模式分类与视觉导航中的分层数据处理研究

（申请清华大学工学博士学位论文）

培 养 单： 计算机科学与技术系

学 科： 计算机科学与技术

研 究 生： 何 英 华

指 导 教 师： 张 钹 教 授

二〇〇五年十月

Studies on the Hierarchical Data Processing in Pattern Classification and Visual Navigation

Dissertation Submitted to

Tsinghua University

in partial fulfillment of the requirement

for the degree of

Doctor of Engineering

By

He Yinghua

(Computer Science and Technology)

Dissertation Supervisor : Professor Zhang Bo

October, 2005

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：清华大学拥有在著作权法规定范围内学位论文的使用权，其中包括：（1）已获学位的研究生必须按学校规定提交学位论文，学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文；（2）为教学和科研目的，学校可以将公开的学位论文作为资料在图书馆、资料室等场所供校内师生阅读，或在校园网上供校内师生浏览部分内容；（3）根据《中华人民共和国学位条例暂行实施办法》，向国家图书馆报送可以公开的学位论文。

本人保证遵守上述规定。

（保密的论文在解密后遵守此规定）

作者签名：			导师签名：	
日　　期：			日　　期：	

摘　要

与分层数据处理相关的理论（粒度计算、多分辨率分析、多尺度分析）是目前的热点研究领域。通过将所研究问题的论域在不同的粒度（分辨率、尺度）下进行表示，有可能使问题的求解变得更容易、求解的计算复杂性降低、或者不可解的问题变得可解。本文以模式分类和视觉导航为背景，讨论采用分层处理的技术对问题求解的帮助，其研究成果有助于大规模数据的分类和图像内容的分析。

论文首先研究了多粒度表示与求解与模式分类问题的关联，取得了以下三个方面的研究成果：(1) 分析了各种风险泛函对求解结果的影响、训练样本的规模同基于结构风险最小化的分类函数求解之间的关联、和模式分类同分层数据处理理论相结合的可能性。这些分析表明通过将训练样本在粗粒度上进行表示以及采用分层次求解策略，可以有效地解决分类问题；(2) 提出了一种区域风险最小化的分类器求解策略。该策略采用超球（超立方体）作为基本的训练样本单位训练分类函数，使用中心点和半径来控制分类边界的位置。实验结果表明，这一方法与其它方法相比较，能够有效地利用区域特征减少训练样本的个数并且降低求解的分类函数的复杂性；(3) 提出了一种基于特征空间划分的多分辨率分类器求解策略。该策略首先将训练样本所位于的区域在粗分辨率下划分成为超长方体，然后以超长方体为单位求解分类边界，将边界所位于的格子逐步细分，重复这个过程直到所求解的分类器满足精度。论文分析了这种求解策略的泛化能力并给出了能够降低计算复杂性的条件。实验结果表明，当数据集合具有聚类性质时，该策略能减少训练和测试阶段的计算量。

论文其次针对校园和城区环境提出了一种基于商空间合成的道路检测算法。该算法首先基于灰度图像求解道路曲率、三个候选边界位置以及路面色彩的计算区域，然后基于彩色图像提取出路面，最后利用商空间合成的结果将候选边界的计算值与区域提取值进行匹配，求解最可靠的道路边界和路面区域。与基于边缘检测和区域分割的方法对比，本文的算法能够鲁棒地、实时地检测城区和校园环境的道路。

关键词：粒度计算；多分辨率分析；模式分类；支持向量机；道路检测

Abstract

Theories related to hierarchical data processing of soft computing are hot topics in the past several years, such as granular computing, multi-resolution analysis, multi-scale analysis, etc. By representing the universe of the problem in different granulation (resolution, scale), it provides a possibility to solve problems more easy, make problems solvable which can not be solved in the original universe, and reduce the computational complexity. In this dissertation, we study the application of hierarchical data processing to pattern classification and visual navigation. Results of the research are helpful to large scale data classification and analysis the content of image.

The first part of this dissertation discusses the relationship between multi-granule representation and processing with pattern classification. Research results can be summarized as follows. (1) By analyzing four kinds of risk minimization principles, the relationship between the size of the training data set and the trained classifier, and the possibility of combining the pattern classification with hierarchical data processing, it shows that the classification problem may be solved efficiently by representing the data set in coarse granulation. (2) An area based risk minimization principle is proposed. By representing the samples set by hyper-sphere (or hyper-cube), the boundary location can be controlled by the center and radius of the area. Experimental results show that the method proposed in this paper can reduce both the number of the training samples set and the support vectors. (3) We propose a multi-resolution classification strategy based on the partition of feature space. The training algorithm locates the boundary between two classes from coarse to fine resolutions by dividing the hyper-cuboids that lie on the boundary step by step. The testing algorithm firstly labels the testing data set by the classifier trained at initial resolution. Then only those lying on the boundary will be labeled at the finer resolution. Theoretical analysis and

experimental results have substantiated the effectiveness of the proposed method.

The second part of this dissertation discusses a road detection algorithm based on quotient space synthesization. This algorithm is composed of two modules: boundaries are first estimated based on the intensity image and road areas are subsequently detected based on the full color image. In the first module, an edge image of the scene is analyzed to obtain the candidates for the left and right road borders and to delimit the area that will subsequently be used to compute the mean and variance of the Gaussian distribution, assumed to be obeyed by the color components of road surfaces. The second module effectively extracts the road area and reinforces boundaries that the most appropriate fit the road-extraction result. The combination of these modules by quotient space synthesization can overcome basic problems due to inaccuracies in edge detection based on the intensity image alone and due to the computational complexity of segmentation algorithms based on color images. Experimental results on real road scenes have substantiated the effectiveness of the proposed method.

Keywords: granular computing; multiresolution analysis; pattern classification; support vector machines; road detection

目 录

第 1 章　引 言

1.1　研究背景

将所遇到的问题采用层次划分的方式进行求解的思想具有较大的研究价值和应用前景，这类计算方法目前也是一个研究热点，多分辨率分析（multi-resolution analysis）、多尺度表示（multi-scale representation）[8][9]、粒度计算（granular computing）[15]等都属于这个范畴。多分辨率分析和多尺度分析概念的提出已经有了一段时间，粒度计算则是近几年才提出的一个概念。

四叉树图像表示（quad-tree）、金字塔图像表示（pyramids）、尺度空间理论（scale space theory）[8]、小波变换（wavelets transformation）、多网格（multi-grid）等都属于多尺度研究领域。多尺度的方法较多地是从计算机视觉领域和图像处理领域发展起来的，用于层次化地表示图像数据，因此大部分多尺度方法都是从信号处理的角度去讨论数据的分层次表示问题，所考虑的信号既可以是连续信号也可以是离散信号。

多分辨率的含义是指从不同的抽象级别去观察同一物体。目前多分辨率的概念已经被应用到许多研究领域中讨论问题的多分辨率表示以及处理，例如多分辨率的地理信息处理、多分辨率的网格表示等等。多分辨率的概念和多尺度的概念是从两个角度去表达同一问题，当尺度增大时，分辨率会降低，当尺度减小时，分辨率增高。

虽然粒度计算也是一个层次化表示的概念，但是它是从信息系统的表示中发展起来的，针对的对象主要是集合。与多尺度（多分辨率）的研究中采用不同尺度的基信号作用于原始信号去获得不同尺度（分辨率）下的信号相比，粒度计算关注得较多的是事物本身之间自然形成的结构关系所引发的层次关系。Zadeh 中初步定义了粒度计算中粒度的概念：信息粒度的含义是将一个对象集合划分成为各个颗粒，其中每个颗粒是由不可区分性，相似性或者功能性所诱导的一组对象的聚集[6]。因此粒度这个概念针对的对象的结构性更强一些。

与分层数据处理密切相关的一个研究领域是软计算（soft computing），其中的许多方法常常作为高层次数据表示的一种形式。软计算方法是能够处理不确定性和不精确性的各种方法相互作用所形成的一类方法的总称。与传统的硬

计算方法不同的是，软计算方法通过在问题求解的过程中容忍不确定性、不精确性以及部分可靠证据，从而使得问题的处理更容易、鲁棒性更强并且求解的计算量更小。虽然目前许多文献将人工神经网络、细胞神经网络、模糊数学、遗传算法、粗糙集、概率推理、混沌理论等算法均归于软计算的领域[1]-[5][36]，但是软计算方法并非是这些求解策略的一个集合，而是上述各种方法相互作用的结果。也就是说，在求解一个问题的过程中，模糊数学、神经网络、遗传算法以及概率理论等方法并非是相互竞争的，而是互补的，如模糊数学同神经网络的结合形成模糊神经网络、模糊数学同粗糙集的结合形成模糊粗糙集和粗糙模糊集[13]、将遗传算法与神经网络结合从而有利于网络的训练等等。当采用分层数据处理的方式去求解一个问题时，高层次上的信息通常存在许多不确定的因素，因此软计算中的模糊数学、粗糙集等等都经常被用来从高层次上表示论域。

模式分类同样是目前受到广泛关注的研究领域，神经网络、支持向量机、决策树等方法都是在实际应用中表现良好的分类算法[37]-[44]。软计算中的方法同分类算法的结合已经有比较长的历史了，如模糊神经网络、模糊支持向量机、基于粗糙集的分类方法等等。这些方法都是讨论不确定表示对求解分类问题的帮助。由于无论是给定的训练样本集合还是要求解的分类器，都可以以层次化的方式表示和求解，因此讨论模式分类问题中的分层数据处理是一个有价值的研究方向，近年来研究人员已经在这方面作了一些尝试[69][72][75][77]-[93]。本文的前半部分将主要围绕与模式分类问题的表示和求解相关的分层数据处理进行讨论，粒度计算、多分辨率分析中的研究成果对于问题的研究起到了重要的理论指导作用。这部分的研究内容包括两个方面：粗粒度下分类器的求解以及采用多分辨率的策略训练分类器。

多媒体的信息处理也是目前的热点研究领域，基于内容的图像检索、基于内容的视频检索、人脸识别、视频监控系统等近十年来受到了研究人员的广泛关注。分层次数据处理的思想同图像处理的结合的历史已经比较久了，例如图像的四叉树表示、小波变换方法在图像处理的应用等等。本文的后半部分针对视觉导航中的道路检测问题，讨论融合图像不同层次的处理结果对问题求解的帮助，其中应用了张铃等提出的商空间理论中的商空间合成的研究成果[27][28]。这部分论文首先给出图像的一些商空间表示以及道路的边缘和区域商空间表示，然后将所得到的商空间结果合成起来，从而检测出鲁棒的道路区域。

本章接下来将介绍和分层次问题求解相关的一些理论。这些理论主要包括

多分辨率信号处理、尺度空间理论以及粒度计算，它们之间的界限并不是很清晰，许多研究成果有相似的地方，例如基于离散元素的多分辨率分析同粒度计算中的研究内容就比较相近。由于本文中所讨论的问题的论域都是离散元素构成的集合，因此下面将主要介绍粒度计算的研究成果，并且将多分辨率分析、多尺度分析相关的研究成果也包括进来。这些介绍是本文的理论基础，在本文后续章节的讨论中这些理论起到了非常重要的指导作用。

1.2　粒度计算

多分辨率分析渗透的范围比较广泛，既有连续信号，也有离散集合，它的应用所覆盖的范围也比较广泛，例如图像处理、图形学、信息系统等等。粒度计算主要应用于信息系统的表示方面，针对集合来考虑问题。多分辨率分析一般很少讨论由事物内在的结构所构成的层次关系，而粒度计算考虑的多粒度之间的层次关系则可以是由事物内在的结构所引起的。下面首先概述粒度计算的研究内容，然后介绍粒度计算的主要研究成果。

1.2.1　研究内容

如果从不确定表示的观点去看粒度计算，那么粒度计算的概念覆盖的范围比较广泛，目前关于不确定表示的一些研究成果，都可以归于粒度表示的范畴，如粗糙集理论、模糊数学、区间计算理论、商空间理论等等。由于 Zadeh 关于粒度计算的概念是定义在集合上面的，因此粗糙集理论是粒度计算的重要组成部分，并且目前关于粒度计算的许多工作都是建立在这一理论基础之上的。Nguyen，Yao 从粗糙集的观点去研究和表示粒度计算问题。他们从划分构成的颗粒化、覆盖构成的颗粒化以及由邻域系统所表示的多层次粒度几个角度研究了由粗糙集所表示的粒度计算问题[16][22]。张铃等研究人员提出的商空间的概念也是描述粒度的一种非常有效的方式。虽然在粒度表示方面也是采用等价关系来划分对象集合，但是他们关于粒度中颗粒之间结构的表示和基于粒度的问题求解方面作了大量的研究工作，深入讨论了分层次求解的问题，并且证明了采用分层次的思想可以提高问题处理的速度的一些条件和结论[27][28]。Kreinovich 等从区间数学的角度研究了粒度计算中的若干问题[31][32]。区间计算具有较长的研究历史了，目前也作为一种比较重要的粒度表示方法之一。区间计算将论域 U

划分成为区间集合，并且在区间集合上面定义了一组运算$\{+,-,\times,\div\}$。将论域划分成为不同的区间可以看作将论域在不同的粒度级别上面进行定义，定义在区间上面的运算可视为定义在粒度级别上面的运算。Yao 等采用决策逻辑语言（DL-语言）来描述集合的粒度（用满足公式 f 元素的集合来定义等价类 m(f)），建立概念之间的 IF-THEN 关系与粒度集合之间的包含关系的联系，并提出利用所有划分构成的格来求解一致分类问题[22]。这些研究为知识挖掘提供了新的方法和角度。另外，Kreinovich 和 Bernadette 研究了粒度计算同模糊逻辑、词计算之间的关系[32]。

1.2.2　研究成果

下面从粒度内的知识表示、不同粒度之间的知识表示以及基于多粒度的问题求解三个方面介绍粒度计算已有的研究成果，并且给出本文所采用的粒度计算的表示形式。

1.2.2.1　粒度内的知识表示

Zadeh、Hobbs、Yao、张铃等关于粒度计算的讨论针对的对象均为集合，通过将集合进行划分来获得粗、细粒度级别上面的研究对象[6][18][22]-[25][27][28]。上述研究成果认定如下的一个粒度定义：给定论域 U 和论域 U 上面的等价关系 R（采用等价关系是一种最为简单的将论域进行划分的方式），关系 R 关于论域 U 的划分为 U/R，为一组等价类集合，称每个等价类为一个颗粒（granule），那么 U/R 为一个颗粒集合。U/R 是对论域 U 的一个颗粒化（granulation）。关系 R 对于论域 U 的划分可以用下面的二元组来表示：

$$S=(U,R) \tag{1-1}$$

1) U：论域，为一个非空集合；
2) $R\subset U\times U$ 为一个等价关系。关系 R 将论域 U 划分成为多个等价类，每一个等价类在该粒度下作为一个基本运算单位，也就是在目前给定的知识库中，该等价类中的元素是无法区分的。关系 R 可由信息系统的属性值所诱导。

对于集合(U,R)中的每个颗粒，可以用一些属性进行衡量，粒度（granularity）即是用来表示颗粒大小的量。有多种衡量标准去表示颗粒的粒度，例如，可以用颗粒的直径表示粒度，也可以用颗粒所对应集合的基数表示该颗粒的粒度。

对于论域 U 在某个等价关系 R 下所构成的颗粒化(U,R)，给定一个用于衡量集合(U,R)内的颗粒的粒度的量 l，求取(U,R)的平均粒度 $\bar{l}$，称(U,R)为论域 U 在粒度级别 $\bar{l}$ 下面的表示。因此当用不同的等价关系作用于论域 U 时，会得到不同粒度级别下面论域 U 的表示。

用于划分论域 U 的关系 R 一般是由论域 U 本身的属性或论域 U 中元素之间的关系所构成的。综合信息系统的表示以及张铃等所采用的关于论域 U 的属性和结构的表示[27]，可将论域 U 原始的性质表示成为一个 3 元组：

$$O=\langle U,\Lambda,\Gamma\rangle \tag{1-2}$$

1) U：非空论域集；
2) Λ：为一个三元组，表示论域中元素的属性性质。

$$\Lambda=(A,V,f) \tag{1-3}$$

- A：非空属性集，$A=\{a_i\}_{i=1}^{N}$；
- V：属性值的非空集合，$V=\{V_{a_i}\}_{i=1}^{N}$，其中V_{a_i}为一个集合，对应于属性a_i的值。属性值所构成的对论域的划分为一个等价划分；
- f：是函数 $f:U\times A\to\bigcup_{a\in A}P(V_a)$，对于任意的$\langle x,a\rangle\in U\times A$，$f(x,a)\in V_a$。

3) Γ：表示的是论域的结构，即论域中的元素之间所满足的关系。对于论域结构的定义已经存在一些研究成果，例如张铃等研究了论域中的元素满足拓扑关系、半序关系时结构的定义[27]。Yao 采用邻域系统来表示论域中的元素之间的结构关系[26]。

式（1-2）的三元组表示了给定的原始问题论域的基本性质，论域 U 是要考察的对象的集合，属性Λ表示了论域中的每个对象的基本性质，结构Γ给出了论域 U 中的各个对象的之间的关系。而当从不同的粒度去观测论域 U 时，这时需要给出不同粒度下论域 U，属性Λ以及结构Γ的表示形式。张铃等提出了一种商空间理论，用来表示不同粒度下面的信息。假设通过等价关系 R 将论域 U 进行划分，得到商集$[U]$，设商集$[U]$中颗粒的平均粒度为 l，称商集$[U]$为粒度 l 下的新的论域，表示为$[U]_l$。通过式（1-4）表示的三元组去表示粒度 l 下论域$[U]_l$的一些性质：

$$O_l=\langle[U]_l,[\Lambda]_l,[\Gamma]_l\rangle \tag{1-4}$$

1) $[U]_l$：粒度 l 下的非空论域集；

2) $[\Lambda]_l$：为一个三元组，表示论域$[U]_l$中元素的属性性质；

$$[\Lambda]_l = ([A]_l, [V]_l, f_l) \tag{1-5}$$

- $[A]_l$：粒度 l 下的非空属性集，$[A]_l = \{a_{l,i}\}_{i=1}^{N_l}$；$[A]_l$ 中的属性既可以由集合 A 中的属性所获得的，也可以不是由集合 A 中的属性所获得的，这时需要附加的知识；
- $[V]_l$：粒度 l 下属性的值的非空集合，$[V]_l = \{V_{a_{l,i}}\}_{i=1}^{N_l}$，其中$V_{a_{l,i}}$ 为一个集合，对应于属性 $a_{l,i}$ 的值。
- f_l：是函数 $f_l : [U]_l \times [A]_l \to \bigcup_{a\in[A]_l} P(V_a)$，对于任意的 $\langle x, a\rangle \in [U]_l \times [A]_l$，$f(x,a) \in V_a$。张铃等从论域有结构和论域无结构两个方面讨论了某一粒度下属性值的计算[27]。例如：

$$\begin{aligned} f_l(X,a) &= \sum_{x\in X} f(x,a) \\ f_l(X,a) &= \sup_{x\in X} f(x,a) \end{aligned} \tag{1-6}$$

3) $[\Gamma]_l$：表示的是粒度 l 下论域的结构。

给出粒度 l 下知识的表示后，关键的问题是如何根据原始的论域 U 的信息求出粒度 l 下的属性和结构。

1.2.2.2　多粒度间的知识表示

分层次的表示与求解是提出粒度计算概念的重要目的之一，因此从不同的粒度上面去表示问题就显得尤为重要，已有的文献关于多粒度间的知识表示的研究较少，大部分的研究都集中在知识的不确定性表示上，也就是某一粒度下的值的表示。但是分层次的表示在过去的研究中已经受到了广泛的关注，例如，起源于生态学的层次化理论（hierarchy theory）较早的就关注到了生态系统的层次化问题[35]；图形学中曲面的网格表示；图像处理和计算机视觉领域中的四叉树、金字塔、小波变换等；机器学习中的决策树方法。而这里讨论的主要是基于粗糙集、区间计算等领域发展起来的用于描述客观世界粒度层次的多粒度表示方式，也就是以集合作为研究论域的分层次表示方法。Yao 在信息表的基础上，构造了一个层次化的多粒度表示，其中不同粒度下的表示对论域的划分构成了一个半序关系[25]。Kreinovich 等从区间数学的角度讨论了多粒度表示问题[31][32]。

在多粒度知识表示问题上面较为系统性的研究成果是张铃等基于商空间的概念所进行的讨论[27][28]。

多粒度间的知识表示分为两个方面。一方面是通过划分粗粒度上的论域从而获得细粒度上的论域所构成的层次关系，如 Yao 构造的层次化多粒度表示以及 Kreinovich 等建立在区间数学基础上的多粒度表示。这种层次关系是在现实生活中经常遇到的，例如地图所构成的层次关系。另一方面是不同粒度下的知识融合。这种表示的含义是，现实世界中很多层次关系并非完全像前一方面那样有着明确的高层次与低层次，粗粒度与细粒度之分，有时候论域 U 在两个不同粒度下的表示是对论域 U 从两个不同方面的观测结果，并且将这两方面的结果合成起来对问题的求解是非常有帮助的。数据融合领域有较多的这方面的研究成果。张铃等较系统的讨论了不同粒度下的论域，属性以及结构的合成技术[27]。

1.2.2.3　基于多粒度的问题求解

提出粒度表示的目的是为了能够更加有效地处理信息，加快问题求解的速度。目前粒度计算领域大部分的研究工作关注的是粒度的表示问题，而关于基于多粒度的问题求解方面的研究工作较少。采用多粒度求解的方法会对问题的求解带来以下的好处：

1. 采用多粒度的方式可以加快问题求解的速度。张铃等讨论了在什么情况下采用多粒度的求解方式可以加快问题求解的速度[27]。

2. 采用多粒度的方式可能使得某些不可解的问题变得可解。Zadeh 指出，如果将问题表示得过细，某些情况下问题是不可解的，成为一个 NP-hard 问题，但是在现实生活中，人们却能从一个较高的抽象层次（粒度）上面使得问题能够较容易得解决[6]。Oscar 等证明了对于 CNF 公式的命题可满足性问题，在某些抽象粒度上可使得有些 NP-hard 问题变得可解[31]。

在基于粗糙集理论的信息系统的表示中，一般属性中有一个决策属性，来决定问题的求解，得出必要的推理规则，但是这种求解并非是基于多粒度的。在基于多粒度求解方面比较系统的工作是张铃等建立的一套基于粒度表示的问题求解理论和相应的算法，并将其应用于启发式搜索、路径规划等问题的求解中[27]。前面描述了两种多粒度表示方式，一般也对应两种求解问题的方式，第一种是逐步缩小解所在的范围，第二种是不同粒度下解的合成。分层次缩小解

所在范围的含义为：首先在粗粒度上进行求解，排除掉那些不可能包含解的论域部分，然后再在更细一层的粒度上进行求解。重复这个过程，逐步缩小解所在的论域。例如，给定初始粒度 s，在粒度 s 下求解的结果为 S_s，解 S_s 一般比较粗糙，下一步可以由 S_s 形成新的论域，得到更细粒度下面的问题描述，从而求得更为精确的解。不同粒度下的解合成的含义为：假设对于原始的问题 P，在粒度 i，j 下的解为 S_i 和 S_j，那么综合解 S_i 和 S_j 一般可以使得得到更为精确的解。

1.3　论文各部分的主要内容

本文主要探讨模式分类与视觉导航中的分层数据处理问题，并且以粒度计算、多分辨率分析和多尺度分析中的某些研究成果作为理论基础。论文从第 2 章到第 4 章介绍了如何在粗粒度下解决模式分类问题以及如何采用多分辨率的方式去求解分类器。论文的第 5 章以视觉导航中的道路检测为例，讨论了如何使用商空间合成的方法鲁棒地检测出来道路边界。各章的主要内容如下：

第 2 章，讨论了模式分类的基本问题、求解策略以及与分层数据处理结合的可能性。本章分析了各种风险泛函对求解结果的影响、训练样本的规模同基于结构风险最小化的分类函数求解之间的关联、和模式分类同分层数据处理理论相结合的可能性。这些分析表明通过将训练样本在粗粒度上进行表示以及采用分层次求解策略，可以有效地解决分类问题。本章是第 3 章和第 4 章的基础。

第 3 章，提出了一种区域风险最小化的分类器求解策略。该策略采用超球（超立方体）作为基本的训练样本单位训练分类函数，使用中心点和半径来控制分类边界的位置。并且证明了在问题求解中，只需要修改 SMO（序贯最小优化算法）的初始条件，即可采用 SMO 获得满足约束条件的分类函数。实验结果表明，这一方法与其它方法相比较，能够有效地利用区域特征减少训练样本的个数并且降低求解的分类函数的复杂性。

第 4 章，提出了一种基于特征空间划分的多分辨率分类器求解策略。该策略首先将训练样本所位于的区域在粗分辨率下划分成为超长方体（超立方体），然后以超长方体（超立方体）为单位求解分类边界，将边界所位于的格子逐步细分，重复这个过程直到所求解的分类器满足精度。论文分析了这种求解策略的泛化能力并给出了能够降低计算复杂性的条件。与基于聚类的分层求解方法对比，这种策略不需要首先对正例样本和负例样本聚类成为层次树，并且对于

不同类样本的混合区域的区分能力较好。实验结果表明，当数据集合具有聚类性质时，该策略能减少训练和测试阶段的计算量。

第 3 章和第 4 章从两个不同的角度分析了模式分类中的分层数据处理问题，两种方法都涉及训练样本集合在高层次（粗粒度、低分辨率、大尺度）上的表示以及由高层次到低层次（粗粒度到细粒度、低分辨率到高分辨率、大尺度到小尺度）的分解。第 3 章主要考虑的是训练样本集合在高层次上的表示对求解分类函数的影响，而第 4 章主要考虑的是如何由粗到细逐步确定分类函数。由于所采用的分类函数是与特征空间中的某些样本点是相关的，因此第 4 章中的方法也是以分解训练样本集合为基础。与第 3 章中获得的分类器不同的是，第 4 章中的方法训练出来了一组分类器，这一组分类器相对于第 3 章中的一个分类器，具有以下的一些优越性：低分辨率下面训练的分类器相对要结构简单，因此使用低分辨率下的分类器进行测试速度相对会比较快；若将测试样本也在低分辨率下进行表示，那么低分辨率下的分类器一次可以对一组测试样本判断出来类别。

第 5 章，提出了一种基于商空间合成的道路检测算法。针对校园和城区环境中道路车道线不清晰、道路边缘模糊以及行人较多等特点，提出一种融合灰度图像和彩色图像的检测结果，获得鲁棒道路边界线的方法。算法首先基于灰度图像给出若干个道路边界线的候选位置以及用于计算路面色彩的区域，然后基于彩色图像提取出路面，最后利用粒度计算领域中商空间合成的结果，将候选道路边界与路面区域的提取值进行匹配，找到最可靠的道路边界线以及路面区域。实验结果表明，采取这种策略可以鲁棒地提取出城区和校园环境中的道路。

第 6 章对整篇论文进行了总结，提出了几个未来可以继续研究的方向。

第 2 章　分类问题的定义与求解

2.1　本章引论

模式分类是目前的热点研究领域，具有广泛的应用价值。本章首先综述模式分类问题的研究内容以及进展，然后分析影响分类函数求解的几个重要因素，并且讨论分层数据处理与这些因素之间的关联。接着分析用于求解分类函数的几种风险最小化原则和训练样本的规模对分类函数求解结果的影响。最后基于所得出的分析结果，提出在模式分类问题的求解中可以采用分层数据处理的几个方面。本文的第 3、4 章设计的算法均是采用分层数据处理的方式来求解分类问题的，因此本章的讨论是第 3、4 章的基础。

本章内容如下：第 2.2 节介绍了模式分类的相关领域以及分类；第 2.3 节介绍了基于区分函数的模式分类问题的定义；第 2.4 节分析用于求解分类问题的几种风险最小化原则对求解结果的影响；第 2.5 节提出了模式分类问题中与分层数据处理相关的几个研究问题；最后是本章小结。

2.2　模式分类概述

2.2.1　模式分类的相关领域

分类能力是人的一种非常重要的认知能力。在过去的几十年中，模式分类问题的研究在人工智能领域占据了非常重要的地位，受到了世界各地研究机构的广泛关注。研究人员提出了许多算法来求解分类问题，如贝叶斯分类器，最近邻分类器、线性和非线性分类器、人工神经网络、RBF 网络、支持向量机等等[37]-[44]。这些分类算法已经成功地应用到了人脸识别、语音识别、手写字符的识别、图像处理、视频检索等领域。

模式分类是一个多学科交叉的研究领域，下面简要概述各个学科对于模式分类算法研究的影响：

1. 数学：数学中的许多分支对模式分类算法的提出和分析都起到了重要的作用，如线性代数、最优化、概率统计、泛函分析、信息理论、计算复杂性等

等。前几年引起广泛关注的支持向量机算法，就是基于统计学习理论，建立在坚实的数学基础上的[43][44]。

2. 心理学和生理学：作为人工智能领域的一个重要分支，模式分类算法同心理学和生理学的发展同样也是密切相关的。许多模式分类算法都是受到心理学和生理学中的研究成果的影响，经验性的设计了在实际应用中具有较好效果的算法。例如，基于连接主义的人工神经网络，就是模拟人脑的神经元连接方式而提出的，虽然从理论方面神经元网络模型还有很多方面没有完全分析清楚，但在过去的几十年中，但是这种模型在解决许多实际问题中显示出了较好的效果。

3. 信息科学：人工智能作为信息科学的一部分，必然使得模式分类算法同信息科学的发展密切相关。信息科学在发展过程中提出的新的需求，能够促进模式识别领域的发展。目前模式分类算法在数据库、软件工程、生物信息学等领域都具有广泛的应用。

4. 信号处理：信号处理领域的图像处理、语音识别都是目前具有较大应用价值的研究领域。许多问题的求解都可以转化为一个分类问题，采用已有的模式分类算法来解决。

数学、心理学和生理学等学科对提出新的分类模型、分析模型的性能起到了关键的作用，信息科学与信号处理学科则主要是采用已有的模式分类算法来解决应用中遇到的实际问题，这些应用又提出了新的需求进一步促进模式分类理论的发展。

以上学科影响到模式分类发展的两个重要方面：理论的突破和应用的驱动。下面简要介绍一下目前与模式分类问题相关的一些热点应用领域：

1. 图象处理和计算机视觉：与图像处理有关的应用目前非常多，许多都是当前的热点研究问题，如基于内容的图像检索[45]-[47]、人脸的检测与识别[48]、视频的检索与分析等等。图像处理中遇到的许多问题都可以首先归结成为识别问题，然后进一步转化成为分类问题，接着采用已有的分类算法去解决。多媒体信息处理的迅速发展使得图像处理是目前模式识别算法主要的应用领域之一。

2. 文字识别：文字识别是模式分类算法比较早也是比较成功的应用领域之一，许多系统目前都进入了商业应用阶段，例如光学字符识别系统（OCR）、手写字符的识别等等。典型的应用包括在邮政系统中自动地识别用户填写的邮政编码、在金融系统中自动读取支票上面的数字。

3. 语音识别：语音识别基本上是一个模式分类的任务，即通过学习，系统能够把输入的语音按一定模式进行分类。目前主流的语音识别技术是基于统计模式识别的，如隐马尔可夫模型。另外像神经网络、支持向量机等也在语音识别领域表现出了较好的效果。

前面三个研究领域都是模式分类技术所应用的传统领域，以下几个领域则是在近几年中蓬勃发展起来的。

4. 数据挖掘：数据挖掘在过去的十几年中是一直是热点研究领域，它是一门受到来自不同学科的研究人员关注的交叉性研究领域[49]。数据分类是数据挖掘中的最重要的技术之一，像决策树分类算法、贝叶斯分类算法、基于关联规则的分类算法等都成功地应用到了各类数据的挖掘中。

5. 生物信息学：生物信息的处理是近几年中的研究热点问题，它同样是一个融合多门学科的领域[50]。目前机器学习的许多算法都可以应用到生物信息的分析中，如序列对比分析、人类基因组的研究、蛋白质的研究等等。其中许多问题都可以归结为模式分类问题，采用已有的分类算法来解决，例如使用支持向量机对蛋白质的结构进行分类、对细胞形态的模式进行分类等等。

