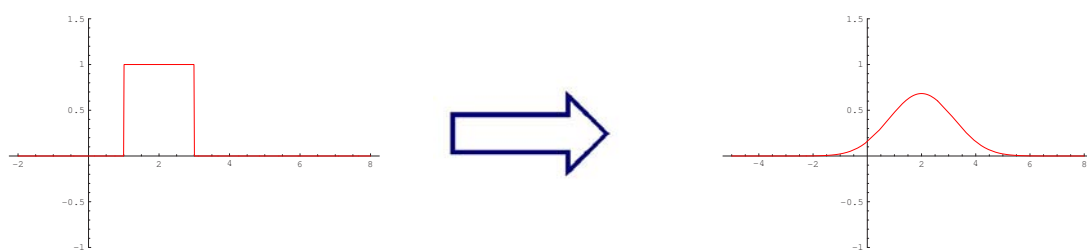
所以, 我们总可以选择合适的 $h(x)$, 使得 $I(x)$ 具有以下形式

$$I(x) = f(x/r) \quad (4.25)$$

$$f(u) = \frac{1}{\pi}(\arccos u - u\sqrt{1-u^2}) \quad (4.26)$$

也就是说, 我们可以构造出分别是由聚焦、光源和物体的边沿引起的三种模糊但具有同样的数学结构, 所以一般的说, 在图像的局部它们是不可区分的。

Elder 给出的模型是理想的, 实际上成像过程中有各种各样的因素。一般的, 将成像过程看作是一个卷积过程, 而卷积的滤波器 (点扩散函数) 是成像过程中这些

图 4.4: 卷积效应: 尺度为 $\sigma = 1.0$ 的高斯卷积滤波的结果图 4.5: 卷积效应: 尺度为 $\sigma = 1.0$ 的高斯卷积滤波的结果

因素的合成。成像过程的点扩散函数可以看作高斯函数，大量的因素的结果往往与高斯函数有关，这是中心极限定理的结论。为了简单，我们讨论一维情况，也就是只考虑垂直于边缘线的截线。由高斯函数的各向同性性质，下面的结论很容易推广到二维。

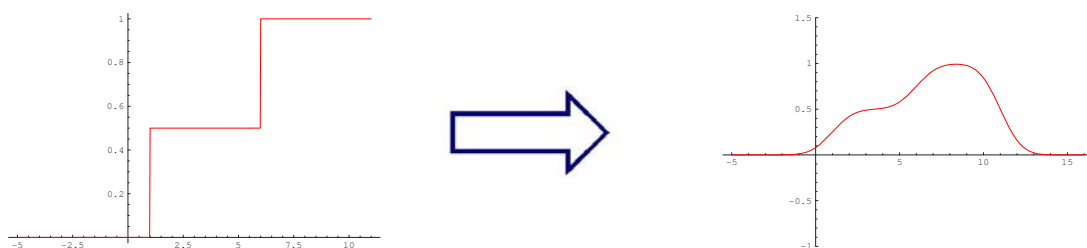
在不考虑光线影响的条件下，分块常数信号经过卷积后不再是分块常数，而变成了一个连续函数，阶梯型边缘消失了。图 4.4 中，一个阶梯形边缘变为平滑的斜坡型边缘；图 4.5 中两个靠得很近的阶梯形边缘变为一个屋脊型边缘；图 4.6 中三个靠得很近的阶梯形边缘变为更加复杂的情形。

可以看出，卷积效应使得图像数据变得光滑。根据卷积与导数的运算规律

$$(I(x) * g(x))' = I(x) * g'(x) \quad (4.27)$$

容易得到这个结论。

因此，阶梯型边缘是一种理想情况，实际图像很少出现，成像过程将边缘模糊

图 4.6: 卷积效应: 尺度为 $\sigma = 1.0$ 的高斯卷积滤波的结果

(blurring)。

§4.2.1.4 成像噪声

除料上述几个因素以外，视觉系统（人类视觉系统，摄像机）在所获得的图像中，还会产生很多噪声。

§4.2.1.5 小结

综上，相机获得的图像要受到成像过程中的噪声、光线不均、物体之间的遮挡、卷积效应等的影响。更重要得是，它们的效果会综合起来形成十分复杂和随机的结果。使得物体之间边界的形态变的十分复杂。

如图 4.7 所示，图 4.7(b) 绘制的是图 4.7(a) 中选定的边界位置处垂直于边缘方向上一条线上的灰度值。从图 4.7(b) 看，图像灰度变化是十分复杂的。

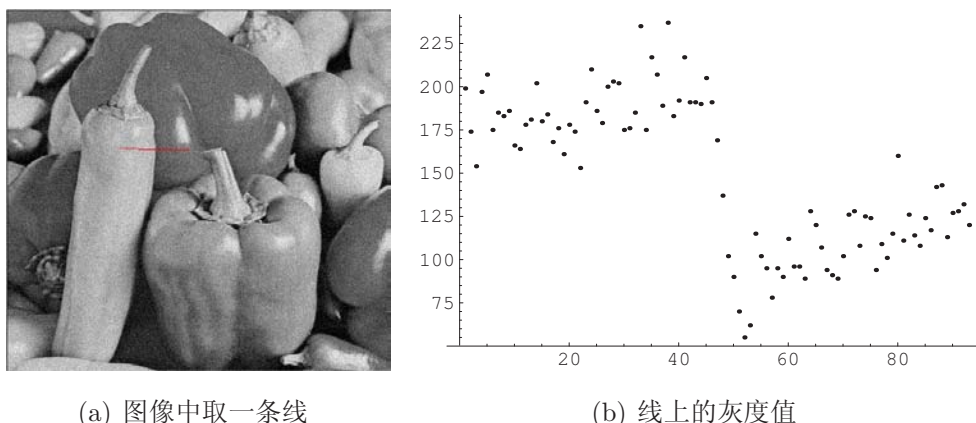


图 4.7: 图像中边界的复杂性

§4.2.2 零交叉点是图像中重要的特征

假设成像系统的可以用高斯卷积来模拟。则给定一个理想的阶梯型边缘 $S(x)$,

$$S(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (4.28)$$

如图 4.8 中左上图所示，实际的边缘位置在 $x = 0$ 点。成像后得到 $\tilde{S}(x)$ ，如图 4.8 中右上图所示，

$$\tilde{S}(x) = \frac{1}{2} \left(1 + \text{Erf} \left(\frac{x}{\sqrt{2}} \right) \right) \quad (4.29)$$

将 $\tilde{S}(x)$ 求微分，我们发现： $x = 0$ 是对应着 $\tilde{S}(x)$ 一阶导函数的极值和二阶导函数的零交叉点，即 $x = 0$ 位置函数值为零，并且在零点左右是一正一负，也称过零点。

也就是说，不考虑噪声和光线的影响，只考虑理想的成像效应，图像中物体的边界对应着二阶零交叉点。

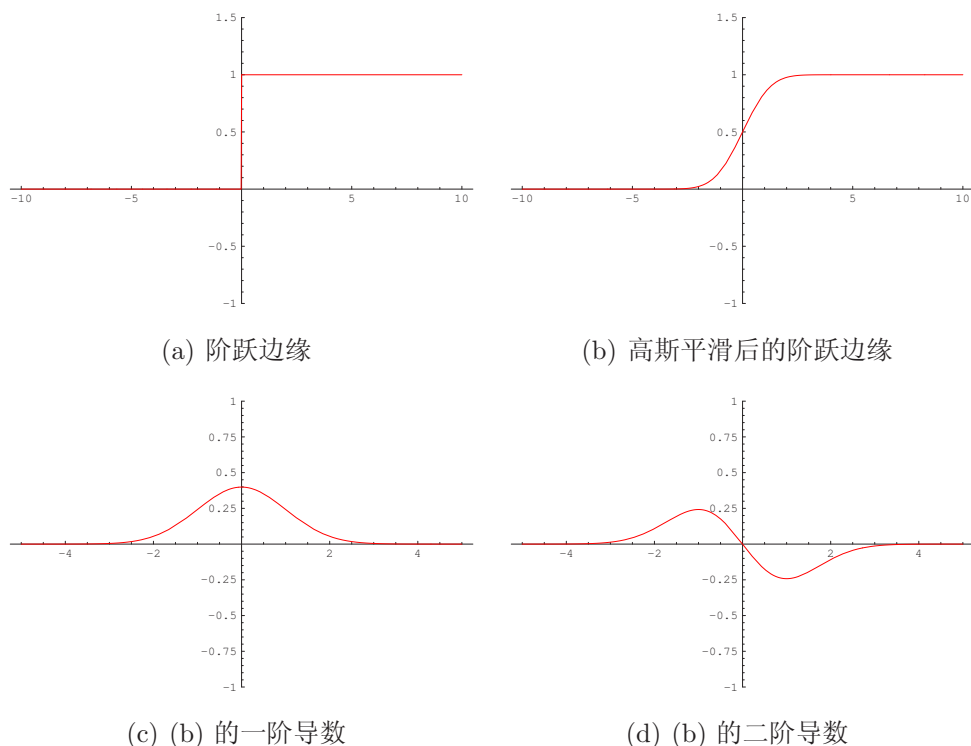


图 4.8: 零交叉点和导数极值点

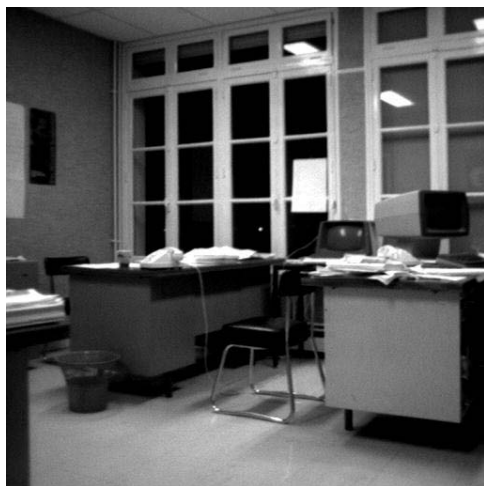
§4.2.3 小结

理想情况下，图像应该是分块常数函数。但是由于光线不均、卷积效应和噪声的复杂作用下，实际上所获得的图像十分复杂。如果只考虑卷积效应，物体之间或者物体与背景之间的边界对应着边界方向上的零交叉点。物体的边界既刻画了物体的形状，而且确定了物体的位置，零交叉点是图像中的重要特征。

这里，我们没有考虑光线不均和噪声的影响。实际上，人类的视觉系统似乎能自动感受光线不均并在图像表象过程中对原始图像进行了补正，而且人的视觉系统能自动抑制噪声。

§4.3 零交叉点的完备性

对于阶跃函数在点扩散函数为高斯函数的情况下，阶跃点的位置对应着阶跃函数被模糊后的二阶零交叉点。由于图像是二维的，所以为了利用二阶零交叉点来检测图像中的边缘，必须确定在那个方向求二阶偏导数；然后找到零交叉点。



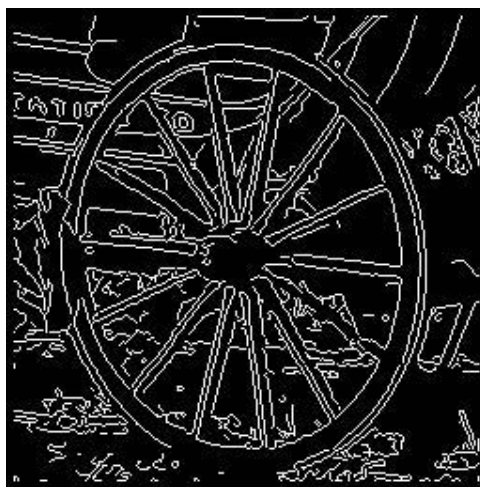
(a) 灰度图像



(b) (a) 的零交叉点图像



(c) 灰度图像



(d) (c) 的零交叉点图像

图 4.9: 边缘对图像信息的刻画

现在得到问题是，图像的零交叉点能多大程度上反映图像中的信息？这是零交叉点的在图像信息描述的完备性问题。如果这个表象不是完备的，那么图像中的信息就会丢失。图 4.9 中给出了一些图像以及对应的零交叉点图像。不难发现，仅仅使用位置信息的边缘图像中也能给出图像中的很多信息。

但是，只有这种例子显然是不够的。我们需要从数学上分析零交叉点的完备性问题：在什么程度上图像可以由其零交叉点重构，这是本节的中心问题。下面这些结论是零交叉点理论发展中的重要结论，但是它们的证明都要使用深刻的数学知

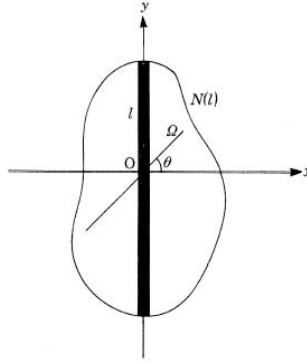


图 4.10: 二阶方向微分和拉普拉斯变换, 选自 [57]

识。

§4.3.1 二阶方向微分和图像的拉普拉斯变换

注意到导致图像中灰度变化的物理原因：例如照明导致的变化或者反射系数导致的变化，表面朝向的变化以及图像中各点与观察点的距离等等在空间上都是连续变化的，并且具有一定的定向。我们选择的求二阶偏导数的方向必须与这些灰度变化的定向一致。具体的说，就是求二阶偏导数的方向应与局部零交叉点的定向垂直。下面给出一个定理 [57]，

定理 4.1. 如图 4.10 所示，令 l 表示一条 y 轴上的线段（不包括两端），并且原点位于 l 上； $N(l)$ 表示 l 的邻域。假设 $f(x, y)$ 是二阶可微的，并且在 l 上满足 $\partial^2 f / \partial x^2 = 0$ ，那么如果在 $N(l)$ 上 $\partial f / \partial y$ 是常数，则垂直于线 l 的方向上的二阶导数的斜率，即 $\partial^2 f / \partial x^2$ 的斜率，大于沿着其它任意方向上的二阶偏导数的斜率。

证明：如图 4.10 所示，假设 $\Omega = (r \cos(\theta) + r \sin(\theta))$ 为给定方向的一条线段，并且 r 足够小，使得 Ω 位于 $N(l)$ 中，则沿着 Ω 的二阶方向导数 $\frac{\partial^2 f}{\partial \Omega^2}$ 可表示为

$$\frac{\partial^2 f}{\partial \Omega^2} = f_{xx} \cos^2(\theta) + f_{xy} 2 \sin(\theta) \cos(\theta) + f_{yy} \sin^2(\theta) \quad (4.30)$$

$$= f_{xx} \cos^2(\theta) + 0 + 0 \quad (4.31)$$

$$= f_{xx} \cos^2(\theta) \quad (4.32)$$

显然，定理成立。

这个定理告诉我们，只要在零交叉点附近平行于零交叉点方向图像的灰度是线性变化的，即方向导数为常数，则二阶零交叉点上二阶导数的斜率最大的方向恰恰就是垂直于零交叉点的方向。这里的这个条件：在边缘位置附近平行于边缘（零交叉点）方向图像的灰度是线性变化的，称为线性变化条件。对于一般的图像，线性变化条件都是近似满足的。

下面再给出一个定理 [57],

定理 4.2. 令 $f(x, y)$ 是平面上二次连续可微函数, l 表示一条 y 轴上的开线段 (不包括两端)。那么当且仅当 $f(0, y)$ 在 l 上是常数或者线性变化时, $\Delta f(x, y) = 0, (x, y) \in l$ 和 $\partial^2 f(x, y)/\partial x^2 = 0, (x, y) \in l$ 是等价的。

证明: 因为 $f(0, y)$ 在 l 上是常数或者线性变化的, 所以 $\partial^2 f(x, y)/\partial y^2 = 0, (x, y) \in l$, 因此

$$\Delta f(x, y) = \partial^2 f(x, y)/\partial x^2 + \partial^2 f(x, y)/\partial y^2 = 0 \Leftrightarrow \partial^2 f(x, y)/\partial x^2 = 0 \quad (4.33)$$

反之, 因为

$$\Delta f(x, y) = \partial^2 f(x, y)/\partial x^2 + \partial^2 f(x, y)/\partial y^2 = \partial^2 f(x, y)/\partial x^2 = 0 \quad (4.34)$$

所以

$$\partial^2 f(x, y)/\partial y^2 = 0, (x, y) \in l \quad (4.35)$$

即, $\partial^2 f(x, y)/\partial y^2 = 0, (x, y) \in l$, 也就是说 $f(0, y)$ 在 l 上是常数或者线性变化的。显然, 定理成立。

所以, 图像中阶跃边缘对应着图像的拉普拉斯变换的零交叉点。

§4.3.2 人眼的辨别力和图像的拉普拉斯变换

在本节以下的讨论中, 都是企图建立某类函数与其零交叉点 (一维时的二阶导数零交叉点或者二维拉普拉斯变换零交叉点) 之间的一一对应关系。如果这种一一对应关系能够建立, 那么零交叉点就可以完全特征化这一类函数。如果把图像的函数看作是图像的拉普拉斯变换 (一维 “图像” 时的二阶微分), 就说明零交叉点完全决定图像的拉普拉斯变换。

$$\text{图像的拉普拉斯变换} \Leftrightarrow \text{零交叉点} \quad (4.36)$$

假设 $f(x, y)$ 和 $g(x, y)$ 是两个函数 $f(x, y) \neq g(x, y)$ 。显然, 有可能

$$\Delta f(x, y) = \Delta g(x, y) \quad (4.37)$$

例如

$$f(x, y) = g(x, y) + x + y + 1 \quad (4.38)$$

我们把拉普拉斯变换相等的两个函数称为是等价的, 即, 如果

$$\Delta f(x, y) = \Delta g(x, y) \quad (4.39)$$

那么 $f(x, y)$ 和 $g(x, y)$ 是等价的, 记做 $f(x, y) \simeq g(x, y)$ 。所有和 $f(x, y)$ 等价的函数 $\{g(x, y) : g(x, y) \simeq f(x, y)\}$ 称为 $f(x, y)$ 的等价类, 记做 $\bar{f}(x, y)$ 。 $f(x, y)$ 的等价类 $\bar{f}(x, y)$ 和拉普拉斯变换的零集合 $\mathcal{N}(\Delta)$ 有密切的关系, 其中拉普拉斯变换的零集合 $\mathcal{N}(\Delta)$ 定义为

$$\mathcal{N}(\Delta) = \{g(x, y) : \Delta g(x, y) = 0\} \quad (4.40)$$

由于拉普拉斯变换是线性的, 所以 $f(x, y)$ 的等价类 $\bar{f}(x, y) = f + \mathcal{N}(\Delta)$

$$f + \mathcal{N}(\Delta) = \{f(x, y) + g(x, y) : g(x, y) \in \mathcal{N}(\Delta)\} \quad (4.41)$$

于是, 这些定理试图说明: (某类) 图像的零交叉点决定了图像关于拉普拉斯变换的等价类, 图像函数的二阶零交叉点和图像自身关于拉普拉斯变换的等价类是一一对应的。

可惜的是, 零交叉点无法区分这个图像类中的两个不同的函数 (图像)。不过, 这并不重要, 经验表明, 人的视觉对于线性梯度变化是不敏感的 [58]。也就是说, 对人类视觉而言, 等价类中的两个函数也是不可区分的。

§4.3.3 Logan 定理

1977 年, IBM 的 Logan [59] 证明了一维信号, 严格的说是带通的一类一维信号的零交叉点构成了原信号的一个完全的表示。

Logan 所说的那类带通函数, 是指指数型整函数 [72] 的子集合 $Z_{\alpha, \beta}$, 对这类集合中的函数, 如果把它看作是实函数, 则

定理 4.3.

$$\text{sgn}\{h_1(t)\} = \text{sgn}\{h_2(t)\}, h_1(t), h_2(t) \in Z_{\alpha, \beta} \quad (4.42)$$

总是意味着

$$h_1(t) = Ah_2(t) \quad (4.43)$$

其中, A 是一个常数, $\text{sgn}(x)$ 为符号函数。

简单的说, $h(t) \in Z_{\alpha, \beta}$ 的零交叉点完全决定 $h(t)$ 。从图像处理的观点而言, 如果把 $h(t)$ 看做是“一维”图像的二阶微分, Logan 的结论说明: 如果图像 $f(x)$ 比较特殊, 其二阶导数 $h(t)$ 是一类特殊的函数 $h(t) \in Z_{\alpha, \beta}$, 那么, $h(t)$ 的二阶零交叉点就可以确定 $h(x)$ 。

正是根据这个结论, 1980 年, Marr 和 Hildreth [57] 推测: 图像可以尺度空间中的零交叉点完全表示。

“由 Logan 定理引出的概念不仅得到了一种令人满意的图像分析方案, 而且对视觉心理学物理学和生理学也具有令人神往的含义。因为看来它们阐明了视觉通路第一级的一些基本性质。特别是这些概念说明了为什么最初对图像要用中心 - 周边型

感受野做滤波处理。它们为‘边缘’检测器提取图像全部基元符号这一见解提供了理论基础,并指明只有用几个滤波器独立的带通滤波器即中心-周边型感受野先行滤波处理才能做到这一点。这些概念还使灵长类动物视觉心理物理学和生理学研究中关于边缘检测器和空间频率通道之间的长期争论得以解决。实际上视觉的第一个阶段在很大程度上是由‘边缘’检测器(其实是零交叉点检测器)而不是由傅里叶分析完成的,但是零交叉检测器要提取有意义的信息,它们就必须对若干独立通道的输出进行运算,其中每一个通道只对某一特殊的空间频率段具有选择性。”

“许多初级视觉处理的心理物理学和生理学结果用这个框架来解释或许是容易的。例如,二阶导数滤波运算结果是由视网膜神经节细胞和外侧膝状体核完成的,视皮层中有一类简单细胞可能是检测有向零交叉线段的,这样一些设想决不是没有道理的。以此而论,理解究竟如何完成上述运算的细节并不是那么重要。Marr 提出的另一个见解是:简单细胞可能从纹状皮层 4C 层小细胞今昔的网格结构中辨认出零交叉的外形轮廓,以各种不同的分辨率重建经滤波处理的图像可能就在这里完成,从而把零交叉点的空间位置精确地确定下来。”

“零交叉计算理论中还有一些空白需要填补。例如,由于零交叉点并不表示图像的完全信息,因此搞清楚其它有用基元的特性是很重要的。在视觉的其他层次上,当然希望得到有利于或者不利于零交叉的实验证据。我深感零交叉试验的成败对 Marr 理论的进一步发展具有特别重要的意义。” [8]

§4.3.4 Carlsson 的数值试验

人的视觉的重要特点就是它对边缘是敏感的 [67]。传统的图像编码算法一般都是利用图像数据的空间冗余,这类算法容易毁坏边缘和轮廓,尤其是当压缩率比较大的时候。

1967 年, Graham [68] 发现如果将图像的分成高频和低频两部分,那么高频部分就对应着图像中的剧烈变换,也就是边缘。而低频部分部分对应着边缘之间的平滑区域。通过轮廓的外在表示,高频部分可以更有效地编码。1983 年, Kocher 等 [69, 70] 通过区域增长获得灰度均匀的像素区域,于是区域之间的边界就是边缘,而区域中的灰度值用低阶多项式逼近。这样就产生了下面一类算法基于边缘的图像压缩编码算法:

1. 提取边缘,对边缘信息进行编码;
2. 对边缘之间的纹理进行编码。

1984 年, Carlsson [60, 61] 提出了一个基于边缘的图像压缩算法证明了只使用边缘的结构和边缘处图像的灰度值就可以重建出视觉上与原始图像没有很大区别的图像。对于纹理图像,借助于图像的金字塔表示进行编码,该算法可以对很多图像进

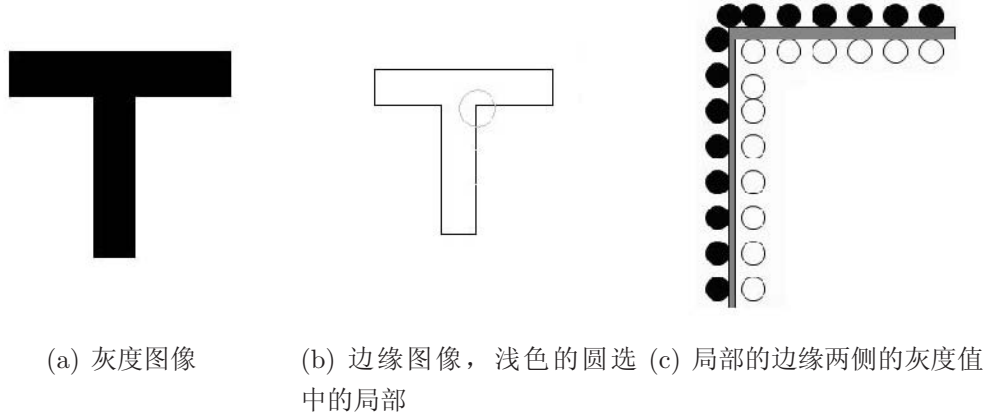


图 4.11: Carlsson: 图像的压缩表示

行大幅的压缩（60 — 70 倍），重构效果在视觉上是可接受的。

Carlsson 的重构算法的核心思想是：图像可以根据基于边缘的填充 (filling-in) 过程恢复。

如图 4.11 所示，检测图像的边缘，然后记录边缘两侧的灰度值。利用合适的数据结构将这些数据进行存储，就可以完成图像压缩。

为了恢复图像，Carlsson 利用轮廓两侧的灰度值将整个图像曲面平滑填充。显然满足要求的平滑连续曲面有非常多，我们希望得到的曲面变化率最小，也就是导函数绝对值的积分要最小化 (这个过程称为正则化过程)，

$$\int \int \left(\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \right) dx dy \quad (4.44)$$

这是一个泛函极值问题。将上式离散化，得到

$$Q = \sum_m \sum_n (I_{m,n} - I_{m,n-1})^2 + (I_{m,n} - I_{m-1,n})^2 \quad (4.45)$$

这是一个最优化问题。为了获得最优解，令

$$\frac{\partial Q}{\partial I_{m,n}} = 0 \quad (4.46)$$

即，

$$\begin{aligned} 0 &= 2(I_{m,n} - I_{m,n-1}) + 2(I_{m,n} - I_{m-1,n}) - 2(I_{m,n+1} - I_{m,n}) - 2(I_{m+1,n} - I_{m,n}) \\ &= I_{m-1,n} + I_{m,n-1} + I_{m+1,n} + I_{m,n+1} - 4I_{m,n} \\ &= \Delta I_{m,n} \end{aligned} \quad (4.47)$$

于是, 这个问题的解满足一个偏微分方程 $\Delta I_{m,n} = 0$, 而这个方程的边界条件恰好就是轮廓点两侧的灰度值。利用偏微分方程中的结论, 可以有下面的迭代公式

$$I_{m,n}^{i+1} = I_{m,n}^i + \frac{1}{4}\omega\Delta I_{m,n}^i \quad (4.48)$$

$$\Delta I_{m,n}^i = I_{m-1,n}^{i-1} + I_{m,n-1}^{i-1} + I_{m,n+1}^{i-1} + I_{m+1,n}^{i-1} - 4I_{m,n}^{i-1} \quad (4.49)$$

其中, ω 是松弛参数。对非轮廓点上的像素, 按照式 4.48 和式 4.49 反复迭代, 直到所有的像素值都稳定为止。

通过数值解法求解这个方程就可以得到光滑连接的图像。从重构的结果看来, 这种稀疏的表示可以近似恢复原始图像, 也就是说它包含了图像中的大部分信息。

Carlsson 的算法可以看作是对零交叉点完全性的数值实验, 从它的结果来看, 算法支持完全性的。虽然有了这样的数值结果, 但是还是缺乏严密的数学理论支持。

§4.3.5 函数的指纹定理

1985 年, Yuill 和 Poggio [62] 证明了一维和二维有限多项式信号在尺度空间上零交叉点的理论上的完全性。

有限多项式的集合记做 $\mathcal{P}$, $F(x) \in \mathcal{P}$ 。对 $F(x)$ 通过高斯函数 $G(x, \sigma)$ 可以得到 $E(x, \sigma) = F(x) * G(x, \sigma)$, 这个过程可以看作是多尺度滤波。令 $E(x, \sigma) = 0$, 得到了一个曲线 $x(\sigma)$, 称为零交叉点曲线, 如图 4.12。

Yuill 和 Poggio 证明, 曲线 $x(\sigma)$ 就决定了 $E(x)$, 最多差一个常数。即, 给定一个零交叉点曲线, 如图 4.12, 就决定了一个函数 F 。所以, Yuill 和 Poggio 把图 4.12 看作是函数 I 的指纹。

定理 4.4. 由 $E(x, \sigma) = 0$ 定义的零交叉点曲线上, 同一尺度上任意两点的导数 (包括零阶导数), 唯一决定了 $\mathcal{P}$ 中的一类函数, 最多差一个常数。

这个结论可以推广到二维多项式。

由于 Weierstrass 定理, 多项式函数具有非常好的逼近性质。理论上说任意函数 (例如图像) 可以用一个有限多项式逼近。即, 任意一个函数 $I(x)$, 记 $f(x) = \frac{d^2 I}{dx^2}$ 为其二阶微分, 总存在 $F(x) \in \mathcal{P}$ 能对 $f(x)$ 充分的近似。

下面给一个具体的例子来说明, 这个例子是文献 [62] 中附录 C 的一个特例。令 $F(x) = 1 - x^2$, 则容易得到

$$E(x, \sigma) = F(x) * G(x, \sigma) = 1 - x^2 - \sigma^2 \quad (4.50)$$

根据定义, 零交叉点曲线由 $x^2 + \sigma^2 = 1, \sigma > 0$ 给出, 这是上半圆周。

按照 Yuill 和 Poggio 的构造性的证明过程, 可以使用这个圆周重构得到函数

$$\hat{F}(x) = \frac{1}{2\sqrt{2\pi}}(x^2 - 1) \quad (4.51)$$

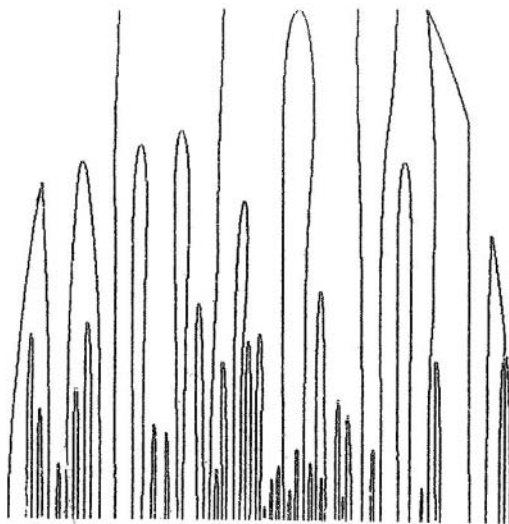


图 4.12: 作为函数指纹的零交叉点

显然, $\hat{F}(x)$ 和 $F(x)$ 差一个常数, $F(x) = -2\sqrt{2\pi}\hat{F}(x)$ 。

§4.3.6 一类有限带宽的二维信号的零交叉点的完全性定理

1987 年, Curtis 等 [58] 证明了, 一类有限带宽的二维信号的零交叉点的完全性, 即信号的零交叉点完全决定信号。

定义 4.1. 不可约函数 令 $V_f = \{(s, w) : f(s, w) = 0\}$ 表示具有两个实 (复) 自变量 s, w 的函数 $f(s, w)$ 的实 (复) 零点集合。如果 $f(x, y)$ 不存在两个整函数 $g(x, y)$ 和 $h(x, y)$, 使得

$$f(s, w) = g(s, w)h(s, w) \quad (4.52)$$

并且

$$V_g \neq \emptyset \quad (4.53)$$

$$V_h \neq \emptyset \quad (4.54)$$

那么函数 $f(x, y)$ 称为不可约函数。

定理 4.5. 如果 $f(x, y)$ 和 $g(x, y)$ 都是二维实值带限函数。如果 $f(s, w)$ 和 $g(s, w)$ 都能分解为有限实不可约函数的积, 并且 $f(s, w)$ 和 $g(s, w)$ 的每个因子都是一重的并且在有限闭区域 $D \subset R^2$ 上取得正值和负值, 且对任意 $(x, y) \in D$ 有 $\text{sgn}\{f(x, y)\} = \text{sgn}\{g(x, y)\}$, 那么 $f(x, y) = cg(x)$, 其中 c 是一个常数。

§4.3.7 Hummel-Moniot 定理

前面这些结论都是支持 Marr 的猜想的。但是, 1989 年, Hummel 和 Moniot [63]

认为前面给出的这些定理在实际中不可能得出稳定的重构结构，而且这些结论很难推广到一般的连续函数。

他们证明了，如果给 N 维信号的多尺度零交叉点补充在零交叉点的梯度数据，那么原始信号理论上就能够被重构。

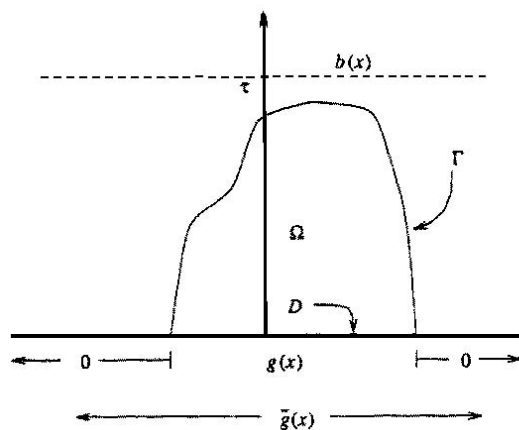


图 4.13: 证明示意图

令

$$v(x, y) = \int_{\mathbb{R}^2} \Delta K(y - x, t) f(y) dy = \int_{\mathbb{R}^2} K(y - x, t) \Delta f(y) dy \quad (4.55)$$