6. 与网络有关的应用：随着互联网技术的发展，对于网络上的信息处理也逐步引起了研究人员的关注，Web 数据挖掘与 Web 信息提取是近几年逐步引起广泛关注的研究问题。随着网络上的信息量逐步增大，对网页信息进行分类有着较大的研究价值与应用前景。目前许多分类算法在处理 Web 信息时都表现出了较好的分类效果。

模式分类算法的发展同以上这些应用是紧密相关的。传统的模式分类算法成功地应用到了这些应用领域，解决了许多实际问题，同时应用中出现的新问题也促进了模式分类算法的进一步发展。比如说，大规模数据对于问题的求解会带来新问题，例如求解速度慢，结果的冗余信息较多等等。传统的模式分类算法一般针对的数据集规模并不是很大，通常训练样本的规模从几十到几千个。而在目前的许多应用中，有时所获的数据集的规模非常大，包含几万到几十万个训练样本。UCI 数据集中的 Adult 和 Anonymous Web Browsing 数据集都包含有几万个数据元组[51]。分层数据处理是大规模数据问题求解的一种有效的方式，利用数据集与分类函数的层次信息必然会对基于大规模数据集的模式分类问题的求解带来一些帮助。本章后面将会讨论这个问题。

2.2.2　模式分类方法的分类

从大的方面来讲，模式分类方法分为结构模式识别和统计模式识别。本文主要关注的是统计模式识别方法，因此下面主要介绍统计模式识别方法的分类。

根据分类器的表达形式，已有的模式分类方法可以分为区分型和生成型。区分型的方法是根据训练样本训练用于区分两类样本的分类函数，也就是在特征空间中寻找超平面或超曲面来对两类样本进行划分。如线性区分函数、神经网络、支持向量机等都是区分型的模式分类方法。而生成型的方法是根据概率依赖关系构造出来分类模型，如混合高斯模型、贝叶斯分类器、隐马尔可夫模型等等。

根据分类算法的思想来源，可以将已有的模式分类算法分为经验型和理论型两种。目前已有的大部分模式分类方法都是经验型的，如神经网络、RBF 网络、最近邻分类器等等。从理论出发的分类算法并不多，贝叶斯分类器虽然有概率论的基本原理作为理论基础，但是由于预先难以得到两类样本的分布函数，因此难以应用。具有较好理论基础的分类模型是基于统计学习理论的支持向量机。统计学习理论从训练样本的个数与期望风险的关系出发，得到了一系列的理论结果，对分类算法的设计有着很好地指导作用。

根据分类算法的求解策略，主要可以分为基于经验风险最小化与基于结构风险最小化两种。早期的分类器求解算法基本上都是基于经验风险最小化原则，例如神经网络的 BP 训练算法、最近邻分类器的求解算法等等。基于经验风险最小化训练的分类器容易产生过拟合、陷入局部最小等缺点，因此近年来基于结构风险最小化的求解策略逐步受到广泛关注，如 RBF 网络、支持向量机等都是采取经验风险与置信范围的折中原则所设计的。

2.3　模式分类问题的定义

2.3.1　问题定义

求解一个模式分类问题通常包括训练和测试两个阶段[37]。除了预处理，特征提取，计算复杂性分析这几个步骤外，训练阶段主要研究如何根据给定的训练样本训练分类器，而测试阶段主要研究如何使用训练出来的分类器对测试样本集合进行测试。在本文中，不考虑预处理与特征提取部分，所研究的对象为

已提取好特征值的样本集合。由于多分类问题通过算法可以转化为二分类问题，因此下面的章节都是基于二分类问题来讨论的。

二分类问题需要解决的问题描述如下：假设给定的训练样本集合 Z 为 l 个独立同分布的满足分布函数 $P(x)$的训练样本，如式（2-1）所示。

$$Z=\{z_1,z_2,\cdots,z_l\}=\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\} \qquad (2\text{-}1)$$

其中 x_i 是训练样本的特征值，y_i 是训练样本的类别标识，y_i 的定义见式（2-2）。

$$y_i=\begin{cases}1, & x_i\in class\ \ 1\\ -1, & x_i\in class\ \ 2\end{cases} \qquad (2\text{-}2)$$

分类算法需要求解的问题是，对于给定的训练样本集合 Z，希望训练分类器 $f(x)$。当测试样本集合 T 同样满足分布函数 $P(x)$时，分类器 $f(x)$能够尽可能地对 T 中的样本给出正确的类别判断。

分类器训练和测试的过程可用图 2.1 表示[44]。

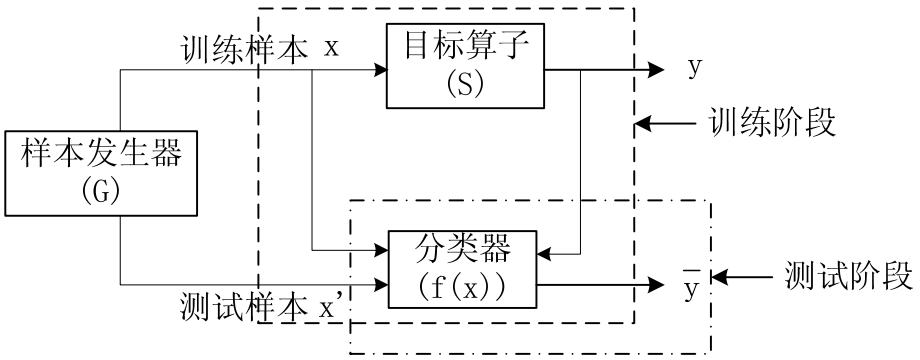


图 2.1 分类器训练和测试示意图

图 2.1 中各项的含义如下：

1. 样本发生器（G）：生成独立同分布的训练和测试样本；
2. 训练样本（x）：具有明确类别标识的样本，用于分类器的训练；
3. 目标算子（S）：对于输入的样本 x，给出 x 正确的类别标识，通常训练样本所对应的类别标识都是用户提前给定的；
4. 分类器（$f(x)$）：训练样本集合所求解得到的。对于输入的样本 x，分类

器 $f(x)$输出一个预测的类别标识，分类器训练的目的是对于 G 所产生的所有样本，$f(x)$预测的类别标识尽可能地反映样本的真实类别标识；

5. 测试样本（x’）：满足与训练样本同样的分布函数，只是测试样本所属的类别是未知的，需要由 $f(x)$来判断。

2.3.2 求解分类问题的几个关键要素

根据给定的样本集合 Z，为了求解分类函数 $f(x)$，需要考虑以下三个关键因素：

1. 分类器的表达形式：研究人员已经提出多种求解分类器的算法，例如贝叶斯分类器、线性区分函数、神经网络、支持向量机等等。不同分类算法通常对应的分类器的表达形式也是不同的，例如贝叶斯分类器与神经网络就具有不同的表达形式。一般选择了分类算法就确定了分类器的大致表达形式。同一个分类算法也可能对应多种分类器的表达形式，如选择不同的神经网络结构时，分类函数的表达形式也是不同的。因此确定分类器的表达形式，并且选择好与训练样本无关的参数是非常重要的。目前存在一些分类算法的分类器的表达形式可以统一起来，文献[38]指出输出层为线性函数的三层神经网络、RBF 网络和支持向量机都可以采用公式（2-3）的表达形式。

$$f(x)=\operatorname{sgn}\left(\sum_{i=1}^{N}\alpha_i\psi_i(x)+b\right) \qquad (2\text{-}3)$$

其中 $\{\psi_i(x)\}_{i=1}^{N}$ 表示的是所选择的基函数集合，同所选择的分类算法以及所针对的应用相关。例如：若选择支持向量机，那么 $\psi_i(x)$ 即为核函数；若选择三层神经网络，那么 $\psi_i(x)$ 为中间层的激励函数；若选择 RBF 网络，那么 $\psi_i(x)$ 即为中间层的半径基函数。关于函数 $f(x)$ 的逼近能力，已经存在丰富的研究成果[38][55]。

公式（2-3）能够表示大部分区分型的分类函数并且对于特征空间中的任意函数具有较好的逼近能力，因此在本文后续章节的讨论中，分类函数均采用公式（2-3）的表达形式。

2. 分类器的选择标准：即使选择了分类算法和分类器的表达形式，由于还有许多未定的参数，因此还有大量的分类器供选择。如何选择出与训练样本拟合得较好的分类器呢？这时候需要有基于给定的训练样本对分类器进行优劣评测的标准。目前关于分类器的选择标准有以下四种：期望风险最小化原则、经

验风险最小化原则、结构风险最小化原则、邻域风险最小化原则。

3. 训练样本的选择：虽然训练样本集合 Z 中的样本是独立同分布的，但是当采用的分类算法不同的时候，并非所有的样本对分类函数的确定均起着同等重要的作用。例如，支持向量机采用的是最大分类间隔超平面的原则，如果多考虑一些位于两类边界上的训练样本，对于获得更为精确的分类器就是显然有利的。

以上的三个因素并非是互相独立的，而是在求解分类问题时相互作用、相互影响。第 2.4.3 节将讨论求解分类函数的结构风险最小化原则同分类函数的选取以及训练样本集合的规模之间的关系。

2.4　风险最小化原则

风险最小化原则是为了判断所获得的分类器关于给定的训练样本的优劣程度的一个准则。目前有以下四种风险最小化准则，下面介绍它们的基本含义[43]：

2.4.1　期望风险最小化原则

设样本(x,y)的联合分布函数为 $F(x,y)$，再假设训练获得的分类器为 $f(x,\alpha)$，其中 α 为分类器的参数值，那么理想的情况下，当 $f(x,\alpha)$ 是最优分类器时，$f(x,\alpha)$ 应该是能够使得公式（2-4）最小的分类器[43]：

$$R(\alpha)=\int L(y,f(x,\alpha))dF(x,y) \qquad (2\text{-}4)$$

其中

$$L(y,f(x,\alpha))=\begin{cases}1, & y=f(x,\alpha)\\ -1, & y\neq f(x,\alpha)\end{cases} \qquad (2\text{-}5)$$

显然采取期望风险作为判断准则是最为理想的，但是由于预先不知道样本的分布情况，并且文献[43]中指出求取样本集合的密度函数的难度不低于求解分类函数，因此在解决实际分类上是难以根据训练样本求取期望风险的。为了判断分类器的优劣，就必须采用其它的标准。

2.4.2　经验风险最小化原则

在早期的分类算法设计中，经验风险最小化原则是最常用的一种判断准则。

它是判断训练样本集合的实际类别标识同分类器 $f(x,\alpha)$ 输出的类别标识之间的差值来确定 $f(x,\alpha)$ 的优劣的，计算公式如下：

$$R_{emp}(\alpha)=\frac{1}{l}\sum_{i=1}^{l}L(y_i,f(x_i,\alpha)) \qquad (2\text{-}6)$$

当 $R_{emp}(\alpha)$ 取值越小时，则说明 $f(x,\alpha)$ 越好。

采取经验风险最小化原则所获得的分类函数往往具有一些缺陷，可以归结为以下几点：

1. 对于训练样本集合容易过拟合：根据概率论中的大数定律，只有当样本的个数 $l\to\infty$ 时，经验风险的值才能够逼近期望风险的值。而在实际问题中，获得的训练样本的个数往往是有限的，因此使得经验风险获得最小值的分类器并不能够保证期望风险获得最小值；

2. 问题的求解是不适定的：采用这种原则来寻找最优分类器本身是一个不适定问题，也就是说求得的解不全部满足适定性问题的三个条件：存在性，唯一性和稳定性；

3. 在迭代过程中容易陷入局部最小点：早期分类器的求解算法通常都采用迭代的方法来更新模型的参数，如神经网络的 BP 训练算法就是不断地根据训练样本的特征值更新网络的权值。这种策略很容易使求得的结果是局部最小，而不是全局最小，即最后所获得的分类器对于训练样本来说，经验风险值不是最优的，这样就难以估计所获得的分类器可能的期望风险值的大小。这是由于在分类器的训练过程中没有考虑给定样本集合的全局信息，而是全部依靠样本的局部特征来确定分类边界的位置，因此所获得的分类器往往泛化能力较差。

虽然经验风险在求解分类器的时候起到了至关重要的作用，但是仅仅只考虑经验风险无法回避上面提到的几个问题。下面讨论近几年引起广泛关注的另外一种风险最小化原则，称为结构风险最小化原则。

2.4.3　结构风险最小化原则

2.4.3.1　结构风险最小化原则基本原理

结构风险最小化原则是在统计学习理论中提出的[43]，在过去的十年中受到了广泛的关注。前面已经指出，根据给定的训练样本训练分类器实际上是一个不适定问题。解决不适定问题的常用方法是采用正则化原则，即在问题的求解

过程中，通过设定一个正则化算子，同时考虑函数的复杂性和函数与给定数据的拟合程度来求解最优函数。结构风险最小化原则就是采用正则化原则来解决分类问题的，通过在分类函数自身的复杂性与经验风险之间求取一种折衷，从而避免获得过于复杂的分类器，使得所获得分类器有较好的泛化性能。正则化的思想较早就被应用到分类器的训练算法中了，如正则化神经网络、RBF 网络都是采用正则化原则来训练分类器的[38]。已有文献证明，基于结构风险最小化原则的支持向量机同 RBF 网络之间具有等价性[52][53]。

同基于经验风险最小化原则的求解条件不同的是，基于结构风险最小化原则的求解条件，除了经验风险项之外，还有一个正则化项。一般通过设定一个用于衡量函数复杂程度的算子 P，采用泛函 Pf 表示函数 f 的复杂程度[43][44][52][53]。基于结构风险最小化的求解表达式如公式（2-7）所示。

$$R_{reg}(\alpha)=R_{emp}(\alpha)+\frac{1}{2}\lambda\|Pf\|^2 \tag{2-7}$$

其中 λ 是一个参数。采用这种原则的分类算法就是求取使得 $R_{reg}(\alpha)$ 最小的函数来作为分类函数。假设分类问题的损失函数如公式（2-5）所示，Vapnik 给出了两个重要的定理如下：

定理 2.1[43]　当训练样本集合的规模为 l 时，对于给定的参数 η，下面的不等式以概率 $1-\eta$ 成立。

$$R(\alpha)\le R_{emp}(\alpha)+\frac{\varepsilon(l)}{2}\left(1+\sqrt{1+\frac{4R_{emp}(\alpha)}{\varepsilon(l)}}\right) \tag{2-8}$$

其中

$$\varepsilon(l)=4\frac{h\left(\ln\frac{2l}{h}+1\right)-\ln\eta/4}{l} \tag{2-9}$$

对于给定的训练样本集合 Z，设 α_l 是使得公式（2-10）最小的参数值：

$$R_{emp}(\alpha_l)=\min_{\alpha\in\Lambda}R_{emp}(\alpha)=\frac{1}{l}\sum_{i=1}^{l}L(y_i,f(x_i,\alpha)) \tag{2-10}$$

定理 2.2[43]　对于分类问题，下面的不等式以概率 $1-\eta$ 成立。

$$R(\alpha_l) < R_{emp}(\alpha_l) + \varepsilon(l)\left(1+\sqrt{1+\frac{4R_{emp}(\alpha_l)}{\varepsilon(l)}}\right) \qquad (2\text{-}11)$$

$\varepsilon(l)$的定义同公式（2-9）。

公式（2-8）可以简单写成公式（2-12）的形式。

$$R(\alpha) \le R_{emp}(\alpha) + \frac{1}{l}\Phi(h) \qquad (2\text{-}12)$$

从定理 2.1 和定理 2.2 可以得到以下结论：

1. 在定理 2.1 中，由于 h 表示的是函数的 VC 维，也就是衡量分类函数的复杂程度的一个指标，因此当训练样本的个数固定时，经验风险和 h 均对控制期望风险的界起作用。结构风险最小化原则，就是将函数按照复杂程序排序，然后选取对训练样本的经验风险和函数复杂程度综合最好的函数作为最优函数。

2. 当样本的个数$l \to \infty$时，公式（2-8）不等式右边的后一项趋向于 0，那么期望风险主要是由经验风险所控制。一般当样本的个数 l 增加的时候，能够提供关于样本分布更多的信息，所获得的分类器的性能应该好于样本个数较少的情况。

3. 定理 2.2 表明，当参数使得经验风险取得最小时，并不一定能够保证基于该参数的期望风险也能够取得最小，公式（2-11）不等式右边的后一项同样影响着期望风险的值。

2.4.3.2　函数选择与结构风险最小化原则

第 2.3.2 节给出了决定分类函数的三个关键因素：分类函数的表示形式，求解策略以及训练样本集合的选择。结构风险最小化原则虽然归于求解分类函数的求解策略，但是它同另外两个因素也是紧密相关的。前面讨论了结构风险最小化原则由两个因素组成：选择的函数的复杂程度与所选择的函数在给定样本集合上面的经验风险。下面讨论函数集合的结构。

设函数$Q(z,\alpha) = L(y, f(x,\alpha))$，$\alpha \in \Lambda$。选择不同的参数$\alpha$时，$Q(z,\alpha)$为不同的函数。将函数按照嵌套子集的形式进行排列，如公式（2-13）所示[43]。

$$S_1 \subset S_2 \subset S_3 \subset \cdots \subset S_n,..., \qquad (2\text{-}13)$$

函数的 VC 维为：

$$h_1 \subset h_2 \subset h_3 \subset \cdots \subset h_n, ..., \qquad (2\text{-}14)$$

公式（2-3）中的函数$\{\psi_i(x)\}_{i=1}^N$不同，所构成的函数集合的复杂程度也是不同的，即 VC 维h会不同。当选择了基函数集合$\{\psi_i(x)\}_{i=1}^N$后，函数集合的范围一般还是相当大的，可以再次按照复杂程度的不同进行结构划分。以支持向量机为例，支持向量机算法并非是在一个较宽函数域上面的结构风险最小化算法。当确定核函数$K(x,x_i)$之后，基函数集合即限定在$\{K(x,x_i)\}_{i=1}^l$，因此分类函数可在公式（2-15）所表示的函数范围内选择。

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{l} \beta_i K(x,x_i) + \beta_0\right) \qquad (2\text{-}15)$$

支持向量机的训练是在上述函数集合中基于给定的训练样本再次根据结构风险最小化原则确定参数$\beta_i (i = 0,1,2,...,l)$。为了能够在更宽的函数范围内选择函数，现在基于支持向量机的分类算法一般都包括模型选择和分类函数的训练两个部分[54]。举例说，若采用高斯核函数，则首先通过算法选择σ的值，然后再训练分类器。综合考虑模型选择算法和支持向量机的训练算法实际上是一个二级的基于结构风险最小化的分类器训练算法：

第一级，按照函数的复杂程度选择函数模型。也就是对于公式（2-13）所表示的嵌套结构，模型选择的任务就是从中选取合适的分类器。

第二级，与训练样本集合相关的函数的复杂程度。当从式（2-13）中选择出一个分类器集合后，每个分类器相对于给定的训练样本的置信范围也是不同的。支持向量机中提出的最大分类间隔的概念就是表示这种差异的，当超平面距离最近的训练样本的间隔越大时，则认为对应的分类器的置信范围越高，在这一级中函数的复杂程度是由公式（2-15）中的x_i和$\beta_i (i = 0,1,2,...,l)$所确定的。

2.4.3.3　训练样本规模与期望风险的界

在结构经验风险化原则受到广泛专注之前，解决分类问题多采用的是经验风险最小化原则。在采用经验风险最小化原则时，一般认为样本的数目越多越能获得更为可靠的分类器。这个结论可由定理 2.1 所证实，当样本的数目$l \to \infty$时，并且选择的分类函数的 VC 维有界时，公式（2-12）右边的后一项趋近于 0，这个时候所选择的分类函数的期望风险基本上是由经验风险所控制的。而在实

际应用中所获得样本的数目一般都是有限的，在这种情况下，公式（2-12）给出了期望风险界的一个重要结论。也就是期望风险的界是由经验风险 $R_{emp}(\alpha)$、样本的规模 l 以及衡量分类函数复杂程度的因素 VC 维 h 三者一起控制的。下面简要分析一下训练样本的规模同期望风险的界的关系：

1．在公式（2-12）中，当训练样本的规模 $l \to \infty$，分类函数的 VC 维 h 有界时，$\Phi(h)/l \to 0$，那么当分类函数 $f(x,\alpha)$ 使得经验风险 $R_{emp}(\alpha)=0$ 时，可以将期望风险 $R(\alpha)$ 的界控制为 0，这时候期望风险达到最小值。因此当训练样本的规模很大时，是可能找到一个分类函数，使得期望风险达到 0。上面的讨论可以解释为什么当采取经验风险最小化原则时，要求训练样本的规模越大越好；

2．当训练样本的规模 l 很小时，由于 h 的值不为 0，$\Phi(h)/l$ 的值也不为 0，因此即使可以选择分类函数使得 $R_{emp}(\alpha)=0$，也无法控制期望风险的界为 0。因此当训练样本的规模比较小时，样本集合提供的关于样本的分布信息也比较少，这时候根据训练样本集合所训练的分类函数实际上是无法使得期望风险的界为 0 的，因此规模较小的样本集合提供的信息量是不够的；

3．对于两个规模分别为 l_1 和 l_2（$l_1 << l_2$）的训练样本集合 Z_1 和 Z_2（$Z_1 \subset Z_2$），下面讨论基于训练样本集合 Z_1 和 Z_2 所训练而得的分类函数期望风险的界。假设给定参数 η，求解的基于训练样本集合 Z_1 的期望风险界的最小值为 R_1^*，假设对应的分类器为 $f(x,\alpha_1^*)$。对于分类器 $f(x,\alpha_1^*)$ 来说，当训练样本的规模 l 增大时，项 h/l 的值是减小的，若经验风险的值增加的并不多，即 $R_{emp}(\alpha_1^*,l_2)+\Phi(h_1^*)/l_2 < R_{emp}(\alpha_1^*,l_2)+\Phi(h_1^*)/l_2$，这个时候期望风险的界会降低。而一般当训练样本增多时会选择更为复杂的分类函数，假设基于训练样本集合 Z_2 求解的期望风险界的最小值为 R_2^*，对应的分类函数为 $f(x,\alpha_2^*)$，当 $R_{emp}(\alpha_1^*,l_1) \approx R_{emp}(\alpha_2^*,l_2)$ 并且 $h_2/l_2 < h_1/l_1$ 时，$R_2^* < R_1^*$。这个条件在训练样本的规模不是很大时是很容易的满足的。当训练样本的规模增大到一定程度时，随着样本规模的增大所求解的分类函数的复杂程度的增加幅度一般远小于训练样本规模的增加幅度，即 $h_2/l_2 < h_1/l_1$，但是经验风险的值会随着样本规模的增加而增加，这时候期望风险界的值即使会降低，降低的幅度也不大，因此当样本的规模大到能够反应出样本的分布信息时，继续增大训练样本的规模的对于获取更好的分类器性能来说意义并不大。通过上面的分析可知，大规模的训练样本集合相对于小规模的训练样本集合来说，是可以降低期望风险的界的，因此讨论大样本集合下的模式分类问题是有意义的。

2.4.3.4　训练样本规模与函数选择

设训练样本集合为 $Z=\{z_i\}_{i=1}^{l}=\{(x_i,y_i)\}_{i=1}^{l}$，将训练样本划分成为增量式的结构，如公式（2-16）表示。

$$Z_1 \subset Z_2 \subset Z_3 \subset \cdots \subset Z_K = Z \tag{2-16}$$

其中 $l_k=|Z_k|$ 表示 Z_k 中的训练样本的个数。

选取训练样本 Z_i 进行分类器的训练，假设最终选择的函数位于函数集合 S_j，将函数的参数表示为 $\alpha_{i,j}$，那么考察公式（2-17）所表示的期望风险泛函的界。

$$R(\alpha_{i,j}) \le R_{emp,i}(\alpha_{i,j}) + \frac{1}{l_i}\Phi(h_j) \tag{2-17}$$

其中 $R_{emp,i}(\alpha_{i,j})$ 为经验风险项，$\Phi(h_j)/l_i$ 为置信范围项。

当训练样本集合的规模增大时，即选择 Z_{i+1} 作为训练样本集合，那么使得结构化风险泛函比较小的分类函数可以有以下几种选择情况：

1. 选择更简单的函数作为分类器，即选择 $S_k(k<j)$ 中的函数。虽然选择简单的分类函数会使得经验风险项的值变大，但置信范围项的值会变小，因此当训练样本增加的时候，选择更为简单的分类器达到结构风险最小化也是可能的。不过，这种情况在样本数目很多和很少的时候都不太容易发生。当样本数量比较少时，虽然训练样本都是独立同分布的，但是新增加的样本会有许多不在原有训练样本的较小邻域内，因此采用更简单的分类函数会使得经验风险的值很大，所以一般当训练样本较少时，随着训练样本的增加，最优分类器一般是更为复杂的分类器。而当训练样本的数目很大时，l 的值是比较大的，这时置信范围项起的作用会比较小，主要是由经验风险项起作用，所以在这种情况下当训练样本的个数继续增加时，为了保证较小的经验风险，不会选择更为简单的分类器。

2. 选择 S_j 中的函数作为分类器。假设基于 Z_{i+1} 所训练的分类器的期望风险的界不大于基于 Z_i 所训练的分类器的期望风险的界。那么若选择参数 $\alpha_{i,j}$ 作为分类器的参数，由于 $Z_i \subset Z_{i+1}$，那么有 $R_{emp,i}(\alpha_{i,j}) \le R_{emp,i+1}(\alpha_{i,j})$，由于 l 的增加使得置信范围项变小，若在函数集合 S_j 中存在以 $\alpha_{i,k}$ 作为参数的分类函数，使得 $R_{emp,i+1}(\alpha_{i,k}) \le R_{emp,i+1}(\alpha_{i,j})$，那么最优分类函数有可能在函数集合 S_j 中选择。

3. 选择更为复杂的函数作为分类器，即选择 $S_k(k>j)$ 中的函数。当训练样本增大时，这种情况发生的可能性是比较大的，即选择更为复杂的分类器来降

低训练样本的经验风险。

上面分析了增加训练样本的个数时，最优分类函数可能满足的一些情况。下面利用公式（2-18）表示的分类函数进一步讨论。

$$f(x)=\sum_{i=1}^{m}\beta_i\Phi(x,x_i)+\beta_0\ , \qquad (2\text{-}18)$$

由公式（2-18）构成的函数集合比由公式（2-3）所构成的函数集合的范围更窄。函数 $f(x)$ 的复杂程度除了同基函数 $\Phi(x,x_i)$ 有关外，还同点集 $\{x_i\}_{i=1}^{m}$、系数 $\beta_i(i=0,...,m)$ 以及参数 m 相关。$\beta_i(i=0,...,m)$ 可以仅与 $\{x_i\}_{i=1}^{m}$ 相关，也可以同时与其它因素相关。例如，在支持向量机中，$\{x_i\}_{i=1}^{m}\subseteq Z$，$\beta_i(i=0,...,m)$ 可由集合 $\{x_i\}_{i=1}^{m}$ 所确定。而对于 RBF 网络，通常是采用聚类等方法得到 $\{x_i\}_{i=1}^{m}$，$\{x_i\}_{i=1}^{m}$ 可以不包含于训练样本集合，参数 $\beta_i(i=0,...,m)$ 由 $\{x_i\}_{i=1}^{m}$ 和训练样本集合所共同确定的。

RBF 网络和支持向量机所求解的分类函数都可以表示成为公式（2-18）的形式。这两种分类算法的置信范围项的计算公式如公式（2-19）表示。

$$\|Pf\|^2=B^{'}GB \qquad (2\text{-}19)$$

其中 $B^{'}=[\beta_1\quad\beta_2\quad\cdots\quad\beta_m]$，G 是一个 $m\times m$ 的矩阵，$G(i,j)=\Phi(x_i,x_j)$。

设最后获得的最优分类器中的系数 $\beta_i\neq 0(i=1,...,m)$。那么是否 m 越大，分类器的复杂程度就越高呢？一般情况下，函数的复杂程度越高，逼近能力就越好。关于函数的逼近能力目前已有许多文献进行讨论[38][55][56]。当集合 $\{x_i\}_{i=1}^{m}$ 中存在过多的邻近点，并且这些邻近点对区分不同类的样本并不起关键作用时，m 的增大并不能显著提高函数的逼近能力。而 m 的值越大，RBF 网络中间层次结点数目越多，SVM 的支持向量的个数越多，使得分类器的结构很复杂，训练和测试的计算复杂性大。在分类器的训练过程中，m 的值通常同训练样本的个数相关，一般训练样本的规模越大，会导致 m 的值越大。因此去除集合 $\{x_i\}_{i=1}^{m}$ 中冗余的点，对分类器的求解和样本的测试均会起到帮助作用。

2.4.4　邻域风险最小化原则

Vapnik 在文献[44]的最后章节提出了邻域风险最小化原则。邻域风险最小化原则与结构风险最小化原则类似，只是在经验风险的计算中采用训练样本点某一邻域内的平均分类值来代替样本点所在位置的单一分类值，它所基于的基本

假设是样本集合所满足的密度函数在特征空间的邻域内的取值是连续的，问题求解如公式（2-20）所示。

$$R_{reg}(\alpha)=\frac{1}{2}\|w\|^2+\lambda\sum_{i=1}^{l}\xi_i \qquad (2\text{-}20)$$

满足公式（2-21）所表示的求解条件。

$$y_i(\frac{1}{N}\sum_{k=1}^{N}f(x_{i_k}))\geq 1-\xi_i,\ \ \xi_i\geq 0,\ \ \ i=1,2,...,l \qquad (2\text{-}21)$$

也就是每个训练样本不仅仅是特征空间中一个孤立的点，而是考虑了这个训练样本周围区域的信息。在采用支持向量机作为基本的分类器时，关于两个样本点的核函数的计算也应该能够考虑到邻域的信息，因此在计算核函数时，是取一个邻域内核函数值得均值，单邻域核与双邻域核的计算如公式（2-22）所示[44]。

$$\begin{aligned}L(x_i,x)&=E_{x'\in v(x_i)}K(x,x')=\int K(x,x')p(x'\mid x_i,r_i)dx'\\ M(x_i,x_j)&=E_{x\in v(x_i)}E_{x'\in v(x_j)}K(x,x')\end{aligned} \qquad (2\text{-}22)$$

其中 $v(x)$ 表示样本点 x 的邻域。

由于在第 3 章中提出新方法时，要与基于邻域风险最小化原则所求解出来的分类器进行对比，因此下面针对基于邻域风险最小化原则的分类器的特点给出以下的性质：

性质 2.1 邻域风险最小化原则是采用给定训练样本周围区域的经验风险的平均值来代替该样本的经验风险的值的。

证明： 给定一个样本点 x，设样本 x 的类别标识为 y，设样本 x 的邻域为 $v(x)$，对于 $\forall x_i\in v(x)$，控制经验风险值的公式为 $yf(x_i)\geq 1-\varepsilon_i$。若 v(x)内共有 N 个样本点，那么在邻域 v(x)内经验风险的平均值为 $\varepsilon=(1/N)\sum_{i=1}^{N}\varepsilon_i$。由于

$$\sum_{i=1}^{N}\varepsilon_i\geq N-y\sum_{i=1}^{N}f(x_i) \qquad (2\text{-}23)$$

可得：

$$y\frac{1}{N}\sum_{i=1}^{N}f(x_i)\geq 1-\varepsilon \qquad (2\text{-}24)$$

因此邻域风险最小化原则是采用一个邻域内经验风险的平均值去代替某一样本

点的经验风险的值的。　　证毕。

当$N \to \infty$时，有：

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \approx \frac{1}{\Omega}\int_{v(x)} f(x')dx' \qquad (2\text{-}25)$$

假设函数 $f(x)$在区域 $v(x)$上是连续的，那么根据多重积分的中值定理，存在 $x_j \in v(x)$，满足：

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) \approx \frac{1}{\Omega}\int_{v(x)} f(x')dx' = f(x_j) \qquad (2\text{-}26)$$

因此，公式（2-24）可以写为：

$$yf(x_j) \geq 1-\varepsilon, \quad x_j \in v(x) \qquad (2\text{-}27)$$

由上述公式可以看出，基于邻域风险最小化原则训练的分类器极有可能穿过划定的区域，而非将该区域作为一个独立的训练单位来考虑。

2.5　模式分类中的分层问题求解

在第 1 章中介绍了分层数据处理的一些基本方法，本章前面讨论了模式分类问题的基本定义以及求解分类函数的基本策略，下面分析研究模式分类中的分层数据处理的意义以及可能采取的一些基本策略。