其中，

$$K(x, t) = \frac{1}{(4\pi t)^{n/2}} e^{-\frac{|x|^2}{4t}} \quad (4.56)$$

于是，式 4.55 建立了 $\Delta f(y)$ 的高斯尺度空间 ($2\sigma^2 = 4t$)。

如图 4.13 所示，令 Ω 表示区域 $\{(x, t) : t \geq 0, \nu(x, t) \neq 0\}$ 中一个有界、连通的区域； $D = \{x \in \mathbb{R}^n : (x, t) \in \Omega\}$ ， $\Gamma = \partial\Omega \cap \{t > 0\}$ ，令 τ 表示满足 $\tau > \sup\{t : (x, t) \in \Omega\}$ ， $g(x) = \Delta f(x)$ ，并且

$$\tilde{g}(x) = \begin{cases} g(x), & x \in D \\ 0, & \text{other} \end{cases} \quad (4.57)$$

$b(x)$ 是对 $\tilde{g}(x)$ 的平滑，

$$b(x) = \int_{\mathbb{R}^n} K(x - y, \tau) \tilde{g}(y) dy \quad (4.58)$$

定理 4.6. 假设 $g(x), x \in D$ 包含在尺度空间函数 ν 的有界零交叉点曲面 Γ 的底部。假设 $g(x)$ 属于广义函数中缓增广义函数 (*tempered distribution*) 的 S' 类，则

$b(x)$ 可以用 Γ 和 Γ 上的 $\nabla\nu$ 重构。

$$b(y) = \int_{\Gamma} K(x-y, \tau-t)(\nabla\nu \cdot \mathbf{n}(x, t))d\sigma \quad (4.59)$$

因为从 $\tilde{g}(x)$ 到 $b(x)$ 的“平滑”变换是 1-1 的，所以理论上说 $g(x), x \in D$ 可以被重构。

但是，这种重构是不稳定的。因为，“去平滑 (deblurring)”过程是典型的不稳定过程。幸运的是，因为我们已经知道 $\tilde{g}(x)$ 是紧支撑的，并且“去平滑”的误差主要发生在高频部分，所以对这部分重构误差，视觉上不敏感的。

§4.3.8 零交叉点的局限性

但是，有很多例子说明不同的连续函数的多尺度零交叉点可能相同。这样的例子非常多，Mayer [64] 给出了一大类这样的例子。下面是 Mallat 在文献 [65] 中的一个简单例子。

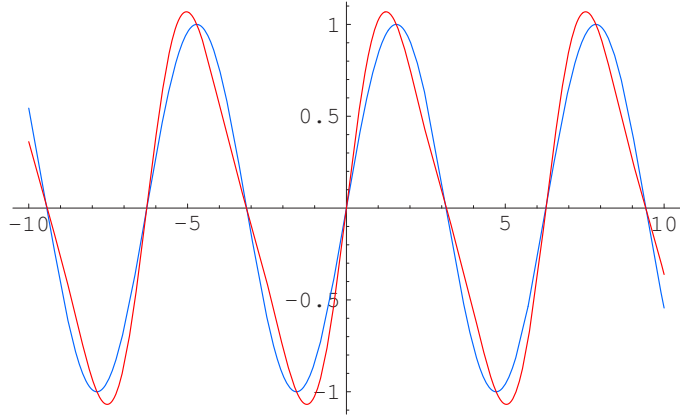


图 4.14: 多尺度零交叉点相同的两个函数

函数 $f_1(x) = \sin(x)$ 和函数 $f_2(x) = \sin(x) + 5\sin(2x)$ (如图 4.14 所示), 它们的高斯尺度空间分别是 $F_1(x, \sigma)$ 和 $F_2(x, \sigma)$,

$$F_1(x, \sigma) = f_1(x) * G(x, \sigma) = e^{-\frac{\sigma^2}{2}} \sin(x) \quad (4.60)$$

$$F_2(x, \sigma) = f_2(x) * G(x, \sigma) = \frac{1}{5}e^{-2\sigma^2} \left(2\cos(x) + 5e^{\frac{3\sigma^2}{2}} \right) \sin(x) \quad (4.61)$$

由于 $e^{-\frac{\sigma^2}{2}}$ 和 $\frac{1}{5}e^{-2\sigma^2} \left(2\cos(x) + 5e^{\frac{3\sigma^2}{2}} \right)$ 都恒大于 0, 所以 $F_1(x, \sigma)$ 和 $F_2(x, \sigma)$ 的零交叉点是完全相同的。

所以，只有零交叉点是不完全的。我们还需要补充一些其它信息。

§4.3.9 Elder 数值试验

2000 年, Elder 等人在长期的研究后 [56,66], 给出了一个利用局部尺度控制算法得到的基于边缘的图像表示和压缩算法, 通过大量的例子说明边缘是“几乎”完全的。

由于我们的目的只是说明边缘在图像信息表达的完全性, 所以本文不关心编码的数据结构和重构的快速实现等, 而主要从理论上考虑理论上(近似)重构的可能性和算法, 并通过大量的例子说明边缘信息确实是近似完全的。Elder [66] 提出的图像编码算法是:

1. 首先计算图像的边缘图像;
2. 计算所有边缘点处的边缘的方向, 精度只需要 $\frac{\pi}{2}$ 就可以, 即大致确定一下方向即可;
3. 计算所有边缘点处位于边缘线两边的灰度值(或者也可以计算亮度和对比度);
4. 计算所有边缘点处的模糊尺度;
5. 将边缘图像和上述计算得到的数据压缩保存。

图像重构算法如下:

1. 利用边缘点的空间结构、边缘方向和边缘两侧的灰度值通求解一个拉普拉斯方程得到一个简单的图像, 这个图像和原始图像已经十分近似;
2. 利用边缘点的空间结构和模糊尺度参数, 通过求解拉普拉斯方程得到原图中各个像素模糊参数的估计;
3. 逐点按照得到的模糊参数模糊简单图像, 就得到最终的重构图像。

图 4.15、图 4.16 和图 4.17 是文献 [66] 中给出的实验结果。

重构结果从视觉上看起来是非常好的。也就是说仅仅使用零交叉点的结构和零交叉点上的部分信息, 可以表示图像中的绝大部分信息(这些信息不只是形状信息, 还包括纹理信息)。

Elder 等使用这个算法进行图像编辑和合成: 首先获取要合成的图像的轮廓(零交叉点), 然后在轮廓域进行合成和编辑; 最后通过求解微分方程重构, 就得到合成后的图像。这个方法启发 Perez [71] 等人使用 Poisson 方程进行图像编辑, 而文献 [73,74] 中的景物提取算法又是受到 Perez 的算法的启发。

§4.3.10 结论: 多尺度零交叉点是合适的形状表象

上一小节最后我们介绍了一种基于多尺度零交叉点的图像表象, 并给出了重构算法, 通过试验验证了算法的有效性。

显然, 根据 Marr 视觉理论, 图像的边缘表示具有一般性、可靠性和精确性。另

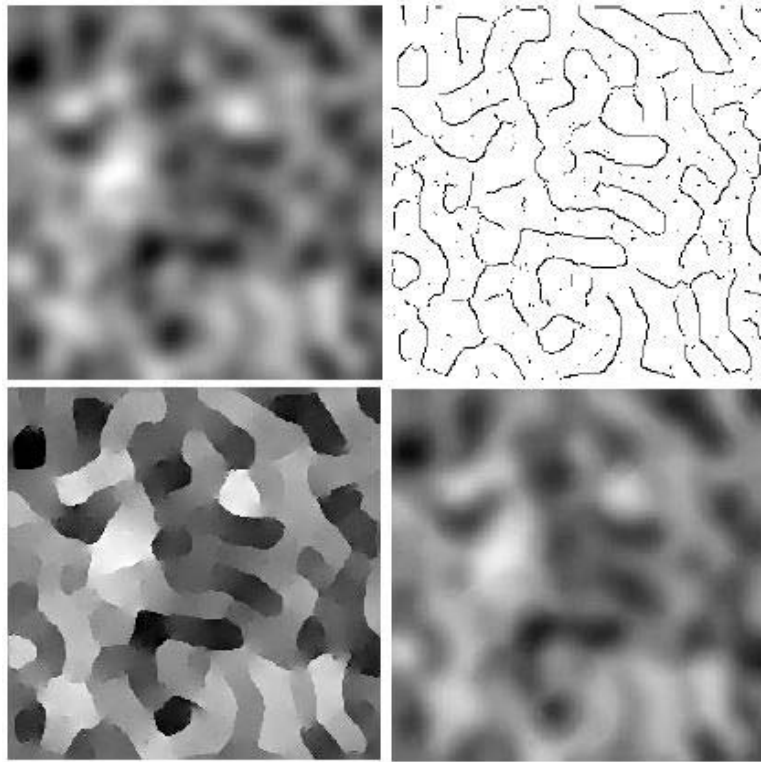


图 4.15: 基于边缘的图像编码及其重构。左上: 原始图像; 右上: 边缘图像 (其它参数无法显示); 左下: 解方程得到的图像; 右下: 重构图像。源自 [66]

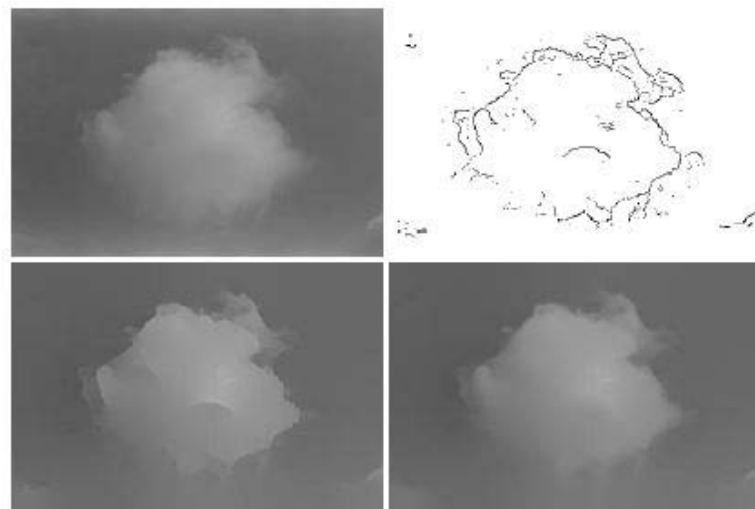


图 4.16: 基于边缘的图像编码及其重构。左上: 原始图像; 右上: 边缘图像 (其它参数无法显示); 左下: 解方程得到的图像; 右下: 重构图像。源自 [66]



图 4.17: 基于边缘的图像编码及其重构。左上: 原始图像; 右上: 边缘图像 (其它参数无法显示); 左下: 解方程得到的图像; 右下: 重构图像。源自 [66]

外, 其简洁性也是不言而喻的。按照本文前面的说法, 边缘对应着可见表面的结构 (物体的边界, 物体与背景的相交线等), 因此这种表示也具有明确性。虽然数学上有例子说明完全的恢复图像是不可能的, 但是通过重构算法视觉上非常近似地恢复原始图像, 所以它也具有视觉上的近似完全性, 也就是说在提供视觉信息方面它是近似完全的。所以这种基于边缘的图像表示能够满足大量的高层视觉领域的需要 (因为它满足上述性质), 能够提供所有的视觉信息 (因为能近似重构)。

所以, 图像中的零交叉点是一般的、可靠的、精确的、简洁的、明确的并且是几乎完全的形状表象。

Elder 提出的表示方法中使用了零交叉点的结构, 边缘点的方向、模糊尺度和两边的灰度值, 从恢复原始图像的角度来说, 它们是几乎完全的。在视觉感知中, 视觉系统并不是为了重构图像, 而是为了获取图像中的“部分”对具体问题有用的信息, 并不是图像中所有信息在具体问题中都能用到。在这里, 灰度值实际上给出了图像的近似颜色表示, 边缘点的方向只是为了在计算过程中决定使用哪一个灰度值。所以, 反应形状的主要是尺度空间中零交叉点。

§4.4 多尺度零交叉点检测算法

形状表象需要考虑多尺度零交叉点的检测, 这个问题归结为边缘检测问题。

§4.4.1 边缘是指具有特定模式的灰度变化的点

图像中的信息是多方面的, 它们都蕴含在灰度变化中。在没有先验知识的条件下, 单一灰度的图像不能给我们任何信息。

可见光成像得到的图像中包含了场景中的信息, 包括场景中的物体、场景中物体之间的关系信息 $R(x, y, \lambda)$ 和场景中的照明信息 $L(x, y, \lambda)$ 。人和动物就是利用这些信息来躲避天敌和寻找猎物的。数据的图形化显示中也包含了被显示的数据中蕴涵的信息。可见光成像得到的图像也包含了镜头和相机的信息: 噪声水平、清晰度、颜色真实感。利用这些信息可以对相机做出评价。图像中还包含了图像的成像模式信息, 利用这些信息可以对图像的来源做出一定的推断 [76]。

由于图像来源以及图像中新产生的物理原因的不同, 灰度变化的形态模式 (以下简称模式) 是多种多样的。下面给出一些图像灰度变化的例子。图 4.18(b) 所示的是由噪声导致的灰度变化模式, 图 4.18(c) 所示的是由光线和纹理导致的灰度变化模式, 而图 4.18(d) 导致的是由不同物体上对光的反射系数不同而导致的灰度变化模式。

灰度变化强度的度量是通过数学模型将灰度变化的程度进行量化。灰度变化是客观的, 但是数学模型是相当主观的, 可以根据个人的喜好设计出各种度量的方法。对某种模式灰度变化的度量得到了这种灰度变化的强度, 利用强度可以将图像

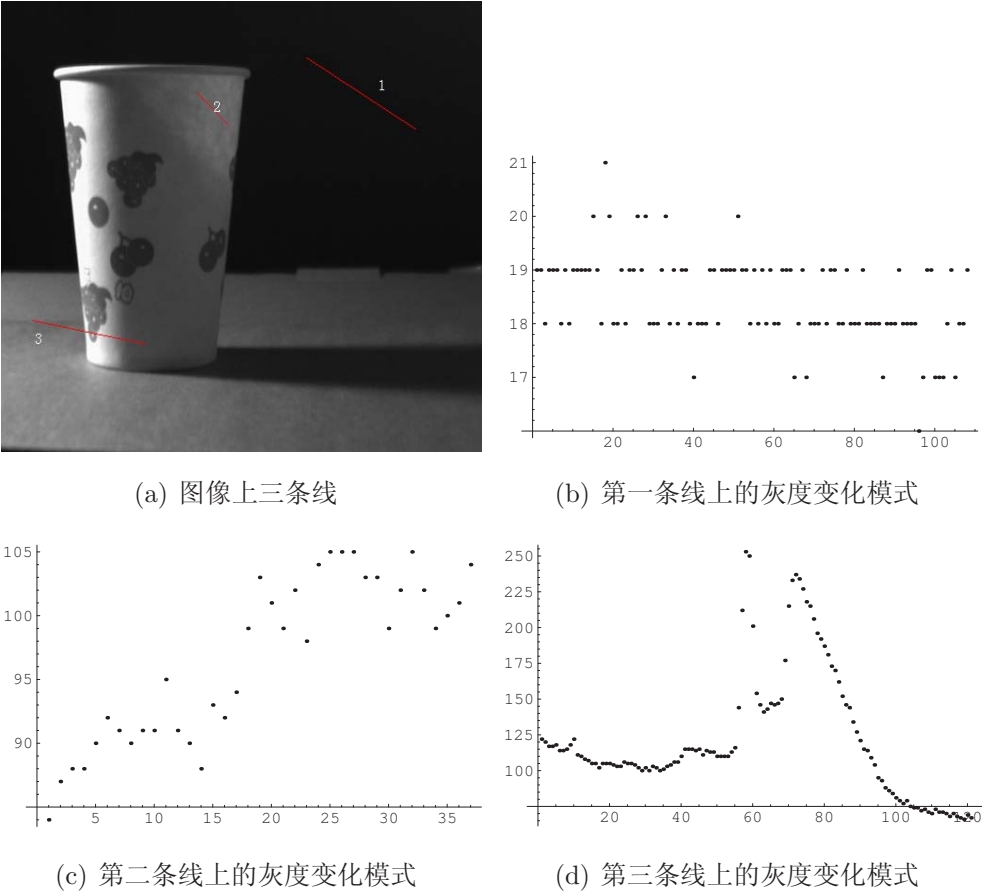


图 4.18: 灰度变化的模式。三条线上灰度变化模式不同。

中这种灰度变化进行排序。一般情况下，我们对强度大的灰度变化有更大的兴趣。

灰度变化有两个内在属性。第一个内在属性是变化发生的方向（角度）。图像是二维的连续曲面（或者二维离散数据），图像中灰度变化可能发生在任何一个方向上，如图 4.19。

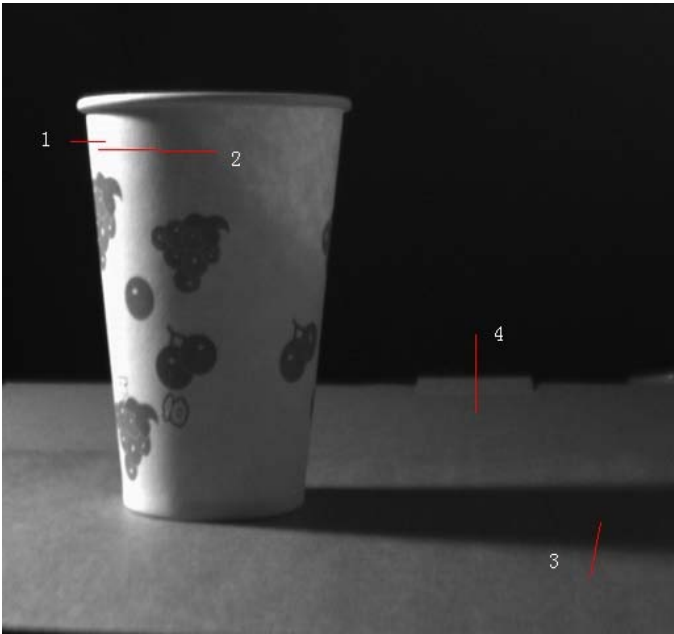


图 4.19: 灰度变化的方向

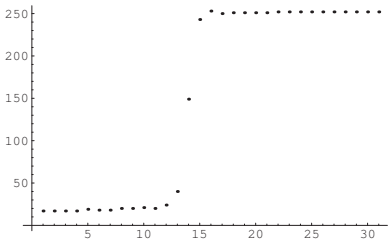
灰度变化的第二个内在属性是变化发生的尺度，即灰度变化的度量与尺度有关。灰度的变化总是拿相邻两个像素的差或者相邻两个小邻域中像素值的（加权）平均值的差来衡量的。尺度即是指，在灰度变化度量中，是拿什么样的邻域大小去衡量灰度变化的。打个比方，“气温随时间变化”这个概念也是有尺度的，相邻两秒的气温变化也是变化，相邻两年的气温变化也是变化，秒或者年就是气温变化的尺度。“气温随空间位置的变化”也是同样的概念，既有厘米尺度的变化，也有公里尺度的变化。图像中的灰度变化根据其产生的物理原因，都有内在的尺度，树要比树上的叶子有更大的尺度，而森林又比树又更大的尺度。内在的尺度通过成像机制转化为灰度变化的尺度。图 4.20 演示的是各种尺度的灰度变化。

灰度变化具有无限多种模式，但是并不是所有的模式都是我们感兴趣的，我们只关心特定的灰度变化模式。

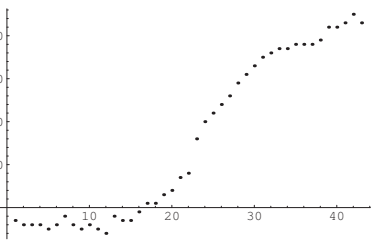
定义4.2. 数字图像的边缘：数字图像的边缘是指数字图像中具有特定模式的灰度变化的点。



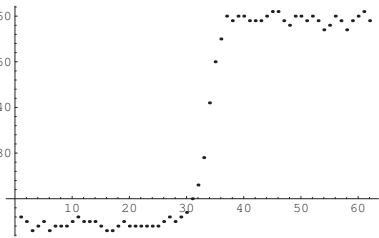
(a) 图像上四条线



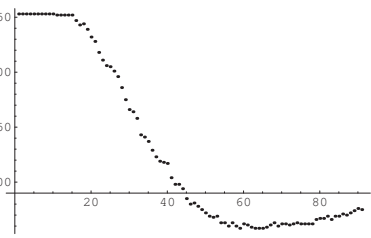
(b) 第一条线上的灰度变化



(c) 第二条线上的灰度变化



(d) 第三条线上的灰度变化



(e) 第四条线上的灰度变化

图 4.20: 灰度变化的尺度

定义4.3. 数字图像的边缘检测：数字图像的边缘检测是指数字图像中具有特定模式的灰度变化的定位、定向和度量。

边缘是图像中特殊的灰度变化模式，“是指图像中周围像素灰度有剧烈变化（阶跃变化或屋顶）等变化的那些像素的集合”，是基于相邻像素灰度的变化形成的区域之间的分界线，具有视觉意义和物理意义。我们对这些边缘模式感兴趣的原因，是因为它们具有明确的物理意义。

边缘是图像中特殊的灰度变化模式。这种灰度变化模式可以是某点处一个方向上灰度变化呈现出来的形态，具有这种形态边缘点一般会聚集形成连续的线条，称作边缘线（edgel）；也可以是某点处若干个方向上灰度变化共同呈现出来的灰度变化模式，这种形态不会在图像中连续的出现，只能是一些孤立的点，称为角点（corner），如图 4.21 所示。

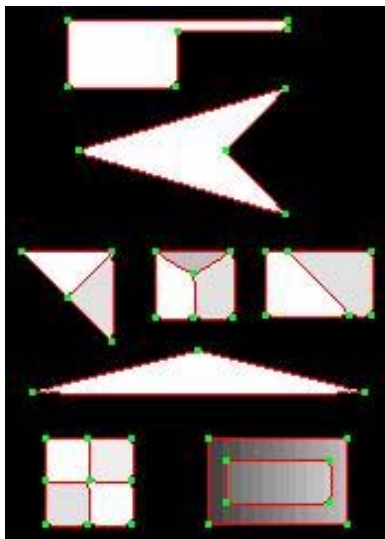


图 4.21: 边缘模式。图像中的边缘线与角点。红色线条是边缘线，它是某一个方向上图像灰度的变化。绿色圈标注的是角点，某点附近多个方向上都有剧烈的变化。源自 [75]

一个方向上的灰度变化模式可以转化为一维灰度变化模式，它体现的图像中灰度变化的一维特征，对这种形态的分析可以在直接使用一维函数（离散信号）进行。对这种一维特征，我们实际上是在图像中的每一个点处、沿着每一个方向检测这种模式。

边缘是主观的。我们可以根据自己的需要选择那些是我们感兴趣的灰度变化模式。并且，可以根据自己的需要和喜好选择灰度变化模式的强度的度量方法。

边缘是一种简单的推断：图像中那些位置存在着我们感兴趣的灰度变化模式。通过这种简单的推断，将数据量庞大的图像数据压缩为稀疏的点和线，实现压缩性的信息表示。

边缘是针对数字图像而言的。在边缘检测过程中，只要是图像中存在的超过一定强度的特定灰度变化模式都检测出来，而不考虑导致这种灰度变化模式的物理原因。在应用中，要根据具体的目的和先验知识对边缘检测的结果进行进一步的推断，以判断这种灰度变化模式是不是真正关心的物理事件导致的。假设希望检测图像中物体的边界，你可以选择阶跃灰度变化模式作为特定的灰度变化模式在图像中进行检测，但是正如 Poggio 等在 [77] 中说那样，实际上检测到的“边缘或许对应着图像中物体 (的边界) 或许并没有对应着图像中物体 (的边界)”，但是通过这样的处理，将搜索物体边界点的范围大大缩小，也就是他说的“边缘具有十分令人满意的性质，它能大大地减少所要处理的信息但是又保留了图像中物体的形状信息。”

§4.4.2 边缘模式

实际上有很多种不同的边缘。很多时候，我们只是通过视觉上的观察，发现某些地方有一些剧烈的变化，所以希望将这些变化检测并定位出来。但是，人并不是仅仅依靠眼睛，很大程度上是通过大脑决定到底那些是真正的“边缘”并对它定位的。在这个过程中，人使用了他在长期学习和训练后得到的所有可用的结果。这是一个相当复杂的智能过程。

§4.4.2.1 边缘产生的物理机制

图像的每一个像素都是对反射光强度 $C(x, y; \lambda)$ 的度量，像素灰度值的变化总是由反射光 $C(x, y; \lambda)$ 的变化引起的。

$$C(x, y; \lambda) = l(x, y; \lambda) \cdot r(x, y; \lambda) \quad (4.62)$$

光源 $l(x, y; \lambda)$ 的变化会导致 $C(x, y; \lambda)$ 的变化，而光源 $l(x, y; \lambda)$ 的改变可能是下面的原因：物体对光源的遮挡；光源本身的不均匀。

物体具有特定的光学性质，同一个物体或者物体上的同一个部分，光学性质是一致的。从这个物体到另一个物体或者从物体到背景，反射系数 $r(x, y; \lambda)$ 发生改变，导致反射光 $C(x, y; \lambda)$ 的变化。

在实际图像中，上述几种情况可能都会出现。图 4.22 画出了一幅图像中的边缘点，从图上看到的边缘有下面的几类：

(1) 空间曲面上的不连续点。如图 4.22 中标为 A 的边缘线，这些边缘线为两个不同曲面的或平面的交线，该点处物体表面的法线方向不连续，在 A 类边缘线的两边，图像的灰度值有明显的不同。

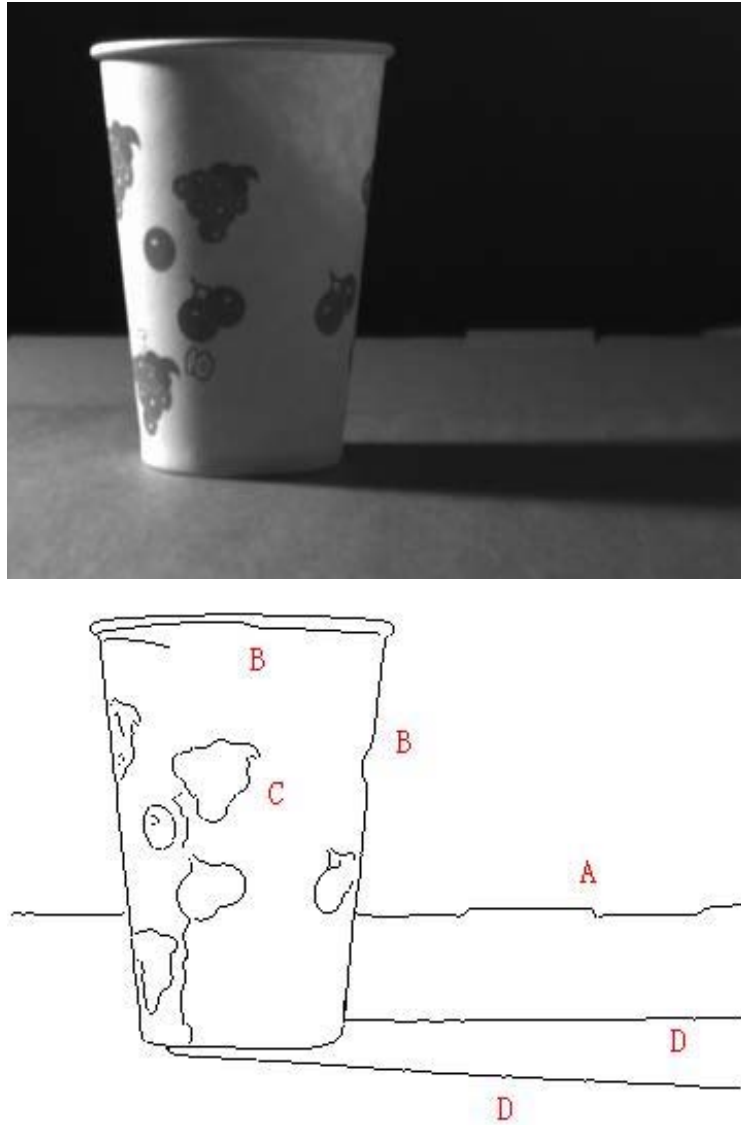


图 4.22: 边缘产生的物理原因

(2) 物体与背景的边界。图 4.22 中 B 类边缘，它是杯子和背景的边界。

(3) 两个不同颜色物体的边界，图 4.22 中 C 类边缘。C 类边缘线是由不同材料或相同材料不同颜色产生的。

(4) D 类边缘。D 是阴影引起的边缘。由于物体表面某一部分被另一物体遮挡，使它得不到光源的照射，从而引起边缘点两侧灰度值有较大的差异。

§4.4.2.2 边缘模式的几何形态

在很多的文献中，作者都是自己提出一个称为“边缘”的数学表达式（曲面），这些模型往往含有刻画边缘强度和方向的参数，然后作者的在给定的噪声的假设下去寻找这些参数。

图 4.23 中通过例子给出了理想情况下几种不同的边缘模式。

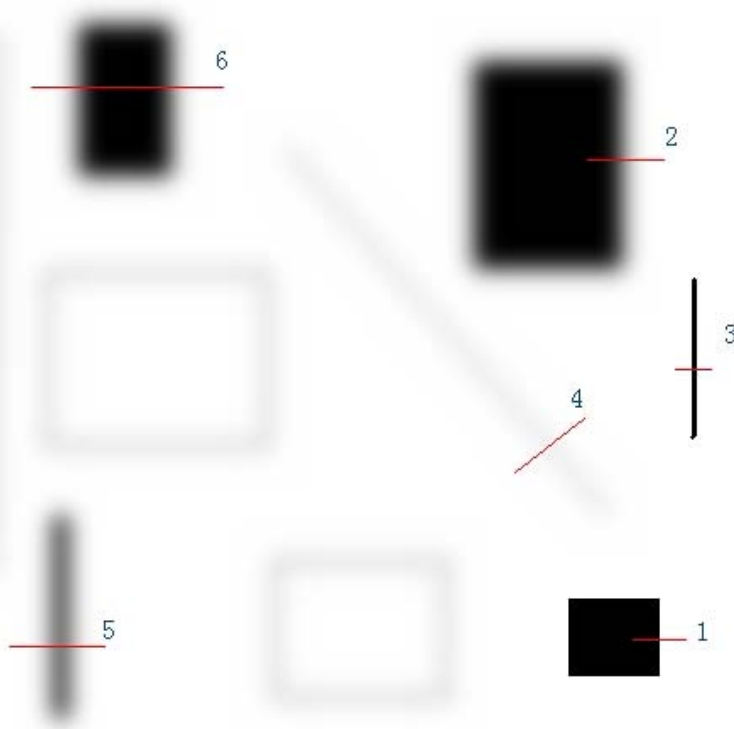


图 4.23: 各种边缘

图 4.24(a) 是理想的阶梯型边缘和带噪声的阶梯型边缘。当相邻的两个物体灰度形成明显对比时，就形成理想的阶梯型边缘模型。因此，阶梯型边缘应该是最常见到发生的边缘，也是最关心的边缘模式，因为它往往对应这物体之间或者物体与背景之间的边界。但是，由于像机的卷积效应，在图像上常常都变成斜坡形边缘，如图 4.24(b)。只是随着分辨率，像机等不同，坡度并不同。

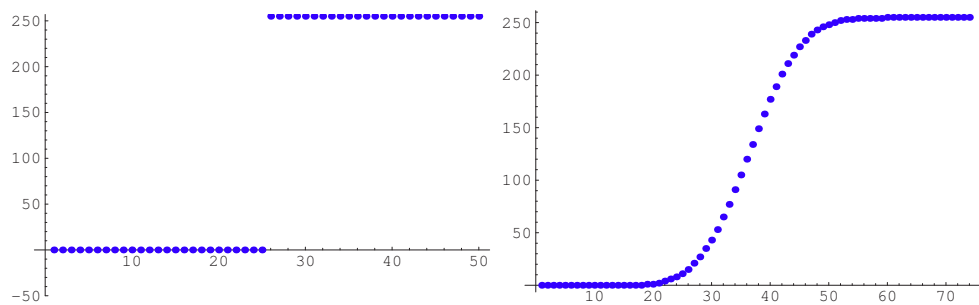


图 4.24: 边缘模式形态: 理想的阶梯形边缘和斜坡型边缘

如图 4.25 所示, 线性边缘也是常见的一种边缘形态。两条靠的近的平行边缘线构成线型边缘。线性边缘非常常见: 例如卫星拍摄到的路面、河流, 医学图像中血管, 文字图像识别中的字符以及一些简单的物体。如果分辨率增大, 线性边缘也可能变成阶梯型边缘。理想的线型边缘在被平滑或者模糊 (相机的卷积效应等造成) 后, 形成屋顶型边缘: 在灰度变化方向上, 灰度先是缓缓的增加 (下降), 到达极值后后立刻缓慢的下降 (上升)。

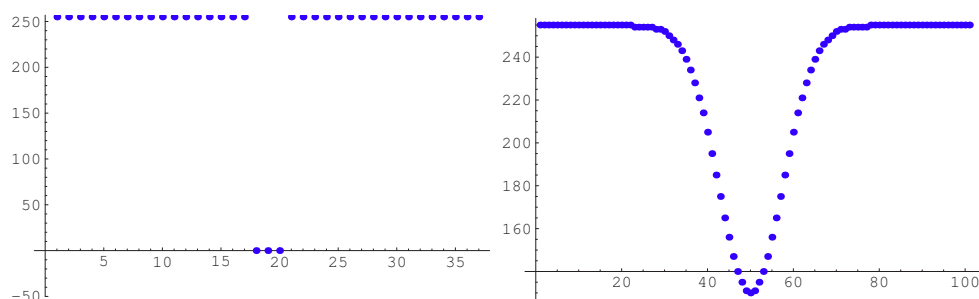


图 4.25: 边缘模式形态: 线型边缘

山脊型边缘 4.26 也是一种边缘模式: 在灰度变化方向上, 灰度先是缓慢的增加 (减小), 到达极值后稳定一定距离, 然后开始还面的减少 (增加)。并且随着分辨率的变大, 山脊型边缘可能变为两个斜坡型边缘; 随着分辨率的变小, 也可能变成屋顶型边缘。

角点常常表现为几条边缘的汇聚点, 常见的角点有 L 型, Y 型, X 型等, 如图 4.27 所示。

§4.4.2.3 边缘模式的数学模型

为了利用数学方法计算得到边缘模式的位置及其强度, 我们首先要对边缘模式进行建模。

阶梯形边缘的数学模型 理想的阶跃边缘用可以用函数 $H(x)$ 表示, 边缘位置对

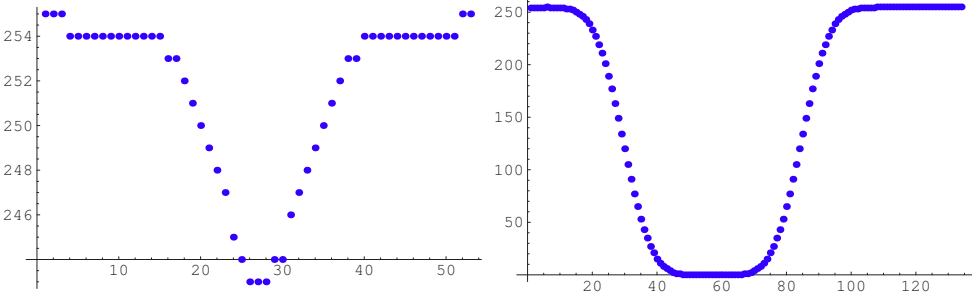


图 4.26: 边缘模式形态：山脊型和屋顶型

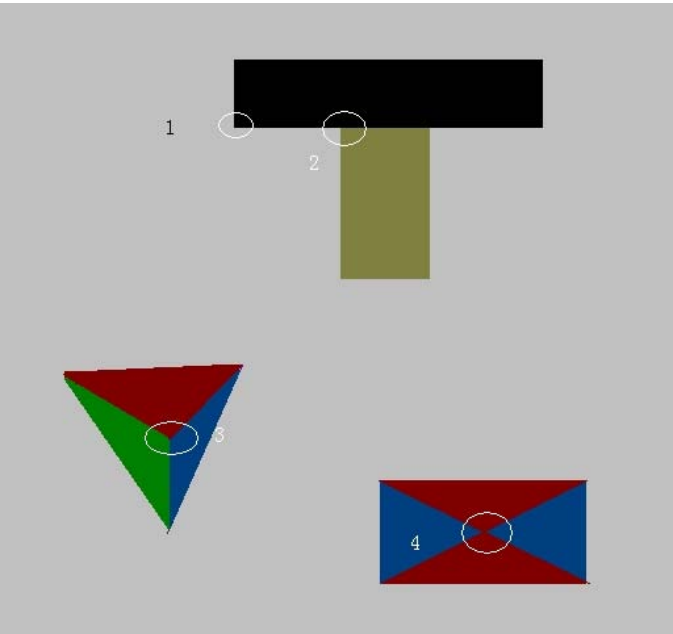


图 4.27: 各种类型的角点

应着 $x = 0$

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (4.63)$$

函数 $H(x)$ 的图像见 4.28。

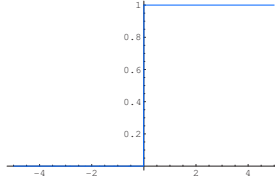


图 4.28: 阶跃边缘

斜坡型边缘的数学模型 斜坡型边缘是实际图像中最常见的边缘，模型有很多。可以将斜坡型边缘看作是阶梯形边缘模糊后得到的，使用高斯函数进行模糊，得到

$$\begin{aligned} R_1(x) &= h(x) * G(x, \sigma) \\ &= \int_{-\infty}^{+\infty} G(x-t, \sigma) h(t) dt \\ &= \int_0^{+\infty} G(x-t, \sigma) dt \\ &= \frac{1}{2} \left(\text{Erf} \left(\frac{x}{\sqrt{2}\sigma} \right) + 1 \right) \end{aligned} \quad (4.64)$$

这可以看作是斜坡型边缘的数学模型，边缘位置对应着 $x = 0$ 的图像见图 4.29(a)。

1991 年，Maria Petrou 和 Josef Kittler [78] 中给出的斜坡型边缘的数学模型，

$$R_2(x) = \begin{cases} 1 - e^{-\sigma x}/2, & x \geq 0 \\ e^{\sigma x}/2, & x < 0 \end{cases} \quad (4.65)$$

边缘位置对应着 $x = 0$ 的图像见图 4.29(b)

在 §4.1 节中，Elder 根据成像的物理机制推导出来的模型如下

$$R_3(x) = f(x/\sigma) \quad (4.66)$$

$$f(u) = \frac{1}{\pi} (\arccos u - u\sqrt{1-u^2}) \quad (4.67)$$

边缘位置对应着 $x = 0$ ， $\sigma = 1$ 时的图像见图 4.29(c)。

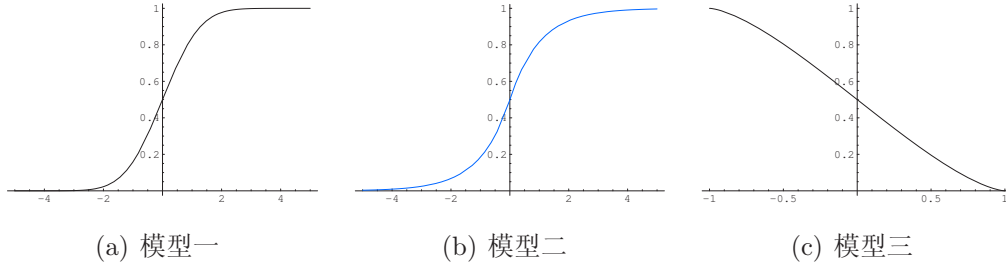


图 4.29: 斜坡型边缘模型

线型边缘的数学模型 理想的宽度为 $2w$ 、强度为 h 的线型边缘的数学模型为

$$L_1(x) = \begin{cases} h, & |x| \leq w \\ 0, & |x| > w \end{cases} \quad (4.68)$$

边缘位置对应着 $x = 0$ 。当 $h = 1, w = 3$ 时的函数 $L_1(x)$ 的图像见图 4.30(a)。

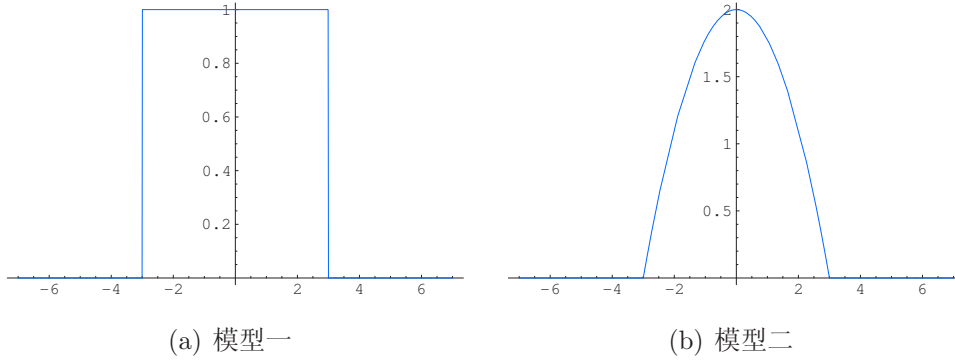


图 4.30: 线型边缘的模型

Carsten Steger [79] 给出了宽度为 $2w$ 、强度为 h 的线型边缘的数学模型

$$L_2(x) = \begin{cases} h(1 - (x/w)^2), & |x| \leq w \\ 0, & |x| > w \end{cases} \quad (4.69)$$

边缘位置对应着 $x = 0$ 。当 $h = 1, w = 3$ 时的函数 $L_2(x)$ 的图像见图 4.30(b)。

角点的数学模型 L 型角点的模型

$$\begin{aligned} & Corner_L(x, y, \theta_0, \theta_1, A_0, A_1) \\ &= (A_1 - A_0) \times H(y \cos(\theta_0) - x \sin(\theta_0)) \times H(y \cos(\theta_1) - x \sin(\theta_1)) + A_0 \end{aligned} \quad (4.70)$$

其中, $-\frac{\pi}{2} < \theta_0 < \theta_1 < \frac{\pi}{2}$, $(\theta_1 - \theta_0)$ 是 L 型角点的角度, $(A_1 - A_0)$ 是 L 型角点的强度, 角点位置为 $(0, 0)$ 。

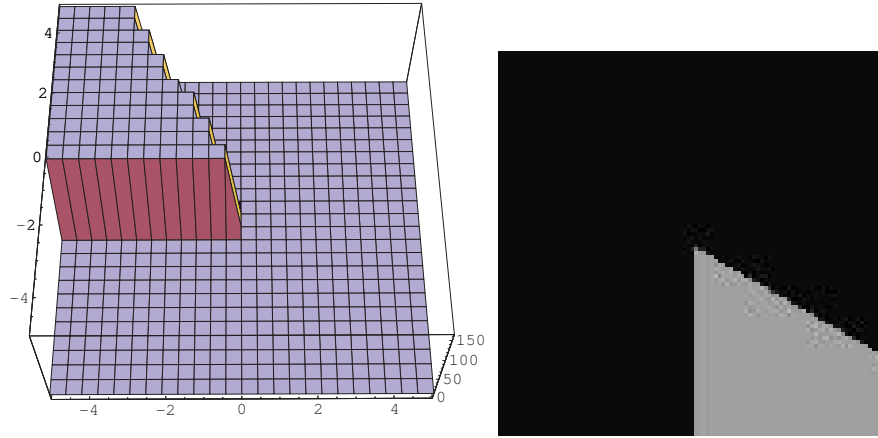


图 4.31: L 型角点的模型的图像

$A_0 = 10, A_1 = 160, \theta_0 = 0, \theta_1 = \frac{2\pi}{3}$ 时的图像为

Y 型角点的模型

$$\begin{aligned} & \text{Corner}_Y(x, y, \theta_0, \theta_1, \theta_2, A_0, A_1, A_2) \\ &= A_0 + (A_1 - A_0) \times H(y \cos(\theta_0) - x \sin(\theta_0)) \times H(y \cos(\theta_1) - x \sin(\theta_1)) \\ & \quad (A_2 - A_1) \times H(y \cos(\theta_1) - x \sin(\theta_1)) \times H(y \cos(\theta_2) - x \sin(\theta_2)) \end{aligned} \quad (4.71)$$

其中, $-\frac{\pi}{2} < \theta_0 < \theta_1 < \frac{\pi}{2}$, $(\theta_1 - \theta_0), (\theta_2 - \theta_1)$ 是 Y 型角点的角度, $(A_1 - A_0), (A_2 - A_1)$ 是 Y 型角点的强度, 角点位置为 $(0, 0)$ 。

$A_0 = 10, A_1 = 160, A_2 = 200, \theta_0 = -\frac{\pi}{4}, \theta_1 = \frac{\pi}{4}, \theta_2 = \frac{3\pi}{4}$ 时的图像为

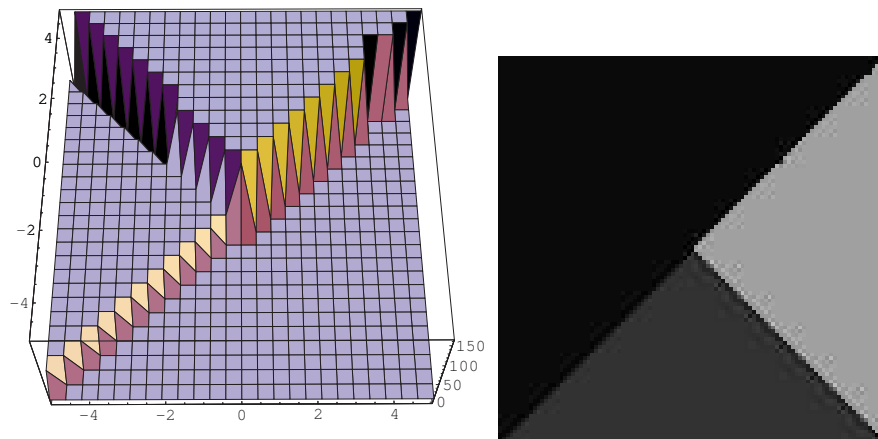


图 4.32: Y 型角点的模型的图像

X 型角点的模型

$$\begin{aligned}
 & Corner_X(x, y, \theta_0, \theta_1, \theta_2, \theta_3, A_0, A_1, A_2, A_3) \\
 = & A_0 + (A_1 - A_0) \times H(y \cos(\theta_0) - x \sin(\theta_0)) \times H(y \cos(\theta_1) - x \sin(\theta_1)) \\
 & (A_2 - A_1) \times H(y \cos(\theta_1) - x \sin(\theta_1)) \times H(y \cos(\theta_2) - x \sin(\theta_2)) \\
 & (A_3 - A_2) \times H(y \cos(\theta_2) - x \sin(\theta_2)) \times H(y \cos(\theta_3) - x \sin(\theta_3)) \quad (4.72)
 \end{aligned}$$

其中, $-\frac{\pi}{2} < \theta_0 < \theta_1 < \frac{\pi}{2}$, $(\theta_1 - \theta_0), (\theta_2 - \theta_1), (\theta_3 - \theta_2)$ 是 X 型角点的角度, $(A_1 - A_0), (A_2 - A_1), (A_3 - A_2)$ 是 X 型角点的强度, 角点位置为 $(0, 0)$ 。

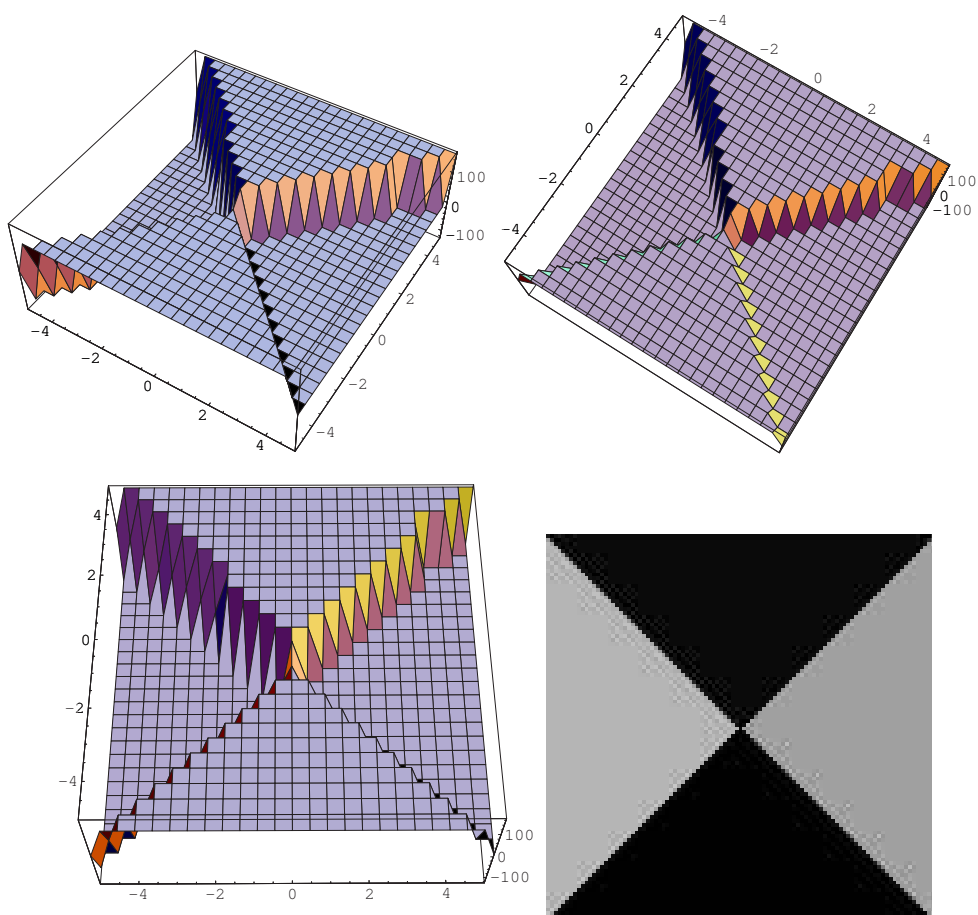


图 4.33: X 型角点的模型的图像

§4.4.2.4 零交叉点是最重要的边缘模式

边缘检测就是在给定一种或者几种边缘模式的前提下, 研究如何设计算法, 在充满噪声中的图像中有效的检测这些边缘模式, 包括边缘发生的位置、方向和强

度。所感兴趣的边缘模式是由具体问题决定的。灰度变化的众多模式中，零交叉点是最重要的灰度变化模式。

首先，零交叉点具有明确的物理意义，理想情况下，零交叉点可能对应着物体之间的边界，这种灰度变化模式在图像中最广泛。

其次，§4.2 节说明，在一定条件下，零交叉点可以近似恢复原始图像，也可以近似恢复其它的灰度变化模式。

另一方面，生物视觉和大部分机器视觉并不需要高精度的零交叉点。例如，就人类而言，对形状、位置的描述都是模糊和粗糙的，并不需要非常准确的零交叉点位置。在这种精度要求较低的条件，很多感兴趣的边缘模式实际上可以不加区分，都可以看作是零交叉点的近似。

只有在机器视觉中需要高精度定位和控制的情况下才需要非常严格的边缘模式模型。

§4.4.3 边缘检测算法

边缘检测是图像处理中相当活跃的研究课题，本节列举一些边缘检测的主要研究人员和研究机构。

1959 年，B.Jule 的 [80]，这个文献是能查到的最早提到边缘检测文章。

1965 年，Roberts [81] 在其博士论文中提出了基于边缘检测的物体匹配算法。Machine Perception of 3-D Solids 是 Roberts 在 1965 年写的 (MIT) 博士毕业论文。这是一篇大家都应该阅读的论文。它是最早分析图像中的边缘、线、模型和图形学的文章。作者提出的系统是第一个 3D 视觉系统，其中有许多后来被大家常用的算子。它使用了三维物体的多边形模型，通过计算图像中的灰度数据寻找图像中的“块”然后对他最适当的表示，最后通过匹配来寻找物体。复杂的物体是由很多“块”组成的，而“块”是由边缘组成的。作者使用他的简单的算子得到了边缘。相邻的边缘点组成线，如果线比较短，那么这条线就被忽略，如果线比较长，那么我们就把它延长直到它与其它的边缘线相交与一点。然后通过对这些线的匹配来识别物体。一旦一个物体识别出来，表示这个物体的边缘就从图像的边缘线表示上删除，然后再对下一个物体进行匹配。

1978 年，David Marr 在总结心理学、人工智能和神经科学等的基础上，提出了著名的视觉计算理论。这个理论的基础是 Marr 和 Ellen C. Hildreth [57,82] 提出的 LOG 边缘检测算法，这个算法结合了多尺度思想和生理学研究的结果，具有非常大的价值。Marr 理论奠定了边缘检测的理论基础。正是在 Marr 工作的基础上，众多数学家研究了多尺度零交叉点和图像的关系，见 §4.2 节。

1983 年，John Canny [85] 在 MIT 硕士毕业时的硕士论文是关于边缘检测的 [83]。他所做的主要工作以论文的形式发表在 1986 年的 IEEE PAMI 上 [84]。Canny

提出了边缘检测中著名的三准则：最优检测准则，最优定位准则和边缘的单响应准则，并在这个准则的基础上，对边缘和噪声作出假设，然后利用数学方法得到了边缘检测的最优滤波器，他发现这个最优滤波器可以用高斯函数的导数很好的近似。因此，高斯函数的导数可以用来检测边缘。这个滤波器，结合非极值抑制以及双阈值二值化就得到边缘检测的 Canny 算法。Canny 算法是边缘检测的标准算法，在边缘检测中一般都使用 Canny 算法来进行算法的比较。Canny 这种思想影响了很多，很多人沿着 Canny 的方向作出了成就，最著名的就是 Deriche 的迭代卷积。

滤波过程是边缘检测中十分重要的部分，它的主要作用是对边缘部分进行增强。滤波过程也是边缘检测中最花时间的环节，一般采用 FFT 进行卷积的加速。在滤波器中，有一类滤波器可以迭代进行，计算量少、运算非常迅速。Deriche [86] 将最优滤波器限制在这类特殊的滤波器之上，得到了最优迭代滤波器 [87,88]。这个滤波器也可以看作是使用这类滤波器对高斯函数的导数进行逼近的得到的近似滤波器，因此，使用数值逼近方法也可以得到类似的结果。迭代滤波大大加速了卷积运算，并且使得滤波的时间与高斯函数的尺度 σ 无关。Deriche 的这个改进是对 Canny 算法的重大贡献，大多数实用的 Canny 算法都是使用 Deriche 卷积算法实现的，因此也称 Canny-Deriche 算法。Deriche 是法国人，他是 INRIA Sophia-Antipolis [89] 的一名研究主管。INRIA Sophia-Antipolis 对边缘检测中的众多问题都进行了深入的研究，研究成果十分丰富，感兴趣的读者可以在它们的主页上找到一些资料。

在 Deriche 思想的启发下，1992 年，Shen Jun [90] 等提出了基于最优平滑滤波器的边缘检测算法。Shen Jun 首先构造出了一个指数型滤波器（作者称为 ISEF 滤波器），和 Deriche 滤波器类似，它可以通过迭代快速的实现。Shen Jun 证明：使用 ISEF 滤波器对图像进行平滑，能在去噪的同时最大可能的保持边缘点的位置；并且这个滤波器与原始图像的差可以非常好的逼近图像的拉普拉斯变换。边缘点包含在这个差值图像的符号（作者称为 BIL）中，于是通过分析 BIL 图像的变符号点就可以得到候选边缘点。

法国人 Demigny [91] 也对边缘检测的最优滤波器理论进行了长期的研究。他认为，Canny 等人所谓的最优滤波器都是在连续域中提出的，这些滤波器要能应用到离散图像上，最终还是要离散为模板去卷积，连续域中所谓的最优在离散域中可能不是最优，他用试验证明了这一点。经过几年的进一步研究，2002 年，他在文献 [92] 提出了边缘检测线性滤波器的最优离散准则，并在此基础上得到了最优离散滤波器。Demigny 另一个贡献是提出了适合硬件实现的快速图像平滑。一般意义的图像边缘检测中，去噪是不可缺少的环节，很多使用是通过高斯滤波器对图像平滑来实现。按照 Deriche 的方法可以快速对图像进行平滑，但是前提是已经获得一幅图像上的全部数据。作为一个实时硬件系统，图像的数据是动态获得的，为了获得高

速度, Deriche 算法不是好的选择, Demigny 给出了解决方案。

尺度是边缘的内在属性, 在没有先验知识时, 需要使用多尺度技术检测边缘。多尺度检测技术首先要构造图像的多尺度表示, 即从原始图像出发, 导出一系列越来越平滑、越来越简化的图像, 这种多尺度表示用来仿真由近到远观测图像中内容的结果。构造多尺度表示的方法有很多, 最简单和最常用的就是高斯滤波。Marr 和 Hildreth 提出的 LoG 实际上就是用时高斯函数构造尺度空间。已经有两本关于尺度空间理论和应用的专著: Tony Lindeberg [94] 的文献 [93] 和 Jon Sporring 等人的文献 [95]。

边缘检测的多尺度算法的基本思路有两种: 将多个尺度中的信息融合在一起或者逐个区域选择合适的尺度进行边缘检测。Bergholm 提出的尺度空间聚焦算法 [98] 就是第一种思路: 大尺度可以检测到显著的边缘, 但是定位精度不如小尺度好。所以, 首先使用大尺度检测显著的边缘, 然后不断将尺度减小, 只要相邻两个尺度的差不是很大, 所检测到的边缘的位置的变化就不会太大, 所以可以不断调整边缘定位以追踪显著边缘的精确位置。Tony Lindeberg 使用的第二种思路。

Zucker 和 Elder [99] 提出了最小可信赖的尺度。1999 年, Elder [56] 提出尺度控制算法。提出了一个基于边缘的图像表示, 可以用作图像编码。但是, 更主要的是它说明了存在一种基于边缘的图像表示, 它是一般的、可靠的、精确的、简洁的、明确的并且是几乎完全的一种图像表示。

图像的多尺度表示中, 图像变的越来越平滑、越来越简化, 这种简化意味着信息的丢失。如果能定量描述每一个尺度中图像的信息, 这对于多尺度表示来说有着重要的作用。Sporring [96] 提出了尺度空间的信息熵概念, 这是直接从香农信息熵的直接推广; 后来又和 Weickert [97] 合作引入了信息阶的概念, 将尺度空间的信息熵定义一般化。使用信息熵概念可以辅助进行尺度选择。王郑耀等在文献 [100] 中, 提出了尺度空间信息量度量的基本要求, Sporring 的度量方法不满足这些直观的要求。他们发现“一维图像”的零交叉点的数量可以作为度量, 并且满足所提出的所有要求, 并把这个度量推广到二维离散情况下。

有关边缘检测的综述和算法比较的算法有很多, 下面列举一些比较著名的。

Michael D. Heath 等 [101, 102] 比较了五种著名的算法: Canny 算法、Nalwa 算法、Iverson 算法、Bergholm 算法和 Rothwell 算法。作者总结了多种边缘检测结果比较的方法, 并提出了自己的方法: 将算法应用到真实图像上, 然后通过人的判断来评估边缘检测算法对物体识别这种特殊任务的优劣。Heath 使用的算法程序都是算法提出者提供的, Heath 给出了大量的测试图像, 使用这些参数和各种不同的参数获得了各种不同的边缘结果, 然后对这些结果进行了分析。在此基础上, 他所在的实验室正在进行更大规模的评价 [103]。

Sheng Tang [104] 给出了九种边缘检测滤波器，并简要介绍了五种著名的边缘检测算法：Canny 算法、Nalwa 算法、Iverson 算法、Bergholm 算法和 Rothwell 算法。

Djemel Ziou 和 Salvatore Tabbone [105] 首先介绍了边缘检测中三大环节：图像平滑、图像微分和边缘标记的基本思路和方法，以及多滤波器算法和多尺度算法。然后作者分析了边缘检测滤波器、边缘自身 (位置、朝向和平滑程度) 对边缘检测结果的影响。阶梯型边缘、线型边缘和角点是三种常见的边缘形态，Ziou 和 Tabbone 对这三种边缘结构的检测方法做了综述。

王郑耀 [106] 在他的学士学位论文 [107] 中总结了十五种左右的边缘检测算法，这些总结是在前面这两篇文献的基础上进行的。与他们不同的是，这篇论文从数学角度总结了 John Canny 最优滤波器的思想及其发展，尤其是 Dimigny 提出的边缘检测最优离散滤波器的设计准则和方法。文章中详细给出了最优离散准则下滤波器的计算过程和 Mathematica® 程序，并给出了使用最优滤波器进行边缘检测的程序和结果。在其主页上可以下载相关算法的可执行程序。附录 1 是 COM 封装的边缘检测算法。

§4.5 多尺度零交叉点对图像视觉信息的表象

§4.5.1 分层次的形状表象

零交叉点检测中，尺度越大，细节越少，此时的零交叉点反映了图像中形状的概貌；可用于物体的粗比较。尺度越小，物体表面呈现出来的细节越丰富，可用于物体精细的比较。

多尺度零交叉点提供了形状的多层次表述，可以根据需要选择合适层次的描述进行识别和匹配，可以大大加速识别和匹配过程。

§4.5.2 零交叉点的封闭性

轮廓线是物体的边界，因此，理论上图像中任何轮廓线要么是封闭的，要么和图像的边界相连通。但是，现存的边缘检测算法，很多无法获得封闭的轮廓。有一些基于拉普拉斯变换的算法，例如 LoG 算法和 Shen 算法，首先计算图像的拉普拉斯变换；然后利用变换后图像的符号，将图像变换为二值图像 (BLI)；最后对 BLI 图像进行边界追踪可以得到满足要求的封闭轮廓。但是，这样的算法实际上很容易出现“虚假轮廓”：也就是在图像中找不到存在的依据。这是一个两难的问题。

本文的观点是：对单幅图像来说，封闭性不是主要问题。因此，首先应该使用合适的算法提取有意义的零交叉点，而不考虑封闭性。正如我们在视觉计算理论中给出的计算模型那样，封闭的轮廓是启发式搜索和跟踪、连接的结果，这应该是一

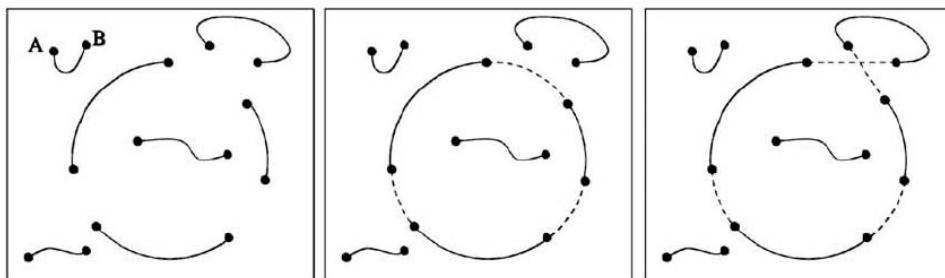


图 4.34: 边缘闭合技术, 源自 [108]

个优化问题。Ratio contour 技术 [108] 研究这个问题。虽然, 人类几乎是在瞬间就做好了这个优化, 但至少在目前, 这还是一个非常难的问题。这应该是目前研究的一个热点问题。

§4.5.3 基于轮廓的颜色和纹理填充

根据 §4.2 的结论, 零交叉点不仅能表示形状, 而且几乎能完备的表示图像中的所有信息 (形状、纹理和灰度)。根据这个结论, 结合 Elder 图像重构算法、视觉心理学和生理学的结论, 我们得到了下面的图像表示方法:

1. 首先提取具有一定强度的零交叉点, 并计算其的尺度, 也就是边缘附近的模糊程度。零交叉点的尺度具有一定的物理意义, 是区分边缘类型的依据之一, 在利用图像表象推导物体表象中具有重要意义, 并可以反应物体与观察者的距离。使用零交叉点表示表示图像中的形状信息。对零交叉点轮廓进行后处理以获得封闭的轮廓, 如图 4.35 所示。

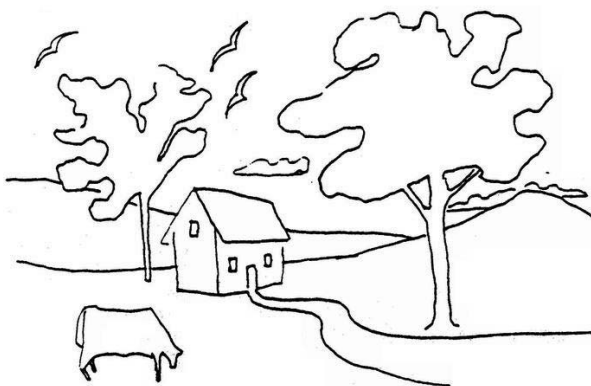


图 4.35: 封闭轮廓线, 源自 [14]

2. 提取零交叉点两侧的灰度值 (颜色), 利用它们表示颜色信息;

3. 通过上述两步，我们给出了图像的形状表示和颜色表示，§4.2 中的理论和实验表明，他们几乎可以完整的重构原图像。但是，实验表明，确实还是有一些信息无法表示，那就是纹理。

与形状相比，对图像中颜色和纹理的表示并不需要很强的。利用形状人类可以区分数以亿计的物体，而且这种才能是天生的。但是，人类的视觉系统能记忆的颜色就相对少的多。人类对形状信息的记忆要比对颜色和纹理信息的记忆细节更丰富。

人类的视觉系统目的并不是在需要的时候恢复先前看到的图像，而是恢复图像中对人类有用的视觉信息。因此，完全没有必要逐点记录图像两侧的颜色值，可以使用更稀疏的数据量记录颜色信息。例如，只是记录零交叉点形成的轮廓线之间所填充区域的主要颜色：主颜色。和颜色信息类似，人类能区分的纹理也非常少。而且，即使视觉能区分两种纹理，人类也很难描述两种纹理的区别。这就说明，人类视觉对纹理的表示是粗糙的和不确定的。