2.5.1　研究意义与研究现状

针对模式分类问题，有多种层次划分的方法，例如决策树可以看作是对样本的属性值的一种层次划分。我们在本文中所研究的模式分类中的分层数据处理针对的是当训练样本集合的规模比较大时，由样本在空间中分布的统计特征所构成的层次性。在第 2.5.2 节中分析模式分类中的分层数据处理方式的时候将会详细介绍将大规模训练样本集合进行层次划分的含义。下面首先介绍一下模式分类中的分层数据处理的意义。总结为两条：

1．由应用所驱动：在第 2.2 节中介绍了与模式分类相关的研究领域，指出了随着计算机硬件性能的逐步提高，计算机应用学科在过去的十年中快速地发展起来，产生了许多热点研究领域，例如图像处理、计算机视觉、文字识别、语音处理、数据挖掘、生物计算以及网络应用等等。模式分类算法在这些应用

中非常有效地解决了许多实际问题，取得了巨大的经济效益。同时这些应用带来的新的问题也进一步促进了模式分类理论的发展。其中一个重要问题就是海量数据所带来的新问题。传统的模式分类算法多是针对小样本集合来研究的，所关注的主要问题是构造错误率较小的分类器。而针对大规模样本集合来说，对分类的速度、分类器的精确度以及分类器的复杂程度等指标综合提出了较高的要求。因此，研究针对大规模数据的分类器的设计是具有实用意义的。我们在第 1 章中介绍了分层数据处理的有关理论，分层数据处理是在大规模论域上进行问题求解的一种有效方式，因此研究模式分类问题中的分层数据处理同样是具有实用意义的；

2．理论意义：在第 2.4.3.3 节中初步分析了训练样本的规模与期望风险的界的关系。从分析的结果可以看出，基于大规模训练样本集合训练分类函数是可以获得较小的期望风险泛函的界的。因此，以分层数据处理的理论为指导，研究模式分类中的分层数据处理具有重要的理论意义。

与本课题相关的国内外的研究成果分为以下几个方面：

1．分层数据处理的理论，如多分辨率分析、多尺度分析和粒度计算等都是当前的热点研究领域。模式分类同样也是目前受到广泛关注的研究课题。虽然目前还没有系统地关注模式分类中的分层数据处理的研究成果，但是研究人员在多分辨率分析、粒度计算同模式分类相结合的方面已经做了较多的研究工作。如将小波理论同神经网络结合构成的多分辨率神经网络[68]，基于稀疏网格的多分辨率神经网络[67]以及基于粗糙集的分类[10][11]等等。在解决实际的应用问题中，将多分辨率的、层次化的思想同模式分类相结合也有较多的研究成果[79]-[91]。

2．在目前的机器学习研究领域，针对规模较大的样本集合所提出的稀疏分类器（sparse classifier）是目前的一个研究热点。基于约减集（reduced set）的方法是其中一类重要方法[94]-[97]。通过减少支持向量的个数可以降低求解的分类函数的复杂程度，有助于提高样本分类时的速度，这同样也是采用分层数据处理希望追求的目标之一。

在第 3、4 章介绍具体提出的算法时将会详细介绍与算法相关的参考文献。

2.5.2　模式分类中的分层数据处理方式

根据所讨论的模式分类问题的基本概念和求解方法，结合第 2.3.2 节中给出的与模式分类问题求解相关的三个要素以及第 1 章中所讨论的分层数据表示和

分层问题求解的的基本方法，简单讨论一下模式分类问题中可能存在的一些分层数据处理方式。

2.5.2.1　粗粒度下论域的划分

对于分类问题来说，有两个论域在求解时是必须要关注的，一个是训练样本集合所构成的论域，一个是候选分类函数集合所构成的论域。在第 1 章介绍粒度计算时讨论了如何将论域在粗粒度下进行表示，以及将问题在粗粒度上进行求解对整个问题求解的帮助。下面针对训练样本集合构成的论域以及分类函数构成的论域讨论在粗粒度上面进行模式分类问题求解有利的一些方面。

1. 样本集合论域：当给定一个训练样本集合 Z 时，将训练样本集合在粗粒度下进行表示是否会对分类问题的求解有帮助呢？设 x 表示一个训练样本，$v(x,r)$ 表示 x 以 r 为半径的邻域。假设两类训练样本的分布函数为 $P_1(x)$ 和 $P_2(x)$，若在邻域 $v(x,r)$内有 $P_1(x) >> P_2(x)\ (x \in v(x))$，则可以简单的记为 $v(x,r) \in class\ \ 1$。这时当训练样本的数目逐步增大时，必然会有越来越多的属于第一类的训练样本落入到 $v(x,r)$内，而在目前大部分的分类算法中，在训练过程中需要考虑每个训练样本对分类器的构成的贡献，即 $v(x,r)$内的样本都要逐一考虑。若对训练样本集合所构成的论域在粗粒度下面进行表示，那么 $v(x,r)$内的训练样本在粗粒度下是可以作为一个训练样本去参与分类器的训练的，这样相对于考虑 $v(x,r)$内的每个训练样本来说，应该能够降低分类器训练的计算时间。本文在第 3 章将详细讨论这个问题。另外，整个训练样本集合也可以采用从粗粒度到细粒度的表示。在第 1 章中指出，粗细粒度的划分可以根据语义概念进行划分，例如地图的粗细粒度划分，但针对训练样本集合来说，这种粗细粒度的划分是基于样本分布的统计信息的。设 $X = v(x,r)$，在粗粒度下，训练样本集合可以被划分成为以下三种方式：

1) 仅属于类别 1：即对于 $\forall x \in X$， $P_1(x) >> P_2(x)$，称 $X \in class\ \ 1$；
2) 仅属于类别 2：即对于 $\forall x \in X$， $P_1(x) << P_2(x)$，称 $X \in class\ \ 2$；
3) 无法判断类别：即对于 $\forall x \in X$，不归于以上两类，称 $X \in undecided$。

在粗粒度下归于上述第三种的样本，无法确定类别，是需要在更细的粒度下进一步细化的。采取这种方式，将训练样本集合从粗粒度到细粒度划分成为层次关系，然后基于所划分的层次进一步讨论分类问题的求解，第 4 章讨论多分辨率分类算法的时候，将详细讨论这个问题。

2. 候选函数集合的论域：结构风险最小化原则对候选分类函数按照复杂程度进行嵌套划分，表示成为层次的关系，可以看作是对候选分类函数由粗到细进行划分。

2.5.2.2　多分辨率分类器求解

第 1 章中简单介绍了多粒度与多分辨率求解所带来的优越性。对于模式分类问题，由于样本集合所构成的论域和候选函数集合所构成的论域都可以实现由粗粒度到细粒度的划分，因此模式分类问题的分层次求解也应该有两种：基于训练样本集合的分层次求解和基于候选函数集合的分层次求解。这两种分层次求解方式实际上是可以结合起来的。下面分别讨论这两种分层次求解的基本含义：

1. 基于训练样本集合的分层次求解

前面讨论了可将训练样本集合根据样本的分布情况在粗粒度下面进行表示，这种表示方式是有可能利于分类问题的求解的。本文中讨论的分类器的解的形式为公式（2-18）所表示的区分函数，前面分析了 $f(x)$ 的求解涉及到集合 $\{x_i\}_{i=1}^m$ 的求解与系数 $\beta_i(i=0,...,m)$ 的求解。对于 $\{x_i\}_{i=1}^m$ 的求解来说，将训练样本在粗粒度上面进行表示是有助于快速找到合适的特征点的。以支持向量机为例，假设要确定最终决定分类边界的那些支持向量，可以在粗粒度下进行寻找，在粗粒度下不位于分类边界周围的那些样本，认为支持向量不太可能在这里面，只考虑那些位于分类边界周围的粗样本，将它们逐步细分，再确定距离分类边界较近的样本，重复这个过程，直至能够精确地确定分类边界线。另外将训练样本在粗粒度上面表示还有助于去掉 $\{x_i\}_{i=1}^m$ 中的冗余点。假设 $u_1,u_2 \in v(x,r)$，并且 $v(x,r)$ 内的样本都来自于同一类，若 $d=\|u_1-u_2\| << r$，那么若 $u_1,u_2 \in \{x_i\}_{i=1}^m$，实际上是冗余的，并且会增大 m 的值。若仅采用若干个点来表示粗粒度上面的样本，实际上是可以解决集合 $\{x_i\}_{i=1}^m$ 中的冗余点问题。

2. 基于候选函数集合的分层次求解

同样候选函数集合也可以根据函数的某些特性划分成为由粗到细的层次。目前的某些模式分类算法实际上都分为两步：模型选择和训练获得模型参数。例如，基于神经网络的模式分类算法先根据样本的一些特性确定网络的结构，支持向量机算法先确定核函数。将候选函数集合根据函数的复杂程度不同划分成为层次模式，并且根据划分的层次选择合适的分类函数，必然会有助于分类

问题的有效解决，是一个值得探讨的研究课题。

2.6　本章小结

本章分析了求解模式分类的几个关键因素，重点介绍了用于求解分类函数的几个风险最小化原则，分析了结构风险最小化原则与分类函数选择、训练样本个数之间的关联，给出了邻域风险最小化原则的一些基本的性质，并在最后提出了分层次理论同模式分类问题可能存在的一些关联。本章是第 3、4 章的基础，第 3、4 章的理论和算法都是建立在本章的分析之上的。本章的分析相对比较简单，若能够采取定量的分析，结果会更有说服力。

第 3 章　基于区域风险最小化原则的分类器设计

3.1　本章引论

第 2 章的最后分析了模式分类问题中可能存在的一些分层数据处理方式，其中指出了可以尝试在粗粒度下考虑分类问题的求解，也就是当训练样本在局部具有聚类的性质时，可以考虑将那些聚集在一起的原始训练样本作为粗粒度下的一个训练样本去训练分类器。本章沿着这个思路提出了一种新的风险最小化原则，称为区域风险最小化原则，算法首先将训练样本集合在粗粒度上面进行表示，然后基于这些以区域为基本单位的训练样本，采用结构风险最小化原则训练分类函数。

本章结构如下安排：第 3.2 节介绍要解决的基本问题和相关工作；第 3.3 节给出基于区域风险最小化原则的模式分类问题求解模型；第 3.4 节介绍了区域的两种表示方式：超立方体和超球；第 3.5 节给出了问题的求解算法，推导出当区域可以采用中心点和半径表示时，只需要在原始的序贯最小优化算法上面稍作改动即可满足要求；第 3.6 节的实验结果表明了本章提出的算法的有效性。

3.2　基本问题与相关工作

3.2.1　问题描述

在实际应用中所接触到的训练样本集合往往局部具有聚类性质，即在特征空间的局部，来自于同一类的样本聚集在一起。某些分类器的构造是在特征空间中寻找能将两类样本区分开的超曲面，例如，支持向量机，神经网络，线性区分函数等；而另外一些分类器的构造则利用样本集合的局部特性，例如混合高斯模型，邻域覆盖算法[98]等等。这两种划分训练样本集合的方式同图像处理领域的区域分割有所相似，图像分割也分为两种方式：寻找不同区域的边缘和利用像素的局部相似性生成分割结果[99]。而模式分类问题同图像处理中的区域分割问题还有所不同，对于区域分割来说，由于自然图像的结构一般较为复杂，很难获得一个表达形式明确的函数来划分出不同的区域，而模式分类问题都可

以转化为区分两类样本。因此基于寻找分类边界的模型具有分类器表达形式简单，泛化能力好，测试过程简单等优点。而利用样本的局部特性构造分类器的方法则利用了更多的样本集合本身的分布特性。因此，将这两种模型结合起来，有助于模式分类问题的解决，已经存在许多研究成果关注这个问题，后面的相关工作中将会提到。下面从求解区分函数的角度去讨论二者的结合问题。在第 2 章中，通过分析得出了以下一些结论：

1. 当训练样本增多时，同类样本聚集在一起的概率增大，因此可以考虑将一组原始训练样本用一个粗粒度下的样本来表示，这种表示有利于减少训练样本的个数；

2. 当训练样本增多时，一般需要更为复杂的分类器去区分两类样本。对于公式（2-18）中的点集$\{x_i\}_{i=1}^{m}$，有可能冗余的点增多，将样本集合在粗粒度上面进行划分，可以易于判断冗余的点并且去除。

在本章考虑如下的一种求解策略，如图 3.1 所示。

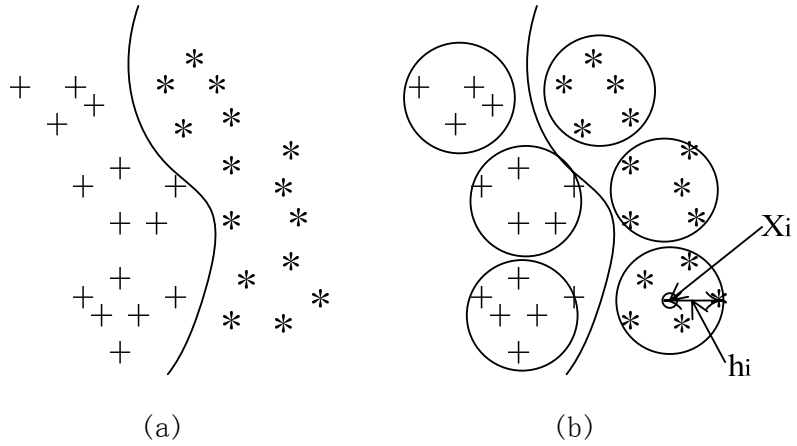
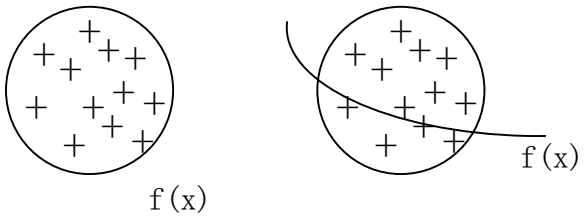


图 3.1　将训练样本进行区域划分示意图

图 3.1(a)表示原始的训练样本集，图 3.1(b)表示的是将训练样本集在粗粒度下的划分结果。虽然第 2 章介绍的邻域风险最小化原则也是在确定分类边界线时考虑一个区域内的训练样本，但它同区域风险最小化原则是有本质不同的。讨论邻域风险最小化原则时分析了采取邻域风险最小化原则来训练分类器实际上还是相当于在计算核函数的值与经验风险的值时采用样本邻域内的某个点去代替计算，而非原始的训练样本点，因此最后求取的分类边界线是很有可能穿过邻域的，如图 3.2(b)所示。而本章所希望达到的区域风险最小化原则是希望在



(a)　　(b)

图 3.2　区域风险最小化与邻域风险最小化在确定分类边界上的对比

粗粒度下能够将整个区域作为一个独立的样本点，因此分类边界线是应该绕过该区域的，如图 3.2(a)所示。

3.2.2　相关工作概述

首先介绍与本章提出的算法相关的研究成果。将特征空间中的训练样本进行划分可以采用聚类技术，已经有一些研究人员在聚类技术同支持向量机的结合方面作了一些研究工作。Wang 等简单的将正例样本集合和反例样本集合分别聚类，实验表明可以减少训练样本的个数，由于算法并未考虑每个聚类的大小等因素，仅仅使用聚类的中心作为训练样本，获得的分类器无法利用大样本数据提供的信息，难以保证为最优分类器[57]。Yu 等提出了一种基于分层次聚类的支持向量机，即首先将正例样本和反例样本聚类成为一个层次结构，然后自顶向下训练分类器，将支持向量机所在的聚类不断地拆分，直到训练的分类器满足要求的精度[74]。Boley 等分析了分析了样本集合的聚类性质以及聚在一起的样本在序贯最小优化算法的求解过程的作用，指出参数的求解常在一个局部空间内进行搜索，Boley 的方法类似于 Yu 的方法，即首先在大的聚类上面求解模式分类问题，然后将非支持向量的部分用较少的点来表示，重点考虑非有界支持向量所在的聚类。这种方法提出的目的为根据样本的局部特性来减少 SMO 算法求解时训练样本的个数，但是求解结果的支持向量的个数一般很难减少[58]。Pavlov 提出的方法首先将训练样本集合采取压碎（squashing）的方法进行处理，减少训练样本的个数，然后再使用支持向量机进行训练[59]。Shih 提出了一种基于统计的数据约简算法，考虑了样本划分的统计信息以及划分规模的大小，该算法是基于 Rocchio 分类模型的，而非本文所讨论的基于区分函数的算法[60]。

前面介绍的都是将训练样本在特征空间中进行处理的方法，方法的研究目的基本上都为减少训练样本的个数，加快分类器的训练。另外还有一类方法同

样涉及特征空间的划分，但是方法是关注分类函数的表示的。如 Garcke 提出的以稀疏网格表示的基函数[67]，Bakshi 提出的小波神经网络[91]，Simpson 提出的模糊最小—最大神经网络等等[66]。这部分研究成果对特征空间的划分是为了表示不能复杂程度的分类函数。

本章提出的方法出发点与上述方法是不同的，本章提出的方法是希望训练样本的表示以及分类器的训练都以特征空间区域作为基本单位，这种表示方法应该不仅能够减少训练样本的个数，同时能够较简单的表示分类函数。因此涉及的一个重要问题就是区域的表示，如图 3.1(b)与图 3.2(a)所示。

3.3　模型描述

在构造基于区域风险最小化原则的分类器时，同传统的分类算法一样需要考虑训练样本的选择，候选区分函数构成的集合，经验风险的计算以及置信范围的计算四个因素，下面从这四个因素出发讨论区域风险最小化原则。假设对于给定的训练样本集合 $Z=\{(x_i,y_i)\}_{i=1}^{l}$。

3.3.1　训练样本

根据前面的讨论，首先要将原始的训练样本集合变换到粗粒度下，即每个训练样本集合为特征空间中的一个封闭区域，例如，比较简单的可以采用超球和超立方体来表示。假设变换后的训练样本集合如式（3-1）所示。

$$Z^{'}=\{z_i^{'}\}=\{(X_i,Y_i)\},\qquad i=1,2,...,n \tag{3-1}$$

其中各个参数的含义如下：

1) X_i：表示的是以区域为单位的训练样本；

2) Y_i：表示的是区域的类别标识，定义如公式（3-2）。

$$Y_i=\begin{cases}1, & x_i\in class\ \ 1\\ -1, & x_i\in class\ \ 2\end{cases} \tag{3-2}$$

显然，这种表示方式是无法直接用于分类函数的训练的，那么必须使用一些参数去表示，若 X_i 是采用超球所表示的区域，那么，可以将训练样本集合表示为式（3-3）。

$$Z^{'}=\{z_i^{'}\}=\{(c(X_i),Y_i,h_i)\},\qquad i=1,2,...,n \tag{3-3}$$

其中各个参数的含义如下：

1) $c(X_i)$：表示的是 X_i 所代表的超球的中心点；
2) Y_i：表示的是区域的类别标识；
3) h_i：$h_i \geq 0$，表示的是以 X_i 为中心的超球的半径；

3.3.2　经验风险

由于每个训练样本是粗粒度下面的一个表示，因此可能包含有多个原始样本空间中的训练样本，因此经验风险部分的表示必须要考虑到区域的大小；若采取超球作为区域的表示形式，超平面作为分类函数的表示形式，那么经验风险部分的表达式可以写为：

$$R_{emp}(\alpha)=\sum_{i=1}^{l}\xi_i$$
$$Y_i f(X_i)\geq 1-\xi_i+h_i\|w\|, \quad i=1,2,...,n \qquad (3\text{-}4)$$
$$\xi_i\geq 0,$$

在可分情况下，分类函数是在条件 $y_i f(x_i)\geq 1$ 的条件下求解的，因此，分类间隔线 $f(x)=1$ 和 $f(x)=-1$ 之间的距离为 $2/\|w\|$。都在 $\|w\|\geq 1$ 的函数集合中进行求解。若一个超平面 $\|w\|<1$，那么通过变换直接会得到 $\|w\|\geq 1$ 的等价超平面。所以 $h\|w\|\geq h$，因此 $1-\xi_i+h_i\|w\|\geq 1-\xi_i+h_i$，由于当不等式约束中为 $h\|w\|$ 时，函数难以求解，因此采用 h 作为 $h\|w\|$ 的估计值，如图 3.3(a)所示。

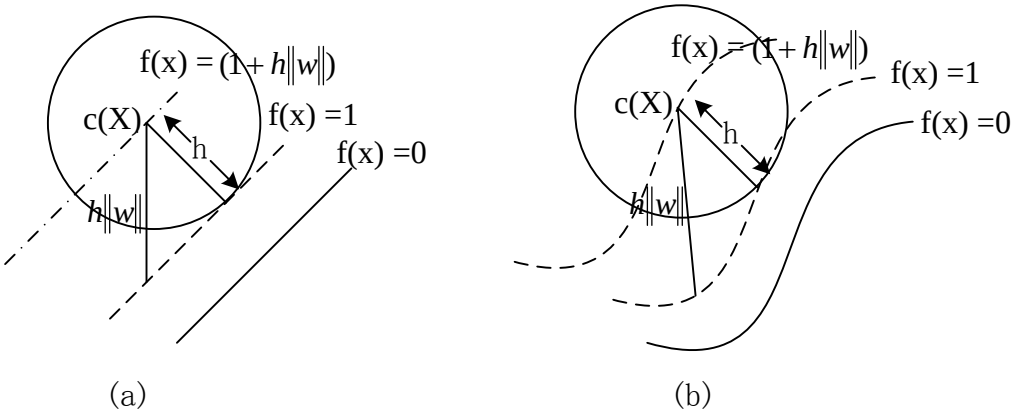


图 3.3　样本到分类边界的距离

那么经验风险采取公式（3-5）所表示的求解条件：

$$R_{emp}(\alpha)=\sum_{i=1}^{l}\xi_i$$
$$Y_i f(X_i)\geq 1-\xi_i+h_i,\quad i=1,2,...,n \qquad (3\text{-}5)$$
$$\xi_i\geq 0,$$

由于算法是在原空间中对训练样本集合进行粗粒度下面的表示，因此用于划分两类的分类曲面在原空间中讨论会更为合理一些。即使在曲面的情况下，也采用 h_i 作为控制经验风险的量，使用公式（3-5）来作为经验风险的求解条件。

图 3.4 中给出了超球和超平面关系的几种表示形式，假设其中 $h\geq 1/\|w\|$。

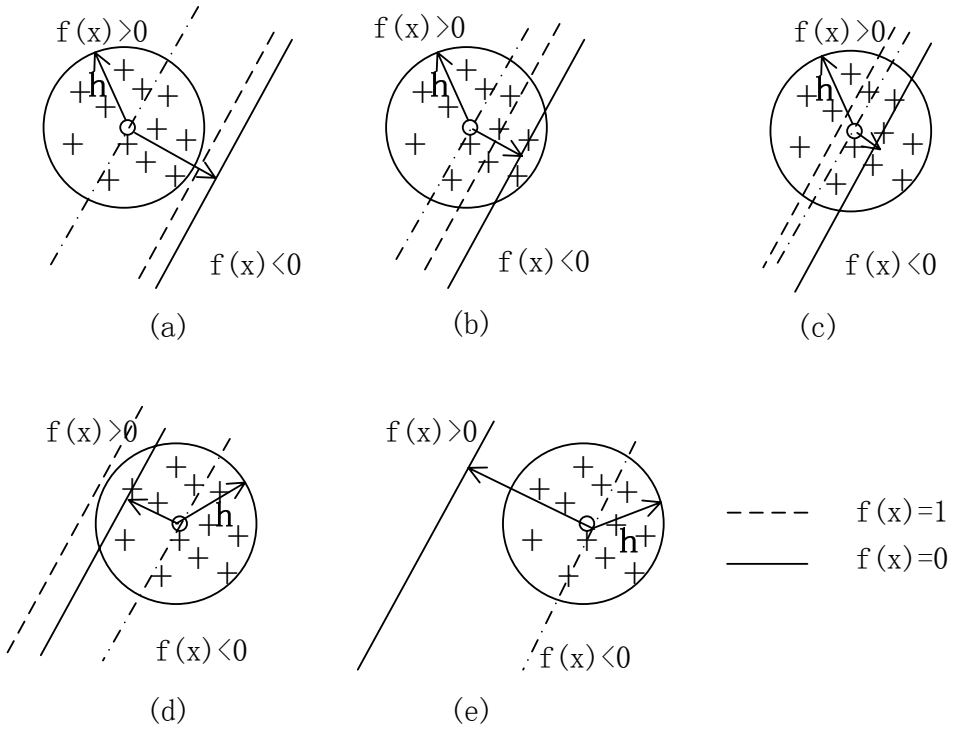


图 3.4　超球与超平面的关系

那么在图中的几种情况下，经验风险的值分别为：

(a) 由于整个超球位于超平面的 $f(x)=1$ 的正方，因此经验风险的值为 $\xi_i=0$；

(b) 虽然超球的中心点落到超平面的 $f(x)=1$ 的正方，但是超球有一部分落入到 $f(x)=1$ 的负方。在这种情况下，经验风险的值不为 0，经验风险的值取为 $\xi_i=1+h_i-f(c(X_i))$，由于 $f(c(X_i))>1$，因此经验风险的取值 $0<\xi_i<h_i$；

(c) 由于超球的中心点落入到 $f(x)=1$ 的下方，经验风险的值取为 $\xi_i=1+h_i-f(c(X_i))$，由于 $0<f(c(X_i))<1$，因此经验风险的取值 $h_i<\xi_i<1+h_i$

(d) 由于超球的中心点落入到 $f(x)=0$ 的下方，经验风险的值取为 $\xi_i=1+h_i-f(c(X_i))$，由于 $f(c(X_i))<0$，因此经验风险的取值 $(1+h_i)<\xi_i<2h_i$

(e) 由于整个超球的落入到 $f(x)=0$ 的下方，经验风险的值取为 $\xi_i=1+h_i-f(c(X_i))$，由于 $f(c(X_i))<-(1+h_i)$，因此经验风险的取值 $\xi_i>2(1+h_i)$。

3.3.3 置信范围

虽然训练样本是采用粗粒度下的表示，但是对于分类函数集合还是使用同结构风险最小化原则一样的函数集合，即在如下的函数集合中去选择分类函数：

$$f(x)=\mathrm{sgn}\left(\sum_{i=1}^{N}\alpha_i\psi_i(x)+b\right) \tag{3-6}$$

置信范围部分，可以仍然使用条件 $1/2\|w\|^2$ 来表示。

综上所述，给定训练样本集合 Z，如下采取区域风险最小化原则来求取分类函数。首先将训练样本样本集合表示成为粗粒度下面的形式 $Z^{'}=\{z_i^{'}\}=\{(c(X_i),Y_i,h_i)\}$，然后采取如下的结构风险最小化原则来求取最优的分类函数：

$$\begin{aligned} &R_{reg}(\alpha)=\frac{1}{2}\|w\|^2+\lambda\sum_{i=1}^{l}\xi_i \\ &Y_i f(c(X_i))\geq 1-\xi_i+h_i, \\ &\xi_i\geq 0, \end{aligned} \qquad i=1,2,...,n \tag{3-7}$$

3.4 区域的表示

区域的表示问题也就是如何在粗粒度下面去表示给定的训练样本集合，首先讨论样本集合以及包含样本集合的特征空间中的区域的可能划分方法。设 $Z=\{(x_i,y_i)\}_{i=1}^{l}$ 表示训练样本集合，$\mathbf{X}_\mathrm{k}=\{x_i\}_{i=1}^{l_k}$ 表示第 k 类训练样本的特征向量集合。将 $\mathbf{X}_\mathrm{k}$ 在特征空间中分解成为粗粒度的形式：$\mathbf{X}_\mathrm{k}=\bigcup_{i=1}^{n}X_i$，$X_i$ 为粗粒度下的训练样本集合。设 $\aleph_k=\{X_i\}_{i=1}^{n}$，对于 $\forall X_i,X_j\in\aleph_k$，若 $X_i\bigcap X_j=\phi$，那么 $\aleph_k$ 对

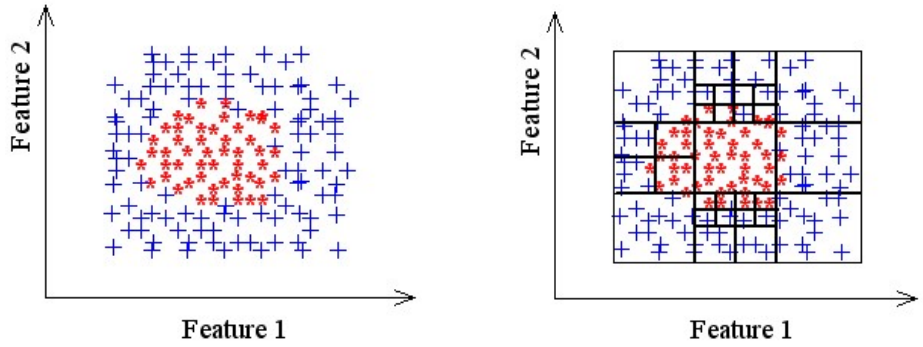
$\mathbf{X}_\mathrm{k}$ 的划分是互不相交的。若 $\exists X_i, X_j \in \aleph_k$，$X_i \bigcap X_j \neq \phi$，那么 $\aleph_k$ 对 $\mathbf{X}_\mathrm{k}$ 的划分是存在相交的。希望对集合 $\mathbf{X}_\mathrm{k}$ 的划分是互不相交的，这样不会重复考虑原始训练样本的作用。前面讨论的是对训练样本集合的划分，而在实际的分类器训练中，要考虑的是采用中心点和半径表示的包含原始训练样本的一个区域，将每个 X_i 所对应的区域表示为 o_i，设 $O_k = \{o_i\}_{i=1}^n$，那么 $\Omega_k = \bigcup_{i=1}^n o_i$ 是包含 $\mathbf{X}_\mathrm{k}$ 的特征空间中的一个区域。若对于 $\forall o_i, o_j \in O_k$，均有 $o_i \bigcap o_j = \phi$，那么 O_k 对 Ω_k 的划分是互不相交的。若 $\exists o_i, o_j \in O_k$，$o_i \bigcap o_j \neq \phi$，那么 O_k 对 Ω_k 的划分是存在相交的。理想的情况下是希望 $\aleph_k$ 对 $\mathbf{X}_\mathrm{k}$ 的划分与 O_k 对 Ω_k 的划分均为互不相交的。并且对于任意的两类 $\mathbf{X}_\mathrm{i}$ 和 $\mathbf{X}_\mathrm{j}$，理想的情况是希望 O_i 中的元素与 O_j 中的元素也互不相交。

虽然有各种各样将训练样本进行划分的方式，有两种方式相对来说问题的求解会变得较为容易些，一种是采用超球，一种是采用超立方体。前面关于粗粒度下的训练样本的定义形式为 $z_i^{'} = (c(X_i), Y_i, h_i)$，其中使用了两个参数去表示粗粒度下的训练样本所在的区域：区域的中心值 $c(X_i)$ 和区域的半径 h_i。采用超球和超立方体对训练样本进行划分可以较容易的采用中心值和半径去表示区域。下面分别介绍这两种表示方式。

3.4.1　采用超立方体表示

采用网格（grid）来划分特征空间在分类问题中也是较为常见的一种方式。通常考虑的网格为超长方体，在前面我们关于区域风险最小化原则的叙述中讲了当所划分的区域相对于中心点在各个方向上的分布比较均匀时，问题容易求解一些。因此本文中考虑基于超立方体的划分。使用超立方体进行划分是对特征空间的一种自顶向下的硬性划分，相比于自底向上的聚类等方法，划分边界更容易计算，不需要对训练样本首先进行聚类，而直接利用样本在空间的位置直接进行划分而获得超立方体；另外使用超立方体对特征空间不断进行划分，可以使得最终的划分结果满足不同类的划分区域之间不相交，同一类的划分区域之间也是不相交的。图 3.5 给出了基于超立方体将样本集合划分的一个示意图。