所以，对图像的形状表示应该是一组封闭的零交叉点轮廓，它是提取颜色表示和纹理表示的基础。在图像表象过程中，视觉系统应该首先提取图像中的轮廓，他们反映了形状信息；封闭的轮廓将图像表面分成了若干个不同的区域，然后粗略提取不同区域中的主颜色和纹理。对颜色的表示，本文不做进一步的分析。对纹理的表象，我们将在下一章中做进一步的讨论。

§4.2 中的理论和实验表明这种方法是可行的，而且这种算法和视觉心理学中基于轮廓的填充理论是一致的。

§4.6 轮廓线的描述和识别

得到零交叉点轮廓后，视觉系统还需要对这些轮廓进行进一步的分析，以获得更具有压缩性的描述，然后将这些描述存储起来形成记忆。然后利用这些记忆进行识别。常见的描述有：多尺度轮廓线表示、曲率尺度空间表示和自由曲线表示等。

基于曲率尺度空间的表示参见 §3.3.3。

§4.6.1 多尺度轮廓线匹配

对待识别的轮廓线建立多尺度描述，尺度越大，细节越少，可以快速的进行粗识别。如果匹配成功，则可以使用更小尺度的曲线进行匹配。多尺度轮廓线的建立可以参考 §3.3.3 节。在每一个尺度上识别的方法有：广义 Hough 变换 [23] 和基于 Hausdorff 距离的识别等，下面简单介绍基于 Hausdorff 距离的形状识别。

Hausdorff 距离度量的是二值参考图像中每一个像素与二值的待识别图像中一些像素靠近的程度。这里的二值图像包括边缘图像、角点图像等等，最常见到的是边

缘。

给定两个有限点集 $A = \{a_1, a_2, \dots, a_p\}$ 和 $A = \{b_1, b_2, \dots, b_q\}$ ，它们之间的 Hausdorff 距离定义为：

$$H(A, B) = \max h(A, B), h(B, A) \quad (4.73)$$

其中

$$h(A, B) = \max \min \|a - b\| \quad (4.74)$$

$\|\cdot\|$ 是点集 A 和 B 所在空间的范数，最常见到的是 L_2 范数。

$h(A, B)$ 称为从 A 到 B 的直接 Hausdorff 距离，它等于点集 A 中与点集 B 的距离最大的点 $a \in A$ 与 B 的距离，而这里的距离度量是 a 与点集 B 中最近的点范数。于是， $h(A, B)$ 实际上评估了 A 中任意一个点与 B 中的点的距离，然后把最大值 (A 中最大的不匹配点) 做为度量值。

Hausdorff 距离 $H(A, B)$ 是 $h(A, B)$ 和 $h(B, A)$ 的最大值，它通过度量点集 A 中与 B 中任意一个点距离最大的点的距离和点集 B 中与 A 中任意一个点距离最大的点的距离而度量了点集 A 和 B 的不匹配的程度。直观上看，如果 $H(A, B) = d$ ，那么 A 中任意一个点都在点集 B 的某个点的距离 d 之内，反之也是如此。

上述定义的 Hausdorff 距离 $H(A, B)$ 只能用来度量两个固定位置的点集之间的不匹配程度。为了能度量任意相对位置下两个点集的不匹配程度。

给定变换群 G ，定义最小 Hausdorff 距离，

$$Measure_G(A, B) = \min_{g_1, g_2 \in G} H(g_1 A, g_2 B) \quad (4.75)$$

这个定义中，我们对两个点集都做了变换。如果变换群 G 满足，对任意 $g \in G$ 和任意点两点 x_1, x_2 ，都有

$$\|gx_1 - gx_2\| = \|x_1 - x_2\| \quad (4.76)$$

则我们只需要对其中的一个点集进行变换，

$$Measure_G(A, B) = \min_{g \in G} H(A, gB) \quad (4.77)$$

满足 4.76 的群是任意范数 $\|\cdot\|$ 时的平移变换和欧式范数时的刚体变换（平移和旋转）。

图 4.36 是文献 [109] 中基于 Hausdorff 距离识别算法的演示。

§4.6.2 自由曲线表示

B 样条 (B-Spline) 或者非均匀有理 B 样条 (NURBS) 是描述曲线的工具，也可以对原始轮廓线拟合得到 B 样条或者 NURBS 曲线，得到一些控制点。得到

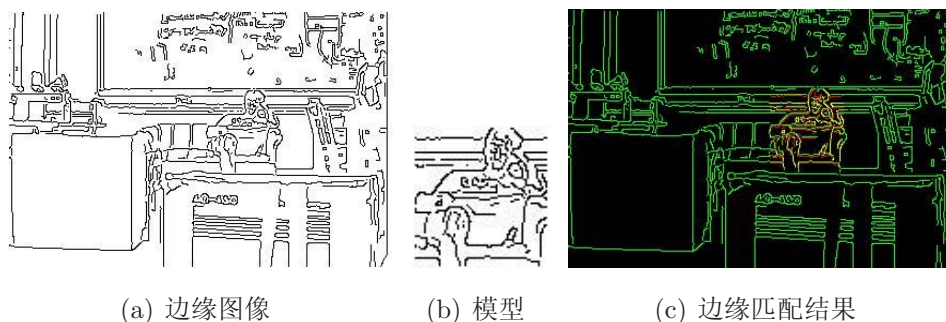


图 4.36: 基于 Hausdroff 距离识别算法的演示, 源自 [109]

的控制点可以作为曲线的描述 [110]。而通过控制点匹配就可以实现曲线形状的匹配 [111, 112]。

§4.6.3 特殊的表示

此外, 在机器视觉应用中, 大多数物体都是人造物, 它们呈现出的形状比较简单: 直线、圆形、椭圆及它们的组合。因此, 可以使这些简单曲线去近似描述一般的轮廓线, 然后进行匹配和识别。值得注意的是, 人的视觉对直线、矩形、圆形和椭圆等是非常敏感的, 初等几何学主要研究的就是这些特殊形状。这就暗示我们, 对一些特殊类型的曲线, 人类的视觉系统似乎具有特殊的描述方式。

另外, 比较常用的是曲线的隐含多项式 (Implicit Polynomials) 表示 [113]。曲线 Γ 的隐含多项式表示是指将曲线用一个多项式 $P(x, y)$ 的零点来表示, 即 $\Gamma = \{(x, y) : P(x, y) = 0\}$ 。

§4.6.4 小结

虽然对轮廓线的描述和识别的方法已经很多, 但是这些描述和识别都是建立在简单曲线的基础上的, 都没有利用轮廓线的拓扑信息。如图 2.3 所示, 封闭的轮廓线具有非常复杂的拓扑结构, 而上面这些描述都没有利用拓扑信息。这是一个值得研究的方向。

第五章 图像的纹理表象

本章讨论如何描述零交叉轮廓线所包围区域的纹理表象。

§5.1 纹理的统计描述

§5.1.1 视觉系统的纹理感知

对纹理的感知是视觉的一项基本认为。远古人类能在森林中生存下来的首要条件就是在各种复杂的环境中识别野兽，例如，能区分狮子的皮毛形成的纹理和草丛的纹理。它们可能具有相近的颜色，也没有明确的形状。

朗文词典中对纹理的解释是：由紧密交织的成分组成的事物；身体或物质的组成微粒构成的组织；某事物可见或可触摸的表面特征及外观。显然，纹理感觉有两种来源：触觉和视觉。我们这里说的是视觉纹理。本文在 §2.3.2.3 和 §2.4.2 中已经给出了 (视觉) 纹理的一些特性和定义。图 5.1 是几幅典型的纹理图像。

纹理的基本特征是移不变性，也即对纹理的视觉感知基本上与其在图象中的位置无关，移不变性可被描述成是确定性的 (结构的)，或是随机的 (不规则的)，但很可能存在着介于这两者之间的类别。我们可以感受到纹理以及两个纹理之间存在差别，但是很难描述具体的纹理。

人们可用来描述纹理的性质有均匀性 (Uniformity)、密度 (Density)、粗细度 (Coarseness)、粗糙度 (Roughness)、规律性 (Regularity)、线性度 (Linearity)、定向性 (Directionality)、方向性 (Direction)、频率 (Frequency) 和相位 (Phase)。图 5.2 是这些描述的理想化图像。

1993 年时, Rao 和 Loshe [116] 对 Brodatz 相册中的 56 幅纹理图像做了研究。他们找到二十个志愿者 (20 — 35 岁之间)，让他们对这些纹理图像进行分组，并详细记录下他们的分组过程，并描述同一组内的图像具有什么相似性。对记录下的数据进行多种统计分析后，Rao 和 Loshe 得到一个结论：人类在纹理识别中最重要的三个维度为周期性、方向性和随机性。1996 年 Picard 和 Liu [117] 也提出：Brodatz 相册中差异性最大的三种纹理为周期性，随机性和方向性的纹理。

纹理的周期性指，纹理可以近似看作是按一定规则对元素 (elements) 或基元 (primitives) 进行排列所形成的重复模式，如图 5.1(c) 所示。这种重复模式具有确定性，人类也可以对这种重复模式进行描述。纹理基元的类型有很多，包括点、斑、格、垅、栅。

方向是纹理感知中很重要得因素。图 5.3(a) 中，感知右边的边界要比感知左边的边界容易。但是，图 5.3(b) 的三个区域中，左边区域和中间区域中纹理基元是不

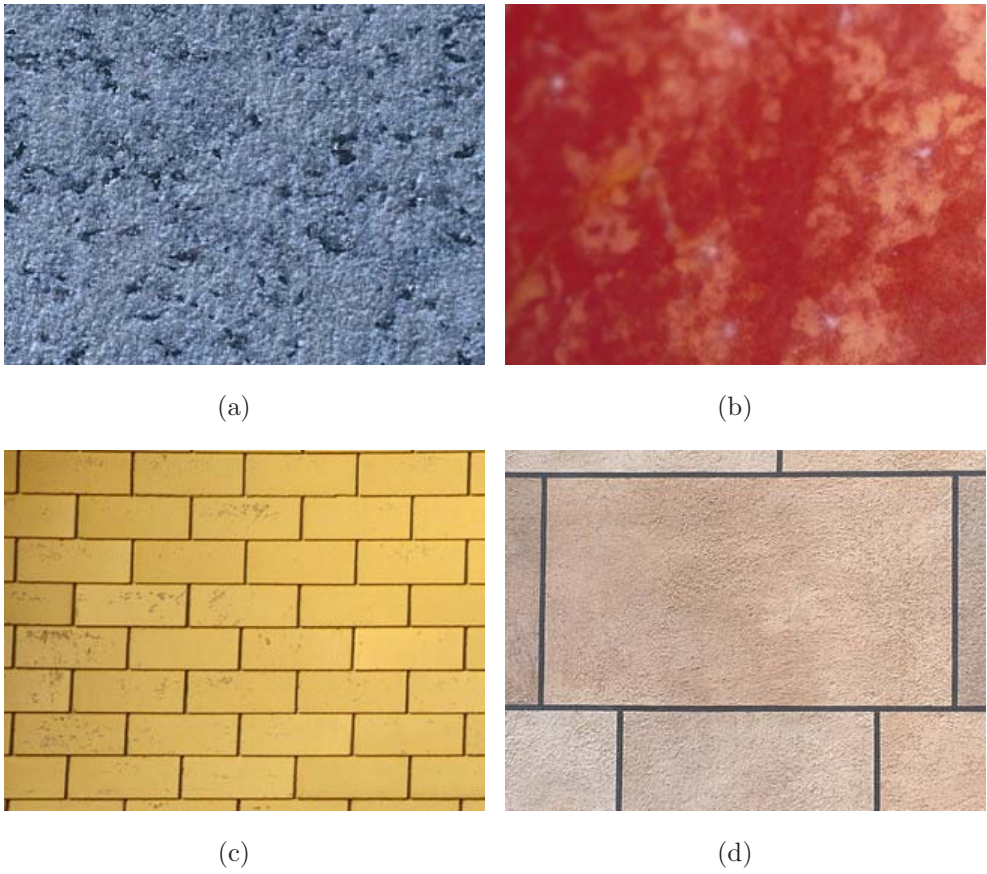


图 5.1: 纹理图像的例子 [115]

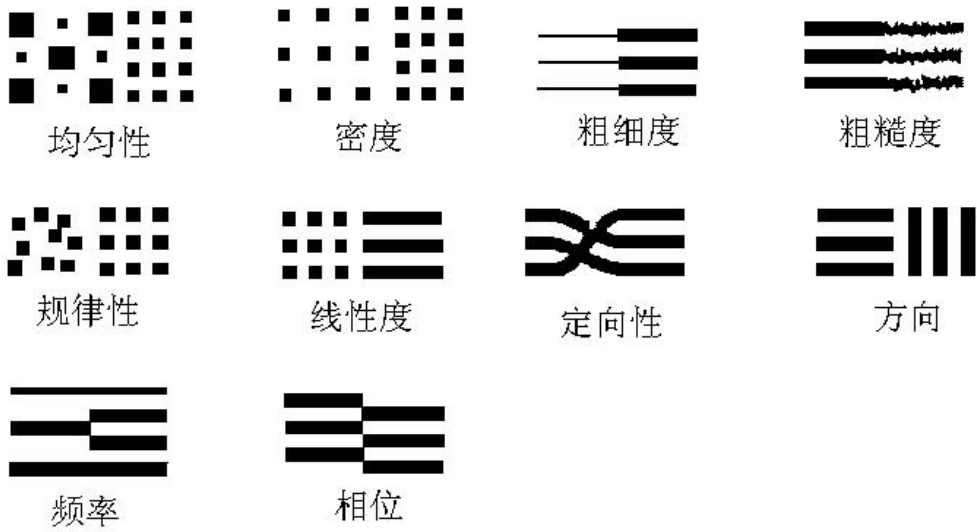
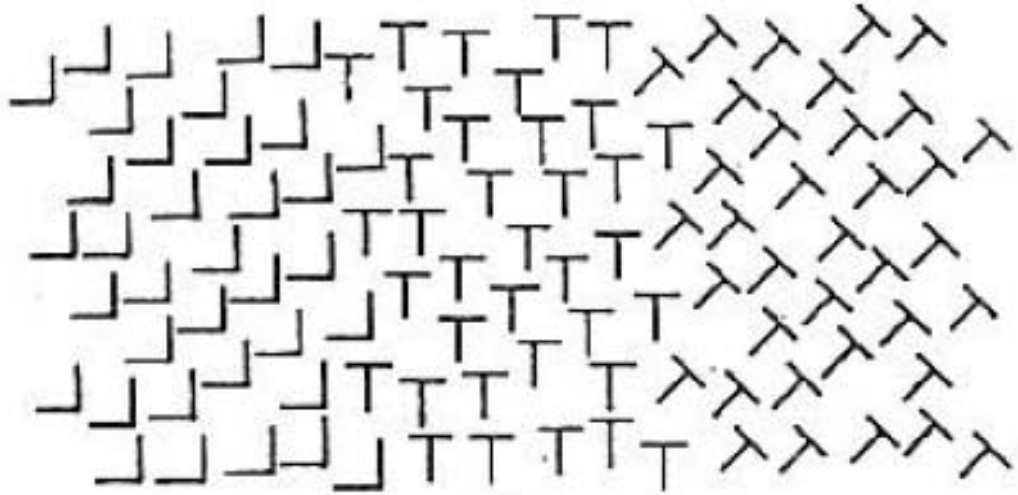
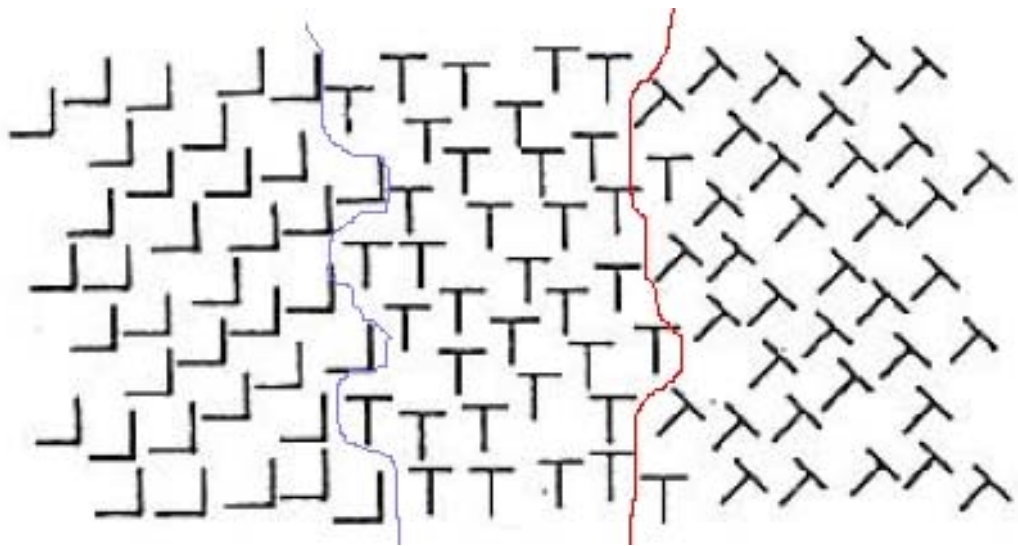


图 5.2: 人类主观对纹理的感受 [114]

同的；而中间区域和右边区域中的纹理基元是相同的，差别仅仅是方向。



(a) 纹理图像



(b) 视觉感知的边界

图 5.3: 纹理感知中的方向 [114]

随机性是纹理的另一个重要特点。在自然纹理中周期性和方向性都是统计意义下的，而不是严格的。只有某些人工生成的纹理才会出现比较严格的重复性和方向性。自然界中绝大多数纹理，都是随机性的，在遥感影像中纹理绝大部分属随机性 [118]。

纹理感知具有分割的作用，利用纹理可以将不同的物体分割开来。零交叉线也具有这种作用。纹理和零交叉线有区别也有联系。图 5.4 中，天与地之间的分割就

是零交叉线，而相邻斑马的分割就需要纹理感知。



图 5.4: 纹理分割, 源自 [125]

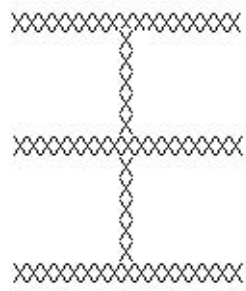
零交叉点分割是灰度不同的两块区域，而根据纹理感知到的分界线是灰度在物体表面上分布模式变化而体现出来的。灰度在物体表面上的分布模式实际上也是零交叉点，只是这些零交叉线在特定的尺度下太小、太弱，所以人的视觉“视而不见”，使用纹理感知的方式进行了感知。实际上，纹理可以用边缘来描述，如 MPEG7 中选用了缘直方图描述纹理 (§2.4.2)。纹理可以看作是小而弱的零交叉线在空间上的分布，这些零交叉线的方向就是决定了纹理的定向。纹理分割线是比零交叉点更高一级的分割。

但是，两者又是有关联的，特定的纹理和零交叉线都只在特定的尺度下才是可见的。如图 5.1 中 (c) 和 (d) 所示，随着尺度的变小，所看到的物体的细节越来越丰富，原来看起来很平滑的区域也出现了复杂的纹理。如果尺度变大，图像中的纹理可能会消失，体现为一定的零交叉线，如图 5.5 所示。

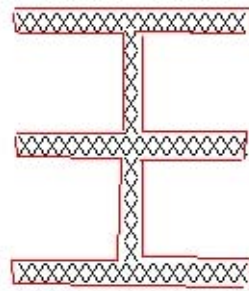
利用纹理的连续变化可以帮助视觉系统恢复三维形状 [121]，这就是所谓的利用纹理三维信息的技术 (Shape from texture)。

一般将纹理分成两大类：周期性的纹理与随机性的纹理。所以，图像处理中也相应的有两类纹理描述和分析的方法：结构性分析法 (structural approach) 及统计性分析法 (statistical approach) [119]。

结构法把纹理视为由基本纹理元按特定的排列规则构成的周期性重复模式，因此常采用基于传统的 Fourier 频谱分析方法以确定纹理元及其排列规律。此外结构元统计法和文法纹理分析也是常用的提取方法。结构法在提取自然景观中不规则纹理时遇到困难，这些纹理很难通过纹理元的重复出现来表示，而且纹理元的抽取和排



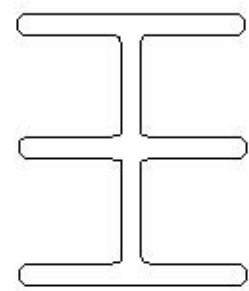
(a) 人造纹理



(b) 纹理感知的边界



(c) 高斯尺度变换后的图像



(d) 零交叉点检测

图 5.5: 纹理、零交叉点与尺度



图 5.6: 个别的基元不称为纹理

列规则的表达本身就是一个极其困难的问题。

§5.1.2 纹理表象与统计和随机

图像的纹理表象就希望通过很少一些参数抓住图像的内在性质，使得我们能够理解图像所表现的社会或者世界。纹理表象的研究就是希望寻找一个多层次的紧凑的描述一般纹理的形式化系统，如果可能的话，我们还希望这个表示是完全的。

每一种纹理实际上都是确定性的和随机性的统一，纹理的随机性具有某些统计规律，这个规律刻画了纹理的内在性质。本章讨论如何对纹理的统计规律进行表象，而纹理确定性将在下一章做简单的讨论。在视觉感知中，纹理的统计描述是十分重要的，这主要基于下面的原因：

人类视觉感知到纹理的一个先决条件是：基本组织（例如纹理基元）的大量出现。如果基本结构只有很少几个，视觉系统一般不会将这个区域视为纹理。图 5.6 是图 5.5(a) 中的基本基元，视觉系统不会将图 5.6 看作是纹理区域，而是根据零交叉线进行形状感知。统计描述不用考虑物体的细节，不考虑区域中存在什么物体，而只是对空间的灰度分布进行描述。大量的基元存在时，每一个都进行形状表象是不现实也是不经济的，应该使用统计描述。

人类视觉感知的经验表明：我们能感知到纹理但是无法完全描述一种纹理；能感知到两种不同纹理之间的差别，但是很多时候我们无法描述它们具体区别是什么。因此，人类视觉中纹理的描述是宏观的而不是精确的，这种描述可能是一些统计量和统计规律。

上一节指出，纹理和零交叉线是与尺度有关的。物体在图像中充分大和充分显著时，视觉器官感知到的物体边界以零交叉线的形式表现出来时，以使得人能从图像中获得足够确切的视觉信息：位置、形状等等。图像中的细节等表现为纹理，这些纹理所表现出来的物体细节特征对人类而言相对不重要，但是完整描述物体表面的细节是十分复杂的并且消耗很多能量和精力。因此可以使用统计性的描述，从整

体描述其统计规律，反应亮度（灰度）的空间变化情况。

从统计意义上讲，纹理图像可以看作是二维随机变量或者二维时间序列。

§5.1.3 统计分析的是什么？

纹理描述方法是统计，那么统计分析的到底是什么呢？如果统计分析的是图像数据本身，那么就说明图像数据本身是纹理表象。但是，很多证据表明，需要分析是图像数据进行某种预处理或者变换后的图像数据。主要依据是：

1. 对视觉感知过程研究的一种结论是：视觉过程是由若干个独立通道并行处理的 [124]，其中每一个通道只对某一特殊的空间频率段具有选择性。这种空间频率选择性必然是对图像数据进行预处理的结果。而且这种频率分解应该具有很好的局部性。

2. 视觉感知的结果是分层次的组织结构。这种分层次的组织结构，可以看作是尺度选择的结果。在大尺度下，图像中的细节消失，可以抓住图像中物体的主要结构；小尺度下能仔细感知图像中的细节。纹理只有在一定的尺度下才能被感知到，说明视觉系统对纹理的感知具有尺度选择能力。这种尺度选择能力，是通过对数据的变换来实现的。

3. 图像数据是十分复杂的，而且数据之间的相关性很强，而且不同类型图像的相关性变化很大。这对视觉系统要实现一般图像的统计特征抽取是十分不利。从统计学本身的需要来说，需要首先对原始图像数据进行特定的变换。

于是，图像的纹理表象基本结构如图 5.7 所示。预处理后的数据作为纹理表象，纹理表象的描述是统计量。

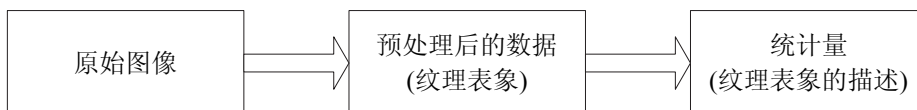


图 5.7: 纹理表象的基本框架

下面要解决的问题是，预处理过程对图像做了什么变换，纹理表象是什么。根据上面的分析，这种变换，应该具有频率分解能力，具有多尺度、多分辨和多层次的特点，而且变换后的数据比原来的数据更容易抽取特征。图像的小波变换系数恰好具有这些能力 [122]。

§5.2 小波变换系数作为纹理表象的优越性

“如果你想画一幅画，画上有天空、云、树和花朵，那么你应该根据特征大小

的不同选用合适的刷子。小波就象是这些刷子。” — Ingrid Daubechies [131]

本节主要说明，小波具有多尺度、多分辨和多层次的特点 (§5.2.1)，小波具有局部化的频率分解能力 (§5.2.2)，而且变换后的数据比原来的数据更容易抽取特征 (§5.2.3)。因此，小波变换后的数据是纹理表象的最佳选择。

§5.2.1 函数的小波变换

本节简要介绍关于小波的知识，具体请参考 [130, 131, 132, 133]。

§5.2.1.1 连续小波变换

连续小波变换是在傅立叶变换的基础上发展起来的，是调和与分析自然发展的结果。

定义5.1 (小波函数). 函数 $\psi(t) \in L^2(\mathbb{R})$ 称为小波函数，如果它满足以下的允许条件：

$$C_\psi \triangleq \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (5.1)$$

其中， $\hat{\psi}(\omega)$ 是 $\psi(t)$ 的傅立叶变换，即

$$\hat{\psi}(\omega) \triangleq \int_{-\infty}^{+\infty} \psi(t) e^{-i\omega t} dt \quad (5.2)$$

如果 $\hat{\psi}(\omega)$ 是连续的，那么允许条件就等价于 $\hat{\psi}(\omega) = 0$ ，即

$$\int_{\mathbb{R}} \psi(t) dt = 0 \quad (5.3)$$

将 $\psi(t)$ 伸缩和平移后，可以得到一族函数 $\{\psi_{a,b}(t) | a, b \in \mathbb{R}, a \neq 0\}$,

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \quad (5.4)$$

a 称为伸缩因子， b 称为平移因子。 $\{\psi_{a,b}(t) | a, b \in \mathbb{R}, a \neq 0\}$ 构成了 $L^2(\mathbb{R})$ 的标准正交基。任意 $g(t) \in \mathbb{R}$ 都可以用这一族标准正交基表出，

$$g(t) = (W_\psi g)(a, b) \psi_{a,b}(t) \quad (5.5)$$

其中，表出系数

$$(W_\psi g)(a, b) \triangleq \langle g, \psi_{a,b} \rangle = |a|^{-\frac{1}{2}} \int_{\mathbb{R}} g(t) \overline{\psi_{a,b}(t)} dt \quad (5.6)$$

定义5.2 (连续小波变换). 式 5.6 实际上定义了一个函数的变换，称为连续小波变换。

定义5.3 (离散参数小波变换). 在连续小波变换中, 将 $a, b \in \mathbb{R}$ 进行离散化, 令 $a = a_0^j, b = lb_0 a_0^j, j, l \in \mathbb{R}$, 得到

$$(DPW_\psi g)(j, l) \triangleq \langle g, \psi_{a,b} \rangle = a_0^{-\frac{j}{2}} \int_{\mathbb{R}} g(t) \overline{\psi_{a_0^{-j}-lb_0}(t)} dt \quad (5.7)$$

称离散参数小波变换。

显然, 上述变换不过定义了 $L^2(\mathbb{R})$ 上一类特殊的函数变换: 将一维函数通过由一个基本函数 (小波基函数) 伸缩和平移得到的一族标准正交基变换到二维平面上。

离散数据的小波变换 由于实际中的信号都是实离散信号, 小波函数也是实函数, 所以需要对连续小波变换公式 5.6 离散化, 令 $f_a(b-t) = |a|^{-\frac{1}{2}} \psi_{a,b}(t)$, 则

$$(W_\psi g)(a, b) = (W_{\psi;a}^g)(b) \triangleq \int_{\mathbb{R}} g(t) f_a(b-t) dt \quad (5.8)$$

式 5.8 实际上就是滤波过程, $f_a(t)$ 就是一个普通的滤波器。在信号处理中有一系列的连续滤波器, 如高斯滤波器、LoG 滤波器等。由于数字图像是离散的, 在实际滤波的时候, 需要考虑连续滤波器与离散数据之间的一致性问题。一般有两种方法: 连续滤波器的离散化; 离散数据的分段逼近方法。

第一种方法: 离散数据的分段逼近方法 文献 [132] 中 §5.3 详细给出了这种算法。这种方法精度高, 但是速度太慢, 实际很难使用。

第二种方法: 连续滤波器的离散化 需要首先将连续的滤波器离散化。下面以高斯滤波器的离散化为例做简单的说明。参数为 σ 的高斯函数 $g_\sigma(x)$ 为

$$g_\sigma(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (5.9)$$

连续滤波器的离散化就是取这个滤波器在整数节点

$$\{\dots, -n, \dots, -2, -1, 0, 1, 2, \dots, n, \dots\}$$

上的值, 使用这些数值作为离散卷积滤波器。令 $\sigma = 1.0$, 此时 $g_{1.0}(x) = e^{-\frac{x^2}{2}}$, 这是一个定义在整个实数轴上的函数, 且在任意有限点上值都不为零, 我们显然无法使用所有整数轴上 $g_{1.0}(x)$ 的值作离散卷积。由于高斯函数当 $x \rightarrow \infty$ 的时迅速趋于零, 所以可以只使用有限项, 使用这有限个数据作为近似的离散卷积。

因为高斯函数关于原点对称的, 因此选取关于原点对称的 $2N+1$ 个数 $\{-N, -(N-1), \dots, -2, -1, 0, 1, 2, \dots, N-1, N\}$ 作为离散节点。当 $N = 10$ 时, 参

表 5.1: 高斯函数及其取值

n	0	-1/1	-2/2	-3/3
$g_{1.0}(n)$	1.0	0.606531	0.135335	0.011109
n	-4/4	-5/5	-6/6	-7/7
$g_{1.0}(n)$	0.000335463	3.7267×10^{-6}	1.523×10^{-8}	2.2897×10^{-11}
n	-8/8	-9/9	-10/10	...
$g_{1.0}(n)$	1.2664×10^{-14}	2.5768×10^{-18}	1.9288×10^{-22}	...