基于超立方体划分特征空间可以采取如下的策略。设 $\Omega \subset R^n$ 是包含样本集合 X 的特征空间中的一个超立方体。对于一个给定的数 l，如果将 Ω 在每一维上面均 l 等分，那么 Ω 将被分解成为 l^d 个小超立方体，即 $\Omega = \cup_{i=1}^{l^d} o_i$，其中 o_i 是一



(a)　　(b)

图 3.5　基于超立方体的特征空间划分示意图
(a)原始的样本集合；(b)采用超立方体划分后的样本集合

个小超立方体。将Ω基于数值 l 的划分记为Ω_l，$\Omega_l = \{o_i\}_{i=1}^{l^d}$。对于$\forall o_i \in \Omega_l$，存在以下三种情况：

1. o_i 是纯的：即o_i中的训练样本全部来自于类别 1 或者全部来自于类别 2，那么o_i即为粗粒度下的一个训练样本；

2. o_i 是空的：即o_i中不包含训练样本，这样的o_i不需要考虑；

3. o_i 是混合的：这些超立方体包含两类样本，这些超立方体是需要进一步划分的。将o_i在每一维上面进行二等分，这样o_i被划分成为包含2^d个更小一些的超立方体$\{o_{i,j}\}_{j=1}^{2^d}$。

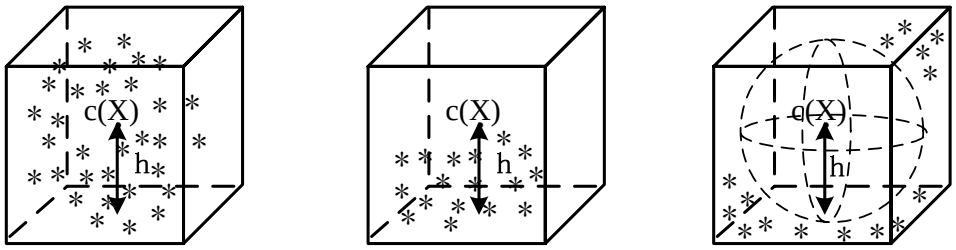
经过不断地划分，包含样本的特征空间中的区域被划分成为一个包含大小不同超立方体的集合，其中的每个超立方体均包含一类样本，这个集合就是特征空间在粗粒度上面的划分。选择超立方体边长的一半作为 h 的值。

上述划分方式存在以下缺陷：

1. 当某个超立方体内的样本个数在超立方体内分布不均匀时，采用中心点和半径去控制分类边界线可能并不是最优分类边界线。如在一个超立方体内，所有的样本都分布在某一半区域中，这时候采用中心点和半径所确定的分类边界线可能并不是最优的，因为在这种情况下实际上样本所分布的区域还可以继续缩小一些，如图 3.6(b)所示；

2. 虽然对特征空间是采用超立方体进行划分的，但是在确定分类边界线时是使用超立方体的中心点和半径进行控制的。那么当训练样本落入到超立方体的角落中的时候，可能会导致所确定的分类函数的错误率较高，如图 3.6(c)所示。

当一个超立方体内的聚类中心偏离超立方体的中心点比较远时，说明样本



(a)　　(b)　　(c)

图 3.6　超立方体作为基本区域的样本分布的几种情况

在超立方体的分布是不均衡的，在这种情况下，可以对超立方体继续划分，直到样本在超立方体内的分布相对比较均衡。

采用超立方体对训练样本进行划分的算法如算法 3.1 所示。

算法 3.1　训练样本基于超立方体的划分算法

HypercubeDivision

1　将训练样本集合在每一维上归一化到[0,1]之间；

2　k=l；% l 是给定的初始值

3　将训练样本集合 X 基于分辨率 k 进行划分，结果表示为 X_l；

4　X_l 中的元素入栈 S；

5　x = pop(S)；

6　如果 x 为空，转到第 11 步；

7　c = x 所位于的超立方体的中心，m = x 的均值；

8　若 x 包含两类训练样本，$k = k \times 2$，在分辨率 k 下划分 x，结果为 x_k，将 x_k 中的元素入栈 S，转到第 5 步；

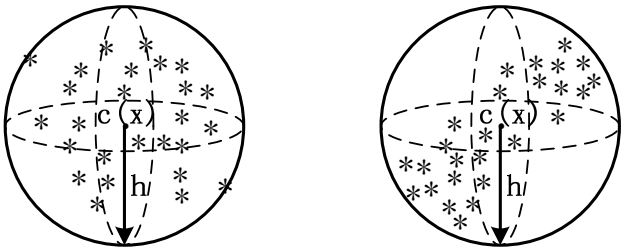
9　若 x 仅包含一类训练样本但是 $\|m - c\| > T_o$，$k = k \times 2$，在分辨率 k 下划分 x，结果为 x_k，将 x_k 中的元素入栈 S，转到第 5 步；

10 转到第 5 步；

11 结束。

3.4.2　采用超球表示

下面介绍另外一种表示方法，基于超球的表示方法。一种比较直观的获得超球的方式是采用聚类的方法，即首先对正例样本和负例样本分别进行聚类，将每个聚类的中心点作为超球的中心点，以该聚类内的样本距离中心点的最大



(a)　　(b)

图 3.7　超球作为基本区域的样本分布情况

距离作为超球的半径。采用聚类方法所获得超球，分布一般都比较均匀。由于预先不能够指定期望聚类的数目，采用 ISODATA（Iterative Self-Organizing Data Analysis Technique）算法对属于同一类的训练样本进行聚类[61]。设 $\mathbf{X}_k$ 表示第 k 类的训练样本集合（这里不考虑样本所属于的类别属性），设采用 ISODATA 算法所得到的聚类结果为 $\aleph_k = \{X_i\}_{i=1}^n$，各个聚类中心的值为 $M_k = \{m_i\}_{i=1}^n$，其中超球的中心点和半径如下计算：

$$\begin{aligned} c(X_i) &= m_i \\ h_i &= \max\left\| x_j - m_i \right\|, \qquad x_j \in X_i \end{aligned} \tag{3-8}$$

首先对属于同一类的训练样本进行聚类，然后将每个聚类按上述公式构成一个超球的划分方式，只能满足不同超球的内的原始训练样本是不相交的，但是同类的超球之间以及不同类的超球之间都可能存在相交。特别是在两类样本混合的区域，如果正例样本和负例样本先单独聚类的话，聚类结果会使得两类样本在混合区域的区分性比较差，难以获得较好的分类边界。

3.5　问题求解

基于区域风险最小化原则的分类问题求解，是在支持向量机中构造软间隔分类超平面的算法上进行改进的。下面首先介绍一下支持向量机算法的求解方法，然后再介绍如何修改算法使得求解的分类函数能够以区域为单位构造分类超平面。

3.5.1　序贯最小优化算法

在支持向量机算法中，假设给定的核函数为 $K(x_i, x_j)$，由核函数导出的从

原始的特征空间到高维特征空间的映射函数为 $\phi(x)$，即 $K(x_i,x_j)=\phi(x_i)\phi(x_j)$。寻找最优超平面可以表示为，寻找参数 α_i，i=1,…,l，使得下列二次型最大：

$$W(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j\phi(x_i)\phi(x_j),\quad i=1,2,...,l$$
$$\sum_{i=1}^{l}\alpha_i y_i=0,\qquad 0\le\alpha_i\le\lambda, \tag{3-9}$$

Platt 提出的 SMO 算法是一种非常有效地求解二次规划的方法[62]。在这个算法中，工作集的大小被限定为 2，首先确定待优化的乘子，不妨设为 α_i，α_j，然后求解如下的子二次规划问题：

$$W(\alpha_i,\alpha_j)=\frac{1}{2}[\alpha_i\quad\alpha_j]\cdot\begin{bmatrix}Q_{ii}&Q_{ij}\\Q_{ji}&Q_{jj}\end{bmatrix}\cdot\begin{bmatrix}\alpha_i\\\alpha_j\end{bmatrix}+[Q_{i,N}\alpha_N-1]\alpha_i+[Q_{j,N}\alpha_N-1]\alpha_j$$
$$\alpha_i=w-y_i y_j\alpha_j,\quad 0\le\alpha_i\le\lambda \tag{3-10}$$
$$Q_{i,N}\alpha_N=y_i f(x_i))+y_i b-\alpha_i Q_{ii},\quad Q_{j,N}\alpha_N=y_j f(x_j)+y_j b-\alpha_j Q_{jj}$$

其中 $Q_{ij}=y_i y_j K(x_i,x_j)$。支持向量方法采取序贯最小优化算法求出来的参数的迭代公式为：

$$\alpha_j^{new}=\max(\min(\alpha_j^{new,unclipped},H),L)$$
$$\alpha_i^{new}=\alpha_i^{old}+y_i y_j(\alpha_j^{old}-\alpha_j^{new}) \tag{3-11}$$

其中

$$\alpha_j^{new,unclipped}=\begin{cases}\alpha_j^{old}+\dfrac{G_i-G_j}{Q_{ii}+Q_{jj}-2Q_{ij}}, & \text{如果} y_i=y_j\\[2ex]\alpha_j^{old}+\dfrac{-G_i-G_j}{Q_{ii}+Q_{jj}+2Q_{ij}}, & \text{如果} y_i\ne y_j\end{cases}$$

$$G_k=y_k(f(x_k)-y_k)(k=i,j) \tag{3-12}$$

$$H=\begin{cases}\min(\lambda,\lambda+\alpha_j^{old}-\alpha_i^{old}), & if\quad y_i\ne y_j\\\min(\lambda,\alpha_j^{old}+\alpha_i^{old}), & if\quad y_i=y_j\end{cases}$$

$$L=\begin{cases}\max(0,\alpha_j^{old}-\alpha_i^{old}), & if\quad y_i\ne y_j\\\max(0,\alpha_i^{old}+\alpha_j^{old}-\lambda), & if\quad y_i=y_j\end{cases}$$

3.5.2　算法改进

本章之所以在定义区域的时候，定义的区域围绕着中心点在各向的宽度是一样，主要是为了方便问题的求解，下面介绍如何采用序贯最小优化算法求解式（3-7）。式（3-7）的最优解是由下面的拉格朗日泛函的鞍点给出的：

$$L(w,b,\xi_i)=\frac{1}{2}\|w\|^2+\lambda\sum_{i=1}^{n}{}_i\xi_i-\sum_{i=1}^{n}\alpha_i(Y_i(w\cdot\phi(c(X_i))-b)-1+\xi_i-h_i)-\sum_{i=1}^{n}\beta_i\xi_i \quad (3\text{-}13)$$

其中 α_i，$\beta_i(i=1,...,n)$ 为拉格朗日乘子。需要把拉格朗日函数关于 w，b，ξ_i 求取最小值，关于 $\alpha_i>0$ 和 $\beta_i>0$ 求取最大值。同支持向量机算法的求解是一样的，在鞍点上，解 w_0，b_0，$\boldsymbol{\xi}^0$，$\boldsymbol{\alpha}^0$，$\boldsymbol{\beta}^0$ 满足下面的条件：

$$\begin{aligned}&\frac{\partial L(w,b,\boldsymbol{\xi}^0,\boldsymbol{\alpha}^0,\boldsymbol{\beta}^0)}{\partial w}=0\\&\frac{\partial L(w,b,\boldsymbol{\xi}^0,\boldsymbol{\alpha}^0,\boldsymbol{\beta}^0)}{\partial b}=0 \qquad i=1,2,...,n\\&\frac{\partial L(w,b,\boldsymbol{\xi}^0,\boldsymbol{\alpha}^0,\boldsymbol{\beta}^0)}{\partial \boldsymbol{\xi}_i}=0\end{aligned} \quad (3\text{-}14)$$

求解可得到：

$$\begin{aligned}&w=\sum_{i=1}^{n}\alpha_iY_i\phi(c(X_i))\\&\sum_{i=1}^{n}\alpha_iY_i=0 \qquad\qquad i=1,2,...,n\\&\alpha_i+\beta_i=\lambda\end{aligned} \quad (3\text{-}15)$$

求解出来的这三个式子同标准的 SVM 的求解式子是一样的。将这三个式子带入到拉格朗日式子中，可得最小化下面的二次型：

$$\begin{aligned}&W(\alpha)=\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_jY_iY_j\phi(c(X_i))\phi(c(X_j))-\sum_{i=1}^{n}(1+h_i)\alpha_i,\\&\qquad\qquad\qquad\qquad\qquad\qquad i=1,2,...,n\\&\sum_{i=1}^{n}\alpha_iY_i=0, \qquad 0\le\alpha_i\le\lambda,\end{aligned} \quad (3\text{-}16)$$

由于 $\phi(c(X_i))\phi(c(X_j))=K(c(X_i),c(X_j))$，那么上式可以写为：

$$\begin{aligned}&W(\alpha)=\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_jY_iY_jK(c(X_i),c(X_j))-\sum_{i=1}^{n}(1+h_i)\alpha_i,\\&\qquad\qquad\qquad\qquad\qquad\qquad i=1,2,...,n\\&\sum_{i=1}^{n}\alpha_iY_i=0, \qquad 0\le\alpha_i\le\lambda,\end{aligned} \quad (3\text{-}17)$$

定理 3.1　采用序贯最小优化算法求解（3-17）的迭代公式为式（3-11）。其中（3-12）中的条件改为：

$$G_k = y_k\left[f(x_k)-(1+h_k)y_k\right] \tag{3-18}$$

证明：

对（3-17）式关于α_i和α_j进行求解，可以写为：

$$\begin{aligned} W(\alpha_i,\alpha_j) = \frac{1}{2}[\alpha_i \quad \alpha_j]\cdot\begin{bmatrix} Q_{ii} & Q_{ij} \\ Q_{ji} & Q_{jj} \end{bmatrix}\cdot\begin{bmatrix} \alpha_i \\ \alpha_j \end{bmatrix} \\ +[Q_{i,N}\alpha_N-(1+h_i)]\alpha_i+[Q_{j,N}\alpha_N-(1+h_j)]\alpha_j \end{aligned} \tag{3-19}$$

（3-19）式满足的约束条件同（3-10）式。

求解下式：

$$\frac{\partial W(\alpha_i,\alpha_j)}{\partial \alpha_j} = 0 \tag{3-20}$$

可得

$$\alpha_j = \alpha_j^* + \frac{Y_j\left[f(c(X_i))-(1+h_i)Y_i - f(c(X_j))+(1+h_j)Y_j\right]}{Q_{ii}+Q_{jj}-2Y_iY_jQ_{ij}} \tag{3-21}$$

因此，$G_k = y_k\left[f(x_k)-(1+h_k)y_k\right](k=i,j)$。　　证毕。

在具体的实现中，将原来的关于支持向量机的程序变成区域风险最小化原则也比较简单，以 LibSVM 为例，只需要将序贯最小优化算法的初始条件$G_i=-1$和$G_j=-1$改为$G_i=-(1+h_i)$和$G_j=-(1+h_j)$即可。因此，当区域可以用中心点和半径两个参数表示的时候，基于区域风险最小化的分类问题的求解是比较容易的，只需要在原来的序贯最小最优化算法上作局部的修改就可以了。

3.6　实验结果

为了说明以上算法的有效性，下面首先给出仿真数据的实验结果，然后是标准数据集上面的实验结果。相比于真实数据，基于仿真数据的样本严格满足独立同分布的假设，并且在实验中可以选择不同规模的数据，使用仿真数据可以对算法的原理进行解释说明。标准数据集上面的实验结果进一步证实了算法的有效性。所有的实验都是在 Pentium IV 1.5G CPU，256M 内存，Windows XP

操作系统，采用 C++实现的，代码没有进行任何优化。

下面的实验均选择支持向量机作为分类器，软件包采用的是 LIBSVM2.4[63]。聚类算法采用的是 ISODATA 算法。

3.6.1　仿真数据的实验结果

生成两组数据，每组包含两类样本，每类中的数据都满足高斯分布，两组数据的分布函数为：

1. $p_1(x)=\exp(-x^2/0.4)$ 和 $p_2(x)=\exp(-(x-4)^2)$。这组数据两类样本在特征空间的重叠部分比较少；
2. $p_1(x)=\exp(-x^2/0.4)$ 和 $p_2(x)=\exp(-(x-3)^2)$。这组数据两类样本在特征空间中有一部分的重叠区域。

在下面的实验中，支持向量机算法中的正则化参数 $\lambda=1$。核函数采用如下的两种核函数：

1. 高斯核函数：$K(x_i,x_j)=\exp(-\sigma|x_i-x_j|)$；
2. 多项式核函数：$K(x_i,x_j)=(r\cdot|x_i-x_j|+b)^d$；

ISODATA 共包括以下参数：

1. M：开始选取的中心值的数目；
2. T：如果一个类内的样本数小于 T，那么丢掉该类；
3. ND：期望的类的个数；
4. sig_s：如果一个类的方差大于 sig_s，那么分裂该类；
5. Dm：两个要合并的类的均值的最大距离；
6. Nmax：一次迭代最多可以合并的类的最大数目；

由于基于超立方体的方法中，训练样本首先要转换到[0,1]之间，因此为了使得以下的比较更具有可信性，在基于原始样本的分类器训练和基于超球的方法中，首先都将训练样本转换到[0,1]之间。

表 3.1~表 3.11 是基于仿真数据的实验结果，各个表内的数据的含义如下：

1. 表 3.1~3.3 是基于支持向量机的实验结果

表 3.1 和表 3.2 中的数据是采用支持向量机进行训练，核函数分别选择高斯核函数和多项式核函数时的实验结果。由于训练样本满足高斯分布，高斯核函数是比较适合训练样本集合的核函数，因此当训练样本较少时采用高斯核函数所能够达到的分类的正确率是比较高的。

表 3.1 *c*-SVM 实验结果（Gaussian Kernel）（$\sigma=80$，$\lambda=1$）

	1		2	
l_1+l_2	s_1+s_2（-1+1）	正确率	s_1+s_2	正确率
10+10	4+9	99.595	6+8	98.135
50+50	8+16	99.745	12+22	98.48
100+100	9+33	99.755	13+33	98.74
1000+1000	19+129	99.79	45+119	98.895
5000+5000	47+229	99.77	200+307	98.865
10000+10000	76+295	99.79	357+505	98.855

表 3.2 *c*-SVM 实验结果（Polynomial Kernel）
（*d*=3，*r*=25，*b*=2，$\lambda=1$）

	1		2	
l_1+l_2	$S_1(-1)+s_2(1)$	正确率	S_1+s_2	正确率
10+10	1+2	99.11	2+1	94.35
50+50	1+2	98.995	4+4	98.645
100+100	2+1	99.335	5+4	98.365
1000+1000	5+5	99.785	32+32	98.895
5000+5000	30+29	99.78	149+149	98.87
10000+10000	41+40	99.795	318+316	98.875

表 3.3 采用 *v*-SVM（Gaussian Kernel）
（$\sigma=80$，*v*=0.5，$\lambda=1$）

	1		2	
l_1+l_2	S_1+s_2	正确率	S_1+s_2	正确率
10+10	6+9	99.355	6+8	98.11
50+50	26+28	99.705	27+31	98.845
100+100	52+57	99.585	50+56	98.73
1000+1000	501+505	99.705	503+506	98.8
5000+5000	2501+2508	99.72	2501+2509	98.845
10000+10000	5002+5007	99.74	5001+5008	98.835

表 3.4 *a*-SVM 实验结果（超球中心点）（Gaussian Kernel）
（$\sigma=80$，$\lambda=1$，M=1，T=1，ND=样本数， Nmax=10）
（sig_s(1)=0.0125，Dm(1)=0.03125，sig_s(-1)=0.025，Dm(-1)=0.0625）

1			**2**		
n_1+n_2	S_1+s_2	正确率	N_1+n_2	S_1+s_2	正确率
5+5	4+5	99.345	5+5	4+5	98.71
22+25	7+16	99.75	25+25	10+19	98.43
29+40	7+25	99.535	33+42	10+23	98.49
75+101	16+55	99.805	71+105	21+54	98.29
99+142	25+68	99.69	103+144	35+77	98.55
120+151	26+62	99.755	125+155	38+84	98.02

表 3.5 *a*-SVM 实验结果（超球中心点）（Polynomial Kernel）
（*d*=3，*r*=25，*b*=2，$\lambda=1$，M=1，T=1，ND=样本数， Nmax=10）
（sig_s(1)=0.0125，Dm(1)=0.03125，sig_s(-1)=0.025，Dm(-1)=0.0625）

1			**2**		
n_1+n_2	S_1+s_2	正确率	N_1+n_2	S_1+s_2	正确率
5+5	2+1	99.475	5+5	1+2	97.465
22+25	1+2	98.625	25+25	2+2	97.225
29+40	2+1	98.99	33+42	5+3	98.535
75+101	3+3	99.745	71+105	12+13	98.26
99+142	13+11	99.735	103+144	22+22	98.335
120+151	10+11	99.765	125+155	25+25	98.09

表 3.3 中的数据是采用 *v*-SVM 算法[64]所得到的实验结果，*v*-SVM 是通过参数 v 来调整经验风险的值同支持向量的数目的支持向量机，求解条件如公式（3-22）所示。

$$R_{ref}(\alpha)=\frac{1}{2}\|w\|^2-v\rho+\frac{1}{l}\sum_{i=1}^{l}\xi_i$$
$$y_i(w^T\phi(x_i)+b)\geq\rho-\xi_i,\quad i=1,2,...,l \qquad (3\text{-}22)$$
$$\xi_i\geq 0,\quad \rho\geq 0$$

表 3.6 *a*-SVM 实验结果（超立方体中心点）（Gaussian Kernel）
（$\sigma=80$，$\lambda=1$，l=8）

1			**2**		
n_1+n_2	S_1+s_2	**正确率**	N_1+n_2	S_1+s_2	**正确率**
7+9	6+9	96.365	5+8	5+8	92.615
8+22	8+22	99.38	12+27	12+26	98.335
9+35	9+33	99.175	17+33	12+31	98.645
14+46	12+41	99.795	77+73	37+61	98.795
101+76	37+64	99.775	383+192	137+161	98.83
118+89	47+72	99.81	745+366	271+293	98.83

表 3.7 *a*-SVM 实验结果（超立方体中心点）（Polynomial Kernel）
（d=3，r=25，b=2，$\lambda=1$，l=8）

1			**2**		
n_1+n_2	S_1+s_2	**正确率**	N_1+n_2	S_1+s_2	**正确率**
7+9	1+2	98.725	5+8	2+2	90.6
8+22	1+2	98.605	12+27	4+4	98.88
9+35	2+1	98.435	17+33	6+3	98.77
14+46	2+3	99.765	77+73	29+28	98.805
101+76	28+25	99.78	383+192	130+128	98.825
118+89	36+34	99.81	745+366	266+267	98.83

实验设置 v=0.5，核函数采用高斯核函数，参数$\sigma=1$。

测试数据是采用相同分布的正例样本和负例样本各 10000 个样本，分类器的正确率是基于测试样本所计算的。

从表 3.1~3.3 可以看出，当训练样本增多的时候，首先分类器分类的正确率会升高，另外分类函数的复杂程度也增大，主要反映在支持向量的数目随着训练样本个数的增多而增多。由于当训练样本的个数比较多时，落在分类间隔线上的样本数目不可避免地也会增多，因此支持向量的数目显然会增多。

2. 表 3.4~3.7 是采用聚类中心进行训练的实验结果

表 3.4 和表 3.5 是采用超球的聚类中心进行训练的实验结果。表 3.6 和表 3.7

表 3.8 *a*-SVM 实验结果（超球）（Gaussian Kernel）
（$\sigma=80$，$\lambda=1$，M=1，T=1，ND=样本数， Nmax=10）
（sig_s(1)=0.0125，Dm(1)=0.03125，sig_s(-1)=0.025，Dm(-1)=0.0625）

	1			2		
l_1+l_2	n_1+n_2 (-1,1)	s_1+s_2 (-1,1)	正确率	N_1+n_2	S_1+s_2	正确率
10+10	5+5	4+5	99.25	5+5	4+5	98.72
50+50	22+25	7+15	99.745	25+25	10+19	98.445
100+100	29+40	8+25	99.53	33+42	11+24	98.49
1000+1000	75+101	16+52	99.805	71+105	22+55	98.3
5000+5000	99+142	24+69	99.69	103+144	35+75	98.56
10000+10000	120+151	25+62	99.75	125+155	36+82	98.04

表 3.9 *a*-SVM 实验结果（超球）（Polynomial Kernel）
（d=3，r=25，b=2，$\lambda=1$，M=1，T=1，ND=样本数， Nmax=10）
（sig_s(1)=0.0125，Dm(1)=0.03125，sig_s(-1)=0.025，Dm(-1)=0.0625）

	1			2		
l_1+l_2	n_1+n_2 (-1,1)	s_1+s_2 (-1,1)	正确率	N_1+n_2	S_1+s_2	正确率
10+10	5+5	2+1	99.535	5+5	1+2	97.775
50+50	22+25	1+2	98.635	25+25	2+2	97.25
100+100	29+40	2+1	99.015	33+42	5+3	98.54
1000+1000	75+101	3+3	99.745	71+105	12+13	98.26
5000+5000	99+142	13+11	99.735	103+144	23+22	98.345
10000+10000	120+151	10+11	99.765	125+155	25+25	98.095

是采用超立方体的中心进行训练的实验结果。在仿真试验中，由于数据分布情况简单，分类边界简单，因此采取聚类中心所得到的实验结果也比较好。

3. 表 3.8~3.11 是采用本章提出的区域风险最小化原则所得出的实验结果

表 3.8 和表 3.9 是首先将训练样本进行聚类，然后采用区域风险最小化原则所获得的实验结果，表 3.10 和表 3.11 是基于超立方体划分的实验结果。从表中

表 3.10 *a*-SVM 实验结果（超立方体）（Gaussian Kernel）
（$\sigma=80$，$\lambda=1$，l=8）

	1			**2**		
l_1+l_2	N_1+n_2	S_1+s_2	正确率	n_1+n_2	S_1+s_2	正确率
10+10	7+9	6+9	96.35	5+8	5+8	92.74
50+50	8+22	8+21	99.425	12+27	12+26	98.33
100+100	9+35	9+32	99.16	17+33	11+31	98.645
1000+1000	14+46	12+39	99.79	77+73	36+60	98.79
5000+5000	101+76	38+64	99.775	383+192	137+161	98.83
10000+10000	118+89	48+72	99.81	745+366	270+293	98.835

表 3.11 *a*-SVM 实验结果（超立方体）（Polynomial Kernel）
（d=3，r=25，b=2，$\lambda=1$，l=8）

	1			**2**		
l_1+l_2	N_1+n_2	S_1+s_2	正确率	n_1+n_2	S_1+s_2	正确率
10+10	7+9	1+2	98.725	5+8	2+2	90.18
50+50	8+22	1+2	98.625	12+27	4+4	98.88
100+100	9+35	2+1	98.475	17+33	6+3	98.79
1000+1000	14+46	2+3	99.765	77+73	29+28	98.805
5000+5000	101+76	29+25	99.78	383+192	130+128	98.825
10000+10000	118+89	36+34	99.805	745+366	266+267	98.83

的数据可以看到，当数据的规模比较大时，通过对数据首先划分区域，再训练分类器，会使得支持向量的数目大大降低，另外训练所得到的分类器的正确率也要好于训练样本较少时的情形。从表 3.8 和表 3.9 的数据可以看出，当训练样本增多时，并且两类训练样本有重叠区域时，首先对两类样本分别聚类，再采用区域风险最小化原则训练所得到的分类器的性能并不好，这是由于在两类样本重叠区域的聚类结果，正例样本和负例样本所位于的超球的相交部分所占的比例会很大，因此聚类得到的超球在这部分区域的可区分程度会很差，这也是由聚类得到的超球不能够满足划分同类的区域之间不相交，不同类的区域之间不相交所造成的。设计更优的聚类算法有待于进一步的研究。

表 3.12　*c*-SVM 实验结果 （Gaussian Kernel）（$\sigma=1$，$\lambda=1$）

数据集	***c*-SVM**			***c*-SVM（随机采样）**		
	l_1+l_2 (-1,1)	S_1+s_2	正确率	l_1+l_2	s_1+s_2	正确率
Balance-scale	288+337	91+86	98.72	96+112	37+38	78.4
Adult-1	1210+395	379+331	83.13	403+131	151+126	82.00
Adult-2	1693+572	526+482	83.22	564+190	207+178	81.72
Adult-3	2412+773	692+628	83.47	804+257	264+233	82.98
Adult-4	3593+1188	993+935	83.92	1197+396	371+339	83.23
Adult-5	4845+1569	1310+1234	83.94	1615+523	487+446	82.77
Adult-6	8528+2692	2171+2074	84.35	2842+897	802+742	83.39
Adult-7	12182+3918	3061+2949	84.59	4060+1306	1116+1057	83.67
Adult-8	17190+5506	4234+4101	84.73	5730+1835	1539+1460	83.97
Adult	24720+7841	5896+5744	84.96	8240+2613	2143+2038	84.34

表 3.13　*a*-SVM 实验结果 （Gaussian Kernel）（$\sigma=1, \lambda=1$）

（M=1，T=1，ND=样本数， Nmax=10，sig_s(1,-1)=0.0125，Dm(1,-1)=0.03125，l=2）

数据集	超球			超立方体		
	n_1+n_2	S_1+s_2	正确率	n_1+n_2	s_1+s_2	正确率
Balance-scale	144+171	55+54	98.24	118+165	66+67	95.2
Adult-1	689+227	236+199	82.55	753+356	347+312	83.072
Adult-2	957+334	338+290	82.59	1076+501	482+441	82.87
Adult-3	1355+454	445+389	82.91	1472+666	638+580	83.31
Adult-4	2012+690	619+553	83.66	2051+1022	930+865	83.47
Adult-5	2729+918	802+741	83.92	2718+1333	1208+1143	83.64
Adult-6	4788+1530	1317+1223	84.09	4498+2224	2009+1907	84.21
Adult-7	6816+2253	1885+1771	84.32	6429+3235	2800+2710	84.33
Adult-8	9529+3107	2576+2420	84.69	8978+4580	3896+3778	84.66
Adult	13223+4429	3535+3443	84.81	12811+6457	5449+5292	84.74

表 3.14　*a*-SVM 实验结果 （中心点）（Gaussian Kernel）（$\sigma=1, \lambda=1$）
（M=1，T=1，ND=样本数， Nmax=10，sig_s(1,-1)=0.0125，Dm(1,-1)=0.03125，*l*=2）

	超球			超立方体		
数据集	n_1+n_2	s_1+s_2	正确率	n_1+n_2	S_1+s_2	正确率
Balance-scale	144+171	54+54	98.08	118+165	62+69	94.72
Adult-1	689+227	242+199	82.58	753+356	346+312	83.1
Adult-2	957+334	345+290	82.49	1076+501	480+441	82.91
Adult-3	1355+454	448+390	82.96	1472+666	628+580	83.25
Adult-4	2012+690	616+553	83.6	2051+1022	929+864	83.47
Adult-5	2729+918	812+740	83.94	2718+1333	1207+1141	83.58
Adult-6	4788+1530	1325+1224	84.06	4498+2224	2003+1905	84.24
Adult-7	6816+2253	1903+1769	84.34	6429+3235	2803+2709	84.3
Adult-8	9529+3107	2577+2420	84.65	8978+4580	3891+3778	84.63
Adult	13223+4429	3753+3435	84.74	12811+6457	5443+5290	84.72

3.6.2　标准数据集上的实验结果

表 3.12~表 3.14 中的结果是在标准数据集合上面进行的，考察了两个数据集：Balance-scale 和 Adult[51]。其中 Adult 被分成为 9 个嵌套的集合，可以比较方便地考察样本数据的规模同分类器的训练之间的关联。随机采样的实验数据是随机选取样本集合规模的 1/3 的数据。从表 3.12 右边的实验结果可以看出，当对一个真实获得的数据集进行随机采样，选取若干个样本训练而获得的分类器的正确率有时是比较低的。仿真数据上随机采样样本所训练而得的分类器的效果之所以还比较好，是因为仿真数据的生成是严格满足独立同分布的，而从真实数据集中随机选取的数据有可能在样本空间中集中在某一区域，不能够代表整个训练样本集合，因此，当给定一个大规模的数据集时，是必须对整个数据集合进行处理的。从表 3.14 的数据可看出，采取中心点进行训练在样本集合规模较大时，正确率是要低于采用区域风险最小化原则进行训练的。而在样本规模比较小时，二者的区别并不大，这是因为在样本规模较小时，每个区域内的样本过于稀疏，采用区域半径去控制分类边界的效果并不明显，也不准确。而当样本规模较大时，区域内的样本相对比较密集，这时候采用区域中心点和半