见表 5.1。可以看出，当 $|n| > 5$ 的时候， $g_{1.0}(n) < 10^{-8}$ ，这么小的量在图像处理中完全可以忽略不计。图 5.8 就是高斯函数的离散表示，显然离散滤波器相当于用分段常数函数 $\hat{g}(x)$ 近似连续滤波器函数 $g(x)$ 。

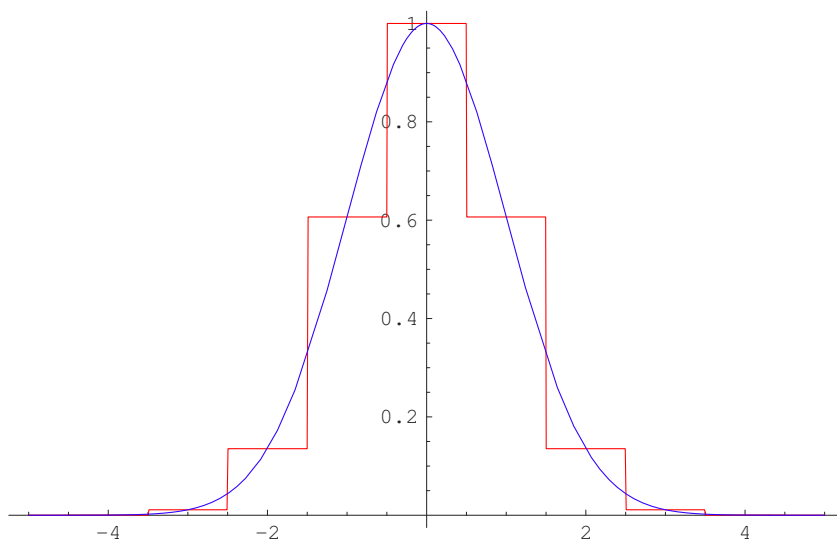


图 5.8: 高斯函数及其离散化表示

图像看作连续函数在定义域和值域的离散化，假设连续图像为 $I(x)$ ，

$$I(x) = \begin{cases} 50.0 \times (1 - e^{-x}), & x \geq 0 \\ 50 \times e^x, & x < 0 \end{cases} \quad (5.10)$$

灰度值相当于用简单的分段常数函数 $\hat{I}(x)$ 近似连续函数 $I(x)$ ，如图 5.9 所示。

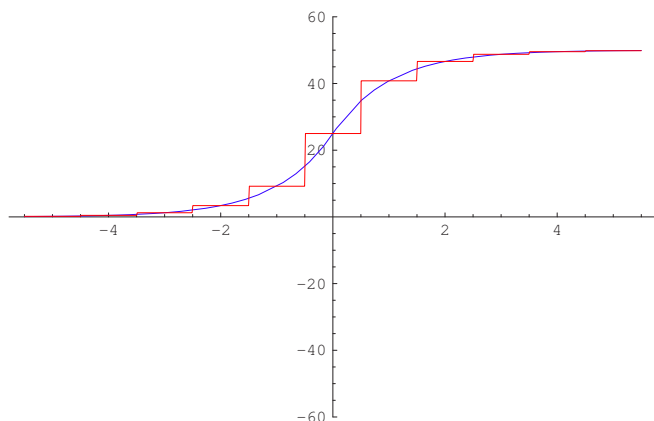


图 5.9: 连续图像及其离散化表示

按照连续卷积的定义，在点 t 处连续卷积定义为 $C(t) = (I(x) * g(x))(t)$,

$$\begin{aligned}
 C(t) &= (I(x) * g(x))(t) = \int_{-\infty}^{\infty} I(y)g(t-y)dy \\
 &\approx \int_{-\infty}^{\infty} \hat{I}(y)\hat{g}(t-y)dy \\
 &= \sum_{n=-\infty}^{\infty} I(n)g(t-n) \\
 &\approx \sum_{n=-N}^N I(n)g(t-n)
 \end{aligned} \tag{5.11}$$

于是根据 5.11, 可以利用整数点上滤波器的值和数字图像灰度值近似计算图像滤波。如图 5.10 所示。

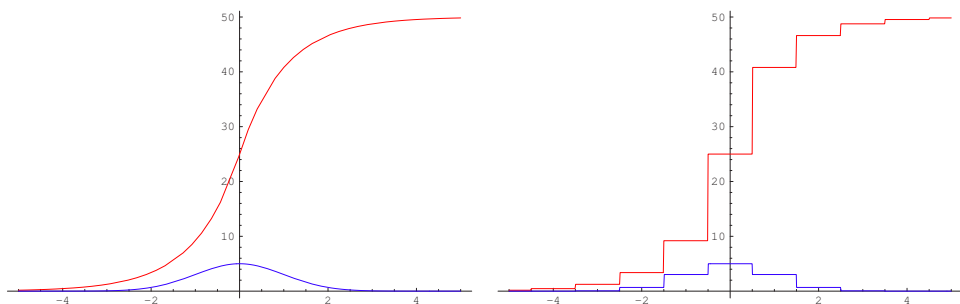


图 5.10: 离散化的意义，为了显示的方便，高斯函数前乘以常数 5.0

如果选择 $N = 5$, 则离散化后的高斯滤波器长度为 $n = 2 \times N + 1 = 11$ 。在图像的每个点处都要进行 n 11 次乘法和 $n - 1$ 10 次加法运算。高斯函数直接对图像滤波的一个问题是，如何选择离散滤波器的长度，也就是如何将无限长滤波器截断为有

限项。选择的 N 越小计算的误差就越大，但是显然计算量就越大；选择 N 越大，计算误差小，但是计算量增加。

连续小波变换的实现参见 [137]。

§5.2.1.2 多分辨分析

小波的广泛应用是建立在 Mallat 多分辨分析理论 [122] 的基础之上的。Mallat 多分辨分析理论正是对视觉中多尺度思想的继承和发展，是对尺度变换思想的严格的数学建模。

视觉是一个根据图象发现周围景物中有什么物体和物体在什么地方过程，也就是从图象得到对观察者有用的符号描述的过程。人在不同的距离下观测同一图像获得的感受是不一样的，远距离看到的是图像轮廓，近距离下看到的是更多细节，这就是尺度效应。物体只在某些尺度下呈现出来，在更小的尺度或者更大的尺度下，它们就消失了。计算机视觉的一个重要的任务就是对物体或者特征进行识别，但是事实上只有在特定的尺度下这些物体或者特征才会出现，因此尺度选择十分重要。

图像多尺度表示的基本思想是：把原图像嵌入到一族从原图像导出的含有一个自由参数的图像中去，使得这个参数是对多尺度观测的一个仿真，如图 5.11 所示。

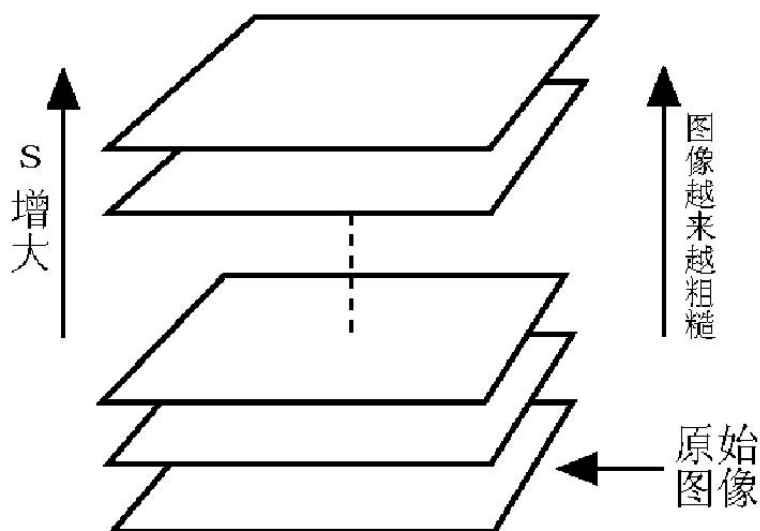


图 5.11: 图像的多尺度表示

在 Mallat 之前，已经基本上建立了完备的高斯尺度空间理论 [126, 127, 128, 129]。Mallat 对信号的尺度变换做了研究后，总结得到尺度变换的几个基本原则 ($I(x)$ 表示信号，假设 T_s 表示尺度参数为 s 的尺度变换):

1. 对任意信号 I , $T_s(T_s(I)) = T_s(I)$;

2. 因果性: 如果 $s_1 > s_0$, 则尺度参数为 s_1 的信号中含有尺度参数为 s_0 的信号中的所有信息;

3. 尺度变换没有优先的尺度参数;

4. 变换应该具有空间平移不变性;

5. 任意尺度的信号中含有的信息应该能用一系列离散数据特征化。

总结上述基本原则, Mallat 建立了一个描述了尺度变换的数学模型。

定义5.4 (多尺度向量空间). $L^2(\mathbb{R})$ 中的函数组成的向量空间 $V_j, j = -\infty, \dots, +\infty$ 称为多尺度向量空间, 如果

- (1) $\forall j \in \mathbb{Z}, V_j \subset V_{j+1}$;
- (2) $\cap_{j=-\infty}^{j=+\infty} V_j = \{0\}$ 并且 $\cup_{j=-\infty}^{j=+\infty} V_j$ 在 $L^2(\mathbb{R})$ 中是稠密的;
- (3) $\forall j \in \mathbb{Z}, f(x) \in V_j \iff f(2x) \in V_{j+1}$;
- (4) $\forall n \in \mathbb{Z}, \forall f(x) \in V_0$, 则 $f(x-n) \in V_0$;
- (5) $\forall j \in \mathbb{Z}, V_j \subset V_{j+1}$; 存在 V_j 映上到 $l^2(\mathbb{Z})$ 的同构。

定义5.5 (尺度变换). 对任意 $j \in \mathbb{Z}$, 函数 $I(x) \in L^2(\mathbb{R})$ 的尺度参数为 $s = 2^j$ 的尺度变换定义为 $I(x)$ 在多尺度向量空间 $V_j \subset L^2(\mathbb{R})$ 的正交投影。

对尺度空间向量的进一步发展 [123], 得到多分辨表示。

定义5.6 (多分辨空间). $L^2(\mathbb{R})$ 中的一串闭子空间序列 $\{V_j\}_{j \in \mathbb{Z}}$ 称为多分辨分析, 如果

- (1) 单调性: $\dots \subset V_{-1} \subset V_0 \subset V_1 \dots$;
- (2) 逼近性: $\cap_{j \in \mathbb{Z}} V_j = \{0\}, \overline{\cup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R})$;
- (3) 伸缩性: $f(t) \in V_j \iff f(2t) \in V_{j+1}$;
- (4) 平移不变性: $f(t) \in V_j \iff f(t-k) \in V_j, \forall k \in \mathbb{Z}$;
- (5) 存在函数 $g(t) \in V_0$ 使得 $\{g(t-k)\}_{k \in \mathbb{Z}}$ 构成 V_0 空间的 Riese 基。

通过对 Riese 基 $g(t)$ 的正交化可以得到 $\varphi(t) \in V_0$, 使得 $\{\varphi(t-k)\}_{k \in \mathbb{Z}}$ 构成 V_0 的规范正交基。由伸缩性和平移不变性, 可以知道 $\varphi_{j,k}(t)_{k \in \mathbb{Z}}$ 构成 V_j 规范正交基,

$$\varphi_{j,k}(t) \triangleq 2^{\frac{j}{2}} \varphi(2^j t - k), j, k \in \mathbb{Z}, t \in \mathbb{R} \quad (5.12)$$

这里, $\varphi_{j,k}(t)_{j,k \in \mathbb{Z}}$ 称为尺度基, 函数 $\varphi(t)$ 称为尺度函数, $\{V_j\}_{j \in \mathbb{Z}}$ 称为尺度空间。

显然, 多分辨表示是 Mallat 尺度变换数学模型的基础, 多尺度和多分辨是同一个客体的两种不同说法而已。信号在 V_j 上的投影实际上就是信号的参数为 j 的尺度变换。由于 $V_j \subset V_{j+1}$, 所以信号在参数为 $j+1$ 的尺度上的信息要包含信号在参数为 j 的尺度上的信息, 多出来的信息属于 V_{j+1} 中, 但不在 V_j 中, 即在 V_j 在 V_{j+1} 的

正交补空间中。令 W_j 表示这个补空间，即

$$V_{j+1} = V_j \oplus W_j \quad (5.13)$$

$\{W_j\}$ 就是小波空间。

如果存在 $\psi(t) \in W_0$ ，使得 $\{\psi(t-k)\}_{k \in \mathbb{Z}}$ 形成空间 W_0 的规范正交基，则 $\psi_{j,k} = 2^{\frac{j}{2}}\psi(2^j t - k)$ 也形成了 $L^2(\mathbb{R})$ 的规范正交基。如果选的好， $\psi(t) \in W_0$ 就可以满足我们对小波的要求，称为小波基函数。

由多分辨分析，

$$\varphi(t) \in V_0 \subset V_1 \quad (5.14)$$

$$\psi(t) \in W_0 \subset V_1 \quad (5.15)$$

所以 $\varphi(t)$ 和 $\psi(t)$ 都可以用 V_1 空间的一个基 $\{\varphi(2t-k)\}_{k \in \mathbb{Z}}$ 线性表出，

$$\varphi(t) = \sum_k h_k \varphi(2t-k) \quad (5.16)$$

$$\psi(t) = \sum_k g_k \varphi(2t-k) \quad (5.17)$$

这就是双尺度方程。

多分辨分析在小波发展过程中具有十分重要的作用：可以对小波有更清晰的理解；给出了一个寻找和构造小波基函数的过程；得到了进行小波变换的 Mallat 算法。

Mallat 算法 根据多分辨分析得到了两个 $L^2(\mathbb{R})$ 上的子空间序列：尺度空间序列 $\{V_j\}_{j \in \mathbb{Z}}$ 和小波空间序列 $\{W_j\}_{j \in \mathbb{Z}}$ ， $\{\varphi_{j,k}\}_{k \in \mathbb{Z}}$ 和 $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ 分别是它们的规范正交基。任意函数都可以在这两列空间上做投影，

$$f_{V_j} = \sum_{k \in \mathbb{Z}} \langle f, \varphi_{j,k} \rangle \varphi_{j,k} \quad (5.18)$$

$$f_{W_j} = \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} \quad (5.19)$$

函数 $f(t)$ 在小波空间的展开为

$$f(t) = \sum_{j \in \mathbb{Z}} f_{W_j} = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k} \quad (5.20)$$

严格的说，小波变换是函数 $f(t)$ 在小波空间上展开的，但是我们不可能无穷和将任意尺度上的小波系数都求出来。因为，对信号来说，我们也不用考虑它所有的

细节信息。假设 $f(t) \in V_J$ (或者使用 $f(t)$ 在 V_J 的投影来代替 $f(t)$) , 则

$$V_J = V_{J-1} \oplus W_{J-1} \quad (5.21)$$

$$= V_{J-2} \oplus W_{J-2} \oplus W_{J-1} \quad (5.22)$$

$$= V_j \oplus W_j \oplus W_{j+1} \oplus W_{j+2} \oplus \cdots \oplus W_{J-1}, (j < J) \quad (5.23)$$

所以,

$$f(t) = \sum_{k \in \mathbb{Z}} \langle f, \varphi_{J,k} \rangle \varphi_{J,k}(t) \quad (5.24)$$

$$= \sum_{k \in \mathbb{Z}} \langle f, \varphi_{j,k} \rangle \varphi_{j,k} + \sum_{j \leq j' < J} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j',k} \rangle \psi_{j',k}(t) \quad (5.25)$$

记 $c_{j,k} \triangleq \langle f, \varphi_{j,k} \rangle$ 和 $d_{j,k} \triangleq \langle f, \psi_{j,k} \rangle$ 分别为尺度系数和小波系数。上式实际上说明, 函数 $f(t)$ 在小波空间中的分解可以一级一级来进行。每一次都利用函数在参数为 $j+1$ 的尺度空间的表示系数 $c_{j+1,k}$, 函数在参数为 $j-1$ 的尺度空间的表示 $c_{j,k}$ 和在参数为 j 的小波空间的表示 $d_{j,k}$ 。

利用双尺度方程得到

$$c_{j,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} h_n c_{j+1,n+2k} \quad (5.26)$$

$$d_{j,k} = \frac{1}{\sqrt{2}} \sum_{n \in \mathbb{Z}} g_n c_{j+1,n+2k} \quad (5.27)$$

这就是 Mallat 的快速小波分解算法, 实际上是一个离散的卷积过程。每一个小波变换都对应两个信号分解滤波器 $\{h_k\} \{g_k\}$, 利用它们可以进行信号的分解。

一维的小波变换和 Mallat 算法都可以推广到二维, 实现对二维图像的多分辨率分析。图像的小波变换的实现参见 [138, 139]。图 5.12、图 5.13 和图 5.14 是用 *Matlab*® 得到的图像小波分解系数。

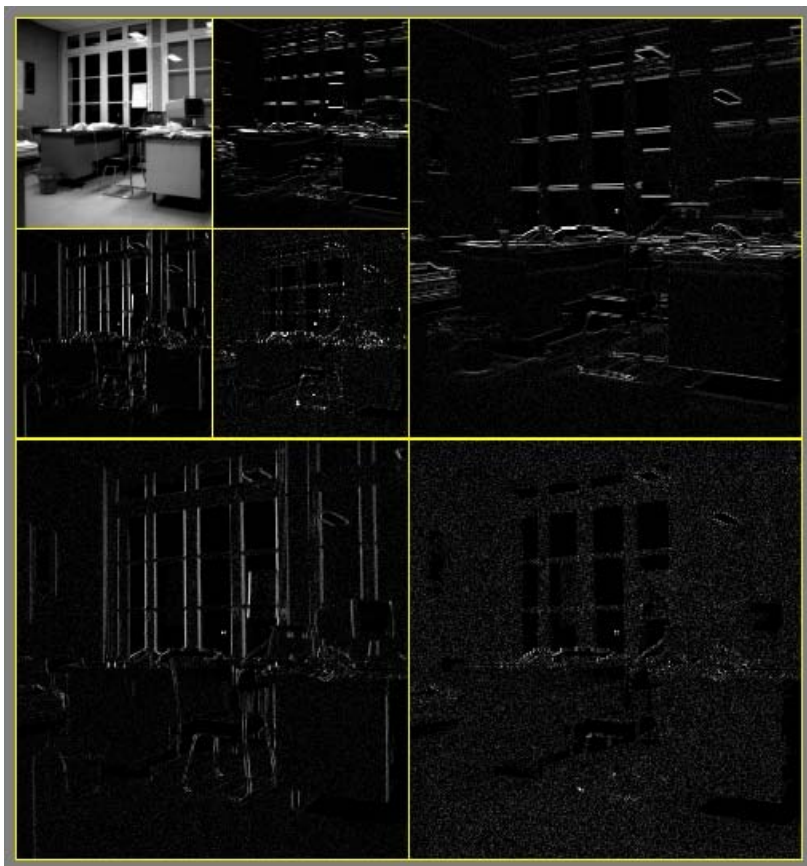
§5.2.1.3 小波的提升格式

1995 年, Sweldens 提出了一种基于空间域的小波构造方法提升 (lifting scheme) 方法 [134, 135, 136]。与传统的多分辨率分析相比, 提升格式具有下面两个特点:

1. 在空间域理解和构造小波, 需要的数学基础少;
2. 计算速度更快, 计算方法更简单, 而且适合于自适应、非线性、非奇异采样和整数到整数的变换。

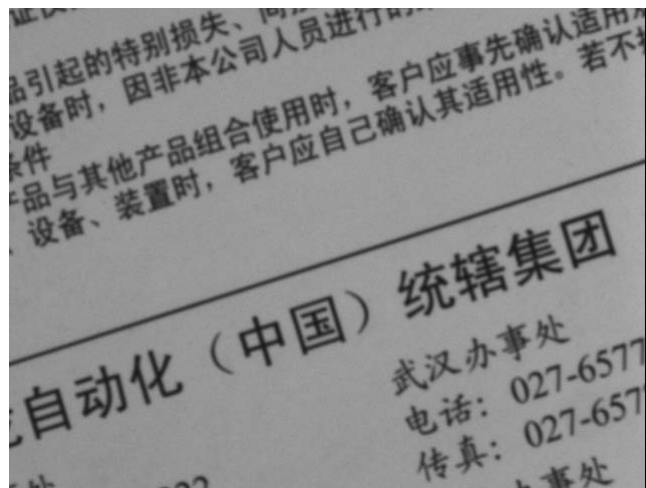


(a) 图像

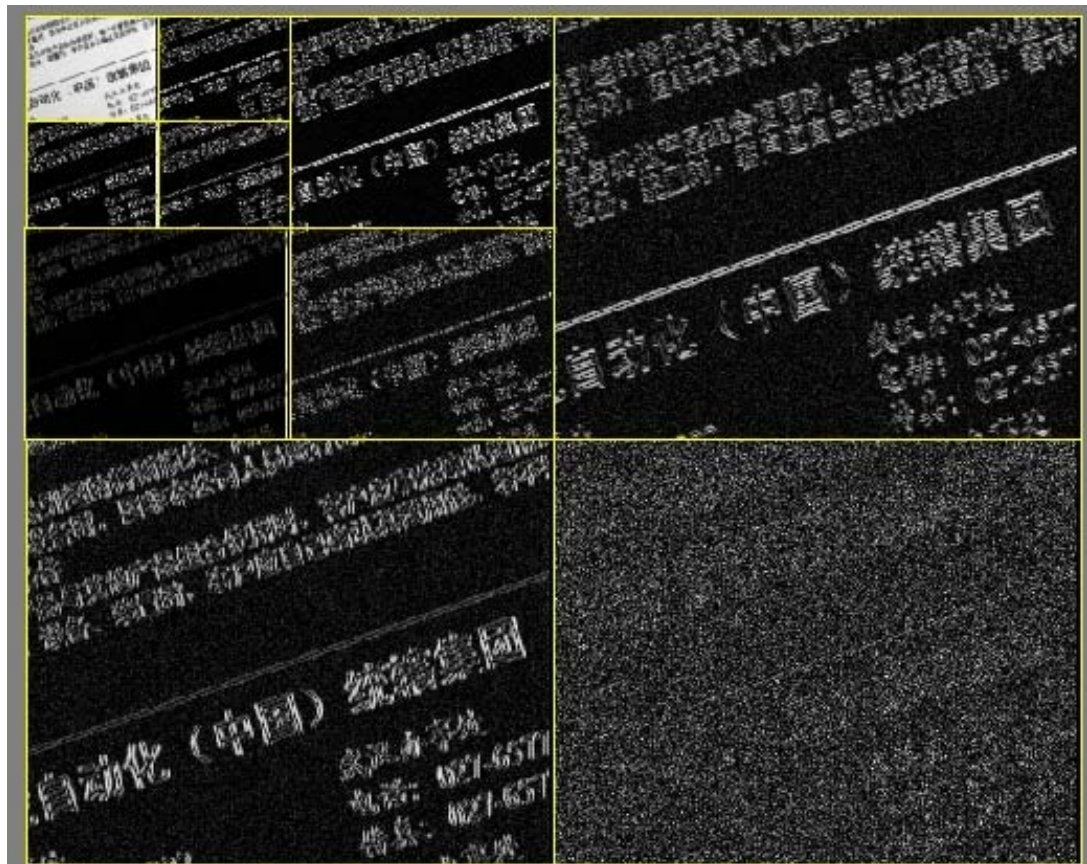


(b) 图像的离散小波变换系数

图 5.12: 图像的 Mallat 分解

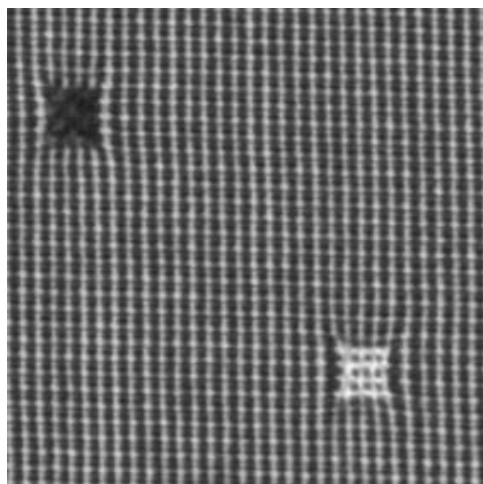


(a) 文字纹理图像

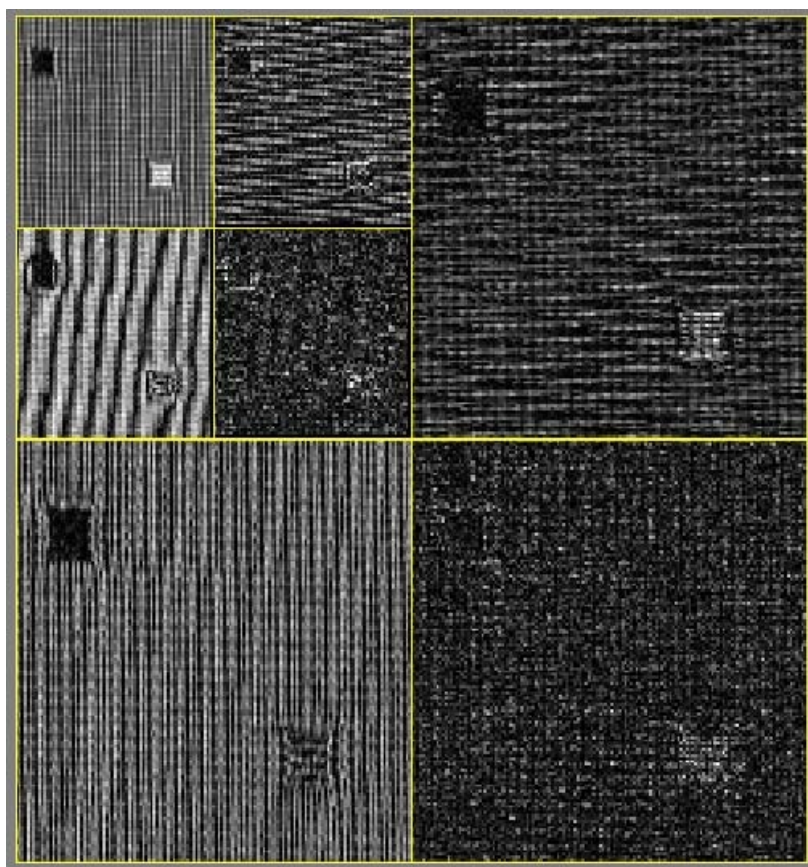


(b) 图像的离散小波变换系数

图 5.13: 文字纹理图像的 Mallat 分解



(a) 纺织物纹理图像



(b) 图像的离散小波变换系数

图 5.14: 纺织物纹理图像的 Mallat 分解

因此，常常称为“二代小波”。但是，它和传统的小波本质上是一样的，是小波的另一理解。

§5.2.2 小波变换的频率分解能力

任意小波基函数 $\hat{\psi}(\omega) = 0$ ，所以小波函数 $\psi(t)$ 是带通的，假设其频率范围为 $[T_0, T_1]$, $T_1 > T_0 > 0$ ，显然 $\psi(t - k)$, $k \in \mathbb{Z}$ 的频率范围也为 $[T_0, T_1]$ ，所以 W_0 空间的频率范围为 $[T_0, T_1]$ 。

令 $F(\omega) = \int_{\mathbb{R}} \psi(t) e^{-i\omega t} dt$ 且 $t' = 2^j t$ ，则有

$$\int_{\mathbb{R}} \psi(2^j t) e^{-i\omega t} dt = 2^{-j} \int_{\mathbb{R}} \psi(2^j t) e^{-i(2^{-j}\omega)(2^j t)} d(2^j t) \quad (5.28)$$

$$= 2^{-j} \int_{\mathbb{R}} \psi(t') e^{-i(2^{-j}\omega)(t')} d(t') \quad (5.29)$$

$$= 2^{-j} F(2^{-j}\omega) \quad (5.30)$$

$\psi(2^j t)$ 的频率范围为 $[\frac{T_0}{2^j}, \frac{T_1}{2^j}]$ ，所以 W_j 空间的频率范围为 $[\frac{T_0}{2^j}, \frac{T_1}{2^j}]$ 。因为小波变换实际上是函数在小波函数空间上的投影，所以，小波变换具有频率分解能力。

§5.2.3 小波变换在统计分析中的优点

对函数的小波变换具有非常好的性质，这些性质使得小波变换在统计分析、时间序列分析和数据挖掘中具有显著的优势。一般的说，这些优势体现在对信号的预处理中，通过小波变换对数据做预处理可以更好的发现数据的特征。

§5.2.3.1 小波变换系数的统计规律

小波分析在数据统计分析中的应用参见 [140, 141]，在数据挖掘中的应用见 [143]。Azimifar [144] 总结了小波系数的统计规律：

1. 因为小波具有局部性、多分辨和压缩的特点，所以可以把小波变换看作是白化过程（whitening process），它使得变换后的小波系数是统计上独立的并且是非高斯的。
2. 系数表示，压缩性，很好一些系数就可以恢复图像；而且这些系数满足下面的规律：如果某一个小波系数比较大或者小，那么与它相邻的小波系数比较大或者小的可能性很大（聚类性，clustering）；小或者大的小波系数试图沿着尺度传播（持续性，persistence）。

此外，Azimifar 总结了四十多篇文献的结果，指出了图像小波变换系数统计规律：这些规律包括同一个子图中小波系数的统计规律、同一个尺度中不同子图间小波系数的统计规律、相邻尺度中小波系数的统计规律。尽管图像是任意复杂的，但是这些规律是一般的。这些规律可以作为视觉系统的先验知识，使得纹理的特征提

取相对简单！

§5.2.3.2 小波变换在时间序列分析中的优点

本节参考了文献 [142]。

时间序列是指和一个有序的独立变量 t 有关的任何观察序列（变量 t 可以是离散值，也可以是连续值）。时间序列的例子包括沿着时间、深度或者距离得到的数据，也就是说时间序列并不必须与时间有关。时间序列的数据往往不能以回归分析的方法来建立模型加以分析，因为回归分析想要建立的是因果模型。而时间序列中之各观测值间通常都存在相关性，时间相隔越短之两观测值，其相关性越大，时间序列并不满足所谓“各观测值为独立”的必要假设。

在物理科学中，数据在不同尺度下的均值和这些均值的变化是非常有意义的。假设要分析学生食堂最近十年来就餐人数的变化情况，我们拥有的数据是这十年来每天记录的就餐人数，数据有 $N = 365 \times 10 = 3650$ 个，记做 $X_i, i = 0, \dots, N - 1$ 。

为了分析人数的变化，可以将每个月的平均就餐人数计算出来，

$$X_m^M = \sum_{j=0}^{29} X_{m+j}, m = 0, \dots, 119 \quad (5.31)$$

然后，计算 $dX_m^M = X_m^M - X_{m-1}^M$ ，从 dX_m^M 就可以分析人数的变化。也可以将每年的平均就餐人数计算出来，

$$X_y^Y = \sum_{j=0}^{364} X_{y+j}, y = 0, \dots, 10 \quad (5.32)$$

计算 $dX_y^Y = X_y^Y - X_{y-1}^Y$ ，从 dX_y^Y 也可以分析人数的变化。

这个过程其实可以看作是 Harr 小波滤波的结果。Harr 小波变换的结果就是相邻两个区间平均值的差；而 Harr 小波本质上就是观察相邻两个区间均值的差，这个均值是抽样的均值 $\frac{1}{N} \sum_{j=0}^{N-1} x_j$ 。而其它各种小波，观察的就是各种各样的加权平均 $\frac{\sum_{j=0}^{N-1} \omega_j x_j}{\sum_{j=0}^{N-1} \omega_j}$ 。因此，小波可以告诉我们一个特定的函数的加权平均是随着所平均的时间不同而如何变化的。

离散小波变换是一个有效的分析工具的基本理由如下：

1. 离散小波变换将时间序列重新表示为与特定的时间 t 和特定的二进尺度 2^{j-1} 的系数的形式。由于可以从它的离散小波变换的系数重构原来的时间序列，所以这些系数就等价于原来的时间序列。

2. 离散小波变换允许我们将时间序列的能量分割为与不同的时间和尺度相联系的片段。能量分解和著名的统计中的方差分析（ANOVA）技术很相似，于是利用离

散小波变换就可以得到基于尺度的方差分析，这与利用能量谱而建立的基于频率的方差分析是十分相似的。

3. 离散小波变换能有效地将物理应用中常见的大量的时间序列去相关。图 5.15 中，图的左半部分是一个信号的及其离散小波变换：最下面是原始信号，最上面是尺度为系数，中间 7 个是不同尺度的小波系数；图的右半部分左边显示了左边对应数据的部分自相关序列与零的偏差的图像。可以发现，尽管原信号自相关性很大，但是小波系数自相关性很小。在一定的假设下，可以推论得到下面的结论：上述数据的自相关性为零。

时间序列统计分析的中心目的是寻找将相关序列表示成不相关序列的组合的方法。离散小波变换能有效地将相关序列（甚至是高度相关的序列）去相关，结合我们可以利用离散小波变换的系数精确重构原来的时间序列，这就是离散小波变换是时间序列分析的一个很有用的工具的最根本原因。

4. 离散小波变换可以通过一种比令人吃惊的快速傅里叶变换更快的算法计算。

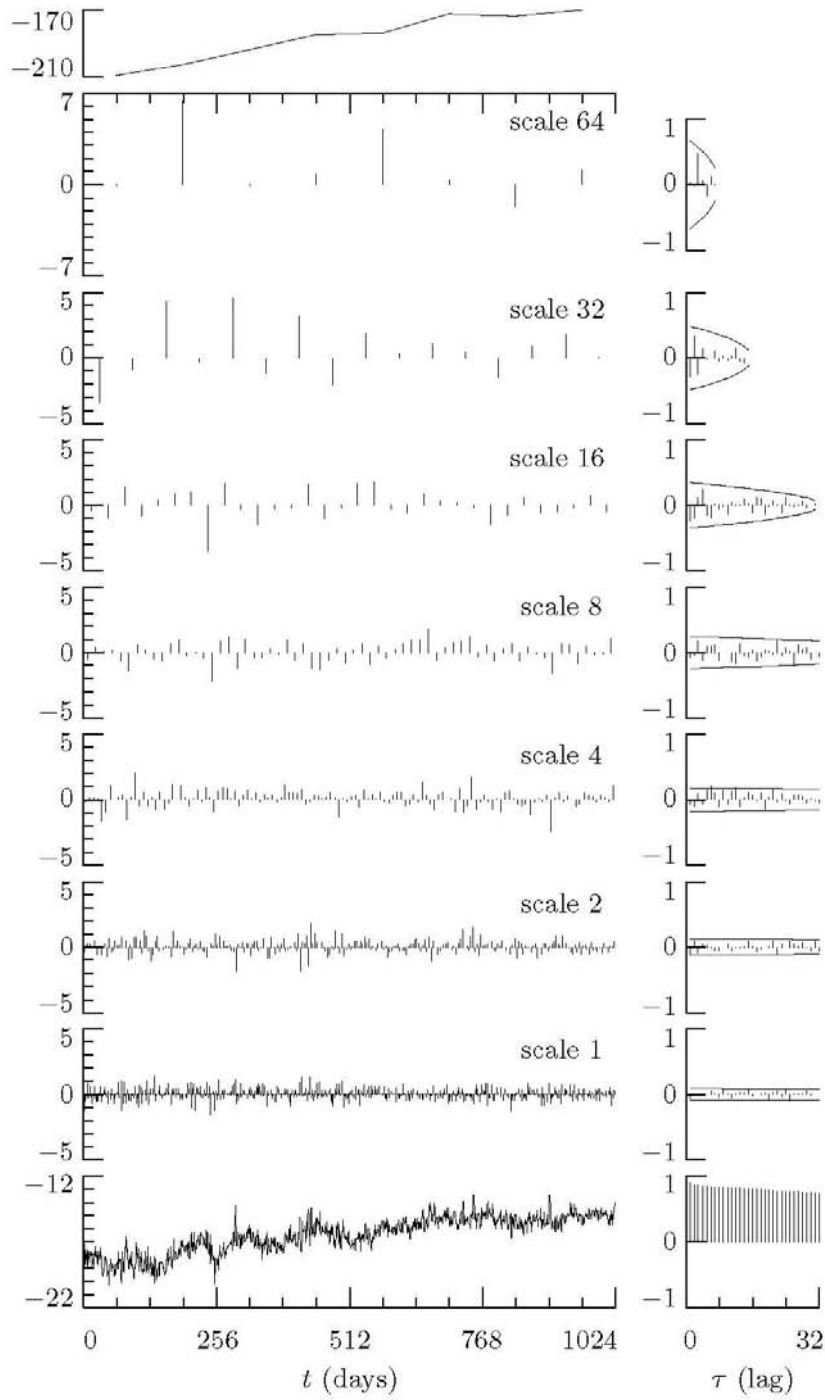


图 5.15: 小波的去相关能力, 源自 [142]