径进行控制能够得到相对比较好的分类效果。Wang 等[57]提出的基于聚类的方法可以看作是采用区域的中心点进行控制，表 3.13 表明，当样本个数较大时，本文提出的方法较 Wang 等提出的方法更为有效。另外，从表 3.13 的实验数据可以看出，采取区域风险最小化原则去训练分类器可以使得训练样本的个数和支持向量的个数均相对有所减少。

3.7　本章小结

本章提出了一种区域风险最小化原则。当给定的训练样本在特征空间中局部具有聚类性质，根据样本的这种性质将训练样本在粗粒度上面进行表示，生成一个粗粒度下的训练样本集合，每个粗粒度下的训练样本可能包含多个原始的训练样本。本章讨论了如何划分训练样本集合以及如何基于区域求解分类器。本章的后续工作包括以下几个方面：

1. 区域的划分。本章在讨论基于超立方体和超球的划分时，指出超球和超立方体两种划分方式都有不足的地方。超球在两类样本混合的区域的区分能力差，超立方体有可能其中的样本分布不均匀。因此讨论更为合理的对训练样本集合的划分方式是非常有必要的。在本章的实验中，对于训练样本的划分的区域个数较多，因此相比于采取中心点的方法，优势不明显，进一步可以探讨如何减少区域划分的数目；

2. 算法的性能分析。从理论上面分析本章提出的将训练样本集合进行划分的方法同其它的方法之间的关联。

第 4 章　基于特征空间划分的多分辨率分类问题求解

4.1　本章引论

本章首先回顾多分辨率分析理论与模式分类问题相结合的一些研究成果，然后具体给出基于特征空间划分的多分辨率分类问题的求解模型和算法描述。本章提出的方法的基本思想是：在分类器的训练阶段，通过逐步缩小训练样本在样本空间中的范围，训练出一组不同粒度级别（分辨率）下面的分类器，当训练样本具有局部聚类性质时，理论分析和实验证明，这种策略可以减少参与分类器训练的样本个数，从而降低分类器训练的计算复杂性；在分类器的测试阶段，对于给定的测试集合，使用训练阶段训练出的一组分类器对样本进行测试，初步的实验结果表明，当给定样本的分布具有聚集性质时，大量样本可以在较粗的分辨率下判别出类别来，从而降低测试阶段的计算复杂性。

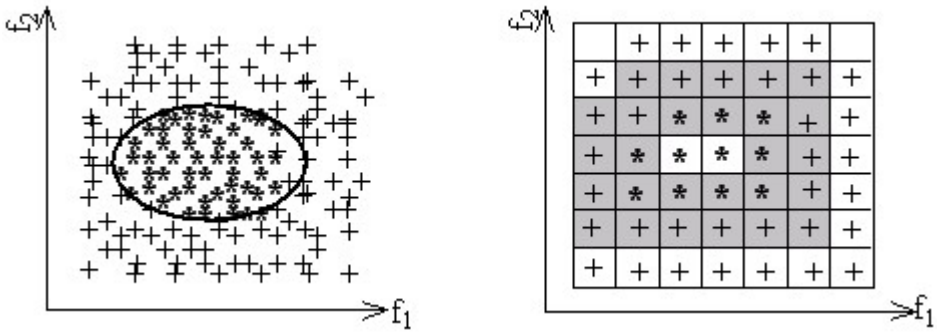
本章内容如下：第 4.2 节介绍要解决的基本问题和已有的相关工作；第 4.3 节给出了基于多分辨率的分类问题求解模型；第 4.4 节描述了分类器训练阶段和测试阶段的算法；第 4.5 节给出了计算模型参数的算法；第 4.6 节设计了一个基于支持向量机的多分辨率分类器；第 4.7 节分析了算法的性能；最后的实验结果表明了本章提出算法的有效性。

4.2　基本问题与相关工作

4.2.1　基本问题描述

多分辨率分析，多尺度分析以及粒度计算是目前受到广泛关注的几个研究领域。这几种研究方法通过将所研究的问题进行粗、细不同层次上面的表示，从而使得问题的求解会变得更容易、计算量更小。对于分类问题，由于在实际的应用中发现，当训练样本集合规模比较大时，许多训练样本在特征空间中往往具有一定的聚类性质，因此对于特征空间中某些样本聚集的区域可以单独作为一个整体去参加分类器的训练，从而不需要单独考虑其中的每个样本。另外人类在解决分类问题时，往往是先粗略确定一个分类边界，然后再逐步细化求

取出较为精确的分类边界，从而加快了分类边界寻找的速度。在第 3 章中重点讨论了如何将一组训练样本作为一个基本单位去训练分类函数，本章则着重讨论第 2 章所提出的另外一个问题，即如何基于多粒度的论域表示，加快分类函数的求解速度。图 4.1 中给出了一个样本分布的示意图。图 4.1(a)表示了训练样本在特征空间的分布，并且给出了基于原始样本的分类边界。图 4.1(b)表示了训



(a)　　(b)

图 4.1 特征空间划分示意图

练样本在低分辨率下的分布属性，其中将训练样本在每维的特征轴上面均匀分为 7 部分，图中的灰粗线表示样本在这个粗分辨率下的粗分类边界，那些不落入到粗分类边界中的训练样本在确定更为精细的分类边界时可以不再考虑。根据上面的想法，在本章中将提出一种基于特征空间划分的多分辨率的分类方法，该方法采取一种逐步求精地策略去确定分类边界，从而降低分类器训练过程的计算复杂性。基于多分辨率的思想去训练分类器和对测试样本集合进行测试的基本含义如图 4.2 和图 4.3 所示。

采取这种多分辨率的策略去训练分类器可以带来如下的好处：

1. 由于所获得的训练样本在特征空间的某些区域通常具有聚类的性质，所以由粗到细的分类方式应该能够降低分类器训练阶段的计算复杂性。相应的问题描述如下：给定训练样本集合 $Z=\{z_i\}_{i=1}^{N}$，该样本集合对应于分布函数 P。假设使用策略 S，求解出分类函数 f，训练出分类函数 f 的时间复杂性为 M_f。希望能够设计一个基于多分辨率的求解策略，该策略通过层次化的求解方式，获得分类函数 g，求得分类函数 g 的计算复杂性为 M_g。分类函数 g 逼近分类函数 f，并且在样本的某些分布情况下 $M_g < M_f$。

2. 在样本的测试阶段也可以采用多分辨率的测试方式，同样希望在某些样

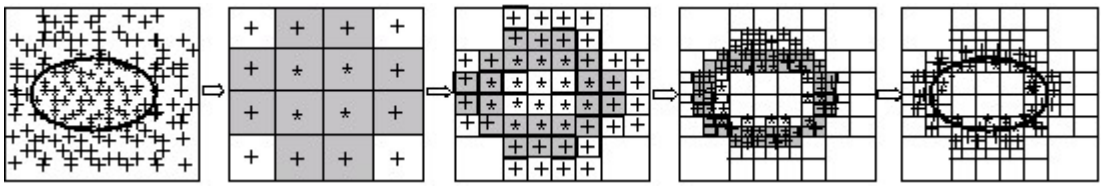


图 4.2　多分辨率分类器训练示意图

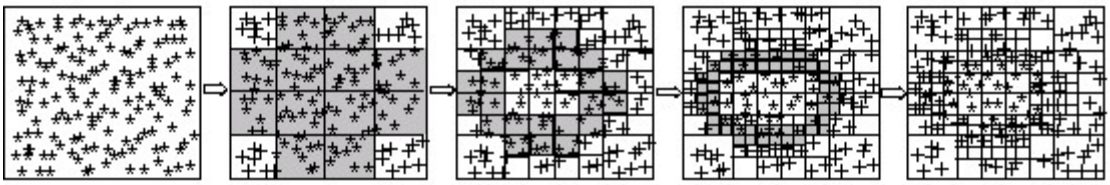


图 4.3　多分辨率分类器测试示意图

本分布的情况下可以降低计算复杂性。相应的问题描述如下：给定测试样本集合 $X=\{x_i\}_{i=1}^{N}$，假设使用原始的分类函数对测试样本集合中的 N 个样本进行测试的计算复杂性为 M_f。而采用多粒度的测试方式的计算复杂性为 M_g。希望在某些样本分布的情况下 $M_g < M_f$。

3. 在实际应用中，并非任何时候都需要一个具有较好泛化能力的、较高分类精度的分类器。有时候分类速度是比较关键的因素，并不要求较高的分类精度。由于粗分辨率下所训练的分类器涉及得训练样本较少，通常采用结构简单的分类器即可以完成分类任务，这样分类的速度较快，但是错误率较高。因此当训练了一组由粗到细的分类器后，可以根据应用所要求的分类时间和分类精度，选择合适分辨率下的分类器去解决问题。

4.2.2　相关工作概述

将多分辨率的思想同分类算法结合已经有长时间的研究历史了。在第二章中简要介绍了将分层次求解理论同模式分类问题相结合的几种方式，本章主要考虑的是如何采用分层次求解的方式获得要训练的分类器。假设分类器的表示形式如公式（4-1）所示。

$$f(x)=\mathrm{sgn}(\sum_{i=1}^{l}\alpha_i\Phi(x,x_i)+\alpha_0) \qquad (4\text{-}1)$$

为了获得函数 $f(x)$，需要求解以下参数：基函数集合 $\{\Phi(x,x_i)\}_{i=1}^{l}$ 和系数集合 $\{\alpha_i\}_{i=0}^{l}$。若函数 $\Phi(x)$ 给定，那么基函数集合是由点集 $\{x_i\}_{i=1}^{l}$ 所确定的。点集 $\{x_i\}_{i=1}^{l}$

的求解可以同训练样本集合关联起来，因此将训练样本集合进行不同分辨率下的表示可以帮助快速求解集合$\{x_i\}_{i=1}^{l}$。若函数$\Phi(x)$未定，那么首先要选择函数$\Phi(x)$。根据上面的分析，可以将多分辨率策略与分类问题结合起来的方式分为三种：训练样本的多分辨率表示、基函数集合的多分辨率表示以及分类函数的多分辨率求解策略。下面从这三方面综述已有的研究成果：

1. 基函数集合的多分辨率表示

早期的多分辨率分类方法大部分指的是由于基函数集合选择的不同，导致最终分类器的复杂程度不同。基函数集合中的基函数的类型和数目能够控制所求得的分类函数可以达到的复杂程度，例如，基于基函数集合$\left\{\exp\left(-\left|x-c_i\right|/\sigma_1\right)\right\}_{i=1}^{N}$和$\left\{\exp\left(-\left|x-c_i\right|/\sigma_2\right)\right\}_{i=1}^{N}$ $\left(\sigma_1\neq\sigma_2\right)$的分类函数所能够达到的复杂程度是不同的。已有的研究成果包括：Simpson 所设计的模糊最小-最大神经网络[66]。Garcke 所设计的基于稀疏网格的神经网络，其中的基函数以稀疏网格作为基本的表示形式，当选择的稀疏网格不同的时候，能够得到的分类器的复杂程度也是不同的[67]。Zhang 等所设计的小波神经网络是采用小波函数作为网络中间层次的基函数，当选择不同的小波基的时候，所得到的分类函数的复杂程度也是不同的[68]。Shao 等设计了一个多分辨率的支持向量机，其中使用了不同尺度的核函数来表示分类函数[92]。

2. 训练样本的多分辨率表示：

本章所考虑的多分辨率分类算法的出发点是希望将特征空间进行划分来获得粗分辨率下面的训练样本。在分类问题中，近期有一部分参考文献涉及到了划分特征空间。Blayavs 提出的多分辨率的分类算法采用树结构来表示不同分辨率下的特征空间，并且在这个树结构上面进行训练和查询[69]-[71]。Singh 所讨论的样本集合的可分性问题同样也是基于将特征空间划分成为超立方体（或超长方体），他的论文从四个角度讨论了样本空间的可分性：纯度、邻域可分性、累积熵以及数据简洁性，所给出的包含样本的超立方体的某些衡量参数可以看作是对样本空间的不确定表示[72][73]。Yu 等设计了一个基于层次聚类的逐步缩小分类边界的分类算法，该算法通过将训练样本在特征空间中进行分层次的聚类，从而将样本集合由粗到细地划分[74]。

3. 分类函数的多分辨率求解策略

在处理回归问题时，Liang和Chan采用了一种多分辨率的策略调整神经网络的权值[75]-[77]。Yu等在训练分类函数时，首先对正例样本和反例样本进行分层次

的聚类，生成两棵聚类树，然后再由粗到细训练分类器[74]。

上述 Blayavs 的算法和 Yu 等的算法都是采用了多分辨率或由粗到细的策略来考虑分类问题的。Blayavs 的算法是基于生成式模型的，训练出来的是一个树结构，每层的超立方格由该立方格内的样本数和样本特征值的平均值来表示[69]-[71]。当对一个样本进行测试时，包含测试样本的最细的超立方格的类别标识即为测试样本的类别标识。Blayavs 的算法有一些不足之处。首先，该算法是顺序扫描训练样本，无论是训练还是测试阶段，均需要考虑到包含训练样本的最细层次，并不能减少计算量。其次，当某些格内的训练样本比较稀疏时，采用这种方式构造的分类器的泛化性能要差。第三，相比较于区分式的模型，需要保存大量的表达分类器的参数。Yu 等提出的算法是采取一种逐步求精的策略来获得区分函数[74]，但是该算法首先要将正例样本和反例样本都聚类成为一个层次树，聚类的过程本身就是耗时的，另外同本文第二章遇到的问题一样，正例和反例单独聚类对于位于分类边界的样本来说区分能力是比较差的。

下面提出一种基于特征空间划分的多分辨率分类方法。划分特征空间的目的是希望能够尽快地训练获得合适的分类函数，并且粗分辨率下的分类器能够有助于加快样本集合测试的速度。

4.3　模型描述

4.3.1　样本集合的多分辨率分解

为了获得不同分辨率下的训练样本，先要将训练样本集合进行多分辨率分解。下面结合多分辨率领域内的四叉树分解策略[78]和粒度计算中的论域划分策略来给出特征空间的分解策略[15]。设 $\Omega \subset R^n$ 为包含样本集合 X 的特征空间中的一个超长方体（或超立方体）。对于一个给定的数 l，如果将 Ω 在每一维上面均 l 等分，那么 Ω 将被分解成为 l^d 个小超长方体（或超立方体），即 $\Omega = \cup_{i=1}^{l^d} o_i$，其中 o_i 是一个小格子。

定义 4.1　o_i 上面的关系 R_l 定义为：对于 $\forall x, y \in \Omega$，如果 $xR_l y$，那么 x 和 y 属于同一个 o_i。

定理 4.1　关系 R_l 构成 Ω 上的一个等价划分。

证明：首先证明关系 R_l 是一个等价关系。显然关系 R_l 满足自反性、对称性

和传递性，是一个等价关系。由于关系 R_l 将样本空间划分为不相交的超长方体，并且由关系 R_l 的定义，易知各个等价类之间不相交，并且等价类的并集为 X，因此 R_l 构成域 X 上面的一个等价划分。　　证毕。

定义 4.2　数值 l 所决定的关系 R_l 对 Ω 所进行的划分，称为对 Ω 基于分辨率 l 的划分。

例如，　图 4.1(b)中 l=7。

定义 4.3 对于给定的整数 l 和 k，$\{\Omega_{2^i l}\}_{i=0}^{k}$ 称为对 Ω 的多分辨率分解，l 和 $2^k l$ 定义为初始和终止分辨率。

定理 4.2　给定数值 l，由数值 $k^n l(n=0,1,2...)$ 所定义的关系 $R_{k^n l}$ 对 Ω 的划分构成一个层次关系。

证明：若能够证明 $R_{k^{n+1}l}$ 对域 U 的划分能够构成 $R_{k^n l}$ 对域 U 划分的商空间 $[U]_{k^n l}$ 的商空间，就可以说明上述定义的划分可以构成一个层次关系。由于 $R_{k^{n+1}l}=R_{k\cdot k^n l}$。因此 $R_{k\cdot k^n l}$ 可以看作将关系 $R_{k^n l}$ 将样本空间划分的超立方格中的每个再 k 等分。因此有 $\forall x,y\in X$，如果 $xR_{k^{n+1}l}y$，那么 $xR_{k^n l}y$。所以 $R_{k^n l}<R_{k^{n+1}l}$。所以域 $[U]_{k^{n+1}l}$ 构成域 $[U]_{k^n l}$ 的商空间。因此由数值 $k^n l(n=0,1,2...)$ 所定义的关系 $R_{k^n l}$ 对域 X 的划分构成一个层次关系。　　证毕。

对于分类问题来说，考虑的是样本集合 X。根据对 Ω 的多分辨率分解，X 同样也被分解成为一个层次化集合 $\{X_{2^i l}\}_{i=0}^{k}$。其中 $X_i=X\setminus R_i$，对于 $\forall x_{i,j}\in X_i$，公式（4-2）成立。

$$x_{i,j}=X\bigcap o_{i,j},\ \ o_{i,j}\in\Omega_i \qquad (4\text{-}2)$$

其中 $o_{i,j}$ 是 $x_{i,j}$ 所对应的超长方体。

4.3.2　训练样本构造

为了获取某一分辨率下面的训练样本，需要对每个超长方体 x_i 构造特征值 $g(x_i)$ 和类别标识 y_i。

1.　特征值

使用 X_l 构造分辨率 l 下面训练样本的特征值。为了能够利用已有的分类算法，对于 $\forall x\in X_l$，使用 Ω 内的一点 $g(x)$ 去表示一个超长方体 x 的特征值以及唯

一的类别标识 $y \in \{-1,1\}$ 来表示 x 所属于的类别。为了计算简单，可以采用分辨率 l 下 x 所位于的超长方体的中心点的特征值作为 x 的特征值。即对于 $\forall x_{l,i} \in X_l$，$g(x_{l,i})$ 的值如公式（4-3）所示。

$$g(x_{l,i}) = center(o_{l,i}) \tag{4-3}$$

如果假设两类样本在核函数的变换下在特征空间中是线性可分的，则同样希望所获得分类样本在高维特征空间上是线性可分的，可如公式（4-4）构造特征值：

$$g(x_{l,i}) = \begin{cases} x_1^*, & \text{如果} y_{l,i} = 1 \\ x_2^*, & \text{如果} y_{l,i} = -1 \end{cases} \tag{4-4}$$

其中 x_1^* 是在超长方体 x 中离中心点最近的第一类样本集合中的一个样本，x_2^* 是在超长方体 x 中离中心点最近的第二类样本集合中的一个样本。

2. 类别标识

而对于类别标识 y，可以采用公式（4-5）进行计算。

$$y_{l,i} = \mathrm{sgn}\left(\frac{1}{|x_{l,i}|} \sum_{x_j \in x_{l,i}} y_j \right) \tag{4-5}$$

使用文献[27]中的理论，可对公式（4-5）做如下解释：假设 $x_{l,i}$ 属于类别 1 的概率为 $p_1(x)$，$x_{l,i}$ 属于类别 2 的概率为 $p_2(x)$，则 $y_{l,i}$ 也可以表示为：

$$y_{l,i} = \begin{cases} 1, & p_1(x) > p_2(x) \\ -1, & p_1(x) \le p_2(x) \end{cases} \tag{4-6}$$

统计 $x_{l,i}$ 中两类样本的个数，分别记为 $N_1(x)$ 和 $N_2(x)$。$p_1(x)$ 和 $p_2(x)$ 的定义如下：

$$p_1(x) = \frac{N_1(x)}{N_1(x) + N_2(x)}, \qquad p_2(x) = \frac{N_2(x)}{N_1(x) + N_2(x)} \tag{4-7}$$

可得到如公式（4-8）所示的关系。

$$p_j(x) = \sum_{i=1}^{2^D} \alpha_i p_j(x_i), \qquad j = 1,2 \tag{4-8}$$

其中
$$\sum_{i=1}^{2^D}\alpha_i = 1, \qquad \alpha_i = \frac{N_1(x_i)+N_2(x_i)}{\sum_{j=1}^{2^D}(N_1(x_j)+N_2(x_j))} \tag{4-9}$$

由公式（4-8）可以看出，$p_j(x)$ 落入到$\{p_j(x_i)\}(i=1,2,\cdots,2^d)$所构成的凸包内。而 $p_1(x)$和 $p_2(x)$ 决定了 $y_{l,i}$ 的值，所以粗分辨率的信息是细分辨率信息的一种统计量，粗分辨率下的分类函数可以用来估计原始分类函数。另外，由于较细的分辨率包含了更多的细节信息，所以采用多分辨率的分类方式可以逐步给出较为精确的分类边界。

分辨率 l 下的训练样本可以表示为 $Z_l = \{(g(x_{l,i}), y_{l,i})\}_{i=1}^{N_l}(N_l \le |\Omega_l|)$。假定已经构造出了分辨率 i 下的训练样本，并且训练得到分类器 $f_i(x)$，那么只需要进一步细化那些位于分类边界的样本去获得更细分辨率下面的训练样本。将分辨率 i 下的超长方体分为两类，类内区域 $I_{\Omega,i}$ 和类边界区域 $B_{\Omega,i}$，如下定义：

定义 4.4 类内区域 $I_{\Omega,l}$ 定义为：对于 $\forall o \in \Omega_l$，如果 $|f_l(g(o))| > T_B$，那么 $o \in I_{\Omega,l}$。

定义 4.5 类边界区域 $B_{\Omega,l}$ 定义为：对于 $\forall o \in \Omega_l$，如果 $|f_l(g(o))| \le T_B$，那么 $o \in B_{\Omega,l}$。

其中 T_B 是控制边界宽度的一个阈值。图 4.4 给出了两种区域的示意图。

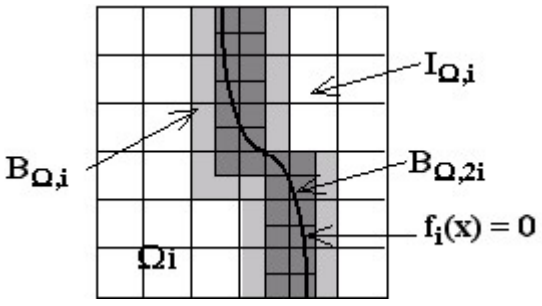


图 4.4　训练样本细分示意图

基于上面的划分，X_i 可以被分为两个部分，类内部集合 $I_{X,i}$ 和类边界集合 $B_{X,i}$。由于类内部集合不位于两类的分类边界上，因此为了获得更细一级分辨率 $2l$ 下面的训练样本，可以仅仅考虑类边界集合 $B_{X,l} \subset X_l$，设 $X'_{2l} = (\cup B_{X,l}) \setminus R_{2l}$，因此分辨率 $2l$ 下的训练样本可以表示式（4-10）的形式。

$$Z_{2l} = \{(g(x_{2l,i}), y_{2l,i})\}_{i=1}^{N_{2l}}, \qquad x_{2l} \in X'_{2l} \tag{4-10}$$

由于 $X_{2l}^{'}$ 是由 $B_{X,l}$ 所生成的，因此下面的结论成立：$\left|X_{2l}^{'}\right| \le \left|B_{X,l}\right| \times 2^d$，并且对于 $\forall x \in X_{2l}$，$\exists y \in B_{X,l}$，满足 $x \subseteq y$。

通过上面的这种层次化的样本构造方式，可以获得从初始分辨率到终止分辨率的训练样本集合 $Z_i\left(l \le i \le 2^k l\right)$，并且训练获得一组分类器 $f_i(x)\left(l \le i \le 2^k l\right)$。

4.3.3　多分辨率分类器

虽然分类函数 $f_i(x)$ 是在分辨率 i 下所训练的，但是它实际上是原始特征空间上的一个分类函数，函数的输入是特征空间中的一点，而非分辨率 i 下的基于超长方体的基本单位。下面给出分辨率 i 下的分类函数 $F_i(x)$ 的定义，$F_i(x)$ 的输入是分辨率 i 下的一个超长方体，$F_i(x)$ 所确定的分类边界是分辨率 i 下两类样本之间的一个粗边界。

定义 4.6 分辨率 i 下的分类器 $F_i(x)$ 如下定义（其中 x 是分辨率 i 下面的一个元素，并且满足：$\exists o_{i,j} \in \Omega_i, x \subseteq o_{i,j}$）：

如果 $i \ne 2^k l$：

$$F_i(x) = \begin{cases} 1: & \text{如果} \quad f_i(g(x)) \ge 0 \,\&\, x \notin B_{X,i} \\ -1: & \text{如果} \quad f_i(g(x)) < 0 \,\&\, x \notin B_{X,i} \\ 0: & \text{如果} \quad x \in B_{X,i} \end{cases} \tag{4-11}$$

其中 0 表示 x 的类别不能够分辨率 i 下的分类器所确定。

如果 $i = 2^k l$：

$$F_i(x) = \begin{cases} 1: & \text{如果} \quad f_i(g(x)) \ge 0 \\ -1: & \text{如果} \quad f_i(g(x)) < 0 \end{cases} \tag{4-12}$$

因此，根据前面所训练的各个分辨率下的分类函数 $f_i(x)\left(l \le i \le 2^k l\right)$，可以获得一系列的分类器：

$$F_l(x), F_{2l}(x), ..., F_{2^k l}(x) \tag{4-13}$$

设 $F(x)(x \in R^d)$ 表示多分辨率分类器，其中输入 x 是特征空间中的一个点：

$$F(x) = \begin{cases} 1: \text{ 如果} x \in class \ \ 1 \\ -1: \text{ 如果} x \in class \ \ 2 \end{cases} \tag{4-14}$$

下面，将使用$\{F_{2^i l}\}_{i=0}^{k}$来表示$F(x)$。

定义 4.7 对于给定的样本$x \in R^d$，设$\omega(x,i)$表示分辨率 i 下 x 所位于的超长方体。

定义 4.8 多分辨率分类器$F(x)(x \in R^d)$定义为：

$$F(x) = F_i(g(\omega(x,i))) \tag{4-15}$$

其中 i 是满足$F_i(g(\omega(x,i))) \neq 0$的最小值。

图 4.5 给出了多分辨率分类训练阶段的流程图。

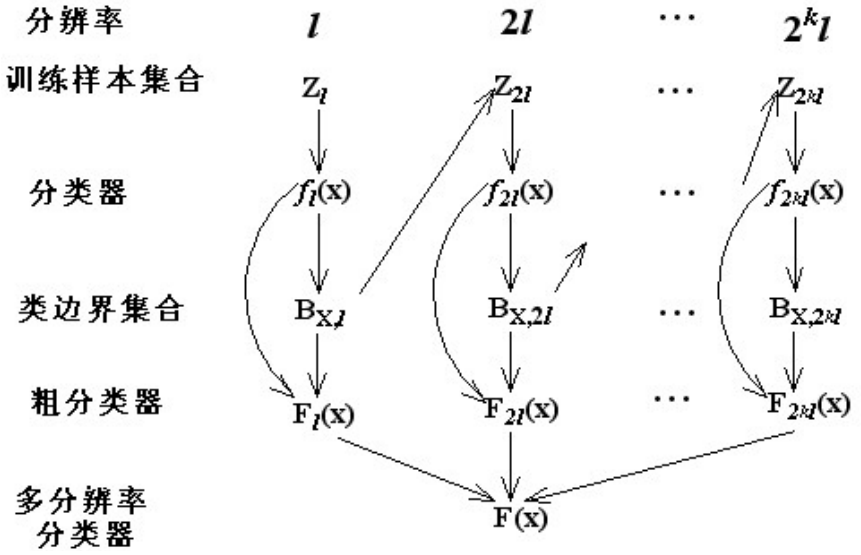


图 4.5　多分辨率分类器训练流程

4.4　多分辨率分类算法

对于给定的训练样本集合 Z，遵循第 4.3.3 所描述的原理，即可获得多分辨率分类器。具体算法描述这里省略。

在样本的测试阶段，由于采用的是多分辨率的策略，因此测试过程同传统的分类器的测试过程是不同的。假设给定的测试样本集为 T，测试算法如算法 4.1。执行完算法 4.1 以后，对于$\forall t \in T$，$Label(t)$表示样本 t 的类别。图 4.6 给出了多分辨率分类器测试阶段的流程图。

注意，$f_i(x)$本身也是原始特征空间上面的一个分类器，并且在理想情况下，

算法 4.1 多分辨率分类器测试算法

MultiClassifierTesting

1 假设 T 内的所有样本均落入到 Ω 内，若不满足，那么修改 Ω 的覆盖范围，并且保持特征空间的划分位置不变；

2 $i=l$，$T_i^{'}=T_i$；

3 使用分类器 F_i 来分类 $T_i^{'}$；

4 对于 $\forall t\in T_i^{'}$，如果 $F_i(t)\neq 0$，那么对于 $\forall u\in t$，$Label(u)=F_i(t)$，否则对于 $\forall u\in t$，$u\in B_{T,i}$；

5 $i=i\times 2$，$T_i^{'}=(\bigcup B_{T,i/2})/R_{2i}$，

6 如果 $i>2^k l$，那么跳到第 7 步，否则到第 3 步；

7 结束。

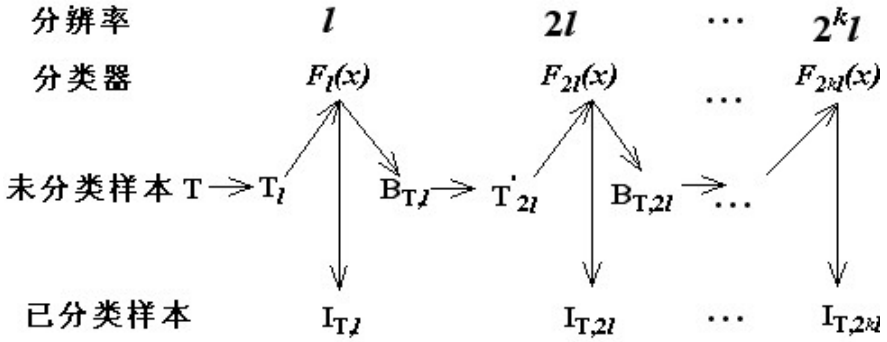


图 4.6　多分辨率分类器测试流程

$f_{2^k l}(x)$ 能够逼近 $f(x)$。$f_i(x)$ 的复杂程度通常要小于 $f_{2i}(x)$，当对测试样本进行测试时，$f_i(x)$ 的计算量会更小一些。另外，由于 $f_{2i}(x)$ 实际上考虑了更多的训练样本的影响，因此一般分类器 $f_{2i}(x)$ 的错误率要小于分类器 $f_i(x)$。基于应用给定的错误率和所要求的分类速度，可以在分类器集合 $\{f_{2^i l}(x)\}_{i-0}^{k}$ 内选择合适的分类器。

4.5　参数选择

4.5.1　初始分辨率

显而易见，多分辨率分类器的有效性是同样本的分布有关的，因此，初始分辨率的选择是至关重要的。当选择的初始分辨率不合适，落入到类内区域的样本比较少时，采用多分辨率的策略就不能够缩减位于边界的训练样本所在的范围，相反还会增加计算量。对样本空间进行划分本身也是一种分层次的聚类方法。实际上，在某一分辨率 i 下，当存在一定比例的超立方体比较“纯”（超立方体中的大部分样本都来自于某一类）时，分辨率 i 下所获得的训练样本在分类器训练阶段才有意义，这样才可能存在一部分样本会落入到类内区域中。下面定义样本集合 X_i 的纯度。

定义 4.9 对于 $\forall x \in X_i$，x 的纯度定义为 x 的熵：

$$H(x) = -(p_1 \log_2 p_1 + p_2 \log_2 p_2) \tag{4-16}$$

其中 $p_1 = n_1(x)/(n_1(x)+n_2(x))$ 并且 $p_2 = n_2(x)/(n_1(x)+n_2(x))$，$n_1(x)$ 和 $n_2(x)$ 表示 x 内来自于类 1 和类 2 的样本数目。若 $H(x)=0$，那么 x 是很纯的，即所有的训练样本都来自于同一类，若 $H(x)=1$，那么说明 x 是由两类训练所组成的，并且两类样本在 x 内所占的比例是一样的。

定义 4.10 X_i 的聚类度 $C(X_i)$ 定义如下：

$$C(X_i) = 1 - \frac{\sum_{j=1}^{|X_i|} H(x_j)}{|X_i|} \tag{4-17}$$

当 $0 \le C(X_i) \le 1$ 时，$C(X_i)$ 越大，就说明 X_i 中较纯的超立方体越多。若 $C(X_i)=1$，那么 X_i 中所有超立方体中的样本都均来自于一类；若 $C(X_i)=0$，那么 X_i 中所有超立方体中的样本都均来自于两类，并且两类样本在每个超立方体中占的比例都是 1/2。

为了获得合适的初始分辨率，设置一个阈值 β_0，当 $C(X_i) > \beta_0$ 时，那么认为分辨率 i 是合适的初始分辨率。当 $C(X_i)$ 的值比较小时，显然在分辨率 i 去训练粗分类器的意义不大。

4.5.2　终止分辨率

当超立方体内的平均样本数比较少时，继续划分超立方体的意义不大，因

为这样做分类边界的改善不大，反而增加了许多计算量。因此，终止分辨率的确定也很重要。下面通过考虑某一分辨率下超立方体内样本的平均数目来确定终止分辨率。

定义 4.11　X_i 的平均密度定义为：

$$Avg(X_i)=\frac{1}{|X_i|}\sum_{j=1}^{|X_i|}|x_{i,j}| \tag{4-18}$$

其中 $|x|$ 表示 x 中元素的个数。

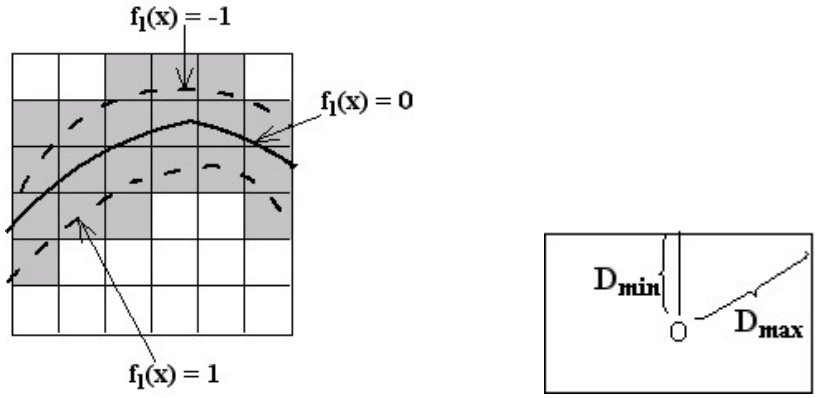
如果分辨率 i 满足下面的条件，那么称分辨率 i 为终止分辨率。

$$|Avg(X_i)-Avg(X_{i/2})|<T_N \tag{4-19}$$

其中 T_N 是控制终止分辨率的一个阈值。

4.6　基于支持向量机的分类器设计

本节选用支持向量机作为基本的分类器来设计一个多分辨率分类器。要描述所设计的多分辨率分类器，关键的问题是确定类边界区域。支持向量机的一个最大的优点是分类间隔线 $f(x)=1$ 和 $f(x)=-1$ 位于分类边界上，并且这两条分类间隔线穿过所有的支持向量。因此，支持向量机提供了一个比较方便的方式去获得粗分类边界。理想的情况下，希望位于 $f(x)=1$ 和 $f(x)=-1$ 之间以及 $f(x)=1$ 和 $f(x)=-1$ 穿过的格子均被认为处于类边界区域内，如图 4.7(a)所示。



(a)　　(b)

图 4.7　粗分类边界计算

假设分辨率 i 下对特征空间划分获得的格子是超长方体，再假设超长方体边长的最大值和最小值分别为 $2D_{i,\max}$ 和 $2D_{i,\min}$，如图 4.7(b)所示，那么按照下面的定义可确定类边界区域：

定义 4.12 分辨率 i 下的边界集合是满足下面的条件的样本构成的集合，对于 $\forall x \in X_l$：

1. 如果 $\left|f_i(g(x))\right| \geq D_{i,\max}+1$，那么 $x \notin B_{X,i}$；
2. 如果 $\left|f_i(g(x))\right| \leq D_{i,\min}+1$，那么 $x \in B_{X,i}$；
3. 如果 $(D_{i,\min}+1) \leq \left|f_i(g(x))\right| \leq (D_{i,\max}+1)$，那么为了简单，认为 $x \in B_{X,i}$；

若划分后的格子均为超立方体，那么 $D_{i,\max} = D_{i,\min}$。

$B_{X,i}$ 包括两种类型的超长方体：(1)支持向量；(2) $f_i(x)=1$ 和 $f_i(x)=-1$ 穿过的格子。为了计算简便，可以只考虑支持向量所位于的那些格子。虽然比全部考虑(1)、(2)两种情况在下一分辨率下所获得的训练样本要少些，但是由于支持向量机算法是采取最大间隔超平面来选择合适的分类函数的，因此只考虑支持向量机所位于的格子对最终的结果的影响不会很大。

4.7　性能分析

提出多分辨率分类器的目的是希望能够在样本数目较大，并且某些样本局部聚在一起的时候，能够通过由粗到细的多分辨率的方式去降低算法的计算复杂性，同时这种策略又保证不增加过多的错误率。因此，算法的性能分析非常重要。在本章中将分析前面提出的多分辨率分类算法的性能，主要是关注计算复杂性和泛化性能两个方面。

4.7.1　泛化性能分析

设计一个多分辨率分类器的目的是降低分类问题训练和测试过程的计算复杂性，并且对比不使用多分辨率策略的分类算法来说这种策略应该不能够带来较大的错误率。下面讨论对于给定的样本集 Z，使用 $F(x)$进行分类的错误率。根据经验风险的表达公式，在分辨率 $i=2^m l(0 \leq m \leq k)$ 下，样本集合 Z 的经验风险可以表示为：

$$R_{emp,i}(F_i(\cdot)) = \sum_{x_j \in (\cup B_{\Omega,i/2}) \cap (\cup I_{\Omega,i})} L(y_j, F_i(\omega(x_j,i))) \qquad (4\text{-}20)$$

其中$\bigcup B_{\Omega,l/2}=\phi$并且$\bigcup I_{\Omega,2^k l}=\phi$。$L(\cdot,\cdot)$如下定义：

$$L(y,F_i(x))=\begin{cases}0:\ \text{如果}\ \ y=F_i(x)\\1:\ \text{如果}\ \ y\neq F_i(x)\end{cases}\qquad(4\text{-}21)$$

因此，$F(x)$的经验风险可以表示为：

$$R_{emp}(F(\cdot))=\frac{1}{N}\sum_{i=0}^{k}R_{emp,2^i l}(F_{2^i l}(\cdot))\qquad(4\text{-}22)$$

由于在给每个超立方体赋予类别值时，对每个超立方体赋予了一个唯一的类别值，因此某些不纯的超立方体内的样本可能错分类的概率比较大。图 4.8 给出了一个例子：对于左下角来自于类别 1 的样本，由于它们所在的每个超立方体内来自于类别 2 的样本均大于来自于类别 1 的样本，因此采用多分辨率的策略时左下角这些来自于类别 1 的样本会被完全错分。

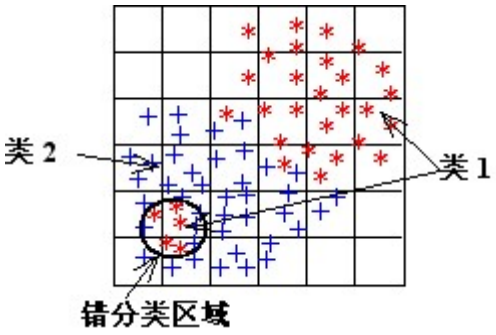


图 4.8　错分区域示意图

4.7.2　计算复杂性分析

计算复杂性分析是本章设计的算法的一个重要方面，下面分析算法各部分的计算复杂性。

1.　求取初始分辨率：

假设给定一个规模为 N 的训练样本集合$Z=\{(x_i,y_i)\}_{i=1}^{N}$。首先为了能够将训练样本在特征空间划分，需要求出样本在各个维上的最大最小值，这需要扫描样本集合一次，计算量为$O(N)$。假设用于计算初始分辨率的起始值为 k，对于一个规模为 n 的样本集合，扫描样本集合并将每个样本放入到对应的超立方体中的计算量为$O(n)$。下面考虑超立方体拆分的计算复杂性：假设共有 m 个超立

方体，其中每个超立方体中的样本数目为 $n_i\ (1 \le i \le m)$，那么将这 m 个超立方体每个拆分成 2^d 个子超立方体的计算复杂性为 $O(n), n = n_1 + ... + n_m$。因此超立方体拆分阶段的计算复杂性是线性的。

综上所述，那么求取初始分辨率部分的计算量为 $O(tN)$，其中 t 为循环迭代的数目，由第 4.5 节介绍的控制初始和终止分辨率值的参数所决定。

2. 分类阶段

设计算出来的初始分辨率为 l，终止分辨率为 $e = 2^k l$，并且分辨率以 2 的倍数递增，从初始分辨率到终止分辨率共训练了 k+1 个分类器。再设样本空间的维数为 d，用 $|X|$ 表示集合 X 中的元素个数，$B_{X,i}$ 表示分辨率 i 下类边界区域中的那些超立方体，X'_{2i} 表示通过细分 $B_{X,i}$ 中的超立方体所获得的分辨率 $2i$ 下的超立方体。$u(x)$ 表示所选用的分类器对数目为 x 的样本集合进行分类的计算量。

每个分辨率下面的计算量由训练样本生成、分类器训练和类边界超立方体拆分三部分组成。由于分辨率 $2i$ 下的训练样本的生成可以由分辨率 i 下类边界区域的超立方体拆分同步获得，所以每个分辨率下面的计算量主要考虑训练分类器和拆分类边界区域的样本两部分。设用于分类器训练的样本总数为 N，多分辨率分类器分类阶段的计算复杂性可以表示为公式（4-23）所示。

$$F(N) = T(N) + C(N) \qquad (4\text{-}23)$$

其中 $T(N)$表示的是训练阶段的计算总量，$C(N)$表示的是拆分阶段的计算总量。

表 4.1 表示了训练阶段和拆分阶段的计算量，其中拆分阶段的计算量同给定的样本数目是线性关系。

表4.1 各阶段的计算量

分辨率	训练阶段的计算量	拆分阶段的计算量
l	$u(\lvert X_l \rvert)$	$O(\lvert \cup B_{X,l} \rvert)$
$2l$	$u(\lvert X'_{2l} \rvert)$	$O(\lvert \cup B_{X,2l} \rvert)$
$2^2 l$	$u(\lvert X'_{2^2 l} \rvert)$	$O(\lvert \cup B_{X,2^2 l} \rvert)$
・・・	・・・	・・・
e	$u(\lvert X'_e \rvert)$	None

假设对$\forall i,l \le i \le e$并且$i = 2^k l$，均有$\left|B_{X,i}\right| \le b\left|X_i^{'}\right|$，即每一层次只有 b 比率的超立方体落入到类边界区域。再设给定的样本总数为 N，初始分辨率 l 下的样本数目为l^d，假设分辨率 i 下每个超立方格内的平均样本数目为$v_i = \left(N / i^d\right)$。$T(N)$的表达式如公式（4-24）所示。

$$T(N) = u(l^d) + u(l^d 2^d b) + u(l^d (2^d b)^2) + ... + u(l^d (2^d b)^k) \qquad (4\text{-}24)$$

$C(N)$的表达式为：

$$\begin{aligned} C(N) &= O\left(bl^d v_l + bl^d 2^d b v_{2l} + bl^d (2^d b)^2 v_{2^2 l} + ... + bl^d (2^d b)^{k-1} v_{2^{k-1} l}\right) \\ &= O\left(bl^d \frac{N}{l^d} + bl^d 2^d b \frac{N}{(2l)^d} + ... + bl^d (2^d b)^{k-1} \frac{N}{(2^{k-1} l)^d}\right) \\ &= O\left(bN + b^2 N + ... + b^k N\right) \\ &= O\left(\frac{1-b^k}{1-b} bN\right) \end{aligned} \qquad (4\text{-}25)$$

下面认为拆分阶段的计算量与训练阶段相比可以忽略不计，仅考虑训练部分的计算量，即 $F(N) \approx T(N)$。

定理 4.3 给出了一个可以降低计算复杂性的条件。

定理 4.3　若$u(x) = cx^r$，那么在以下几种情况下，$F(N) < u(N)$：

1. $d < \min(\frac{1}{r}\log_2 \frac{1-b^r}{b^r}, -\log_2 b)$；
2. $b^r < \frac{1}{2}$并且$d = -\log_2 b$；

证明：给定的训练样本数目为 N，假设初始训练样本数目 n 与 N 的关系也满足 $n=bN$。那么在初始样本下每个超立方体中包含的样本的平均数目为 N/n。设当样本格中的平均样本数降为 1 时算法结束，那么结束需要的层次数 k+1 满足条件$(N/n)\cdot(1/(2^d)^k) = 1$，求解得到$k = \lceil 1/d \log_2 (N/n) \rceil = \lceil 1/d \log_2 (1/b) \rceil$。

因此 $F(N)$可以 表示为：

$$\begin{aligned} F(N) &= u(n) + u(bn2^d) + \cdots + u(b^k n(2^d)^k) \\ &= cn^r (1 + b\cdot 2^d + (b\cdot 2^d)^{2r} + \cdots + (b\cdot 2^d)^{kr} \end{aligned} \qquad (4\text{-}26)$$

由于当$b\cdot 2^d > 1$时，求解出的$k < 1$，那么层次数等于 1，即并没有采取多分辨率的策略，因此不考虑这种情况。下面分$b\cdot 2^d < 1$，$b\cdot 2^d = 1$两种情况进行讨

论：

1. 若$b\cdot 2^d<1$

那么$d<-\log_2 b$，$k=\left\lceil \frac{1}{d}\log_2\frac{1}{b}\right\rceil>1$。那么$F(N)$的计算公式为：

$$\begin{aligned}F(N)&=u(n)+u(bn2^d)+\cdots+u(b^k n(2^d)^k)\\&=cn^r\frac{1}{1-(b\cdot 2^d)^r}\left[1-(b\cdot 2^d)^{r(\left\lceil \frac{1}{d}\log_2\frac{1}{b}\right\rceil+1)}\right]\end{aligned}\qquad (4\text{-}27)$$

希望$F(N)<u(N)$。因此

$$cn^r\frac{1}{1-(b\cdot 2^d)^r}\left[1-(b\cdot 2^d)^{r(\left\lceil \frac{1}{d}\log_2\frac{1}{b}\right\rceil+1)}\right]<cN^r\qquad (4\text{-}28)$$

即
$$\frac{b^r}{1-(b\cdot 2^d)^r}\left[1-(b\cdot 2^d)^{r(\left\lceil \frac{1}{d}\log_2\frac{1}{b}\right\rceil+1)}\right]<1\qquad (4\text{-}29)$$

由于
$$\frac{1}{1-(b\cdot 2^d)^r}\left[1-(b\cdot 2^d)^{r(\left\lceil \frac{1}{d}\log_2\frac{1}{b}\right\rceil+1)}\right]<\frac{1}{1-(b\cdot 2^d)^r}\qquad (4\text{-}30)$$

因此若下面的条件成立，就有$F(N)<u(N)$：

$$\frac{b^r}{1-(b\cdot 2^d)^r}<1\qquad (4\text{-}31)$$

求解得$d<\frac{1}{r}\log_2\frac{1-b^r}{b^r}$。又因为$b\cdot 2^d<1$，即$d<-\log_2 b$。因此当$d<\min(\frac{1}{r}\log_2\frac{1-b^r}{b^r},-\log_2 b)$时，有$F(N)<u(N)$。

2. 若$b\cdot 2^d=1$

首先可以计算得到 k=1

$$F(N)=cn^r(1+\left\lceil \frac{1}{d}\log_2\frac{1}{b}\right\rceil)=2cn^r$$

所以若 $2cn^r < cN^r$，就有 $F(N) < u(N)$。求解得 $b^r < \frac{1}{2}$。

因此当 $b^r < \frac{1}{2}$ 并且 $d = -\log_2 b$ 时，有 $F(N) < u(N)$。

证毕。

上面简单地给出了一个计算复杂性可以降低的条件。从上面的分析可以看出，当大规模数据构成的样本集合存在很多样本聚集在一起时，如果每层均能有效地缩小分类边界所在的位置，是可以降低算法的计算复杂性的，下面将通过实验进一步证实这个结论。

4.8　实验结果

下面简单的采用仿真试验来说明实验结果的有效性。生成两类样本，样本的分布情况如图 4.9 所示。选择支持向量机作为分类器进行训练和测试，其中高斯核的参数选为 0.1，正则化因子 λ 选择为 1。

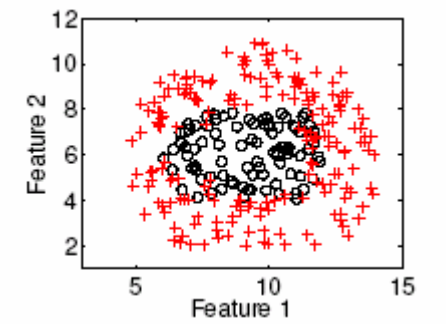
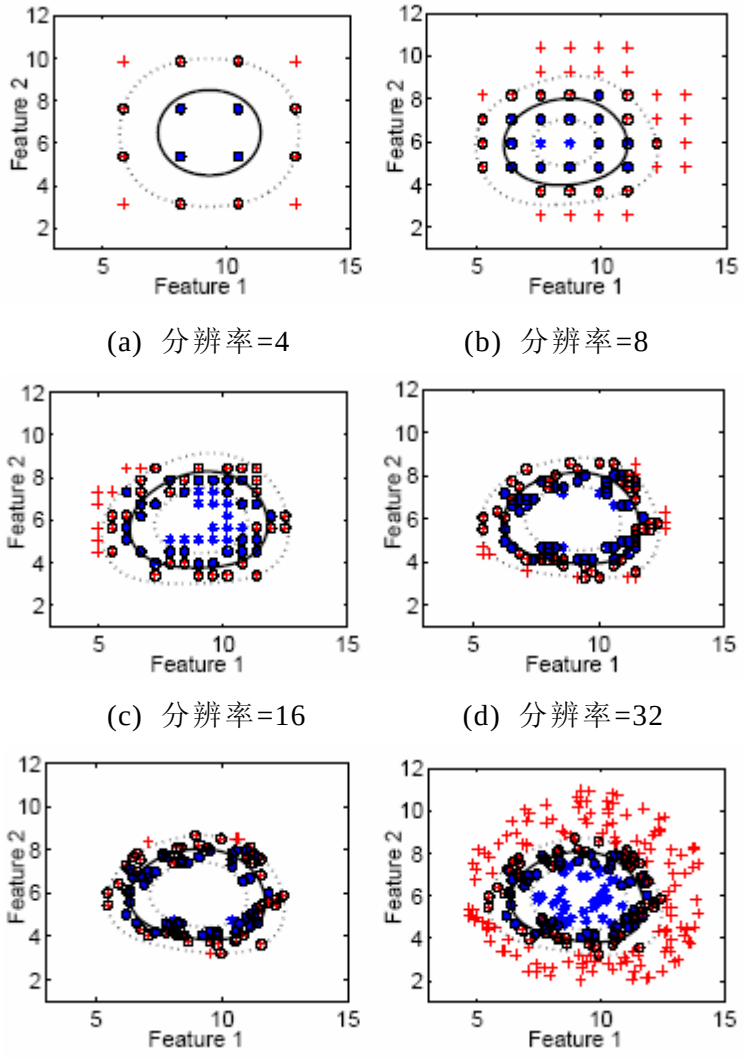


图 4.9　样本分布示意图

图 4.10 和图 4.11 说明了仿真数据的学习和分类的过程。类别 1 和类别 2 中的样本数分别为 100 和 178。设置控制初始分辨率和终止分辨率的参数的值是 $\beta_0 = 0.67$，$T_N = 1$，执行第 4.5 节的算法，计算出来的初始和终止分辨率为 $l = 4$ 和 $2^k l = 64$。采取第 4.3 节和第 4.4.节的算法训练分类器，得到 5 个分类器，表示为 $f_{2^i}(x)(i = 2,3,4,5,6)$。图 4.10(a)~图 4.10(e)是从分辨率 4 到 64 的 5 个分类器。图 4.10(f)是使用原始的训练样本所获得的分类器。图 4.11(a)是用于测试的训练样本。图 4.11(b)~图 4.11(e)是使用分类器 所获得的分类结果。在第 i 步，用“+”表示的样本表示被判断属于类别 1 的样本，用“*”表示的样本表示被判断属于类别 2 的样本，用“o”表示的样本表示通过当前的分类器还无法判断类别的样



(a) 分辨率=4　　(b) 分辨率=8

(c) 分辨率=16　　(d) 分辨率=32

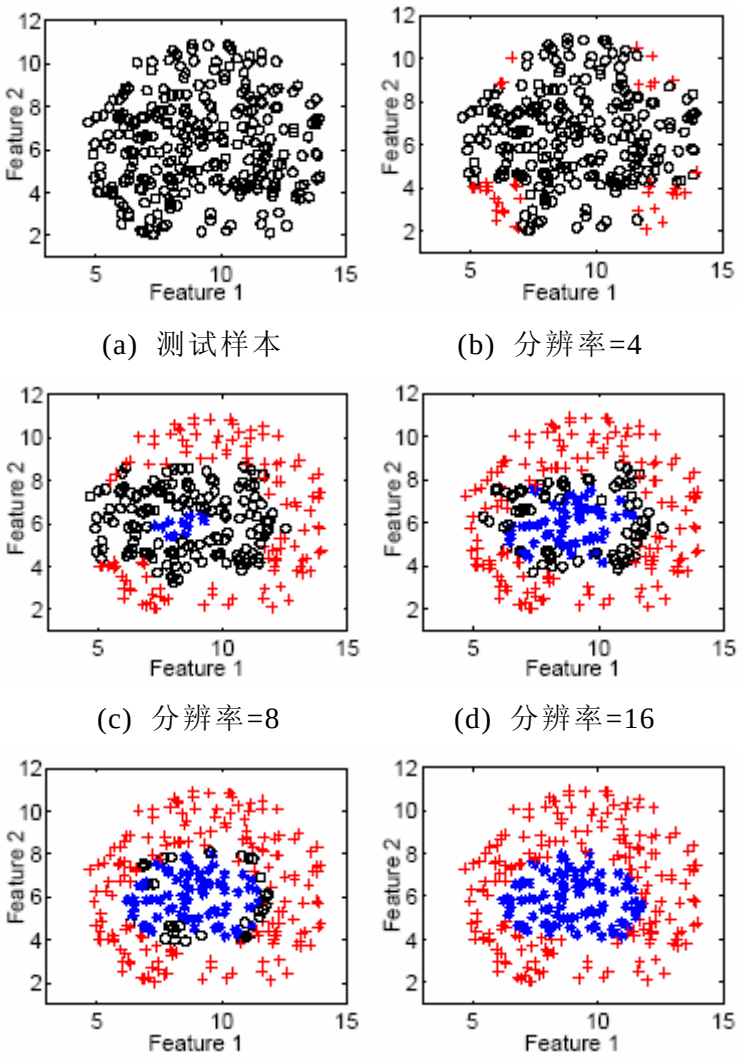
(e) 分辨率=64　　(f) 基于原始训练样本

图 4.10　训练阶段实验结果

本。

从图 4.10 和图 4.11 可以看出，通过逐步缩小分类边界，可以减少训练严格的个数并且逐步逼近分类边界。

表 4.2~表 4.4 表示的是另外一个例子，其中训练样本和测试样本的数目都比较多。满足的分布函数同例 1 的分布函数一样。用于训练类 1 和类 2 的训练样本为 10000 和 19175。其中参数 $\beta_0 = 0.67$，$T_N = 10$。通过执行训练算法，可以计算得到初始粒度和终止粒度为 $l = 4$ 和 $2^k l = 128$。训练得到 6 个分类器，表示为 $f_{2^i}(x)(i = 2,3,4,5,6,7)$，由这些分类器所获的粗分辨率下的分类器为 $F_{2^i}(x)(i = 2,3,4,5,6,7)$。表 4.2 表示的是各个分类器的参数。表 4.2 的最后一行是



(a) 测试样本　(b) 分辨率=4

(c) 分辨率=8　(d) 分辨率=16

(e) 分辨率=32　(f) 分辨率=64

图 4.11　测试阶段实验结果

基于原始样本的分类器的参数。给定测试样本集合 T 的规模是 28376，表 4.2 中最后一列的分类器的错误率是基于测试样本集合计算的。表 4.3 表示的是对样本集合 T 进行测试的过程。基于第 4.4 节所提出的算法，使用所训练出来的分类器 $F_{2^i}(x)(i=2,3,4,5,6,7)$ 来判断测试样本的类别。表 4.3 表示了每一阶段所判断出来的样本数目占总样本数目的比率。从表 4.3 的实验结果可以看出，很大比例的测试样本都可以在较低的分辨率下判断出来类别。

表 4.4 给出了采用传统的 SVM 算法和采用本章提出的算法的计算时间对比。结果表明，对于给出的数据集合来说，无论是训练阶段和测试阶段均可以大大

表 4.2 分类器参数

分类器	分辨率	平均值	样本个数	SVs 个数	错误率
$f_4(x)$	4	1822	4+12	4+8	18.63
$f_8(x)$	8	470	14+34	12+14	14.67
$f_{16}(x)$	16	124	51+53	31+34	11.87
$f_{32}(x)$	32	36	148+112	87+87	11.72
$f_{64}(x)$	64	10	312+384	275+276	11.32
$f_{128}(x)$	128	3	864+1169	844+849	11.98
$F(x)$					12.76
$F(x)$			10000+19175	3722+3721	11.31

表 4.3 分类结果

分类器	$f_4(x)$	$F_8(x)$	$f_{16}(x)$	$f_{32}(x)$	$f_{64}(x)$	$f_{128}(x)$
分类样本比率	15.76	34.38	1.07	0.44	1.19	47.16

表 4.4 计算时间对比

	$f(x)$	F(x)
训练阶段	202(s)	51(s)
测试阶段	136(s)	16(s)

降低计算复杂性。

4.9 本章小结

本章提出了一种多分辨率（多粒度）的分类器求解策略。首先将训练样本所在的特征空间在粗分辨率下划分成为相等的超长方体，然后以这些超长方体为基础求解出来分类边界，将分类边界所位于的超长方体继续划分，求解出来更细的分类边界，重复这个过程直到求解出来最细分辨率下的分类边界线。本章分析了这种求解策略的泛化能力以及计算复杂性，给出了计算复杂性会降低的条件。这一章的研究内容的后续工作包括以下几个方面：

1. 本章同第 2 章对样本集合的划分是不同的，第 2 章获得的粗粒度下的每个训练样本都是“纯”的，即一个区域内的样本都来自于同一类。而本章所获得超长方体不一定是纯，但是超长方体的类别采用的是明确的类别标识。进一步的工作可以考虑以概率值作为类别标识来训练分类器；

2. 进一步给出粗分辨率下类边界定义的算法。在本章中，仅对基于支持向量机的算法给出了类边界的精确定义，进一步可以考虑给出更为广泛的算法，能够计算以其它算法作为基本算法的类边界；

3. 分析工作还可以进一步深入。本章中给出的计算复杂性的分析并没有将训练样本的分布同计算复杂性的降低直接结合起来。因此进一步的工作可以考虑给出更为具体深入的分析。

第 5 章　基于商空间合成的道路检测

5.1　本章引论

道路检测是自主导航系统的一个关键模块，城区和校园环境中的自主导航系统是目前的一个研究热点。本章中的研究对象是城区和校园环境中的半结构化道路。由于在这种路面环境下，道路的边缘和路面区域均对最终确定道路边界线起到了一定的作用，因此可以将这两个方面的检测结果看作道路区域的不同的商空间。本章采用商空间合成的方法建模道路检测问题，不仅给出了较好的检测结果，而且还对“边缘+区域”这样的经验型方法给出了一个较好的理论上面的解释。

本章按照如下方式组织：第 5.2 节综述了已有的道路检测方法；第 5.3 节介绍了商空间合成的基本概念，并且给出了几种图像的商空间表示；第 5.4 节建立了基于商空间合成的道路检测的基本模型；第 5.5 节和第 5.6 节分别给出了基于边界和区域的道路求解模型；第 5.7 节描述了区域和边界的合成算法；第 5.8 节给出了实验结果与分析；最后是本章小结。

5.2　道路检测综述

在过去的二十年中，在道路检测方面已经提出了许多方法[100][101]。大部分方法针对的是高速公路或其它结构化性质较好的道路，通过检测路面上的车道线来进行侧向定位和导航[102]-[110]。如意大利 Parama 大学的 GOLD 系统[102][103]，CMU 大学的 RALPH 系统[105]和 YARF 系统[109]等等。由于这些系统的道路跟踪模块都是建立在左-中-右车道线检测的基础上，显然它们不适于那些路面没有清晰车道线的道路环境。关于非结构化道路，也存在一些研究结果[111]-[123]，CMU 大学的 SCARF 和 UNSCARF 针对公园中的小路使用彩色图像来检测道路[111][112]，路面与周围环境的色彩差异是区分道路和非道路的关键因素。该算法首先要对路面进行聚类，然后还要对多个模型进行匹配，因此时间复杂度比较大，难以满足城区和校园环境所提出的实时性要求。C.Rasmussen 使用纹理特征在乡村道路上进行导航[114][115]，但一般的道路并没有很规则的纹理特征；K.Onoguchi

和 N. Takeda 采用立体平面投影（Planar Projection Stereopsis）的方法来解决道路提取问题[116]，但是在大部分的城区环境下，路肩的高度太低了，不足以通过 PPS 算法提取出道路区域。以上这些检测非结构化道路的方法一般都针对较为特定的道路环境。文献[117][118]中的方法在图像的特定区域中搜索道路，虽然试验结果可以在城区环境下工作，但是方法基本上还是基于较清晰的道路和非道路边界线的检测。F.Paetzold 和 U. Franke 给出了检测某些城区环境中道路的一些想法[119]，但是他们只考虑了灰度图像。由于车辆和行人的影响，使得他们的算法很难提取出正确的道路边缘。

城区和校园环境中的道路通常具有不同于高速公路和校园环境的特征：部分道路没有车道线或车道线不清晰；道路区域和周围环境的分界通常是路肩，并且在某些部分分界线模糊不清；路面上的车辆和行人要多于高速公路和公园小路。为了解决城区和校园环境中的道路检测问题，可以尝试将基于灰度图像的边缘提取和基于彩色图像的路面提取结合起来。由于边缘提取模块可以提供一些道路边界信息，而区域提取模块可根据路面颜色的一致性抽取出道路区域，因此两个过程的交互和结合可给出较好的检测结果。

在已有的针对各种道路环境的道路检测方法中，有一些结合边缘和区域的研究成果，但是它们基本上都没有考虑城区环境下的道路特征[120]-[123]。SCARF 和 UNSCARF 算法虽然也有区域分割和模型匹配两个模块，但是并没有利用边界特征，只是使用预先设定好的模型和区域分割的结果进行拟合。F.W.J. Gibb 融合边缘检测、路面分割以及车道线跟踪三个模块的结果来检测路面区域，该方法的问题是：由于三个模块是独立执行的，因此耗时；由于车道线、行人车辆等的影响，边缘检测模块的结果有可能相距真实道路边缘很远，从而使得融合结果偏离于真实结果；在缺乏先验知识的前提下难以选择合适的权值去融合三个模块[120]。M.Lutzeler 提出的方法是在图像的下半部分采用边缘检测，而在图像的上半部分采用区域分割，他这种划分无法解决边缘模糊的问题[122]。

道路的边界线和道路区域可以说是从两个不同的角度去观测道路。由于这两个方面的观测都有一定的可靠性，因此可以结合两方面的信息去获得更为可靠的关于道路的知识。本章提出了一种结合边缘检测和区域提取的方法。该方法首先通过灰度图像获得道路候选边界以及用于计算路面颜色的区域，然后假设路面色彩满足高斯模型，根据计算出来的参数值提取出道路区域，最后根据提取出的道路区域去判断那一对候选道路边界置信度最高，作为最终的检测结