§5.3 纹理的小波表象的描述

对图像进行小波变换后，视觉系统需要提取图像中的纹理区域的统计信息。对统计信息的内容目前还不清楚，这也是纹理分析中的研究点。下面是一些视觉系统中采用的统计信息，可以用来进行图像信息描述或者纹理分析。下面我们使用 D_j, H_j, V_i, L 表示离散小波变换后的各子图，参见图 5.16。

L	H_3	H_2	H_1
V_3	D_3		
V_2		D_2	
V_1			D_1

图 5.16: 图像的离散小波变换得到的子图

§5.3.1 子图内的统计量

子图内的统计量 [145] 包括每个子图的一阶矩和二阶矩。为了方便，这里将所有的子图不作区分，都记作 $Graph$ 。这里假设 $Graph$ 的大小为 $M \times N$ ，并且都已经量化为 $0, 1, \dots, G-1$ 之间的整数值。

子图 $Graph$ 的一阶统计量观察的是子图中取值的分布信息，它们都可以从子图的统计直方图 $\{H(g)\}$ 计算得到，

$$H(g) = \frac{n_g}{N}, g = 0, 1, \dots, G-1 \quad (5.33)$$

表 5.2: 小波系数的一阶统计量

名称	计算方法
均值 (Mean)	$P_1 = \sum_{g=0}^{G-1} gH(g)$
标准偏差 (standard)	$P_2 = \sqrt{\sum_{g=0}^{G-1} (g - P_1)^2 H(g)}$
Skew	$P_3 = \sum_{g=0}^{G-1} (g - P_1)^3 H(g)$
能量 (Energy)	$P_4 = \sum_{g=0}^{G-1} H^2(g)$
熵 (Entropy)	$P_5 = -\sum_{g=0}^{G-1} H(g) \log_2(H(g))$
一阶矩 (1st Moment)	$P_6 = P_1$
二阶矩 (2nd Moment)	$P_7 = \sum_{g=0}^{G-1} (g - P_6)^2 H(g)$
三阶矩 (3rd Moment)	$P_8 = \sum_{g=0}^{G-1} (g - P_6)^3 H(g)$
四阶矩 (4th Moment)	$P_9 = \sum_{g=0}^{G-1} (g - P_6)^4 H(g)$

表 5.3: 小波系数的二阶统计量

名称	计算方法
能量 (Energy)	$F_1 = \sum_{g=0}^{G-1} \sum_{g=0}^{G-1} C^2(i, j)$
熵 (Entropy)	$F_2 = -\sum_{g=0}^{G-1} \sum_{g=0}^{G-1} C(i, j) \log_2(C(i, j))$
对比度 (contrast)	$F_3 = \sum_{g=0}^{G-1} \sum_{g=0}^{G-1} (i - j)^2 C(i, j)$
Homogeneity	$F_4 = \sum_{g=0}^{G-1} \sum_{g=0}^{G-1} \frac{C(i, j)}{1 + (i - j)^2}$

常用的一阶统计量及其计算方法见表 5.3。

子图 *Graph* 的二阶统计量观察的是子图相邻数据的关系。首先定义共生函数 (cooccurrence function) $C(i, j)$,

$$C(i, j) = \arg\{C_{\vec{d}_k}(i, j); \vec{d}_k \in \Delta\} \quad (5.34)$$

其中,

$$\begin{aligned} C_{\vec{d}}(i, j) &= C_{+d}(i, j) + C_{-d}(i, j) \\ C_{+d}(i, j) &= \frac{|\{(r, s), (t, v); (r, s) = i, (t, v) = j, (t, v) = (r + dx, s + dy)\}|}{(M - dx)(N - dy)} \end{aligned} \quad (5.35)$$

以上均假设 $(r, s), (t, v) \in M \times N$ 。

则常用的二阶统计量及其计算方法见表 5.3。

§5.3.2 其它统计量

除了子图内的统计量外, 还应该涉及同一尺度不同子图间的统计量和不同尺度的子图间的统计量。但是这个方面现在还很少有人研究, 只有 Hany Farid 和 Siwei Lyu 在文献 [146] 给出了一组涉及尺度间子图的统计量。

对竖直子块 $\{V_i\}$, 有下面的预测公式

$$\begin{aligned} |V_i(x, y)| &= w_1|V_i(x-1, y)| + w_2|V_i(x+1, y)| + w_3|V_i(x, y-1)| \\ &\quad + w_4|V_i(x, y+1)| + w_5|V_{i+1}(x/2, y/2)| + w_6|D_i(x, y)| \\ &\quad + w_7|D_{i+1}(x/2, y/2)| \end{aligned} \quad (5.36)$$

这是一个线形的预测公式, 未知的只有 $\vec{w} = (w_1, w_2, \dots, w_7)^T$, 可以写成

$$\vec{V} = Q\vec{w} \quad (5.37)$$

不难知道, $\vec{w} = (Q^T Q)^{-1} Q^T \vec{V}$ 。令

$$\vec{E} = \log_2(\vec{V}) - \log_2(|Q\vec{w}|) \quad (5.38)$$

则 $\vec{E}$ 的均值 (Mean)、方差 (variance)、斜度 (skewness) 和峰度 (kurtosis) 就是四个统计量。

对水平子块 $\{H_i\}$, 有下面的预测公式

$$\begin{aligned} |H_i(x, y)| &= w_1|H_i(x-1, y)| + w_2|H_i(x+1, y)| + w_3|H_i(x, y-1)| \\ &\quad + w_4|H_i(x, y+1)| + w_5|H_{i+1}(x/2, y/2)| + w_6|D_i(x, y)| \\ &\quad + w_7|D_{i+1}(x/2, y/2)| \end{aligned} \quad (5.39)$$

按照同样的方法, 可以得到四个统计量。

对对角子块 $\{D_i\}$, 有下面的预测公式

$$\begin{aligned} |D_i(x, y)| &= w_1|D_i(x-1, y)| + w_2|D_i(x+1, y)| + w_3|D_i(x, y-1)| \\ &\quad + w_4|D_i(x, y+1)| + w_5|D_{i+1}(x/2, y/2)| + w_6|H_i(x, y)| \\ &\quad + w_7|V_{i+1}(x, y)| \end{aligned} \quad (5.40)$$

按照同样的方法, 可以得到四个统计量。

第六章 边缘检测和纹理特征提取的统一

根据前面两章的结论，视觉系统可以使用特定的多尺度零交叉点表象形状，使用小波变换后的系数来表象纹理。实际上他们可以统一在小波分析的框架下。主要基于下面两个原因：首先，离散小波变换本身来源于视觉的多尺度模型，所以小波变换本身提供了图像的多尺度描述 (§5.2.1.2)；第二，使用小波分解可以进行边缘检测，更重要的是，常见的用于零交叉点检测的线性滤波器本质上都可以看作广义的小波函数 (§6.1)。

§6.1 边缘检测的微分滤波器一般是小波函数

使用小波进行边缘检测的方法可以大致分为两类，一类是基于连续小波变换 (CWT)；第二类是基于离散小波变换 (DWT)。研究使用小波分析进行边缘检测的两篇硕士论文见文献 [147, 148]。但是，目前看到的绝大多数算法都是使用连续小波变换 [132, 133]。

称 $\theta(x)$ 为磨光函数，且二阶可导。如果

$$\int_{-\infty}^{+\infty} \theta(x) dx = 1 \quad (6.1)$$

$$\lim_{x \rightarrow \pm\infty} \theta(x) = 0 \quad (6.2)$$

设 $\theta(x)$ 二阶可导，定义 $\psi^{(1)}(x) \triangleq \frac{d}{dx}\theta(x)$, $\psi^{(2)}(x) \triangleq \frac{d^2}{dx^2}\theta(x)$ ，则

$$\int_{-\infty}^{+\infty} \psi^{(1)}(x) dx = 0 \quad (6.3)$$

$$\int_{-\infty}^{+\infty} \psi^{(2)}(x) dx = 0 \quad (6.4)$$

所以， $\psi^{(1)}(x)$ 和 $\psi^{(2)}(x)$ 都是小波函数。假设对函数 ξ_x 的尺度变换记为 $\xi_s(x) = \frac{1}{s}\xi(\frac{x}{s})$ ，则 $\psi_s^{(1)}(x)$ 和 $\psi_s^{(2)}(x)$ 为两个尺度为 s 的小波滤波器，相应的尺度为 s 的连续小波变换定义为

$$CWT_s^{(1)}f(x) = \psi^{(1)}(x) * f(x) = \int_{-\infty}^{+\infty} \psi^{(1)}(t)f(x-t)dt \quad (6.5)$$

$$CWT_s^{(2)}f(x) = \psi^{(2)}(x) * f(x) = \int_{-\infty}^{+\infty} \psi^{(2)}(t)f(x-t)dt \quad (6.6)$$

利用连续小波变换进行边缘检测，就是进行连续小波变换后，通过检测 $CWT_s^{(1)}f(x)$

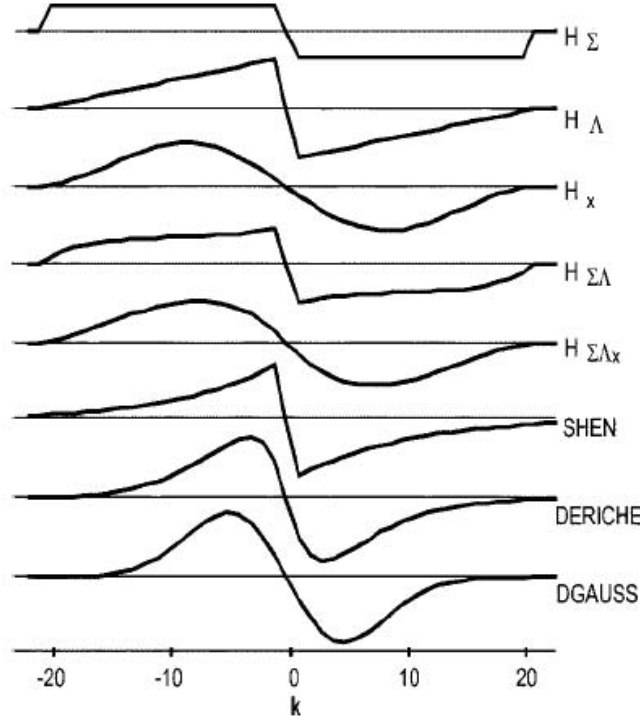


图 6.1: 常见的边缘检测滤波器, 源自 [92]

的局部极值或者 $CWT_s^{(2)}f(x)$ 的零交叉点对边缘位置进行定位。

实际上, 由于

$$CWT_s^{(1)}f(x) = f(x) * (s \frac{d\theta_s}{dx})(x) = s \frac{d}{dx}(f * \theta_s)(x) \quad (6.7)$$

$$CWT_s^{(2)}f(x) = f(x) * (s^2 \frac{d^2\theta_s}{dx^2})(x) = s^2 \frac{d^2}{dx^2}(f * \theta_s)(x) \quad (6.8)$$

所以, 上述基于连续小波变换的边缘检测算法等价于这样的过程: 首先使用尺度为 s 的磨光函数进行图像平滑, 然后计算一阶微分和二阶微分; 最后检测局部极值或者零交叉点。事实上, 文献 [107] 中大多数边缘检测算法都是这样的同样的过程, 一般选用高斯滤波器平滑图像以抑制高频噪声, 不同的只是如何估计去噪后图像的梯度。特别地, 在上述算法中, 磨光函数常常取高斯函数或者规范 B 样条函数 (规范 B 样条可以看作是高斯函数的近似 [132])。此时, 基于连续小波变换的边缘检测算法等价于 Canny 算法和 LoG 算法。在这个意义上说, 基于连续小波变换的边缘检测算法只是对已经发现的边缘检测算法在小波框架下的理解。

实际上, 常见的用于零交叉点检测的线性滤波器本质上都可以看作广义的小波函数。图 6.1 给出了常见的一些滤波器的函数图像, 这些函数都是反对称的, 所以在 $\mathbb{R}$ 上的积分必然是零, 所以它们都是小波函数。

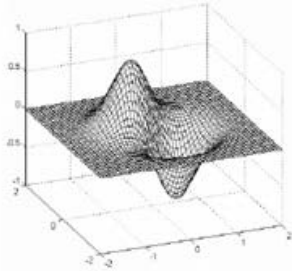
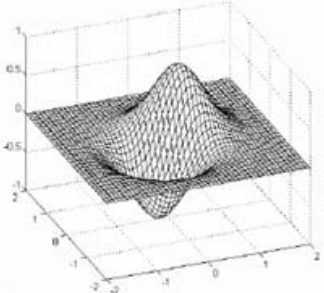
名称	模板	对应小波函数	小波函数的图像
Robert	$s = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$	$(y - x)e^{-\frac{x^2+y^2}{2}}$	
	$t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}$	$(x + y)e^{-\frac{x^2+y^2}{2}}$	

图 6.2: Roberts 模板在小波框架下的理解, 源自 [147]

文献 [107] 中还有一些离散模板形式给出的边缘检测滤波器: Robert 算子、Sober 算子、Prewitt 算子和 Frei-Chen 算子。LiJun [147] 中也给出了相应的小波函数, 这些离散模板可以看作对应的小波函数的抽样。图 6.2 给出了 Robert 算子的小波函数逼近, 图 6.3 给出了 Sober 算子的小波函数逼近, 6.4 给出了 Prewitt 算子的小波函数逼近。Frei-Chen 算子的小波逼近见 Lijun 的论文。

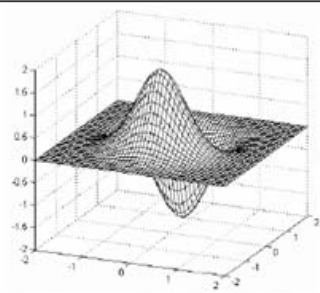
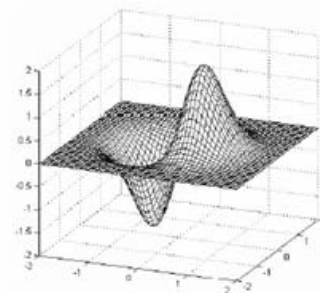
名称	模板	对应小波函数	小波函数的图像
Sober	$s = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$	$-xe^{-\frac{x^2+y^2}{2}}$	
	$t = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$	$-ye^{-\frac{x^2+y^2}{2}}$	

图 6.3: Sobel 模板在小波框架下的理解, 源自 [147]

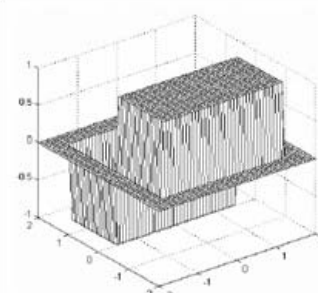
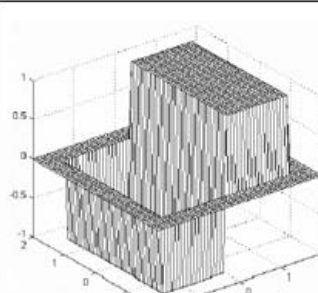
名称	模板	对应小波函数	小波函数的图像
Prewitt	$s = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$	$\begin{cases} -1, & -1.5 \leq x < 0; -1.5 \leq y \leq 1.5, \\ 1, & 0 \leq x < 1.5; -1.5 \leq y \leq 1.5, \\ 0, & \text{otherwise} \end{cases}$	
	$t = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$	$\begin{cases} -1, & -1.5 \leq y < 0; -1.5 \leq x \leq 1.5 \\ 1, & 0 \leq y < 1.5; -1.5 \leq x \leq 1.5 \\ 0, & \text{otherwise} \end{cases}$	

图 6.4: Prewitt 模板在小波框架下的理解, 源自 [147]

§6.2 形状和纹理的提取统一为小波分析

根据前面的结论，小波变换下多尺度零交叉点可以作为形状的表象，而图像的小波变换系数可以作为纹理的表象。形状表象和纹理表象都可以在小波分析的框架下给出。于是，我们得到了如图 6.5 所示的视觉系统表象的统一模型。

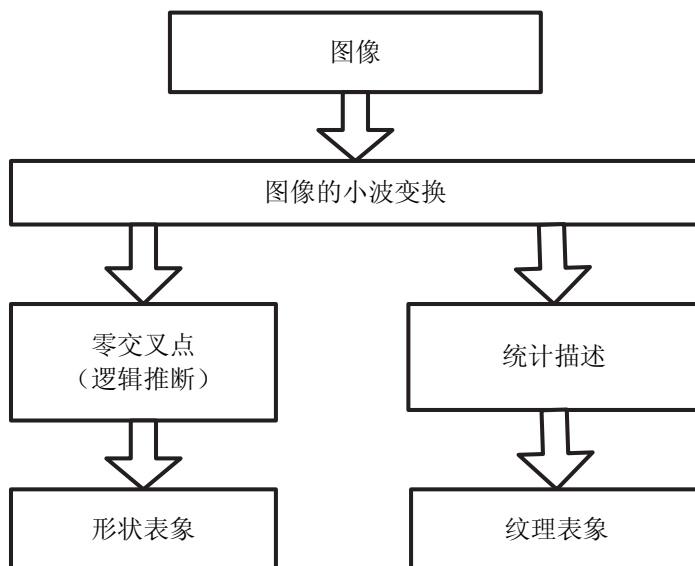


图 6.5: 获取形状表象和纹理表象的统一

对这个模型的解释如下：

1. 首先进行小波变换，小波变换系数提供了图像的多尺度描述；然后视觉系统在合适的尺度中提取形状和纹理描述。

2. 对小波系数的逻辑推断以提取具有一定强度的零交叉点。首先，零交叉点具有物理意义，视觉系统希望获得外界的信息，就必须分析和研究这些零交叉点。零交叉点的强度越大，它是偶然噪声的可能性就越小，对应的物理过程越重要，所以视觉系统越感兴趣。在这个过程中，视觉系统必须结合先验知识，对小波变换系数进行逻辑推断，以提取这些零交叉点，它们可以表象图像中的形状信息。如果尺度相对比较大，物体边界对应的零交叉点就变的模糊，此时对小波系数的逻辑分析也无法准确的零交叉点。此时，感知系统认为这是相对平滑的区域。

3. 但是，视觉系统的形状感知能力不是无限的。如果图像中的纹理丰富，那么零交叉点就非常多，全部进行详尽的形状描述是视觉系统不允许。此时，视觉系统使用统计方式描述这些相对微弱零交叉点所在区域（也就是一般说的纹理区域）的特征，作为区域的描述。因为图像的小波变换具有十分好的性质 (§5.2)，所以这种统

计描述仍然抓住了区域的某些特征,可以用来区分不同的纹理模式。

这个模型说明,纹理是视觉系统的能力与图像复杂性之间的平衡的产物。本质上,利用零交叉点可以决定(重构)任何纹理模式,纹理只是零交叉点的统计描述。§2.4.2 中, MPEG7 使用边缘空间分布来描述直方图,体现的就是这种思想。

这个模型可以解释很多现象:因为对纹理区域的描述只是统计性的,所以无法视觉系统无法还原纹理,所以尽管能感知到两种不同的纹理(统计量不同)但是我们无法说清到底两种纹理的不同。

§6.2.1 图像区域的复杂性度量

在形状表象和纹理表象的统一模型中,还存在一个重要的问题,即视觉系统如何判断图像区域在特定尺度下的复杂性。在特定尺度下,如果某一区域复杂,就需要使用纹理表象;如果区域简单,就使用形状表象。如果把复杂性理解为视觉信息的多少,则这个问题就是图像多尺度空间中的信息量估计问题。

Sporring [96, 97] 提出了尺度空间的信息熵概念;后来又和 Weickert 合作引入了信息阶的概念,将信息熵定义一般化。这些概念都是直接从二维离散情形下得到然后推广到连续域中。

定义6.1. 二维离散图像记作 $P = (P_1, P_2, \dots, P_N)$, 这里用简单的一维指标表示二维像素。相应的尺度空间记作 $P(t) = (P_1(t), P_2(t), \dots, P_N(t))$, 其中 t 是尺度参数。

定义6.2. t 尺度下, 图像的随机分布函数 $p(t) = \{p(n; t)\}$ 定义为 $p(n; t) := \frac{P_n(t)}{\sum_n P_n(t)}$ 。

定义6.3. t 尺度下, 图像的 α 阶信息熵 $S_{\alpha, t}(p)$ 定义为

$$S_{\alpha, t}(p) = \frac{1}{1 - \alpha} \log \sum_{i=1}^N (p_i(t))^\alpha, \alpha \neq 1 \quad (6.9)$$

$$S_{1, t}(p) = - \sum_{i=1}^N p_i \log p_i \quad (6.10)$$

$S_{1, t}(p)$ 就是 Shannon-Wiener 熵。

下面先给出一个定义。

定义6.4. 尺度空间 $f(x, t)$ 的信息量记作 $I_t, I_t = I[f(x, t)], t \geq 0$ 。

王郑耀等 [100] 从视觉直观出发, 认为 I_t 应该有以下的性质:

- A. 非负性: 尺度空间中图像的信息量是正值泛函。即 $\forall t \geq 0$, 有 $I_t \geq 0$;
- B. 因果性: 根据尺度空间的因果性理论, 小尺度是大尺度的平滑和简化, 所以尺度空间中图像的信息量随着尺度的增加是减少的: 如果 $t_1 > t_2$, 则 $I_{t_1} \leq I_{t_2}$;
- C. 灰度反转不变性: 灰度的反转并不改变信息。这个是直观的, 在 256 色灰度图像中, 如果将灰度值变为 255 减去原来的灰度, 对视觉来说图像的内容没有发生改变。用公式表达为 $I[f(x, t)] = I[M - f(x, t)]$;

D. 局域可加性：整体信息是局部信息之和。公式表达为如果 $\Omega = \Omega_1 \cup \Omega_2$, $\Omega, \Omega_1, \Omega_2$ 均为开集, 并且 f_Ω 表示 $f(x; t)$, $x \in \Omega$ 则 $I_t[f_\Omega] = I_t[f_{\Omega_1}] + I_t[f_{\Omega_2}] - I_t[f_{\Omega_1 \cap \Omega_2}]$

E. 形态不变性：图像的灰度的线性变换不改变信息的度量。假设 $p(\cdot) = \cdot + c$ (c 为常数) 或者 $p(\cdot) = k \cdot$ (k 为常数), 则 $I[p[f]] = I[f]$;

F. 几何不变性：图像信息的度量与图像的坐标系没有关系。

G. 规范性：灰度一致图像 $f(x; t) \equiv \text{常数}$ 的信息度量为零。

通过定义和 [96] 中的引理一, 二, 三可以知道因果性是满足的。容易知道一般的说区域可加性、灰度反转不变性、形态不变性是不满足的。这种尺度空间的信息度量方法有很大的局限性。

假设：图像中极值点和拐点的个数是有限的。因为给定一幅图像, 它总是在一个特定的尺度下得到的, 比这个尺度小的细节是不可能获得的, 所以这个假设是合理的。

下面对固定的尺度 t 下考虑信息量的定义。以下设 $f(x) = f(x, t)$ 。

定义6.5. 一维图像 $f(x)$ 的第一、二类特征点的集合 $E_1(f)$ 、 $E_2(f)$ 分别定义为

$$\begin{aligned} E_1(f) = & \{x \in \Omega : f'(x) = 0; \exists \varepsilon > 0, \\ & st. \forall y \in ((x - \varepsilon, x) \cup (x, x + \varepsilon)) \cap \Omega, f'(y) \neq 0\} \end{aligned} \quad (6.11)$$

$$\begin{aligned} E_2(f) = & \{x \in \Omega : f''(x) = 0; \exists \varepsilon > 0, \\ & st. \forall y \in ((x - \varepsilon, x) \cup (x, x + \varepsilon)) \cap \Omega, f''(y) \neq 0\} \end{aligned} \quad (6.12)$$

定义6.6. 一维图像 $f(x)$ 的特征点的集合 $E(f)$ 定义为 $E(f) = E_1(f) \cup E_2(f)$ 。

定义6.7. 一维图像 $f(x)$ 的信息量 $I_t(f)$ 定义为 $I_t(f) = |E(f(x))|$, 其中, $|\cdot|$ 表示集合中元素的数量。

下面本文就以高斯尺度空间中证明这一信息度量的合理性。图像 $f(x)$ 的高斯尺度空间为 $f(x, t)$, 其中 $f(x, t) = S_t(f(x)) = f(x) * g(x, t) = (2\pi t^2)^{-\frac{1}{2}} \cdot f(x) * \exp(-\frac{x^2}{2t^2})$ 。于是, $f(x, t)$ 构成了半平面 $\Theta = \{(x, t) \mid x \in (-\infty, +\infty), t \in (0, +\infty)\}$ 上的曲面。

引理6.1. 高斯尺度空间 $f(x, t)$ 如上定义, 则 $(\frac{\partial}{\partial t} - t \frac{\partial^2}{\partial x^2})f(x, t) = 0$ 于是, 进一步有 $(\frac{\partial}{\partial t} - t \frac{\partial^2}{\partial x^2}) \frac{\partial^n f}{\partial x^n}(x, t) = 0, n = 1, 2, 3, \dots$ 。

引理 1 是 [149] 中定理一的推广, 利用卷积的性质直接就可以证明。

定义6.8. $f(x, t)$ 的零交叉点是满足 $f(x, t) = 0$ 的点的集合, 称为 $f(x, t)$ 的零阶零交叉点; $f(x, t)$ 的 $n = 1, 2, 3, \dots$ 阶零交叉点是满足 $\frac{\partial^n f}{\partial x^n}(x, t) = 0$ 的集合。

定理 6.2. $f(x, t)$ 的任意 n 阶零交叉点 ($n = 0, 1, 2, \dots$) 随着尺度 t 的增加可能会减少, 但是绝对不会增加。注：本定理是 [149] 上定理 2 的推广。

证明：以下记 $f(x, t) := \frac{\partial^0 f}{\partial x^0}(x, t)$; $f^{(n)}(x, t) := \frac{\partial^n f}{\partial x^n}(x, t)$;

对任意 $n = 1, 2, 3, \dots$, 方程

$$\frac{\partial^n f}{\partial x^n}(x, t) = 0 \quad (6.13)$$

决定了 $f(x, t)$ 的零交叉点曲线 $t_n(x)$ 。把 t 看作是 x 的函数, 对 (6.13) 两边关于 x 求导数, 得

$$0 = f_x^{(n)}(x, t) + f_t^{(n)} t_n'(x) \quad (6.14)$$

$$\begin{aligned} 0 = & f_{xx}^{(n)}(x, t) + f_{xt}^{(n)}(x, t) t_n'(x) + [f_{tx}^{(n)}(x, t) t_n'(x) + f_{tt}^{(n)}(x, t) (t_n'(x))^2] t_n'(x) \\ & + f_t^{(n)}(x, t) t_n''(x) \end{aligned} \quad (6.15)$$

曲线 $t_n(x)$ 的极值点满足 $t_n'(x) = 0$, 在这些点上

$$\begin{aligned} 0 = & f_{xx}^{(n)}(x, t) + f_{xt}^{(n)}(x, t) t_n'(x) + [f_{tx}^{(n)}(x, t) t_n'(x) + f_{tt}^{(n)}(x, t) (t_n'(x))^2] t_n'(x) \\ & + f_t^{(n)}(x, t) t_n''(x) \end{aligned} \quad (6.16)$$

$$= f_{xx}^{(n)}(x, t) + f_t^{(n)}(x, t) t_n''(x) \quad (6.17)$$

从而有 $t_n''(x) = -\frac{f_{xx}^{(n)}(x, t)}{f_t^{(n)}(x, t)}$ 由引理 1, $t_n''(x) = -\frac{1}{t} < 0$, 也就是说 $t_n(x)$ 的极值点总是上凸的, 所以定理成立。□

A.因果性 由定理 6.2, 图像 $f(x)$ 的第一类特征点和第二类特征点的数目随着 t 的增加可能是减少的, 但是一定不增加。所以因果性成立。

B.形态不变性 令 $\tilde{f}(x, t) = k \cdot f(x, t)$, 则 $\tilde{f}_x(x, t) = k \cdot f_x(x, t)$ $\tilde{f}_{xx}(x, t) = k \cdot f_{xx}(x, t)$ 因此, 在灰度的线性变换下, 图像的信息量是不变的。

C. 由于本文采用微分的零点定义了信息度量, 所以区域可加性、灰度反转不变性、灰度为常数的图像信息为零等是显然的。

真正实用的信息量度量应该是二维离散情形下定义的, 下面将前述理论直接推广到二维离散情形。

定义6.9. 二维离散图像记作 $P(m, n) : 1 \leq m \leq M, 1 \leq n \leq N$ 。

任意 $1 < m_0 < M, 1 < n_0 < N$, 将图像在 (m_0, n_0) 点处的可以看作如图 6.6 所示的 4 个方向。在每一个方向上都可以将图像看作是一维的图像 (以下称为 (m_0, n_0) 像素的四条邻接边), 则可以利用一维图像信息量的度量来定义二维离散图像的信息量。定义二维离散图像的第一类特征点为四条邻接边上图像均为第一类特征点的像素点; 定义二维离散图像的第二类特征点为四条邻接边上图像均为第二类特征点的像素点。即

$$\begin{aligned} E_1(P) = & \{(m, n) : 1 < m < M, 1 < n < N; (P_{m,n} - P_{m-1,n})(P_{m,n} - P_{m+1,n}) > 0; \\ & (P_{m,n} - P_{m,n-1})(P_{m,n} - P_{m,n+1}) > 0; (P_{m,n} - P_{m-1,n-1})(P_{m,n} - P_{m+1,n+1}) > 0; \\ & (P_{m,n} - P_{m+1,n-1})(P_{m,n} - P_{m-1,n+1}) > 0\} \end{aligned}$$

在一维时第二类特征点也就是一次导数取得极值的点, 为了求得第二类极值点, 就要先计算各个方向的一次离散导数。本文使用中心差分来计算一次导数。即

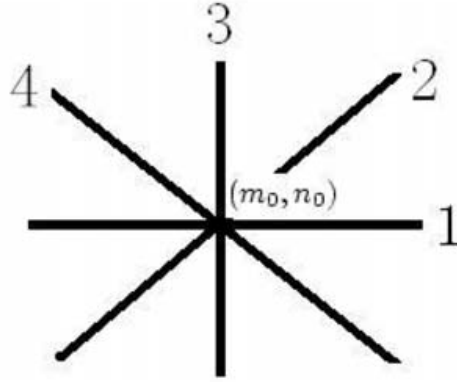


图 6.6: 在一点处图像像素的结构

(m, n) 像素在第一条邻接边上的一次导数为 $P(m+1, n) - P(m-1, n)$, 其余方向上一次导数类似的计算。于是

$$\begin{aligned}
 E_2(P) = \{ & (m, n) : 2 < m < M-1, 2 < n < N-1; \\
 & ((P_{m+1,n} - P_{m-1,n}) - (P_{m+2,n} - P_{m,n}))((P_{m+1,n} - P_{m-1,n}) - (P_{m,n} - P_{m-2,n})) > 0; \\
 & ((P_{m,n+1} - P_{m,n-1}) - (P_{m,n+2} - P_{m,n}))((P_{m,n+1} - P_{m,n-1}) - (P_{m,n} - P_{m,n-2})) > 0; \\
 & ((P_{m+1,n+1} - P_{m-1,n-1}) - (P_{m+2,n+2} - P_{m,n})) \\
 & ((P_{m+1,n+1} - P_{m-1,n-1}) - (P_{m,n} - P_{m-2,n-2})) > 0; \\
 & ((P_{m+1,n-1} - P_{m-1,n+1}) - (P_{m+2,n-2} - P_{m,n})) \\
 & ((P_{m+1,n-1} - P_{m-1,n+1}) - (P_{m,n} - P_{m-2,n+2})) > 0 \}
 \end{aligned}$$