果。本章采用前面给出的粒度计算中的商空间模型去建模图像和道路，将灰度、二值等图像作为原始图像的商空间，边缘和区域作为道路图像的商空间。这种建模方法不仅能够可靠地检测出道路边界，而且能够利用已有的商空间合成的研究结果较好地解释本章提出的方法。

5.3　商空间基本理论

5.3.1　商空间合成

张铃等提出的商空间模型是粒度计算领域的一个重要的组成部分[27][28][30]。商空间合成是商空间理论的一个重要研究成果。根据第一章中给出的粒度计算模型，商空间合成的数学模型可以简要描述如下：假设论域 U 原始的性质为 O，

$$O = \langle U, \Lambda, \Gamma \rangle$$

再假设 O 在两个等价关系 R_i 和 R_j 所确定的粒度 i 和 j 下的表示为 $O_i = (U_i, \Lambda_i, \Gamma_i)$ 和 $O_j = (U_j, \Lambda_j, \Gamma_j)$。如果合成 O_i 和 O_j，生成一个新的粒度表示 O_k，那么 O_k 的定义如下：

$$O_k = \langle U_k, \Lambda_k, \Gamma_k \rangle$$

1) 论域 U_k 是 U_i 和 U_j 的合成。$U_k = \{a_i \cap b_j \mid a_i \in U_i, b_j \in U_j\}$；显然，$U_k$ 中的元素的个数要大于 U_i 和 U_j 中的元素的个数。
2) 属性 Λ_k：属性合成的方法有多种，张铃等给出了属性函数的合成原则以及一些合成示例[28]；
3) 结构 Γ_k 的合成：合成方法也比较多，如拓扑结构的合成，半序结构的合成等等。

参考文献[28]的第三章“合成技术”中较详细地讨论了论域的合成、结构的合成以及属性函数的合成。

商空间表示是粒度表示的一种基本形式。虽然下面关于图像以及道路的标识采用的是第一章中的粒度计算的模型，但是由于本章主要的理论基础是商空间的合成理论，因此在本章中的表示以及讨论采用“商空间”这个术语来代替“粒度”。

5.3.2　图像的商空间表示

根据本章研究的需要，下面定义原始图像以及原始图像的几种商空间表示。

1. 原始彩色图像：

原始的彩色图像为一个像素集合，如下定义：

$$O_o = \langle U_o, \Lambda_o, \Gamma_o \rangle$$

1) U_o：为一幅图像，是一个像素构成的集合；

2) Λ_o：为一个三元组，表示集合 U 中每个像素的属性性质。

$$\Lambda_o = (A_o, V_o, f_o)$$

- $A_o = \{R, G, B, X, Y\}$，其中 R，G，B 分别对应的是红，绿，蓝三色通道，X 和 Y 对应的是该像素在 X 和 Y 轴的坐标；
- $V_o = \{V_R, V_G, V_B, V_X, V_Y\} = \{[0,1],[0,1],[0,1],[1, Width],[1, Height]\}$，其中 R，G，B 三通道的采样值归一化到[0,1]之间，Width 和 Height 为采样图像的宽度和高度。
- $f_{o,i}(x)(i = R, G, B, X, Y)$ 分别为元素 x 所对应的属性值。

图像中的像素之间是满足一定的邻域关系的，因此 Γ_o 是有定义的，像素的邻域关系可以由像素的位置属性所诱导出来。

2. 灰度图像：

灰度图像可以说是彩色图像的色彩属性的商空间所构成的对应图像的新的表示，即当色彩属性{R,G,B}使用商值 I 去表示时，使用下面的商空间模型去表示灰度图像：

$$O_G = \langle U_G, \Lambda_G, \Gamma_G \rangle$$

1) U_G：是一个表示 U 所对应的灰度图像的像素构成的集合；

2) Λ_G：为一个三元组，表示集合 U_G 中每个像素的属性性质。

$$\Lambda_G = (A_G, V_G, f_G)$$

- $A_G = \{I, X, Y\}$，其中 I 对应的是光的强度，X 和 Y 对应的是该像素在 X 和 Y 轴的坐标；
- $V_B = \{V_I, V_X, V_Y\} = \{[0,1],[1, Width],[1, Height]\}$；
- $f_{G,I}$ 可以有多种定义形式，如 $f_{G,I}(x) = (f_R(x) + f_G(x) + f_B(x))/3$ 或

$f_{G,I}(x)=0.3f_R(x)+0.6f_G(x)+0.1f_B(x))$；

3) 由于结构Γ_G是由元素的位置属性所确定的，所以$\Gamma_G=\Gamma$。

3. 二值图像：

对灰度图像O_G做边缘提取操作，生成边缘图像O_E（为灰度图像），然后二值化，生成一幅二值图像，则该二值图像可以说是将灰度图像的强度属性二值化构成的原论域的商空间，即当强度属性 I 使用二值{0,1}去表示时。如下表示所生成的二值图像：

$$O_B=\langle U_B,\Lambda_B,\Gamma_B\rangle$$

1) 根据前面的叙述，可以知道强度图像对于边界线的检测U_B为一幅图像，是一个像素构成的集合；

2) Λ_B：为一个三元组，表示集合U_B中每个像素的属性性质。

$$\Lambda_B=(A_B,V_B,f_B)$$

- $A_B=\{Bi,X,Y\}$，其中 B_i 为二值属性，X 和 Y 对应的是该像素在 X 和 Y 轴的坐标；
- $V_B=\{V_{Bi},V_X,V_Y\}=\{\{0,1\},[1,Width],[1,Height]\}$；
- $f_{B,Bi}(x)=0$ 或 1，$f_{B,X}(x)=f_X(x)$、$f_{B,Y}(x)=f_Y(x)$分别为元素 x 所对应的属性值。

3) 由二值属性和位置属性给出论域U_B中的一个结构，如下定义：

$$s(x,y)=\begin{cases}1, & \text{如果}x\text{和}y\text{的关系满足}C_1\\ 0, & \text{否则}\end{cases} \qquad (5\text{-}1)$$

条件C_1为：

$$\begin{aligned}&(f_{B,Bi}(x)=1\ \ and\ \ f_{B,Bi}(y)=1)\\ and\ \ &(|f_{B,X}(x)-f_{B,X}(y)|\le 1)\ \ and\ \ (|f_{B,Y}(x)-f_{B,Y}(y)|\le 1)\end{aligned} \qquad (5\text{-}2)$$

上面的结构定义实际上是当一个像素在另一个像素的八邻域内，那么这两个像素之间存在结构关系。

4. 直方图图像：

直方图图像指的是由如下的关系 R 构成的对于二值图像的商空间表示：

$\forall x,y\in O_T,xRy$ 当且仅当 $f_{T,X}(x)=f_{T,X}(y)$ and $f_{T,Bi}(x)=1$ and $f_{T,Bi}(y)=1$

直方图图像的商空间定义如下：

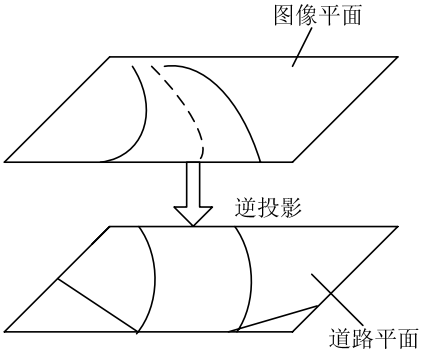


图 5.1　投影关系

$$O_P = \langle U_P, \Lambda_P, \Gamma_P \rangle$$

1) $\forall x \in U_P$，x 是一个像素构成的集合；
2) $\Lambda_P = (A_P, V_P, f_P)$
   - $A_P = \{Num, Pos\}$，属性 Num 指的是像素个数，Pos 指的是像素所位于的列的位置；
   - $V_P == \{N, [1, Width]\}$；
   - $f_{P,Num}(x) = x$ 中元素的个数，$f_{P,Pos}(x) = f_{T,X}(u)$，$\forall u \in x$。
3) 由二值属性和位置属性给出论域 U_T 中的一个结构，如下定义：

$$s(x,y) = \begin{cases} 1, & \text{如果} x \text{和} y \text{的关系满足} C_1 \\ 0, & \text{否则} \end{cases} \qquad (5\text{-}3)$$

条件 C_1 为：$|f_{P,Pos}(x) - f_{P,Pos}(y)| = 1$。

5.3.3　变形图像的商空间表示

前面讨论的是原始图像的几种商空间表示。在图像处理和计算机视觉等领域中，经常会对图像做一些变形处理。下面介绍与本章的道路检测相关的两种图像的变形表示：逆投影图像（Reprojection Image）和偏移图像（Shifting Image）。

1. 逆投影图像：

在地面上，道路边界通常都是一对平行线，因此可以将摄像机所采集的图像投影到地平面上寻找道路边界，如图 5.1 所示。假设道路是平坦的，那么道路平面（世界坐标系中假设 z=0）上的点(x,y)与图像平面上的点(u,v) 之间的投影关系可如下描述[102][105][115]：

$$\rho[x \quad y \quad 1]^{Y} = M[u \quad v \quad 1]^{T} \qquad (5\text{-}4)$$

其中 M 为摄像机标定的参数。假设给定投影矩阵 M，给定原始图像 O，给定逆投影图像的尺寸 IWidth 和 IHeight，那么如下表示所生成的逆投影图像：

$$O_N = \langle U_N, \Lambda_N, \Gamma_N \rangle$$

1) U_N 是一个像素构成的集合，U_N 中像素的个数可以不同于 U 中的像素个数， 它由尺寸 IWidth 和 IHeight 决定，即 $|U_N| = IWidth \times IHeight$；

2) Λ_N：为一个三元组，表示集合 U_N 中每个像素的属性性质。

$$\Lambda_N = (A_N, V_N, f_N)$$

- $A_N = A_o$；
- V_N：其中 $V_X = [1, Width]$，$V_Y = [1, Height]$，其他部分同 V_o 中的定义；将属性映射修改为 $f_{T,X}(x) = f_X(x') + h(x')$，其中 x' 为 x 在原始图像中的对应点，$h(x')$ 是由矩阵 M 所决定的原始图像与逆投影图像对应像素的偏移值。

3) Γ_N 可由 U_N 中像素的位置属性所诱导出来。

2. 偏移图像：

下面考虑图像的偏移图像表示。定义平移函数 $h(x,d)(x \in U)$ 表示的是像素 x 所在的行的像素的平移量，$d = -1$ 表示向左平移，$d = 1$ 表示右平移。给定一副图像 O，以及如下表示所生成的偏移图像：

$$O_T = \langle U_T, \Lambda_T, \Gamma_T \rangle$$

1) U_T：是一个像素构成的集合；

2) Λ_T：为一个三元组，表示集合 U_T 中每个像素的属性性质。

$$\Lambda_T = (A_T, V_T, f_T)$$

- $A_T = A_o$；
- $V_T = V_o$；
- 将属性映射修改为 $f_{T,X}(x) = f_X(x) + h(x,b)$。

逆投影图像和偏移图像与变换前的原始图像的类型的是一样的，如果使用彩色图像进行变换，那么变换后的图像也是彩色图像。

上面定义的这几种图象都是我们在下面讨论道路检测问题时会使用到的，可以看出，采用商空间的模型可以非常清楚描述出这几种图像之间的关系。

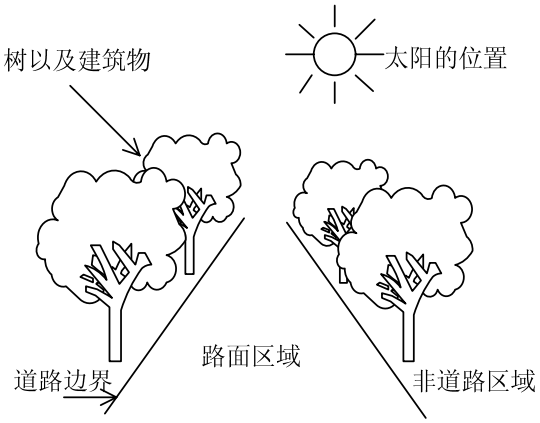


图 5.2　道路场景模型

5.4　基于商空间合成的道路模型

5.4.1　道路场景模型

为了设计出一个鲁棒的道路检测系统，下面首先分析城区环境中影响道路检测的一些因素，在图 5.2 中：

1. 道路区域：路边不是很宽，并且可能没有车道线，路面上可能有一些离摄像机较近的行人和车辆。

2. 周围环境：通常是由路肩或不同的色彩（如草坪，土地）与路面区别开来；

3. 树和建筑物：一般在道路的两侧，由于光照可在路面上生成阴影，但是路和建筑物同样也可作为区分道路和边界的证据；

4. 太阳：决定着入射光线的强弱以及入射光谱；

5. 智能车：可以提供一些速度和位置信息，从而缩小道路的检测范围。

上述第 1 到第 4 个因素与所采集到的图像的色彩有关。接下来考虑现实场景中的一点与图像中的对应点的光测定关系。假设图像中位于位置(u,v)的一点同三维场景中位于位置(x,y,z)的一点相对应，那么由一个 CCD 摄像机所采集的彩色图像包括 R，G，B 三个传感器的输出值。那么第 i 个传感器的输出能量可以如下表示$i \in \{R,G,B\}$[124]：

$$\begin{aligned} E_i(u,v) &= \int_{\lambda} L(\lambda,x,y,z)\rho(\lambda,x,y,z)c(\varphi,\theta,\gamma,x,y,z)S_i(\lambda)d\lambda \\ &= c(\varphi,\theta,\gamma,x,y,z)C_i(x,y,z),\ \ i \in \{R,G,B\} \end{aligned} \tag{5-5}$$

其中 $c(\varphi,\theta,\gamma,x,y,z)$ 依赖于反射的几何属性，$C_i(x,y,z)$ 与物体表面的反射属性相关。位于位置(u,v)的象素点的总能量可表示为：

$$\begin{aligned} I(u,v) &= \frac{1}{3}(E_R(u,v)+E_G(u,v)+E_B(u,v)) \\ &= \frac{1}{3}C(\varphi,\theta,\gamma,x,y,z)(C_R+C_G+C_B)(x,y,z) \end{aligned} \qquad (5\text{-}6)$$

许多道路检测的方法都是基于灰度图像的[102]-[110][113][116]-[119]。但是公式（5-6）说明当路面上没有清晰的车道线并且交通稍微繁忙些，只使用灰度图像就很难得到满意的检测结果。这是因为阴影的边缘、某些路面上不全的车道线（经常出现在城区环境中）、车辆的边界等在灰度图像中均可以混淆道路边缘的检测，而且仅依靠灰素级别有时路面与周围环境很难区分开。因此本章采用彩色图像来提取道路，同时参考 Birchfield 所使用的互补集合理论[125]，将所采集的图像中的道路分为两个集合：边界集和内部点集，其中边界指的是道路边缘，内部点集指的是边缘内的道路区域。图像中的道路使用一个四元组来表示：

$$R=(l,r,c,u) \qquad (5\text{-}7)$$

其中 l 和 r 是左和右道路边界，c 道路的曲率，u 是路面的色彩平均值。为了得到 R 的参数，可使用的信息有：

1. 实际的道路边界：通常是指路肩或道路区域和周围环境的边界线。等式（5-6）表明尽管路面和路肩的颜色一样，但是不同的几何性质使得路肩可在灰度图的边缘图中凸现出来。因此可以使用灰度图来估计道路边缘。

2. 路面，也称作内部点：通常具有与周围环境不同的色彩。就像上面讨论的一样，使用彩色图像来得到内部区域。基于彩色的道路区域提取可以补偿道路边界部分的不足。

上面所描述的划分是下面提出的算法的基础。

5.4.2　道路的两种商空间表示

本文中主要关注的是道路的检测。给定一幅包含有道路的图像，有两种方式可将道路与非道路区分开来，一种方式是通过寻找区分道路与非道路的边界线特征（例如路肩等）来区分，另一种方式是通过路面与周围的区域的色彩的不同来区分。本章将这两种方式看作是从不同的角度去观察道路。

1. 基于边缘的道路表示

假设确定的道路的左右边缘线为 l，r，由左右边缘线所确定的等价关系 R 将图像分为两个部分：道路和非道路。如下表示：

$$O_D = \langle U_D, \Lambda_D, \Gamma_D \rangle$$

1) $U_D = \{Road, Non\_road\}$，$Road$ 为道路像素点所构成的集合，Non_road 为非道路像素点所构成的集合；

2) Λ_D 为道路部分和非道路部分的属性，如下定义：

$$\Lambda_D = (A_D, V_D, f_D)$$

- $A_D = \{Num, Cur, CAvg, CVar, Left, Right\}$；
- $V_D == \{N, R, R^3, R^3R, R\}$；
- $f_{D,Num}(x) = x$ 中元素的个数，若 x 为道路，则 $f_{D,Cur}(x)$ 表示道路的曲率。

$$f_{D,CAvg}(x) = \frac{1}{f_{D,Num}(x) - 1} \sum_{x_i \in x} f_{o,c}(x_i)$$

$$f_{D,CVar}(x) = \frac{1}{f_{D,Num}(x) - 2} \sum_{x_i \in x} (f_{o,c}(x_i) - f_{D,CAvg}(x))(f_{o,c}(x_i) - f_{D,CAvg}(x))^T$$

其中 $f_{o,c}(x_i) = (f_{o,R}(x_i), f_{o,G}(x_i), f_{o,B}(x_i))$。

$f_{D,Left}(x) = l$，$f_{D,Right}(x) = r$。

3) 由于不考虑道路像素点与非道路像素点之间的关系，因此结构 Γ_G 为空。

2. 基于路面区域的道路表示

假设提取出来的道路区域为 Ω，由区域 Ω 所确定的等价关系 R 将图像分为两个部分：道路和非道路。如下表示：

$$O_A = \langle U_A, \Lambda_A, \Gamma_A \rangle$$

1) $U_A = \{Road, Non\_road\}$，$Road$ 为道路像素点所构成的集合，Non_road 为非道路像素点所构成的集合；

2) Λ_A 为道路部分和非道路部分的属性，如下定义：

$$\Lambda_A = (A_A, V_A, f_A)$$

- $A_A = \{Num, CAvg, CVar\}$；
- $V_A == \{N, R^3, R^3\}$；
- $V_A == \{N, R^3, R^3\}$ 中元素的个数，

$$f_{A,CAvg}(x)=\frac{1}{f_{A,Num}(x)-1}\sum_{x_i\in x}f_{o,c}(x_i)$$

$$f_{A,CVar}(x)=\frac{1}{f_{A,Num}(x)-2}\sum_{x_i\in x}(f_{o,c}(x_i)-f_{A,CAvg}(x))(f_{o,c}(x_i)-f_{A,CAvg}(x))^T$$

其中 $f_{o,c}(x_i)=(f_{o,R}(x_i),f_{o,G}(x_i),f_{o,B}(x_i))$。

3)　$\Gamma_A=\phi$。

5.5　基于边界的道路求解

基于边界的道路求解有两个任务：(1)确定道路曲率；(2)在图像平面中求出计算道路色彩的区域；(3)给出道路边界的一些估计值。对于公式（5-7）中所给出的道路模型来说，要确定 c，估计 l 和 r 的值。

假设给定的用于检测道路的原始图像为 $I_O(x,y)$，先采用第 5.4.2 节给出的表示所获得的原始图像的逆投影图像为 $I_G(u,v)$。然后使用基本的形态学的腐蚀和膨胀算子提取图像 $I_G(u,v)$ 的边缘图像，为了减少噪声的影响以及简化计算，给定一个阈值 T_B，获得 $I_G(u,v)$ 图像的二值图像 $I_B(u,v)$。基于边界的道路求解模块都是基于二值图像 $I_B(u,v)$ 进行的。

假设弯路的几何形状可表示为圆弧的一部分，则在左(右)道路边界上的点之间的关系可表示为：

$$(x-O_x)^2+(y-O_y)^2=R^2 \qquad (5\text{-}8)$$

其中 O_x 和 O_y 是圆弧的圆心，R 是半径。

5.5.1　曲率估计

关于道路曲率的估计已经存在一些研究成果[105][109][114][120]。K.Klage 和 C.Thorpe 提出的方法基于已经定位出的可靠的车道线来精确地估计道路的曲率[109]。D.Pomerleau and T. Jochem 提前给出了五种曲率的投影图像，然后使用这五种曲率对图像进行“纠正”，找出“纠正”后最直的图像所对应的曲率即认为是道路的曲率[105]。上面所提出的方法针对的都是高速公路，在高速公路上，有非常清晰的左、中、右车道线，并且这些车道线一般都能够被鲁棒地提取出来，因此道路曲率的估计都是以这些鲁棒提取的车道线为基础的。但是在城区

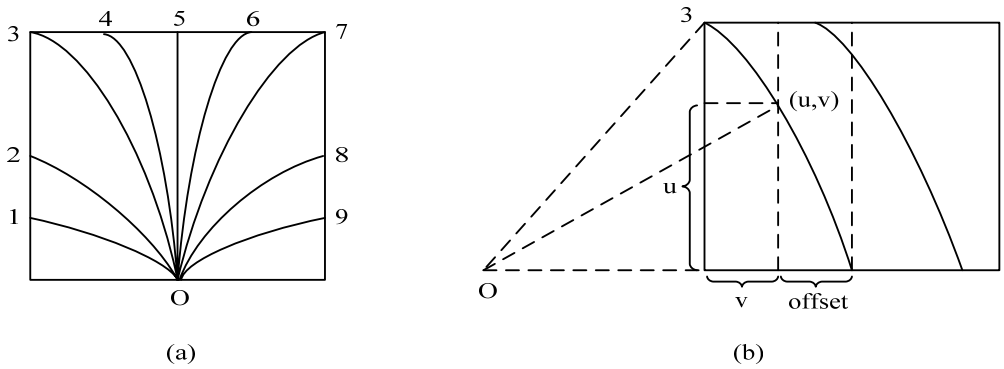


图 5.3　曲率模型

(a)9 种道路曲率；(b)曲率 3($v = g_X(x)$，$u = g_Y(x)$)。

环境中，路面状况要复杂一些，因此道路曲率的估计相对也要困难一些。D.Pomerleau 所提出的方法只能够粗略地估计出道路的曲率[105]，下面给出我们的道路曲率估计方法。

首先针对在城区环境中路面上很少存在清晰的车道线、道路曲率的变化更为复杂、道路上的行人和车在投影图像上的边缘图上能够混淆真正的道路边界等问题，我们给出九种道路曲率，如图 5.3(a)所示。结合道路形状圆弧的假设，对于每种曲率均可算出投影图像的偏移值。将这些偏移值作用于二值图像，即可构成 9 个偏移图像。对于 9 种曲率，偏移图像的偏移函数可如下计算(图像的宽度和高度均设为 W)：

1. $h(x,1) = \sqrt{f_X(x)^2 + f_Y(x)^2 + 3Wf_X(x)/4} - f_Y(x)$
2. $h(x,1) = \sqrt{f_X(x)^2 + f_Y(x)^2} - f_Y(x)$
3. $h(x,1) = \sqrt{f_X(x)^2 + (f_Y(x) + 3W/4)^2} - (f_Y(x) + 3W/4)$
4. $h(x,1) = \sqrt{f_X(x)^2 + (f_Y(x) + 13W/8)^2} - (f_Y(x) + 13W/8)$
5. $h(x,1) = 0$
6. $h(x,-1) = \sqrt{f_X(x)^2 + f_Y(x)^2 - 21Wf_Y(x)/4 + 441W^2/64} - (21W/8 - f_Y(x))$
7. $h(x,-1) = \sqrt{f_X(x)^2 + f_Y(x)^2 - 7Wf_Y(x)/2 + 49W^2/16} - (7W/4 - f_Y(x))$
8. $h(x,-1) = \sqrt{f_X(x)^2 + (W - f_Y(x))^2} - (W - f_Y(x))$

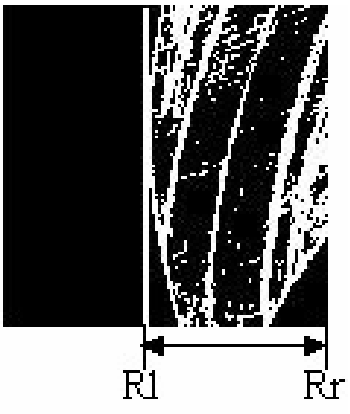


图 5.4　有意义计算区域

算法 5.1 计算道路曲率

CurvatureDetermining

1　对于 $\forall x \in U_{Pi}$，求出 $f_{Pi,Num}(x)$ 的最大值，记为 $M(i)(1 \le i \le 9)$；

2　求出 $M(i)(1 \le i \le 9)$ 对应的 3 个最大值，设对应的曲率为 $k1$，$k2$，$k3$，那么 $k1$，$k2$，$k3$ 是最可能的候选曲率；

3　设定逆投影图像的有意义区域，这一步的目的是在确定曲率时，去掉直方图图像 I_{Pi} 中的无意义部分。图 5.4 给出了有意义区域的基本含义，图左边的黑色区域在原图上并不存在对应点，因此在以后的计算中将这部分区域排除出去，称 left 和 right 之间的区域为有意义区域。针对上步确定出的三个最可能候选 $k1$，$k2$，$k3$，对每个曲率，计算相应直方图图像 I_{Pi} 有意义区域的左右边界，记为 R_{li} 和 R_{ri}；

4　设

$b_i = Min(f_{Pi,Num}(x))$，$R_{li} \le f_{Pi,Pos}(x) \le R_{ri}$，$i = 1,2,3$，

取 $b = Min(b_1, b_2, b_3)$；

5　道路的曲率可由以下方式定义：

$$curvature = \begin{cases} k1 & if \quad b = b_1 \\ k2 & if \quad b = b_2 \\ k3 & if \quad b = b_3 \end{cases}$$

6　结束。

9.　$h(x,-1)=\sqrt{f_X(x)^2+f_Y(x)^2+3Wf_X(x)/4-2Wf_Y(x)+W^2}-(W-f_Y(x))$

接着用设定的九种曲率分别对图像进行“纠正”。图 5.3(b)中给出了一个采用曲率 3 对一幅图像进行纠正的一个例子，其中 *offset* 即为点(*u*,*v*)所要偏移的量。

将图像 I_B 分别用 9 种曲率拉直之后，得到 9 幅偏移图像，记做 $I_{T1},\ldots,\ I_{T9}$，最后确定这 9 幅图像中哪幅图像使得边缘最与图像底边垂直，从而估计出道路的大概曲率。图像中具有一些与道路边缘平行的特征，如道路边界、车辆边缘、某些车道线等。如果第 i 种曲率与道路曲率接近，那么在图像 I_{Ti} 中这些与道路平行的特征将更垂直于图像底部。求图像 $I_{T1},\ldots,\ I_{T9}$ 的 9 幅直方图图像，记为 $I_{P1},\ldots,\ I_{P9}$。将 I_{Pi} 采用商空间的表示为：$I_{Pi}=\langle U_{Pi},\Lambda_{Pi},\Gamma_{Pi}\rangle$。算法 5.1 根据 9 幅直方图图像确定道路的曲率。

5.5.2　选取候选边界

选取候选边界的任务并非精确地估计出道路边界，而是给出一些左右边界的候选位置，并且选取一块区域用来计算路面的颜色。假设道路的曲率是第 i 个曲率，那么就从直方图图像 I_{Pi} 来计算道路的左右边界各三个候选位置，分别为：

$$bl_1=f_{pi,pos}(x_1),\quad bl_2=f_{pi,pos}(x_2),\quad bl_3=f_{pi,pos}(x_3)$$

和

$$br_1=f_{pi,pos}(y_1),\quad br_2=f_{pi,pos}(y_2),\quad br_3=f_{pi,pos}(y_3)$$

图 5.5 的(a)和(b)显示了用于寻找左右边界候选值的区域。

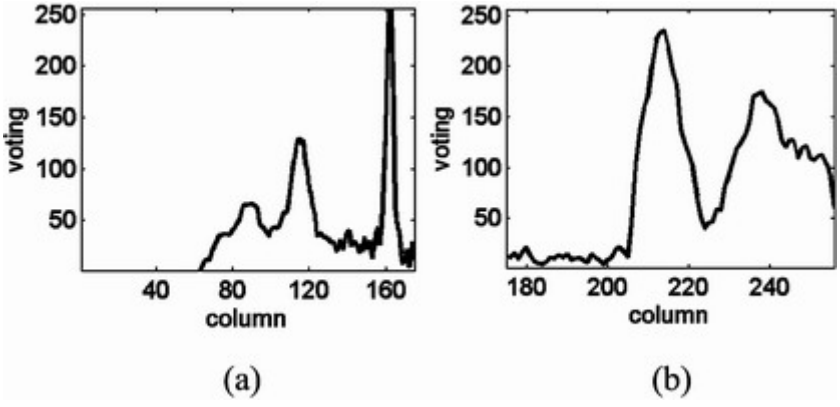


图 5.5　由于寻找候选边界的左右区域

可通过在直方图图像 I_{Pi} 中寻找属性 $f_{Pi,Num}(x)$ 最大的几个值来定位左右候选

边界，但是定位的候选边界必须满足以下关系：

$$\begin{aligned} &|bl_j - bl_k| \geq T_W \\ &|br_j - br_k| \geq T_W ,\quad j,k \in \{1,2,3\} \end{aligned} \tag{5-9}$$

其中T_W是保证选出的三个候选边界之间有一定距离的阈值。

下面将使用左边界的估计值和右边界的估计值来获取计算道路色彩的区域。就上面给出的候选边界来说，对于左边界选取偏右的一条边界，对于右边界则选取偏左的一条边界，同时为了保证所选取的区域不至于太小，选取的左右边界 *left* 和 *right* 应满足下面的条件：

$$|left - right| \geq T_b \tag{5-10}$$

其中T_b 是一个保证选取的左右边界有一定距离的阈值。

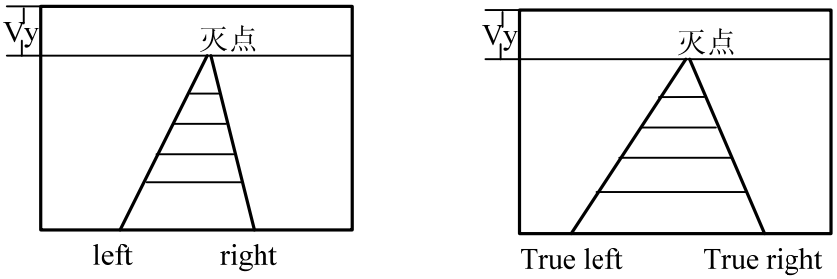
左、右各三个候选边界对应九种道路的粒度表示，记为$O_{D,i,j} = \langle U_{D,i,j}, \Lambda_{D,i,j}, \Gamma_{D,i,j} \rangle (1 \leq i \leq 3, 1 \leq j \leq 3)$，其中 i 对应左边界的候选序号，j 对应右边界的候选序号，*left* 和 *right* 左右边界组合对应的道路的粒度表示为$O_{D,left,right} = \langle U_{D,left,right}, \Lambda_{D,left,right}, \Gamma_{D,left,right} \rangle$。

5.6　基于区域的道路求解

下面工作基于的假设是路面的色彩基本上是均匀的。由于高于地面的车辆和行人在逆投影图像$I_N(u,v)$中会被拉长，所以我们使用原始图像$I_o(x,y)$来提取道路区域。尽管在路面上会存在一些车辆和行人，但是在城区和校园道路环境中路面会在图像中占据较大的比例。通常使用第 5.5 节的算法所获取的计算道路色彩的区域边界要窄于真实的道路边界，如图 5.6 的(a)和(b)所示。首先我们利用图 5.6(a)的中的阴影区域来计算路面色彩，再提取出和路面色彩一致的区域。然后依次将第 5.5.2 节中给出的候选边界和路面区域进行匹配，从而确定可靠的道路边界。

假设第 n 帧的路面色彩满足均值为μ_n，方差为Σ_n的高斯分布，如下表示：

$$P(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_n|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_n)^T \Sigma_n^{-1} (x-\mu_n)} \tag{5-11}$$



(a)　　(b)