于是二维离散图像信息量 $I[P]$ 为 $I[P] = |E_1(P)| + |E_2(P)|$ 。

图 6.7 是一幅纹理图像的多尺度表示以及在各个不同尺度下的视觉信息度量, (a) 为原始图像, (b)-(d) 为从原图像得到的几个尺度下的表示 (高斯多尺度表示, 本文只选取三幅作为代表), 尺度越来越大! (e) 为原图像在各个尺度下信息量, 其中横坐标是尺度参数, 本文使用尺度参数的对数作为横坐标, 纵坐标是视觉信息量大小; (f) 为相邻两尺度的信息的减少。

从图 6.7(e)、(f) 可以看出随着尺度的指数增加, 视觉信息量先是平缓的递减, 然后是剧烈递减, 最后又平缓的递减。在实验中尺度参数是指数增加的, 由于图像中视觉特征的尺度大多集中在一定的范围之内, 一开始虽然尺度也在指数增加, 但是这些尺度仍然要比图像中大部分视觉特征的尺度要小, 所以视觉特征的丢失比较慢; 然后当尺度参数增加到图像中大部分视觉特征所在的尺度范围时, 随着尺度的增加图像的视觉特征开始大幅减少, 于是视觉信息量开始大幅减少; 最后尺度参数增大到大于图像中绝大多数的视觉特征, 此时图像的视觉特征变的非常少, 这时随着尺度的增加图像的视觉特征减少也变慢。

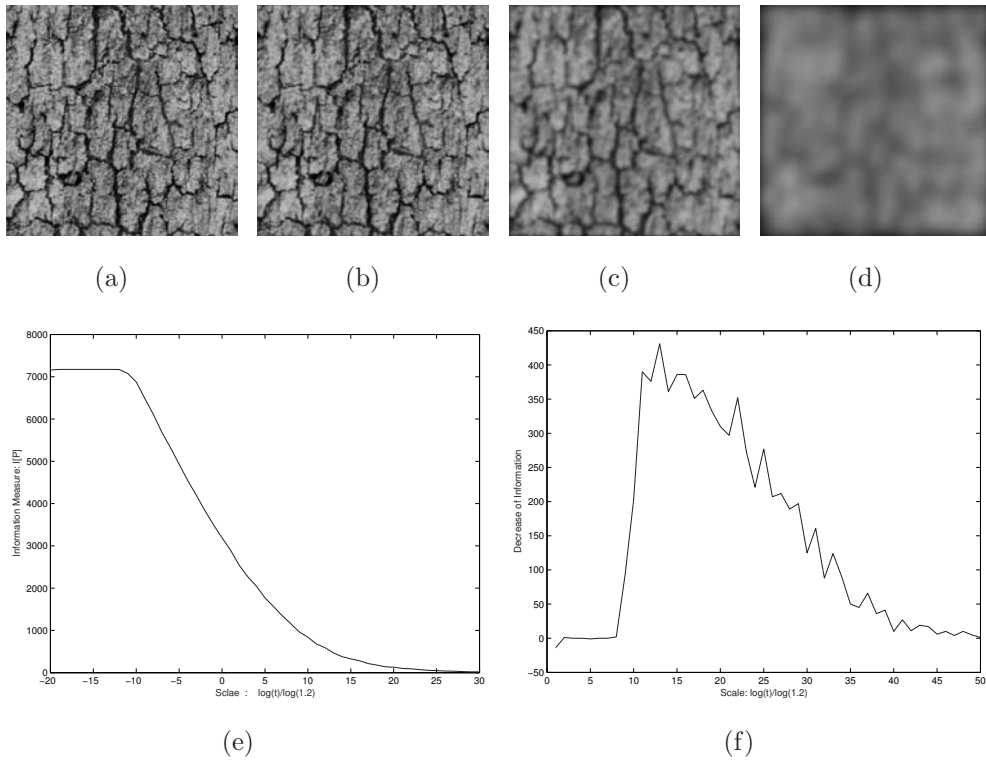


图 6.7: 图像尺度空间中视觉信息度量: 上左一为原始图像, 其余为原图像的多尺度表示的三个代表, 尺度越来越大; 下左一为原图像在各个尺度下的信息量, 其中横坐标是尺度 (对尺度取了对数), 纵坐标是视觉信息量大小; 下右图为相邻两尺度的信息的减少。

理论和实验都已证明, 这种信息量的定义满足上面给出的要求, 因此是合适的图像信息量度量。也就是说, 图像的零交叉点的数量实际上就可以作为图像复杂性的度量。

§6.2.2 视觉系统的 Gabor 模型

1985 年, 中国科学院生物物理研究所、中国科学院视觉信息加工开放实验室的齐翔林, 朱舜山和汪云九等人提出了初级视觉的广义 Gabor 模型, Gabor 函数见式 6.18。随着二十世纪九十年代初小波理论的建立, 他们又于 1996 年提出了初级视觉信息的 Gabor 小波表达 [150, 151, 152], 它们的主要依据是感受野特性和利用 Gabor 小波能重建图像信息。他们所说的初级视觉信息表达和本文中所说的视觉信息表象含义是类似的。因此, 他们的研究表明, Gabor 小波可以作为初级视觉系统的模型, 而 Gabor 小波系数可以作为视觉表象。

$$g(x) = \frac{1}{2\pi\sigma} \exp\left(-\frac{x^2}{2\sigma^2} + jux\right) \quad (6.18)$$

可以认为, 他们的研究解决的是视觉系统的建模问题。本文可以看作这一理论的补充, 说明了为什么视觉系统选用小波作为视觉表象。

实际上, 使用 Gabor 小波进行边缘检测 [154] 和纹理分析 [155, 156, 157, 158, 159] 的算法非常多。尤其是纹理分析中, Gabor 小波已经成为标准算法。MPEG7 (§2.4.2) 中有两个纹理描述子利用使用了 Gabor 小波。

第七章 图像表象的简单应用：基于内容的特定图像检索

设计 §2.5 中一般意义的图像检索引擎是十分有用的，但也是十分复杂的，在现在的技术条件下很难实现。本文做一个相对简单可行的问题：基于上面的图像表象理论，设计从互联网上检索含有特定内容的图像的软件（Software of Special Image Search Based On Famous Search Engin），以下简称 SSIS。SSIS 可以作为表象和算法验证平台，其测试图像集是互联网络上日益增长的海量图像。

§7.1 需求分析

一般而言，含有下面内容的图像是比较感兴趣的：

1. 文字：文字对图像中的内容具有提示作用，而且它本身也是图像内容的一部分：书籍封面上的标题、图像上路牌上的字符或者图像拍摄时间。有些信息，为了欺骗监控系统的需要，也常常通过图像发送：反动垃圾邮件等。

2. 特殊符号：条形码或者二维码，由于每个条码总对应着一些字符，所以条码实际上等价于图像中的字符。另外，矩形、圆形或则椭圆形等也常常作为特殊符号放置在图像中。

3. 人脸：从图像中寻找人脸，可以进一步对人脸进行分析，进行人脸识别，人脸分析；例如，可以从庞大的网络上寻找犯罪人踪迹；搜索自己的照片有没有被侵权使用。

4. 黄色图像检索：网络上充斥着大量的不良信息，黄色图像就是其中之一；如果能自动搜索黄色图像，就可以自动寻找黄色网站，封禁黄色网站的 IP 地址。

5. 特定内容的图像库搜集：网页 Logo、商标、国旗、图章、台标等等；搜集整理数据库进行分析和研究。

如果能自动在网络上检索含有上述内容的图像，就可以给我们的工作和生活带来很多便利。另一方面，为了实现检索上述目前内容的目的，需要用到形状、颜色和纹理表象；而互联网是一个非常庞大并且日渐增长的测试图像库，可以帮助测试视觉表象的有效性和完备性，以设计更合理的图像表象及其描述。

对含有这些内容的图像进行分析和比较的算法研究已经有了很长时间，打下了相当好的基础。OCR 识别技术已经非常实用化；条码图像的识别已经在实际中得到广泛的应用；人脸跟踪和识别也比较成熟，欧姆龙公司在上海新国际展览中心展出过一些相关的技术：自动人脸检测，自动人脸跟踪，根据人脸对知识层次和女性、男性化程度排序等等；色情图像分析技术在网吧管理软件中应用广泛；商标、国



(a) 含有字符和人脸的图像



(b) 书的封面



(c) 含有条形码的图像



(d) 含有商标的图案



(e) 色情图像



(f) 台标

图 7.1: 含有感兴趣内容的图像

旗、图章和台标等的分析比较已经形成了产品。

但是，在没有任何先验知识下，寻找图像中的这些感兴趣内容仍然是十分困难的。例如，传统的 OCR 技术总是假设图像中存在很多字符而且背景相对简单。但是，实际图像中很多情况是：个别汉字出现在复杂背景上，这对文字定位和倾斜校正带来困难。

要实现上述所有这些方面的内容检索是相当困难的，需要长期的工作，主要难点就在于图像内容分析的算法设计十分困难。本文首先搭建 SSIS 的框架，然后从最简单的检索开始，目前已经实现了圆（环）的检索和 QR 码检索，下一步准备设计检索含有条形码和色情内容图像的算法。

§7.2 基于现有搜索引擎的特定内容图像检索方案设计

§7.2.1 基于现有搜索引擎的检索

为了实现上面的需求，首先需要获得网络上所有图像的网络地址（URL），然后逐一下载到本地后进行图像分析。获得互联网上网页或者图像链接的工作一般是由“网络机器人”来实现的。“网络机器人”是网页收集器，负责收集互联网上所有的网页。但是网络机器人的有效运行需要专业的服务器支持才能实现其功能，对我们这个系统来说是很不经济的。本文设计的基于内容的特定图像检索是建立在现有的搜索引擎基础上的。首先使用与特定内容相关的一系列关键字检索优秀的检索引擎（Web 检索或者图像检索），利用搜索引擎返回的结果获得相关的网页，然后在这些网页中寻找感兴趣的图像。

这种方法的优点有两个：首先，成本低。第二，速度快。由于现在的基于文字的网页检索和文字检索引擎已经发展的相对成熟，可靠性高。如果选好合适的一组关键字，能找到绝大多数与感兴趣内容相关的网页，遗漏的网页相对很少，搜索引擎的作用相当于粗过滤。由于图像分析过程相对比较慢，使用搜索引擎做粗过滤可以大大加快检索的进程。

另一方面，这种方法也是可靠的。随着互联网的发展，互联网上的搜索引擎已经成为必不可少的信息获取工具，很少有互联网上找不到的网页。如果有，这种网页也是没有太多意义的。

§7.2.2 SSIS 中使用的搜索引擎

SSIS 中使用使用的搜索引擎包括：GoogleImage, GoogleWeb, BaiduImage, BaiduWeb 和 AltaVistaImage。

Google 是目前最强大、使用最广泛的搜索引擎，提供了很多与搜索有关的免费服务，网页检索（GoogleWeb）和图像检索（GooleImage）是 Google 的其中两个服

务。Google 对英文关键字的搜索结果最全，避免遗漏很多图像。首选 Google 的另一个重要原因是 Google 已经尝试推出了 Google Web API 服务 [160]。利用 Google Web API 服务，软件开发人员可以通过他们自己的程序直接查询超过 30 亿的文档，从而更好的进行图像检索。但是，现在的 Google Web API 还处于测试阶段，对使用者的限制太多，因此目前还没有在 SSIS 中使用。一旦 Google 完全开放这项服务，软件可以立刻升级。

Baidu 是最强大的中文搜索引擎，提供了大量的免费服务，包括网页检索（BaiduWeb）和图像检索（BaiduImage）。Baidu 搜索引擎适合检索一些中文关键字。另外，与 Google 相比，Baidu 搜索得到的图像比较少，而且结果中与主题无关的图像太多；但是，Baidu 搜索的结果一般都是国内网站中的图像，下载速度快。

AltaVista 是一个专业的图像检索引擎（AltaVistaImage），它对图像已经做了初步的内容分析：不仅按照图像的颜色数将图像分为两类：彩色图像（Color）和黑白图像（Black and White），而且能自动区分照片（Photo）、图形（Graphics）和按钮/标题（Buttons/Banners）。但是，还是缺乏深入的图像内容的分析。

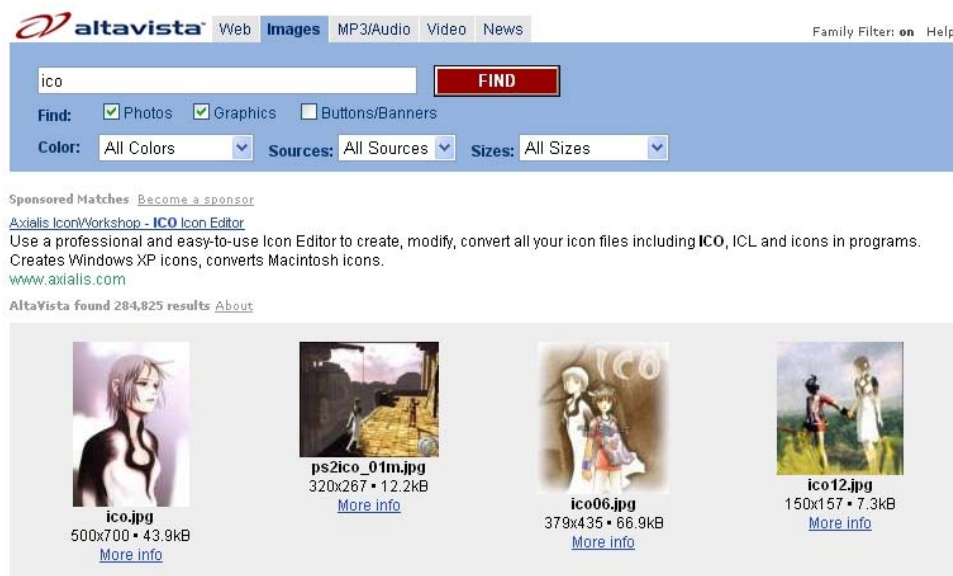


图 7.2: AltaVista 搜索引擎界面

另外，在网络上重复下载图像十分缓慢，很难做实验。因此，进行实验时可以先收集一些图像放在本机上某个文件夹下，然后通过对文件夹下图像的遍历实现图像检索。这种模式称为 LocalMachine 模式。

§7.2.3 总体方案设计图

根据 Web 方式检索搜索引擎返回的是网页，需要进一步对这些网页进行解析

以获得图像链接，而使用 Image 方式检索搜索引擎返回的是直接就是图像。后者比前者更简单。如图 5.2.1.2.1 和图 7.4 所示，对这两种情况分别设计了程序实现的流程。

其中，特定内容的图像检索部分作为它们共有的部分，也是最核心的部分，实现具体如图 7.5 所示。

§7.3 SSIS 中各模块实现的简单说明

SSIS 使用 VC++6.0 编写程序。基于这六种搜索引擎检索图像的过程是类似的，下面利用 GoogleImage 模式说明实现过程。

§7.3.1 GoogleImage 关键字检索网页的获取

通过分析，可以通过链接 *Url* 直接访问 Google 图像搜索引擎

$$Url = str1 + KeyWords + str2 + Number + str3 \quad (7.1)$$

其中，

$$str1 = \text{http://images.google.com/images?q=} \quad (7.2)$$

$$str2 = \&svnum=10\&hl=en\&lr=\&newwindow=1\&start= \quad (7.3)$$

$$str3 = \&sa=N \quad (7.4)$$

KeyWords 表示要检索的关键字，*Number* 表示返回的网页中是从 Google 数据库中第 *Number* 个图像开始显示图像的。按照现在 Google 图像检索引擎的规则，访问这个 *Url* 返回的网页中最多包含二十个图像（共五行，每行四个）。因此，为了遍历当前关键字下的所有图像，*Number* 取 $20 \times i, i = 0, 1, 2, \dots$ 直到 Google 图像检索搜索引擎返回的网页中没有图像为止。

§7.3.2 网页分析

通过 HTTP 协议访问网址 *Url*，得到网址对应的网页内容，保存在字符串 *Html* 中。网页内容是用 HTML 语言描述的，所以必须按照 HTML 语言的规则解析字符串 *Html*。对 Google 图像检索返回的网页分析后，发现网页中图像链接都是以 *StrStart* 开始，以 *StrEnd* 结束的。因此，使用字符串匹配完成图像链接的寻找。

$$StrStart = <ahref = imgres?imgurl = \quad (7.5)$$

$$StrEnd = \&imgrefurl = \quad (7.6)$$

假设找到的每一个图像链接记作 *ImageUrl*。

§7.3.3 服务器和图像链接的数据库管理

获得图像链接 *ImageUrl* 后，就要访问 *ImageUrl* 以获取图像数据。为了避

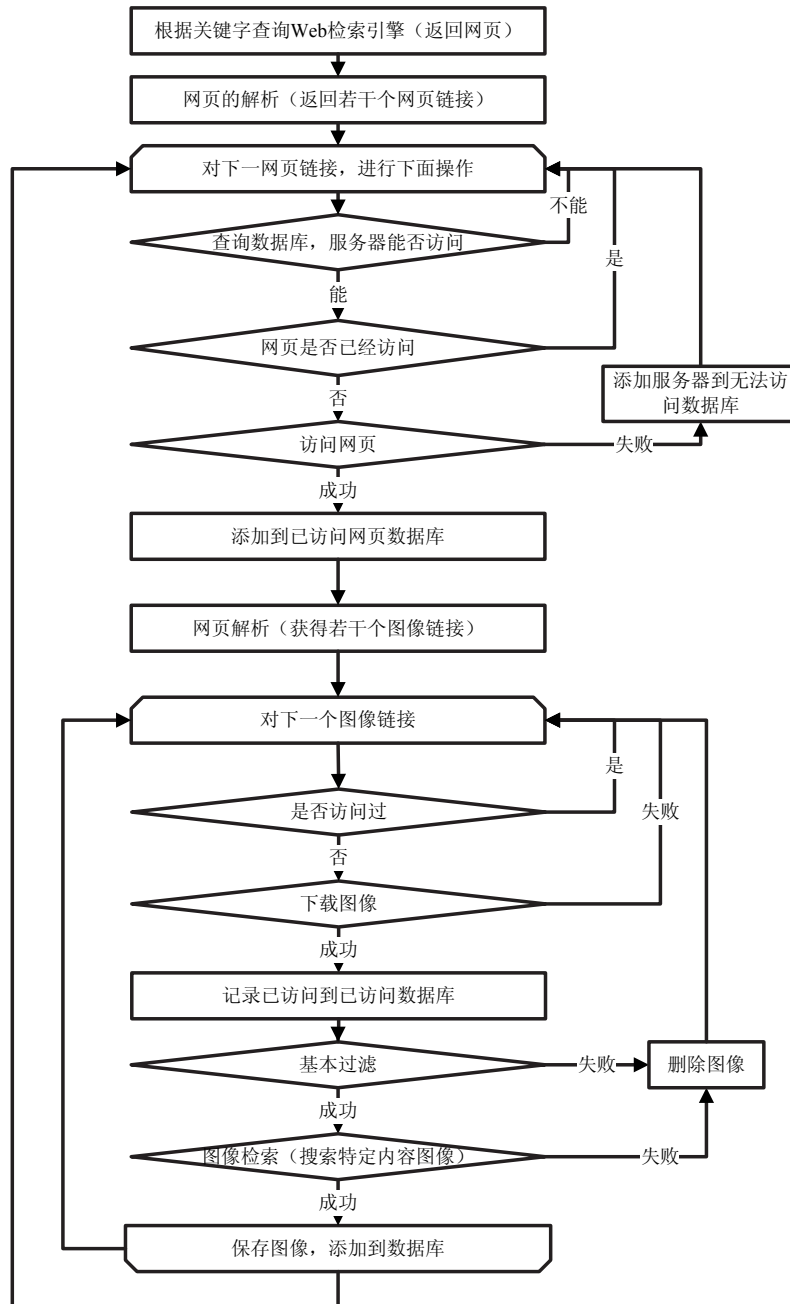


图 7.3: SSIS 的总体方案合计: 基于 Web 检索

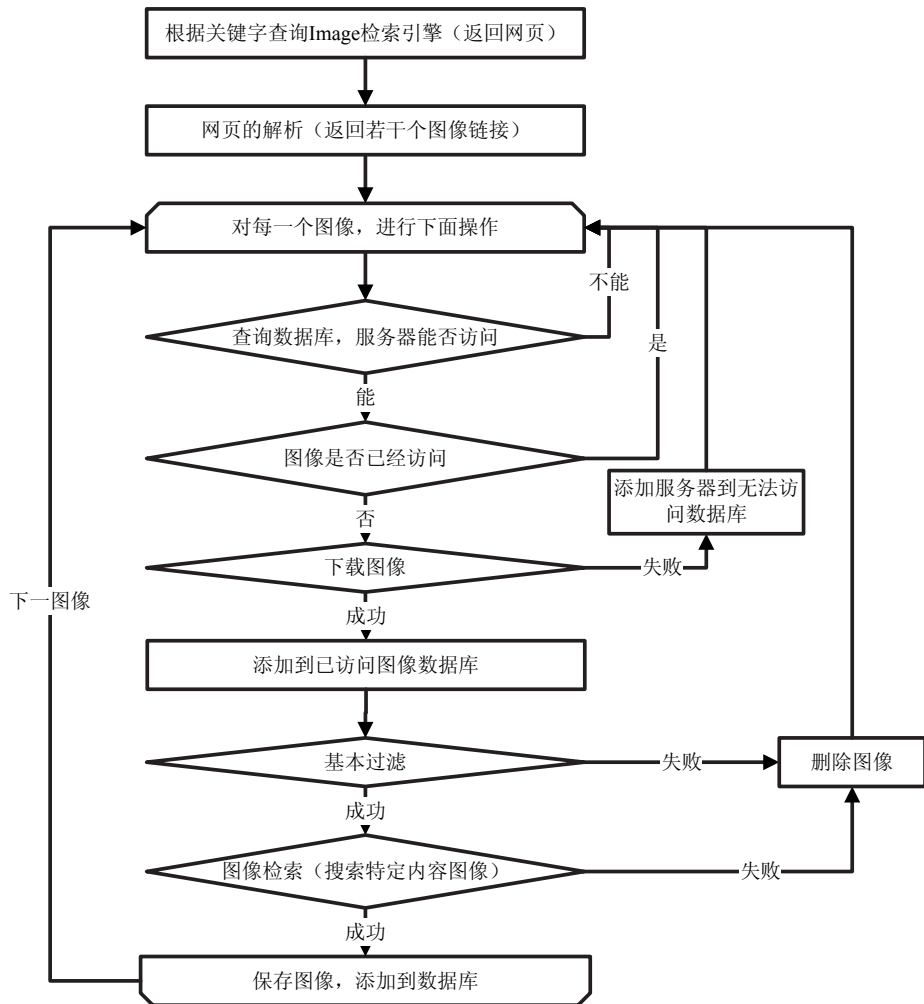


图 7.4: SSIS 的总体方案合计: 基于 Image 检索

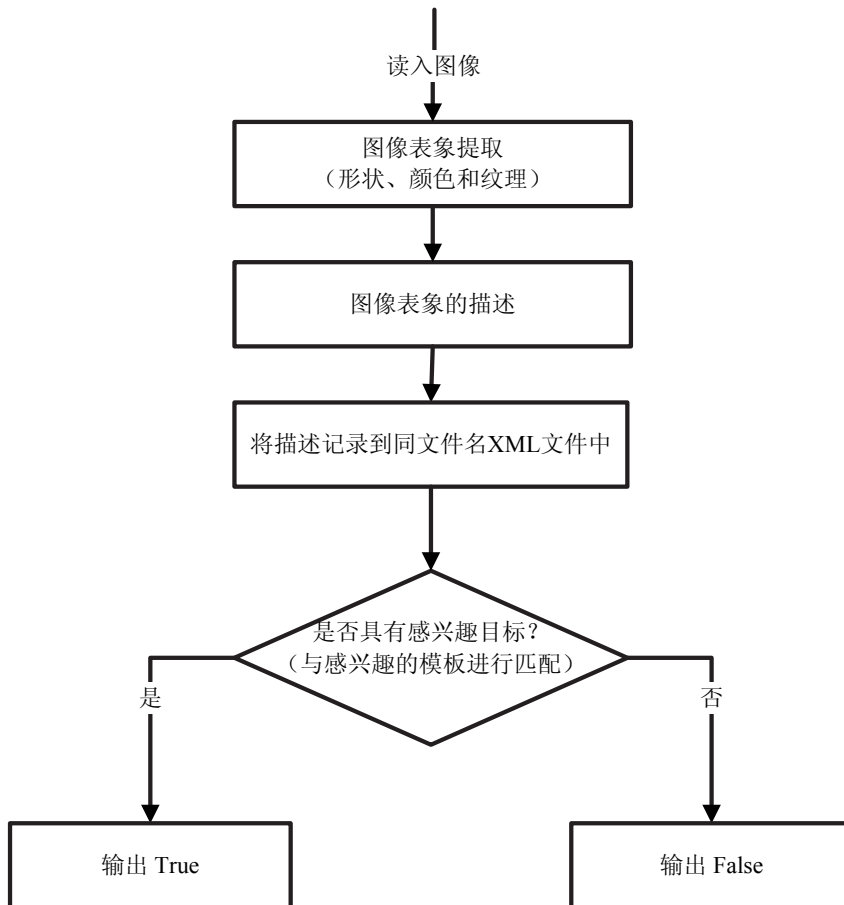


图 7.5: SSIS 中特定内容的图像检索模块

免反复访问已经不存在的服务器，使用数据库记录无法访问的服务器名称。获得连接后，首先检查图像所在的服务器是否属于难访问的服务器列表中。例如，图像链接 `http://www.quzhi.net/images/wzy_s.jpg` 中，服务器地址为 `http://www.quzhi.net`。如果发现这个服务器无法访问，就将这个服务器添加到数据库中相应位置。

对同一个图像，搜索引擎的结果中可能反复出现。为了避免这种情况，每一次找到图像地址，都检查这个地址是否已经访问过，如果没有，就访问并下载图像，并将已经访问过的图像链接添加到数据库中；如果已经访问过，则跳过访问并分析下一个链接。

§7.3.4 图像下载

HTTP 协议 (HypertextTransferProtocol, 超文本传输协议) 是用于从 WWW 服务器传输超文本（包括图像）到本地浏览器的传送协议，它规定了客户端与网络服务器中间进行超文本传输的方式和方法。具体请参考相关的书籍。

SSIS 目前使用 VckBase 网站上 querw 提供的 http 动态链接库 [161] 进行图像下载。这个库的缺点是，它使用了单线程下载，下载速度慢。为此，将来需要优化这一模块，使用多线程下载图像，以提高程序的运行速度。

§7.3.5 图像读写库

网络上图像的可能格式很多，常见的就有 JPEG 格式、GIF 格式、BMP 格式、PNG 格式和 TIF 格式等等。所以，通过上面方法找到图像，其格式有很多种：“.gif”、“.jpg”、“.png”和“.bmp”等。通过比对 `ImageUrl` 中最后三个字符就可以判断图像格式。

SSIS 使用了 FreeImage 图像读写函数库 (开源) [161] 提供的动态链接库。使用这个库能打开网络上常见的 BMP, JPG, GIF, PNG, TIF 等五种格式的图像。

§7.3.6 基本过滤

搜索引擎上获得图像种类很多。

形式上的图像分类。从图像的颜色来分类，有彩色图像灰度图像和黑白图像。从图像的大小来分，有巨型图像（100 万像素以上）、中型图像（1 万像素以上、100 万像素以下）和小型图像（1 万像素以下）。

从图像的来源分类，有照片（相机等成像设备采集到的对应着真实场景的图像或者它们修改）、图形（计算机设计出来的图形，例如 CAD 设计图、网站 Banner 和网站上的广告等）。

我们感兴趣的图像应该是中型以上的图像，所以应该能根据需求对图像进行初步的过滤。

1. 滤掉网站 Banner 和网站上的广告。主要依据是网站 Banner 和网站上的广告图像的特点：一般都很小，使用 GIF 格式保存，长宽比很大并且颜色种类少 [24]。

2. 根据需求定制过滤要求。

§7.3.7 图像表象的提取模块

图像表象是基于内容的图像检索中的核心问题，也是难点，绝大多数算法都相当复杂。SSIS 中目前已经实现了形状表象的相关功能：包括边缘检测库的开发 (EdgeCom)，边缘描述和识别算法的部分开发（边缘的链码描述、边缘的圆拟合）。

对颜色和纹理的提取和描述将是今后重点研究的模块。

§7.3.8 表象的 XML 语言描述

XML 即 Extensible Markup Language(可扩展标记语言) 的缩写。XML 实际上是 Web 上表示结构化信息的一种标准文本格式，它没有复杂的语法和包罗万象的数据定义。MPEG7 中使用 XML 来表达形状、颜色和纹理描述子。

SSIS 中使用 XML 的优点有：首先是可扩展性，XML 允许自己创建和使用标记，因此，描述中可以自由使用 Shape, Color, Texture, Edge ,points 等自然语言标记描述视觉信息的项目。其次是灵活性，XML 提供了一种结构化的数据表示方式，使得用户界面分离于结构化数据，有利于数据在不同平台上的交互。第三是自描述性，XML 文档通常包含一个文档类型声明，因而 XML 文档是自描述的。不仅人能读懂 XML 文档，计算机也能处理。XML 表示数据的方式真正做到了独立于应用系统，并且数据能够重用。这些优点，可以方便 SSIS 向分布式计算扩展。

表象的 XML 语言描述模块的主要功能是将图像信息转化为结构化 XML 文件，方便下一次检索。表象的 XML 语言描述功能还不是目前工作的重点，所以这一模块还没有实现。图 7.6 是对 XML 结构和内容的一种设想。

§7.3.9 图像中特定内容检测算法

由于还没有一般意义的内容检索方法，所以需要针对每一类感兴趣的图像设计特定的检索算法。SSIS 已经实现了以下两类图像的自动检索：含有圆及圆环的图像；含有 QR 码的图像（QR 码是在法视特上海图像科技有限公司实习期间完成的）。下面主要介绍圆以及圆环的检测算法。

§7.3.9.1 圆及圆环检测

如图 7.7 所示，圆环和圆检测的基础都是实现圆的检测，然后判断圆与圆之间的关系，以判断圆与圆之间相交、相离、同心和包含关系。如果同心，则为圆环。

圆检测的常用方法是 Hough 变换，但是 Hough 变换速度慢，在 SSIS 中很难实用。SSIS 中使用了一种更有效地算法，本文称之为圆检测的 EAF 算法：

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- ImagePresentation of Image -->
- <ImagePresentation>
- <Shape>
- <Edge>
  <length>50</length>
  <points>(50,51),(50,52),(50,53),(51,54),...,(51,49)</points>
  <type>Circle</type>
  <description>N0</description>
</Edge>
+ <Edge>
</Shape>
- <Color>
- <region>
  <roi>3,(5,6),(70,89),(65,126)</roi>
  <maincolor>[255,0,0]</maincolor>
</region>
- <region>
  <roi>3,(45,3),(23,45),(38,50)</roi>
  <maincolor>[255,158,14]</maincolor>
</region>
</Color>
- <Texture>
- <region>
  <roi>3,(5,6),(70,89),(65,126)</roi>
  <edgehist>...</edgehist>
</region>
+ <region>
</Texture>
</ImagePresentation>

```

图 7.6: 视觉表象的 XML 语言描述

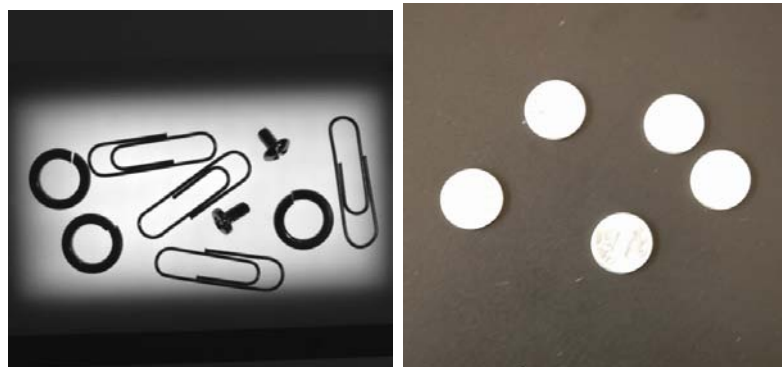


图 7.7: 圆环检测算法

1. 首先使用 EdgeCom 控件提取图像的形状表象一边缘；EdgeCom 对边缘做了连接处理，使得边缘成为一条条的线，成为边缘片段；
2. 对边缘片段进行分析，过滤掉不是圆弧形的边缘片段；
3. 对类似圆弧的边缘片段使用圆拟合算法，得到圆心 (x_i, y_i) 和半径 r_i ；
4. 对圆心和半径进行聚类，以避免重复；聚类后，得到可能的 N 个圆的圆心和位置 $\{(x_i, y_i), r_i\}, i = 0, \dots, N - 1$ ；
5. 假设上面得到的每一个圆心为 (x_i, y_i) ，对应的半径为 r_i 。令 $R_{min} = 0.75r_i, R_{max} = 1.25r_i$ 。如图 7.8 所示，在以 (x_i, y_i) 为圆心，以 R_{min} 和 R_{max} 为半径的圆环中，寻找所有可能的边缘点拟合圆。这个过程可以反复进行多次，以求得更高精度的圆心和半径。

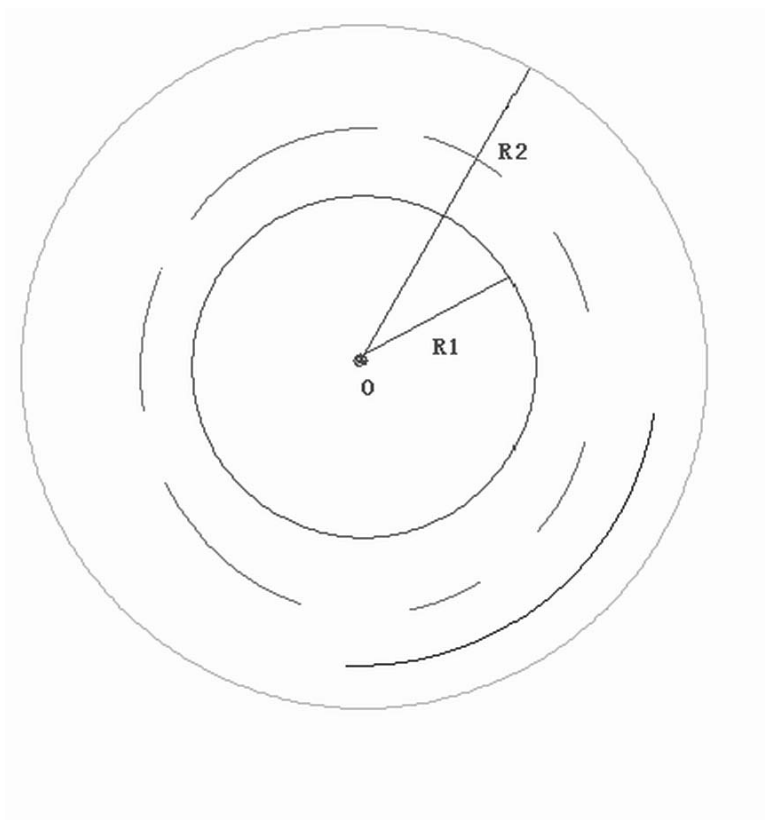


图 7.8: 圆精确定位

6. 得到一个圆后，将对应的边缘片段从形状表象中删去，然后寻找下一个圆。

§7.3.10 界面设计

SSIS 的可执行程序可以在我的主页 [163] 上下载，这里只作简单的介绍。

如图 7.9 所示，界面分六个区。左上区域是图像显示区，用来显示正在处理的图



图 7.9: SSIS 的界面

像；左下区域是状态显示区，用来显示当前进行正在进行的操作、当前的网络连接状况和任务完成进度条；右边第一个区域的作用是搜索设定，选择要搜索的引擎、感兴趣的内容和要搜索的图像数，然后单击搜索后程序开始运行；右边第二个区域可以修改存储目录；右边第三个区域两个按钮用来停止检索和关闭程序；右边第四个区域是软件 LOGO。

第八章 结论与展望

本文从视觉计算理论、表象学、图像识别和分析技术、多媒体内容描述界面 / 接口 (MPEG-7) 和基于内容的图像搜索引擎技术中抽取出它们共同的部分, 提出了图像的视觉信息表象理论, 并研究了其数学基础问题。

本文说明了, 多尺度零交叉点是合适的形状表象, 图像的小波变换是合适的纹理表象, 并在小波分析的框架下给出了形状和纹理提取、描述的一般模型。本文提出了基于现有搜索引擎的特定图像搜索软件的框架, 并实现了这个软件的基本流程, 这个软件可以作为图像表象理论的测试平台。

图像的视觉信息表象理论是一个理论性和应用性都非常强的领域, 本文只是对部分基础问题进行了研究。下面是今后需要深入研究的问题:

1. 小波变换的神经网络实现。人的视觉系统的计算过程是通过神经网络来实现的, 如果小波变换是人类视觉系统的一部分, 那么这个过程是如何用神经网络实现的呢?
2. 图像表象的提取和描述函数库的开发。C++ 中类 (Class) 和表象理论具有一定的相似性, 是实现图像表象描述和提取算法的好工具。这是一个十分庞大的项目, 需要长期的积累才能完成。
3. SSIS 的优化和实用化。

致谢

本文是在程正兴教授的悉心指导下完成的。正是在程老师的引导下，我开始了边缘检测的研究，并逐渐拓展到计算机视觉领域，一直到这一系列的研究题目。感谢程老师从本科三年级以来在学业上给予我的指导和帮助，他严谨的治学态度和渊博的知识给了我很多的启示。

研究学习期间的最后一年多我是在法视特（上海）图像科技有限公司度过的。在公司的一年多时间里，在同事们的帮助下，我学会了编程、学会了如何将理论与实践结合起来解决实际问题。计算机视觉是一个实践性很强的专业，需要理论水平和编程能力两方面都有比较高的水平才能取得做出成果。在公司实习期间，通过参与具体应用项目的讨论和开发，使我对计算机视觉有了比较深刻的理解。感谢公司所有员工对我的关心和帮助，特别是朱翊斌、贾波、刘伟、马领、徐可佳、施俊和宁兆斌。

感谢 Demigny, Deriche, Ellen C. Hildreth 和 James Elder 他们热情地回答我在 Email 中提出的问题, 甚至给我发来了一些文章和资料; 感谢李雪梅, 她在澳大利亚墨而本读书期间, 帮我寻找我迫切需要的文献资料; 感谢中国视觉网艾小洋和讨论视觉技术应用中的问题, 并给我很多宝贵的资讯; 感谢 BingJian 和我讨论一些关于图像处理和计算机视觉的问题; 感谢孙雪青和我讨论 VC++ 编程的问题; 感谢中国科学技术大学的龚敏敏帮我解决了 COM 开发中的一些疑难问题。

谨此向所有帮助过我的老师和同学们致以由衷的感谢!

感谢我的女朋友, 没有她的热爱和支持, 本论文就无法完成。

本论文奉献给我父亲和母亲。

西安 2006 年 2 月 27 日

参 考 文 献

- [1] RAMESH C. Jain. Ignorance, Myopia, and Naivete in computer vision systems, CVGIP: image unstanding[J], 1991, 53(1): 112-117.
- [2] Jonas Gomes, Luiz Velho, Silvio Levy. Image Processing For Computer Graphics[M]. New York: Springer-Verlag, 2000. 3-125.
- [3] 阎洪禄, 于秀敏. 眼生理学 [M], 北京: 人民卫生出版社, 2001 年. 1-67.
- [4] 寿天德. 视觉信息处理的脑机制 [M]. 上海: 科学教育出版社, 1997 年. 1-120.
- [5] 蔡浩然, 马万禄. 视觉的分子生理学基础 [M]. 北京: 科学出版社, 1978 年. 1-55.
- [6] 弗莱德 陶逊斯, 感觉系统, <http://jokocjiang.blogchina.com/1200320.html>.
- [7] [美]托马斯, 贝纳特. 感觉世界 — 感觉和知觉导论 [M]. 旦明译. 北京: 科学出版社, 1983 年. 20-30.
- [8] D.Marr. 视觉计算理论 [M]. 姚国正等译. 北京: 科学出版社, 1988. 1-97.
- [9] 徐光佑. 计算机视觉. 清华大学计算机系讲义. 1999.
- [10] Hubel and Wiesel. Receptive fields, Binocular interaction and functional architecture, In Cat's visual cortex, J. Physiol[J], 1962, 166: 106-154.
- [11] Hubel and Wiesel. Receptive fields and functional architecture of monkey striate cortex, J. Physiol[J] 1968. 195: 215-243.
- [12] Campbell, F. W. C. and Robson, J., Application of Fourier analysis to the visibility of gratings, J. Physiol[J]. 1968, 197: 551-566.
- [13] Application of Gestalt Theory, <http://tds.ic.polyu.edu.hk/vc/>.
- [14] [英]弗朗西斯克里克. 惊人的假说 - 灵魂的科学探索 [M]. 汪云九、齐翔林、吴新年、曾晓东等译. 湖南: 湖南科学技术出版社, 1999. 第四章.
- [15] Elder, J.H., Zucker, S.W.. Evidence for boundary-specific grouping in human vision. Vision Research[J]. 1998, 38(1): 143-152.
- [16] 袁红章. 表象学 [M]. 北京: 科学技术文献出版社, 1995.
- [17] Markus Ulrich, Carsten Steger. Performance Evaluation of 2D Object Recognition Techniques[R]. Technical Report PF-2002-01, Lehrstuhl für Photogrammetrie und Fernerkundung, Technische Universität München, 2002.
- [18] R. M. Haralick. Statistical and structural approaches to texture. Proceedings of the IEEE[J]. 1979, 67(5): 786-804.

- [19] W. K. Pratt, O. D. Faugeras and A. Gagalowicz. Visual discrimination of stochastic texture fields. IEEE Transactions on System, Man and Cybernetics[J], 1978, 8: 796-804.
- [20] A. K. Jain and K. Karu. Learning texture discrimination masks. IEEE Transactions on PAMI[J]. 1996, 18(2): 195-205.
- [21] Jiahua Wu. Rotation Invariant Classification of 3D Surface Texture Using Photometric Stereo[D]. Heriot-Watt University, 2003.
- [22] Chen, Y. Q. , Nixon, M. S. ,Thomas, D. W. . Statistical Geometrical Texture Description[R]. EPrints, 1994.
- [23] 章毓晋. 数字图像工程—图像处理和分析 [M]. 北京: 清华大学出版社, 1999 年.
- [24] 章毓晋. 基于内容的视觉信息检索 [M], 北京: 科学出版社, 2003 年.
- [25] 吴朝相. 多媒体之旅—从 MPEG1 到 MPEG7[N], 《电脑报》32 期, 2000 年.
- [26] MPEG-7 Overview (version 10), <http://www.chiariglione.org/MPEG/>, 2005 年 12 月 12 日.
- [27] 搜索引擎诞生纪年, <http://www.sowang.com/zhuanjia/XZHY/20050124.htm>, 2005 年 11 月 12 日.
- [28] AllTheWeb, <http://www.alltheweb.com/>, 2005 年 11 月 12 日.
- [29] AltaVista, <http://www.altavista.com/sites/search/advimage>, 2005 年 11 月 12 日.
- [30] Cobion, <http://www.visoo.com/default.htm>, 2005 年 11 月 12 日.
- [31] DINO online, <http://www.dino-online.de/dino/Suche/Bilder/>, 2005 年 11 月 12 日.
- [32] Ditto, <http://www.ditto.com/>, 2005 年 11 月 12 日.
- [33] Google, <http://images.google.com/>, 2005 年 11 月 12 日.
- [34] Picsearch, <http://www.picsearch.com/>, 2005 年 11 月 12 日.
- [35] Tiscali, <http://search.tiscali.it/images/>, 2005 年 11 月 12 日.
- [36] Rob Fergus 的主页. <http://people.csail.mit.edu/fergus/>, 2005 年 11 月 12 日.
- [37] Rob Fergus. Visual Object Category Recognition[D], University of Oxford:2005.
- [38] Rob Fergus, Fei-Fei L. , Perona, P. and Zisserman. Learning Object Categories from Google's Image Search, Proc. of the 10th International Conference on Computer Vision[C], ICCV 2005.
- [39] 李唐. 解读网络图像搜索引擎 [N]. 计算机世界报, 第 39 期.

- [40] Th. Gevers and A. W. M. Smeulders. Content-based Image Retrieval: An Overview, Survey on content-based image retrieval[M]. Prentice Hall, 2004.
- [41] T. Kadir and M. Brady. Scale. Saliency and image description. IJCV[J]. 2001, 45(2): 83 - 105.
- [42] B. Sankur, M. Sezgin. A Survey Over Image Thresholding Techniques And Quantitative Performance Evaluation. Journal of Electronic Imaging[J]. 2004, 13(1): 146-165.
- [43] Miquel Ralló. Wavelet based techniques for textile inspection, On Internet. <http://www.imub.ub.es/wavelets/Rallo.pdf>, 2006 年 2 月 23.
- [44] 法视特图像处理库函数. 网络资源. <http://www.fast-corp.com.cn/product.asp>, 2006 年 2 月 27.
- [45] M. Bober. Mpeg-7 visual shape descriptors. IEEE Transactions on Circuits and Systems for Video Technology[J], 2001, 11(6): 716 - 719.
- [46] Julien Ricard, David Coeurjolly, Atilla Baskurt. Generalizations of angular radial transform for 2D and 3D shape retrieval. Pattern Recognition Letters[J]. 2004, 26(14): 2174-2186. <http://liris.cnrs.fr/docs/RR-2004-023.pdf>.
- [47] Sun-Kyoo Hwang, Whoi-Yul Kim. Fast and Efficient Method for Computing ART, IEEE Transactions On Image Processing[J]. 2006, 15(1).
- [48] Mokhtarian, F.; Mackworth, A. K.. A theory of multiscale, curvature-based shape representation for planar curves, IEEE Transactions on PAMI[J]. 1992, 14(8): 789-805.
- [49] Mokhtarian, F., Bober, M.. Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization[M]. Series: Computational Imaging and Vision(25). 2003.
- [50] Search for similar shapes in the SQUID system: Shape Queries Using Image Databases. <http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html>.
- [51] Homepage of Farzin Mokhtarian, <http://www.ee.surrey.ac.uk/Personal/F.Mokhtarian/>.
- [52] M. Ren, J. Yang, H. Sun. Tracing boundary contours in a binary image. Image and Vision Computing[J]. 2002, 20: 125-131.
- [53] H. Blum. A transformation for extracting new descriptors of shape, in Models for the Perception of Speech and Visual Form[M] (W.Wathen-Dunn, ed.), Cambridge MA: MIT Press, 1967. 362-380.
- [54] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, Irwin Sobel. A Variational Framework for Retinex. International Journal of Computer Vision[J]. 2003, 52(1): 7-23.

- [55] Yu-Li You, Wenyan Xu, Tannenbaum, A., Kaveh, M.. Behavioral analysis of anisotropic diffusion in image processing,IEEE Transactions on Image Processing[J]. 1996, 5(11): 1539-1553.
- [56] James H. Elder, and Steven W. Zucker. Local Scale Control for Edge Detection and Blur Estimation, IEEE Transactions On PAMI[J]. 1998, 20(7):699-716.
- [57] D. C. Marr and E. Hildreth. Theory of Edge Detection.Proc Roy Soc London[J]. 1980, B275: 187-217.
- [58] S. R. Curtis, S. Shitz, and A. V. Oppenheim. Reconstruction of nonperiodic two-dimensional signals from zero crossings.IEEE Trans on Acoust Speech, Signal Processing[J]. 1987, 35: 890-893.
- [59] Logan. Information in the zero-crossings of bandpass signals. Bell System Tech J[J]. 1977, 56: 487-510.
- [60] S. Carlsson. Sketch based image coding. Proc of Premier Colloque Image Biarritz,France[J]. 1984, 71-77.
- [61] S. Carlsson. Sketch based coding of gray level images.Signal Processing[J]. 1988, 165: 57-83.
- [62] A. L. Yuill and T. Poggio. Fingerprints theorems for zero crossings. Journal of the Optical Society of America[J]. 1985, 2(5): 683-692.
- [63] R. Hummel and R. Moniot. Reconstruction from zero crossings in scale space. IEEE Trans on Acoust Speech,Signal Processing[J]. 1989, 37(12): 2111-2130.
- [64] Y. Meyer. Un contre-exemple a la conjecture de Marr et a celle de S. Mallat. preprint, 1991.
- [65] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. IEEE Transactions on PAMI. 1992, 14(7): 710-732.
- [66] James H. Elder. Are edges incomplete?. International Journal of Computer Vision[J], 1999, 34(2/3): 97-122.
- [67] T. Cornsweet. Visual Perception[M].New York: Academic Press, 1970.
- [68] D. N. Graham. Image transmission by two-dimensional contour coding. Proc IEEE[J]. 1976, 55(3): 336-346.
- [69] M. Koher, M. Kunt. Image data compression by coutour texture modelling.SPIE Int. Conf. on the Applications of Digital Image Processing[C], Geneva, 1983. 131-139.

- [70] M. Kunt, A. Ikonomopoulos and M. Kocher. Second generation image coding techniques, Proc IEEE[J]. 1985, 73(4): 549-574.
- [71] P. Perez, M. Gangnet, and A. Blake. Poisson image editing. ACM Transactions on Graphics (SIGGRAPH'03)[C]. 2003, 2(3): 313-318.
- [72] Eric W. Weisstein. Entire Function. MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/EntireFunction.html>. 2005.12.15.
- [73] Jian Sun, Jiaya Jia, Chi-Keung Tang and Heung-Yeung Shum. Poisson Matting. SIGGRAPH 2004. 2004, 23(3).
- [74] 管理, 数字图象中基于 α 值的景象提取算法研究 [D], 浙江大学, 2004.
- [75] Smith SM, Brady J M. SUSAN-A new approach to low level image processing. International Journal of Computer Vision[J]. 1997, 23(1): 45-78.
- [76] 伪造照片有克星, 识别系统已问世. http://www.chinadaily.com.cn/gb/doc/2003-09/23/content_266689.htm.
- [77] T. Poggio, H. Voorhees and A. Yuille. A Regularized Solution to Edge Detection[R], Tech Report AIM-833, MIT Artificial Intell Lab 1985.
- [78] Maria Petrou and Josef Kittler. Optimal Edge Detectors for Ramp Edges. IEEE Trans on PAMI[J]. 1991, 13(5): 483-491.
- [79] C. Steger. Subpixel-Precise Extraction of Lines and Edges. International Archives of Photogrammetry and Remote Sensing[J]. 2000, XXXIII(B3): 141-156.
- [80] Julez. A Method of Coding TV Signals Based on Edge Detection. Bell System Tech. 1959, 38(4):1001-1020.
- [81] L. G. Roberts. Machine Perception of Three-Dimension Solids. J. t. Tippett, et al., Ed. Optical and Electro-Optimal Information Processing[C]. Cambridge, MA: MIT Press, 1965.
- [82] Hildreth 的主页, <http://www.wellesley.edu/CS/ehildreth.html>
- [83] John Canny. Finding Edges and Lines in Images[R]. AI-TR-720. Cambridge, MA: MIT Artif. Intell. Lab. 1983.
- [84] John Canny. A Computational Approach to Edge Detection. IEEE Trans. PAMI[J]. 1986, 8(1):679-697.
- [85] John Canny 的主页. <http://www.cs.berkeley.edu/~jfc/>.
- [86] Deriche 的主页. <http://www-sop.inria.fr/odyssee/team/Rachid.Deriche>.

-
- [87] R.Deriche. Using Canny's Criteria to Derive a Recursively Implemented Optimal Edge Detector. Int. J. Comput. Vision[J]. 1987, 1(2).
 - [88] R. Deriche. Fast Algorithm for Low-Level Vision. IEEE Trans on PAMI[J]. 1990,12: 78-87.
 - [89] INRIA Sophia-Antipolis 的主页, <http://www-sop.inria.fr>.
 - [90] J. Shen. An Optimal Linear Operator for Step Edge Detection.CVGIP: Graph. Models Image Process[J]. 1992, 54: 112-133.
 - [91] D.Demigny 的主页. <http://www-etis.ensea.fr/~demigny/>
 - [92] D.Demigny. On Optimal Linear Filtering for Edge Detection. IEEE Trans. Image Processing[J]. 2002, 11(7): 728-737.
 - [93] Tony Lindeberg. Scale-Space Theory in Computer Vision[M]. Dordrecht, Netherlands: Kluwer Academic Publishers, 1994.
 - [94] Lindeberg 的主页. <http://www.nada.kth.se/~tony/>.
 - [95] Jon Sporring, Mads Nielsen, L. M. Florack, Peter Johansen. Gaussian Scale-Space Theory[M]. Norwell, MA, USA: Kluwer Academic Publishers, 1996.
 - [96] Jon Sporring, Joachim Weickert. Information Measures in Scale-Spaces. IEEE Transactions on Information Theory[J], 1999, 45(3): 1051-1058
 - [97] Jon Sporring. The entropy of scale-space[C]. Proceedings of the 13th International Conference on Pattern Recognition. 1996, 1: 900-904.
 - [98] Fredrik Bergholm. Edge Focusing. IEEE Transactions on PAMI[J]. 1987, 9(6): 726-741.
 - [99] Elder 的个人主页. <http://elderlab.yorku.ca/~elder/>.
 - [100] 王郑耀, 程正兴, 汤少杰. 基于视觉特征的尺度空间信息量度量. 中国图象图形学报[J].2005,10(7):922-928.
 - [101] Michael D. Heath 的主页—边缘检测评价.
http://figment.csee.usf.edu/edge/edge_detection.html
 - [102] Michael D. Heath. Robust Visual Method for Assessing the Relative Performance of Edge-Detection Algorithms. IEEE Transactions on PAMI[J]. 1997, 19(12): 1338-1359.
 - [103] Edge Detector Performance Evaluation Study at USF.
<http://figment.csee.usf.edu/edge/>.
 - [104] Sheng Tang, Survey of Edge Detection.
http://www.cs.unr.edu/~tang_s/research/image/surv.ps.

- [105] D. Ziou and S. Tabbone. Edge Detection Techniques - An Overview. International Journal of Pattern Recognition and Image Analysis[J], 1998, 8(4): 537-559.
- [106] 数字图像的边缘检测.
<http://www.quzhi.net/chinese/vision/edge/edge.htm>.
- [107] 王郑耀, 程正兴 (指导). 数字图像边缘检测 [D], 西安交通大学, 2003.
- [108] Song Wang, Toshiro Kubota, Jeffrey Mark Siskind, Jun Wang. Salient Closed Boundary Extraction with Ratio Contour. IEEE Transactions On PAMI[J]. 2005, 27(4).
- [109] D. P. Huttenlocher, G. A. Klanderman, W. A. Rucklidge. Comparing Images Using the Hausdorff Distance. IEEE Transactions on PAMI[J]. 1993, 15(9): 850-863.
- [110] Tat-Jen Cham, Roberto Cipolla. Automated B-Spline Curve Representation Incorporating MDL and Error-Minimizing Control Point Insertion Strategies. IEEE Trans On PAMA[J]. 1999, 21(1): 49-53.
- [111] Tet Toe, Tang Van To. Curve Matching by Using B-spline Curves.
http://wscg.zcu.cz/wscg2004/Papers_2004_Poster/O19.pdf
- [112] O. Sukmarg and K. R. Rao. B-Spline Curve Representation of Segmented Object in MPEG Compressed Domain.
<http://www-ee.uta.edu/dip/paper/wpmc00-1018.pdf>.
- [113] Amir Helzer, Meir Barzohar, and David Malah. Stable Fitting of 2D Curves and 3D Surfaces by Implicit Polynomials.IEEE Transactions ON PAMI[J]. 2004, 26(10).
- [114] Texture Analysis and Synthesis. <http://www.mcislab.org.cn/course/cv/2005/cv2005-ch08.pdf>.
- [115] Free-pictures-photos 图像库. <http://www.free-pictures-photos.com/texture/>.
- [116] A. R. Rao and G. L. Lohse. Towards a Texture Naming System: Identifying Relevant Dimensions of Texture. Vision Research[J]. 1996, 36(11):1649-1669.
- [117] Liu, F., Picard, R.W.. Periodicity, Directionality and Randomness: Wold Features for Image Modelling and Retrieval.IEEE Transactions on PAMA[J]. 1996, 18(7):722-733.
- [118] 遥感图像信息提取方法综述. <http://www.7way.com.cn/wxtx3.htm>.
- [119] N. Sebe, M. S. Lew. Principles of Visual Information Retrieval[M]. Springer. 2001. 51-86.
- [120] R. M. Haralick. Statistical and structural approaches to texture. Proceedings of the IEEE[J]. 1979, 67(5): 786-804.

- [121] Gibson. The Perception of the Visual World[M]. Boston: Houghton Mifflin. 1950
- [122] SG Mallat. Scale change versus scale space representation. ICCV[C]. 1989. 592-596.
- [123] SG Mallat. A theory for multiresolution signal decomposition : the wavelet representation. IEEE Transaction on PAMI[J]. 1989, 11: 674-693.
- [124] Campbell and Robson. Application of Fourier analysis to the visibility of gratings, J Physiol[J]. 1968, 197: 551-566.
- [125] Michael S. Landy. Norma Graham, Visual Perception of Texture.
<http://www.cns.nyu.edu/~msl/papers/landygraham02.pdf>
- [126] A.P.Witkin. Scale Space Filtering[C]. in Proc. Int. Joint Conf. Artificial Intell,CA ,1983: 1019-1022.
- [127] A.P.Witkin, Jean Babaud, Michel Baudin. Uniqueness of the Gaussian Kernel for Scale-Space Filtering. IEEE Trans on PAMI[J]. 1986, 8(1): 26-33.
- [128] Yuille, Poggio. Scaling theorem for zero crossing.IEEE Trans on PAMI[J]. 1986, 8(1): 15-25.
- [129] Tony Lindeberg, Harr Romeny. Linear scale-space[C]. Geometry-Driven Diffusion in Computer Vision. Dordrecht, Netherlands: Kluwer Academic Publishers, 1994.1-77.
- [130] C. K. Chui. An introduction to wavlets[M]. 程正兴译. 西安: 西安交通大学出版社,1995.
- [131] Daubechies I. Ten lectures on wavelets[M]. 李建平, 杨万年译. 湖南: 国防工业出版社,2004 年.
- [132] 唐远炎, 王玲. 小波分析与文本文字识别 [M]. 北京: 科学出版社,2004 年.
- [133] 程正兴. 小波分析算法与应用 [M], 西安: 西安交通大学出版,1998 年.
- [134] W. Sweldens and P. Schr. Building your own wavelets at home[R]. in Wavelets in Computer Graphics. ACM SIGGRAPH Course notes. 1996: 15-87.
- [135] Homepage of Wim Sweldens. <http://cm.bell-labs.com/who/wim/>.
- [136] Arne Jensen and Anders la Cour-Harbo.Ripples in Mathematics-The Discrete Wavelet Transform. Springer Verlag, 2001.
- [137] CWT Library, <http://www.sourceforge.net/cwtlib>.
- [138] Wavelet, a C++ class library for Wavelet transforms on images.
<http://sourceforge.net/projects/wavelet/>.
- [139] 靳济芳.VisualC++ 小波变换技术与工程实践 [M]. 人民邮电出版社.2004.

- [140] Abramovich F., Bailey T., Sapatinas T.. Wavelet analysis and its statistical applications. The Statistician - Journal of the Royal Statistical Society[J]. 2000, D(49): 1-29.
- [141] R. T. Ogden. Essential wavelets for statistical applications and data analysis[M]. Boston, 1998.
- [142] Percival, Donald B.; Walden, Andrew T. . Wavelet Methods for Time Series Analysis[M]. Cambridge University Press, 2000.
- [143] T. Li, Q. Li, S. Zhu, and M. Ogiwara. A survey on wavelet applications in data mining. <http://www.cs.rochester.edu/~zsh/pub/li.pdf>.
- [144] Zohreh Azimifar, Seyedeh-Zohreh. Image Models for Wavelet Domain Statistics[D]. in Systems Design Engineering, Faculty of Engineering, 2005.
- [145] Matjaz Bevk, Igor Kononenko. A Statistical Approach to Texture Description of Medical Images: A Preliminary Study. <http://lkm.fri.uni-lj.si/xaigor/slo/clanki/BevkICML2002.pdf>.
- [146] Hany Farid, Siwei Lyu. Higher-order Wavelet Statistics and their Application to Digital Forensics. In IEEE Workshop on Statistical Analysis in Computer Vision(in Conjunction with CVPR)[C], 2003.
- [147] Jun Li. A Wavelet Approach to Edge Detection[D]. The Department of Mathematics and Statistics, Sam Houston State University, 2003.
- [148] Chaganti, venkata ravikiran.Edge detection of noisy images using 2-d discrete wavelet transform[D], Department of Electrical and Computer Engineering, Florida State University, 2005.
- [149] Toshikazu WADA, Makoto SATO. Scale-Space Tree and Hierarchy. 10th International Conference on Pattern Recognition[C]. 1990, 2: 103-108.
- [150] 齐翔林, 汪云九, 朱舜山. 初级视觉信息的 Gabor 小波表达研究. 自然科学进展—国家重点实验室通讯 [J],1996 年.
- [151] 齐翔林, 朱舜山, 汪云九. 基于视觉编码的图象处理研究 (I): 原理、算法及实现. 生物物理学报 [J].1996 年,12(2).
- [152] 齐翔林, 朱舜山, 汪云九. 基于视觉编码的图象处理研究 (II): 计算机模拟结果及分析. 生物物理学报 [J].1996 年,12(2).
- [153] Lee, T.S.. Image Representation Using 2D Gabor Wavelets. IEEE Trans On PAMI[J]. 1996, 18(10): 959-971.

- [154] Mehrotra R, Namuduri KR, Ranganathan R. Gabor filter-based edge detection. Pattern Recognition[J]. 1992, 52: 1479-1494.
- [155] Chen, C. C., Chen, D. C.. Multi-resolutional Gabor Filter in Texture Analysis. PRL. 1996, 17(10): 1069-1076.
- [156] Pichler O, Teuner A, Hosticka B J. A comparison of texture feature extraction using adaptive Gabor filtering, pyramidal, and tree structured wavelet transforms. Pattern Recognition[J]. 1996, 29(5): 733-742.
- [157] 贾天旭. 初级视觉的多频率通道信息处理与非线性自适应平滑方法研究 [D]. 西安交通大学,1996 年.
- [158] Wei-Ying Ma, BS Manjunath. EdgeFlow: a technique for boundary detection and image segmentation. IEEE Transactions on Image Processing[J]. 2000, 9(8): 1375-1388.
- [159] 李学杰著, 齐春指导. 基于边界流矢量的边缘检测与图像分割算法研究 [D]. 西安交通大学,2003.
- [160] Google Web APIs (beta). <http://www.google.com/apis/>.
- [161] querw, 直接用 socket 实现 HTTP 协议 (下载专用).
<http://www.vckbase.com/document/viewdoc/?id=1208>.
- [162] Freeimage Project. <http://freeimage.sourceforge.net/>.
- [163] SSIS 的主页.<http://www.quzhi.net/chinese/vision/ssis/index.htm>.

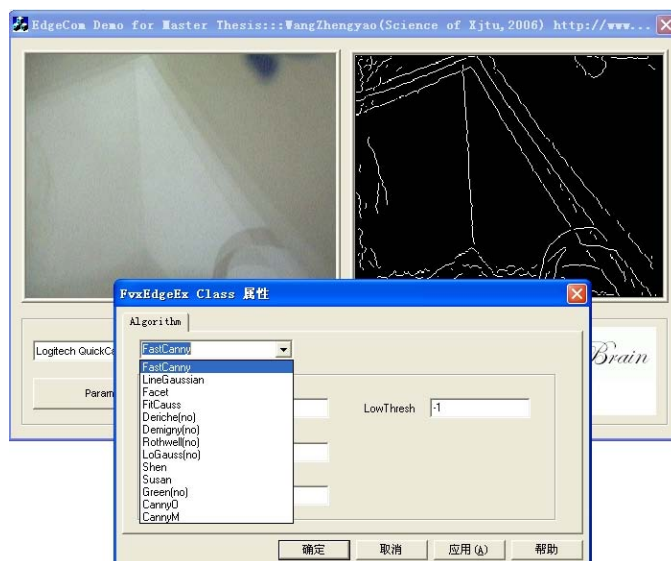
附录 1 边缘检测 COM 组件: EdgeCom

COM 是组件对象模型 (Component Object Model) 的简称, 是由微软公司开发的规范和执行操作, 是为集成组件提供的一组框架。每一个 COM 组件提供一系列接口, 以实现特定的目的。COM 组件可以在 VB 和 VC 等语言中方便的使用。

EdgeCom 是我在法视特 (上海) 图像科技有限公司实习期间和同事施俊合写的进行边缘检测的 COM 组件。我程序实现并初步优化了其中的边缘检测算法, 由施俊封装为 COM 组件。EdgeCom 的目标是尽可能多的实现现有的边缘检测算法、优化后以相对一致形式 (输入和输出一致) 封装到一起, 方便在实际中选用边缘检测算法。

到目前为止, 已经实现了 10 种算法: Canny 算法 (CannyO), Matlab 中的 Canny 算法 (CannyM), 优化后的 Canny 算法 (FastCanny), Steger 算法 (LineGaussian), Haralick 算法 (Facet), 拟合算法 (FitCauss), Shen 算法 (Shen), Susan 算法 (Susan)。已经代码实现, 准备封装到 EdgeCom 中的算法有: Green 滤波器算法 (Green), Deriche 算法 (Deriche), Demigny 算法 (Demigny), Rothwell 算法 (Rothwell) 和 LoG 算法 (LoGauss)。

下面是使用 EdgeCom 编写的演示程序, 使用 USB 摄像头做图像输入, 动态检测图像中的边缘, 在运行中可以切换不同的算法和参数。



附录 2 攻读硕士期间的主要成果

- [1]王郑耀, 程正兴, 汤少杰. 基于视觉特征的尺度空间信息度量. 中国图象图形学报 [J].2005,10(7):922-928.
- [2] 王郑耀, 汪宏武, 杨敏, 戴永红. 目标定位最优布站的研究. 数学的实践与认识 [J].2005,35(7):27-33.
- [3] 白翔, 刘文予, 王郑耀. 基于最优离散滤波器的边缘检测算法. 红外与激光工程 [J].2005,34(6):737-740.