图 5.6　真实道路边界与用于计算道路色彩的边界的关系

若点 $x \in U_o$ 属于道路区域，那么该点的色彩与 μ_n 相似。假设点 x 属于道路区域的概率为 $P(X)$，其中 $X = (f_{o,R}(x), f_{o,G}(x), f_{o,B}(x))$。如果 $P(X) \le T$（其中 T 为阈值），那么该点属于道路区域的概率较低，被认为是属于非道路区域，否则认为点 x 属于道路区域。为了计算简便，假设 R，G，B 三通道独立。因此，可以使用 3σ 法则来设定阈值，

$$f_{D,CAvg}(x) = \frac{1}{f_{D,Num}(x) - 1} \sum_{x_i \in x} f_{o,c}(x_i)$$

$$f_{D,CVar}(x) = \frac{1}{f_{D,Num}(x) - 2} \sum_{x_i \in x} (f_{o,c}(x_i) - f_{o,CAvg}(x))(f_{o,c}(x_i) - f_{o,CAvg}(x))^T$$

$O_{D,left,right} = \langle U_{D,left,right}, \Lambda_{D,left,right}, \Gamma_{D,left,right} \rangle$ 中的属性 $f_{D,left,right,CAvg}(Road)$ 和 $f_{D,left,right,CVar}(Road)$ 表示的是高斯模型的均值和方差。所提取出来的道路区域表示为：$O_{A,0} = \langle U_{A,0}, \Lambda_{A,0}, \Gamma_{A,0} \rangle$。

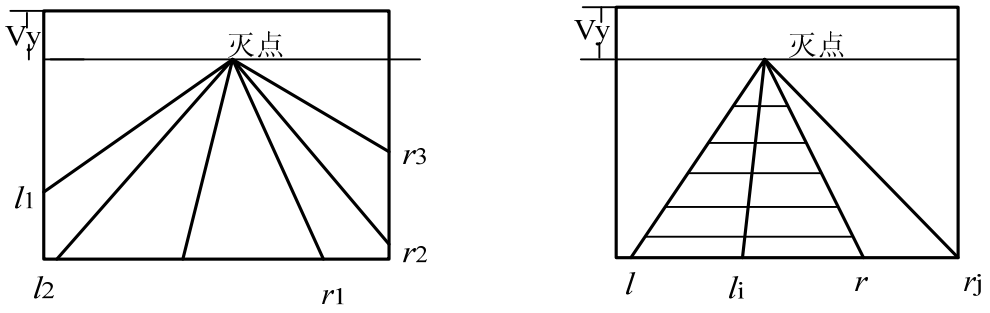
5.7　区域和边界结果合成

在这一节讨论如何根据边界的检测结果和道路区域的提取结果，以商空间合成的理论为指导，得到可靠的道路边界以及路面区域。设边界模块所得到的道路的商空间的表示为：

$$O_{D,i,j} = \langle U_{D,i,j}, \Lambda_{D,i,j}, \Gamma_{D,i,j} \rangle \qquad (1 \le i \le 3, 1 \le j \le 3)$$

区域模块所得到的道路的商空间的表示为：：

$$O_{A,0} = \langle U_{A,0}, \Lambda_{A,0}, \Gamma_{A,0} \rangle$$



(a)　　(b)

图 5.7　边界确定

为了能够融合这两方面的求解结果，从而获得一个更为鲁棒地求解结果，需要合成$O_{D,i,j}=\langle U_{D,i,j},\Lambda_{D,i,j},\Gamma_{D,i,j}\rangle$)和$O_{A,0}=\langle U_{A,0},\Lambda_{A,0},\Gamma_{A,0}\rangle$为一个新的道路提取结果。设合成结果为：

$$O_{i,j,0}=\langle U_{i,j,0},\Lambda_{i,j,0},\Gamma_{i,j,0}\rangle)$$

1)　$U_{D,i,j}=\{Rd_D,NRd_D\}$，　$U_{A,0}=\{Rd_A,NRd_A\}$。

$U_{i,j,0}=\{a_m\bigcap b_n\mid a_m\in U_{D,i,j},b_n\in U_{A,0}\}$，共有 4 类：

2)　$\{Rd_D\bigcap Rd_A,Rd_D\bigcap NRd_A,NRd_D\bigcap Rd_A,NRd_D\bigcap NRd_A\}$

$\Lambda_{i,j,0}$为道路部分和非道路部分的属性，如下定义：

$$\Lambda_{i,j,0}=(A_{i,j,0},V_{i,j,0},f_{i,j,0})$$

- $A_{i,j,0}=\{Num\}$；
- $V_{i,j,0}==\{N\}$；
- $f_{i,j,0,Num}(x)=x$ 中元素的个数；

3)　结构$\Gamma_{i,j,0}$不定义。

图 5.7 给出了两个模块进行合成的示意图。图 5.7(a)表示了左、右边界的各三个候选值，图 5.7(b)表示由 l 和 r 所确定的道路区域的提取结果与候选边界 l_{i} 和 r_{j} 进行合成的示意图。合成结果$O_{i,j,0}$与真实的道路区域的匹配结果的置信度是由以下三个因素所确定的：

首先定义第一个比率，这个比率表示的是边界和区域粒度均被认为是道路的图像部分占仅边界粒度被认为是道路的图像部分的比例：

$$ratio_{i,j} = \frac{f_{i,j,0,Num}(Rd_D \cap Rd_A)}{f_{D,i,j,Num}(Rd_D)} \tag{5-12}$$

其次定义另外一个比率为区域粒度被认为是道路但边缘粒度被认为不是的部分占整个区域提取出来的道路的比例：

$$ratio\_non_{i,j} = \frac{f_{i,j,0,Num}(NRd_D \cap Rd_A)}{f_{A,0,Num}(Rd_A)} \tag{5-13}$$

另外候选边界在直方图图像中的投票值也影响最终地判断，投票值越大，则认为该位置的边界线为真实道路边界线的概率也相对比较大。因此，定义下面的比率：

$$vote_{i,j} = \frac{f_{P,Num}(x) \times f_{P,Num}(y)}{N_\Lambda} \tag{5-14}$$

其中 $f_{P,Pos}(x) = bl_i$，$f_{P,Pos}(y) = br_j$。N_V 是一个归一化因子，它保证 $\sum_{i,j=1}^{3} vote_{i,j} = 1$。

$O_{i,j,0}$的置信度计算如下：

$$CF_{i,j,0} = \beta[\alpha \times ratio_{i,j} + (1-\alpha) ratio\_non_{i,j}] + (1-\beta) vote_{i,j}$$
$$1 \le i \le 3,\ 1 \le j \le 3 \tag{5-15}$$

其中α和β通常分别选取为 0.7 和 0.9。

5.8　实验结果与分析

实验是使用THMR-V（清华移动智能车五代）在校园环境下进行的。以上的这些算法同时在Matlab和C++上实现，没有进行任何代码优化。计算机的配置是：CPU为PentiumIV 1.5GMHz，内存为256M。

5.8.1　离线参数计算

在所有的实验中，将车辆前方 50 米，水平方向 20 米的区域投影到 256 ×256 地平面图像上。 在运行道路检测系统之前，用户可以先离线地设置一些参数。

表 5.1 场景 1 和 2 的候选边界估计值

Sn	左候选边界（位置值）			右候选边界（位置值）		
	l_1	l_2	l_3	r_1	r_2	r_3
#1	58	81	0	0	106	120
#2	0	64	75	90	102	107

表 5.2 九种边界组合的置信值

Sn	$P(S=\Lambda \mid L=l_i, R=r_j)$								
	l_1&r_1	l_1&r_2	L_1&r_3	l_2&r_1	l_2&r_2	l_2&r_3	l_3&r_1	l_3&r_2	L_3&r_3
#1	0	0.56	0.54	0	0.46	0.43	0	0	0
#2	0	0	0	0.55	0.63	0.60	0.46	0.53	0.50

这些参数包括：

1. VanishY：是灭点垂直方向的坐标，可由摄像机的标定结果离线地计算出来；
2. I2G 和 G2I：是两个查找表。公式 I2G(u,v)=(x,y)的意思是图像平面中的点(u,v)与道路平面中的点(x,y)对应，G2I(x,y)=(u,v)是道路平面中的点(x,y)与图像平面中的点(u,v)相对应。首先将道路平面与图像平面点之间的对应关系计算出来，可以加快实时处理时的速度；
3. Cur1~Cur9：是九个查找表。存储在 Curi(x,y) 中的值是点(x,y)在曲率 i 下的偏移值。

以上的这些参数值都是可以由摄像机的参数、要采集的场景图像的大小、逆投影图像的大小以及给定的道路曲率模型所提前确定的。

5.8.2　在线计算

下面给出两个场景来证明算法的有效性，一个是道路为直线道路的情形，另外一个是道路是曲线道路的情形。场景的原始图如图 5.8(a)和图 5.9(a)所示。

对于场景#1，图 5.8(a)显示的是摄像机所采集到的智能车前方的场景图像，道路检测的任务即是从这幅图像中鲁棒地提取出道路区域。图 5.8(b)显示的是图

5.8(a)的逆投影图像。

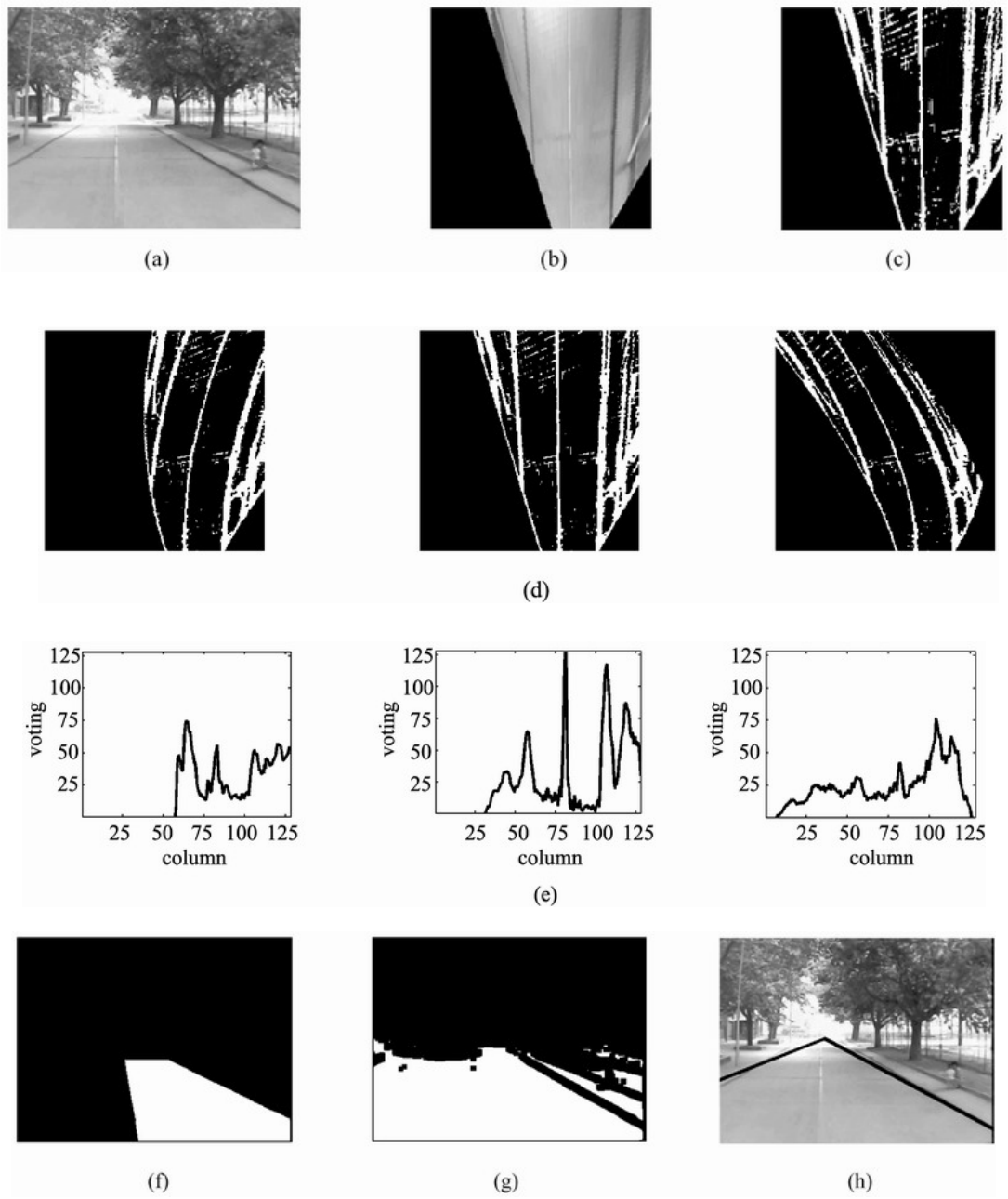


图 5.8　实验结果

(a)原始图像；(b)逆投影图像；(c)边缘图像；　(d)　三个候选曲率的边缘图(对应于曲率 4，5 和 6)；(e)　(d)中图像的投票直方图；(f)　用于计算路面色彩的区域；(g)　道路提取的结果；(h)　最终结果。

可以看出道路边界在逆投影图像中是一对平行线，寻找平行线比在原始图像中寻找道路要简单一些。

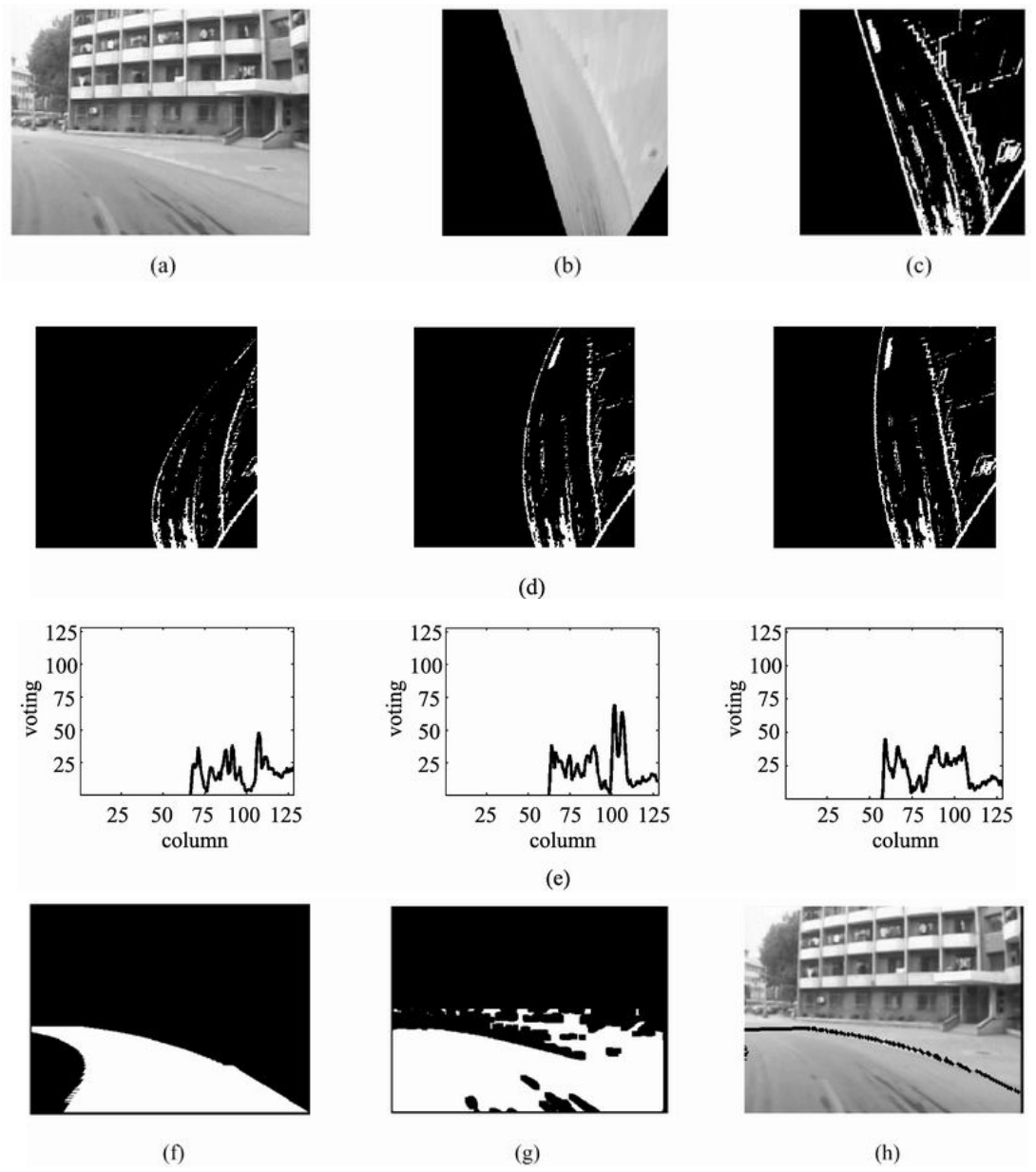


图 5.9　实验结果

(a)原始图像；(b)逆投影图像；(c)边缘图像；　(d)　三个候选曲率的边缘图(对应于曲率 2，3 和 4)；(e)　(d)中图像的投票直方图；(f)　用于计算路面色彩的区域；(g)　道路提取的结果；(h)　最终结果。

图 5.8(c)是将图 5.8(b) 通过腐蚀膨胀算子所提取的边缘图像。使用图 5.3(a)给出的九种曲率将图 5.8(c)中的图像做偏移，求出 9 个偏移图像。图 5.8(d)显示了最可能反映道路曲率的三个偏移图像。图 5.8(e)是 5.8(d)所对应的直方图。表 5.1 给出左、右边界的三个候选位置。接下来图 5.8(f)给出了用于计算道路色彩的区域，图 5.8(g)是使用公式（5-11）得到的道路区域提取结果。表 5.2 给出了不同候选左右边界与提取出来的道路区域进行合成所得到的置信度的值。图 5.8(h)是最终的边界检测结果。

图 5.9(a)~(h)是针对弯路环境的实验结果。从上述实验结果可以看出，本章所提出的算法能够得到较好的检测结果。

5.8.3　性能对比与分析

下面将本章中提出的方法与另外的三种方法进行对比，这三种方法是基于边缘的方法、基于区域的方法以及边缘和区域结合的方法。用于提取道路区域的图像的尺寸为 192 ×144。距离智能车前方长 50 米、宽 20 米的区域被投影到尺寸为 128 ×128 的图像上，花费的计算时间是基于 MATLAB 代码的。

1.　与基于边缘的方法进行对比：

图 5.8 和图 5.9 表明在城区和校园环境中，由于路面状况的结构化不是很好，因此有的地方有车道线，有的地方没有车道线。此外，边缘检测图上面有许多容易引起混淆的边缘线，这些线将影响道路边缘的定位，例如图 5.10 中的 A，B，C 和 D。还有，由于在城区和校园环境下，路面不是很宽，如果仅基于边缘图像，就可能会给出较窄的检测结果，这也是难以接受的，如图 5.8(g)和图 5.9(g)所示。因此，显然仅使用文献[102]-[110]中的方法难以给出鲁棒的道路边缘检测结果。

2.　与基于区域的方法对比：

这里将根据色彩值判断图像中的每个像素属于道路和非道路的方法均称为基于区域的方法。一些基于有监督学习的方法通常训练和测试所耗费的时间过长，并且并没有利用城区和校园环境下道路所具有的一些特征[126][127]。这里的对比仅考虑基于聚类的图像分割方法。我们实现了 UNSCARF 算法[112]，将灭点下面的像素点聚为 6 类，每个像素使用一个 5 维的特征向量，表示如下：

$$x = [red, green, blue, row, column]$$

UNSCARF 中使用的聚类算法为 ISODATA。除了 ISODATA 聚类算法外，

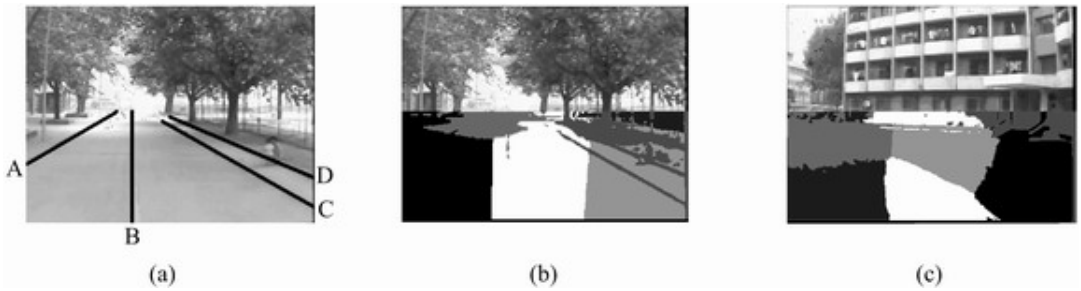


图 5.10　聚类结果

(a) 图像中可能的道路边界；(b)直线道路聚类结果；(c) 曲线道路的聚类结果。

表 5.3 计算时间对比

算法	聚类阶段	花费时间(s)
UNSCARF	K 均值聚类（6 类）	5.02
	ISODATA（6 类）	110.8
SCARF	道路区域聚类（k 均值 4 类）	1.32
	非道路区域聚类（ k 均值 4 类）	0.9210
本文算法	边界模块	0.7310
	区域模块	0.7110

我们也给出了使用 k 均值方法的计算时间。另外，为了模拟 SCARF 算法，下面的实验结果首先预测道路区域的位置，然后将道路部分和非道路部分均分别聚为 4 类，每个像素点使用如下的一个三维特征向量表示：

$$x = [red, green, blue]$$

从表 5.3 中的值可以看出，SCARF 和 UNSCARF 算法仅仅是聚类阶段计算时间就过长难以满足实时要求。此外，SCARF 和 UNSCARF 算法的一个基本假设是道路区域和周围环境可以从色彩上面区分开来。图 5.10 的(b)和(c)分别是图 5.8(a)和图 5.9(a)的聚类结果，这个实验结果表明实际上难以单单从颜色上面区分开道路和非道路，这是因为有时候二者是由路肩所区分的。还有，由于 SCARF

和 UNSCARF 的模型匹配阶段是将可能的所有模型与区域分割结果进行匹配，而不利用任何边缘信息，因此计算起来是非常繁琐费时的。同样，如果不使用边缘信息的话，在图像的限定区域上面进行分割的方法同样也很难获得较好的分割结果 [117]。

3. 与边缘和区域结合的方法进行对比：

大部分的“边缘+区域”的方法是由独立的边缘检测模块和图像分割模块所组成[120]-[123]，因此算法的计算量通常是比较大的，而且当边缘模块所给出的边缘结果距离正确边缘结果较远时，两个模块的融合会得到不可靠的检测结果，例如，使用文献[120][121]的算法，图 5.10(a)中位于 C 和 D 之间的部分可能会被认为属于道路区域。这是由于基于区域的方法认为这部分属于道路，融合后的结果也会认为这部分属于道路的可能性比较大。而实际上 C 和 D 之间的这部分区域是由清晰的道路路肩所隔开的，而且路肩能够很清晰地反映在边缘图像上。本章提出的方法认为最终的道路与非道路的分割线位于给定的候选边界上，并且当两组候选边界的置信度相似时，选择距离中心点较近的候选边界能够较好地去除那些由路肩所确定的非道路区域。另外其他的边界和区域的组合方法，在选择边界结果时，只考虑投票值最大的那些边界，因此有可能选择出图 5.10(a)中的 A 和 B 组合，这显然不是可靠的可行驶的道路区域，本章提出的方法则是通过限定左右候选边界的范围来克服这个问题。

关于道路检测问题，实际上无法设计出来一个算法能够适应各种环境，因此往往是针对特定的环境去设计特定的算法。基于边缘的方法适于高速公路和结构化性较好的道路。基于图像分割的方法适于公园小路，其中往往路面区域和环境能以色彩所区分。本章提出的方法适于城区和校园环境，在这种环境中道路是半结构化的，道路边缘和路面区域对最终确定道路的可行驶区域都会起到一定的影响。

5.9　本章小结

本章提出了一种基于商空间合成的道路检测算法。该算法针对校园和城区环境中道路车道线不清晰，道路边缘模糊以及行人较多等特点，采取了合成灰度图像和彩色图像的检测结果获得鲁棒道路边界线的方法。本章提出的算法首先基于灰度图像给出若干个道路边界线的候选位置以及用于计算路面色彩的区

域，然后基于彩色图像提取出路面，最后利用商空间合成的一些结果，将候选道路边界与路面区域的提取值进行匹配，找到最可靠的道路边界线以及路面区域。

本章研究的后续工作可以包括以下几个方面：

1. 由图像序列帧与帧之间的关系对道路边界以及路面色彩进行预测，这部分工作可以采用卡尔曼滤波来进行；

2. 消除路面上的阴影以及水渍的影响，这部分工作可以通过分析 HSV 色彩空间的色彩分量来进行。

第 6 章 结 论

6.1 结论

本文讨论了模式分类和视觉导航中的分层数据处理问题，粒度计算、多分辨率分析、多尺度分析的理论成果在问题的讨论中起到了非常重要的指导作用。从本文的理论分析以及实验结果可以看出，采用分层数据处理的方法的确能够更加有效地解决所面临的问题。本文的主要创新点有以下几个方面：

1. 选择区分函数作为分类器，分析了期望风险最小化、经验风险最小化、结构风险最小化以及邻域风险最小化原则在求解模式分类问题中的优缺点，讨论了给定的训练样本的个数在模式分类问题求解中的影响，指出了模式分类中可以采用分层处理策略的几个方面。分析表明，通过将训练样本集合在粗粒度上进行表示以及采取多分辨率的策略去求解分类器，可以更加合理有效地解决分类问题。

2. 提出了一种区域风险最小化的分类器求解策略。该策略将训练样本根据分布情况采用超球或者超立方体在粗粒度上进行表示，以超球或超立方体为单位进行分类器的训练。论文中还给出了采取序贯最小优化算法的求解方法。实验结果表明，这种求解策略可以减少训练样本的个数以及降低求解的分类函数的复杂性。

3. 提出了一种基于特征空间划分的多分辨率分类器求解策略。该策略首先将训练样本所在的特征空间在粗分辨率下划分成为相等的超立方体，然后以这些超立方体为基础求解出分类边界，再将分类边界所位于的超立方体继续划分，求解出来更细的分类边界，重复这个过程直到求解出的分类器的精度满足要求。本文分析了这种求解策略的泛化能力以及计算复杂性，给出了计算复杂性会降低的条件。

4. 提出了一种基于商空间合成的道路检测算法。该算法针对校园和城区环境中道路车道线不清晰，道路边缘模糊以及行人较多等特点，采取了融合灰度图像和彩色图像的检测结果获得鲁棒道路边界线的方法。算法首先基于灰度图像给出若干个道路边界线的候选位置以及用于计算路面色彩的区域，然后基于彩色图像提取出路面，最后利用商空间合成的一些结果，将候选道路边界与路

面区域的提取值进行匹配，找到最可靠的道路边界线以及路面区域。

6.2 展望

模式分类问题在过去的十年中受到了广泛的关注，关于模式分类理论方面的成果也越来越多。随着计算机应用技术的发展，计算机越来越多地要处理海量数据，因此近年来大规模数据的模式分类问题逐步成为引起研究人员关注的热点问题。研究模式分类问题中的分层数据处理有着理论价值和实际的应用意义的。

另一方面，图像处理和计算机视觉的应用目前也非常多。图像处理要处理的数据量一般是比较大的。虽然目前在图像处理中存在许多分层次处理的方法，但是许多方法都是基于经验性的并且大部分方法都是关于图像表示的，将理论上比较完善的方法应用上去的并不多，目前比较成功的应用是小波分析理论。讨论图像数据的分层数据表示以及对分层次求解有助于对图像内容的分析。

作为理论指导的粒度计算和多分辨率分析本身是正在发展中的研究方向，虽然它们在经验上解决了不少实际问题，但在理论上许多方面还有待进一步完善。粒度计算的理论目前才初步形成了一个研究方向，只是将相关的一些研究成果纳入到一个大的框架内，系统性的理论还有待进一步发展。它们的发展和完善对于模式分类和图像分析中的分层数据处理的研究会起到进一步的推进作用。

针对本文主要的研究内容，下面给出可以进一步开展的工作：

1. 本文在第 2 章所讨论的模式分类中的分层数据处理相对比较简单，缺乏理论性的分析，进一步可以完善这个工作。这个工作可以基于统计学习理论来开展；

2. 粗粒度下训练样本的表示问题。第 3 章简单地讨论了如何以超球和超立方体作为基本的表示单位、采用中心点和半径来表示粗粒度下的训练样本。当一个超球或超立方体内的样本不纯的时候，即并非所有的样本均来自于同一类时，那么可以尝试引入概率的类别标识来进行表示。第 3 章采用超球对样本进行划分时，由于正例样本和反例样本是单独聚类的，因此在正例样本和反例样本混合的区域，聚类的结果会导致训练出的分类器区分能力较差。而采用超立方体进行划分的方式容易使得超立方体内的样本分布不均匀，从而影响求解的

分类边界线，进一步可以设计更好的算法来解决这些问题；

3. 对于多分辨率分类问题来说，性能分析是非常重要的一个部分。本文的第 4 章只是初步地给出了一些性能分析的结果，进一步的工作可以讨论样本在什么分布情况下采取所选择的算法可以降低算法的计算复杂性，即将样本的分布情况同计算复杂性的降低联系起来；

4. 采取多分辨率的策略训练分类器，所获得的一组分类函数的泛化性能以及对最优分类函数的逼近程度还有待进一步分析。

本文只是在模式分类的分层数据处理方面进行了初步的探讨，随着计算机应用的快速发展，海量数据的处理必然会引起越来越多的研究人员的重视，探讨针对海量数据的模式分类算法具有较大的研究价值，研究模式分类中的分层数据的表示和处理可以对设计新的算法起到指导作用。

参考文献

[1] Smith B R. Soft computing: art and design. Addison-Wesley publishing company, 1984

[2] Zadeh L A. Soft computing and fuzzy logic. IEEE Software. 1994, 11: 48-58

[3] Aliev R A., Aliev R R. Soft computing and its application. World scientific publishers, 2001

[4] Shanahan J G. Soft computing for knowledge discovery: introducing Cartesian granule features. Kluwer academic publishers, 2000

[5] Forrtuna L, Rizzotto G, et al. Soft computing. Springer-Verlag London Limited, 2001

[6] Zadeh L A. Information granularity, fuzzy logic, and computing with words. unpublished talk at the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based systems (IPMU'96), Granada, Spain, July 1, 1996

[7] Zadeh L A. fuzzy logic=computing with words. IEEE Transactions on Fuzzy Systems. 1996, 4(1):103~111

[8] Lindeberg L. Scale-space theory: a basic tool for analyzing structures at different scales. Journal of applied statistics. 1994, 20(2): 225-270

[9] Lindeberg L. Scale-space for discrete signals. IEEE Transactions of Pattern Analysis and Machine Intelligence. 1990, 12(3): 234-254

[10] Pawlak Z. Rough Sets: Theoretical aspects of reasoning about data. Dordrecht: Kluwer Academic Publishers, 1991

[11] Polkowski L. Rough sets: mathematical foundations. New York : Physica-Verlag, 2002

[12] Dumitrescu D, Lazzerini B, Jain L C. Fuzzy sets and their application to clustering and training. Fla.: CRC Press, 2000

[13] Dubois D, Prade H. Rough fuzzy sets and fuzzy rough sets. International Journal of General Systems. 1990, 17(2):191~209

[14] Pal S K, Skowron A. Rough fuzzy hybridization: a new trend in decision-making. New York: Springer, 1999

[15] Bargiela A, Pedrycz W. Granular computing: an introduction. Boston: Kluwer Academic Publishers, 2003

[16] Nguyen H S, Skoeron A, Stepaniuk J. Granular computing: a rough set approach. Computational Intelligence. 2001, 17: 514-544

[17] Demri S P, Orlowska E S. Incomplete information: structure, inference, complexity. New York: Springer, 2002

[18] Zadeh L A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets and Systems. 1997, 90(2):111~127

[19] Zadeh L A. Some reflections on soft computing, granular computing and their roles in the conception, design and utilization of information intelligent systems. Soft Computing. 1998, 2(1):23~25

[20] Sunaga T. Theory of an interval algebra and its application to numerical analysis. Gaukutsu Bunken Fukeyu-kai, Tokyo, 1958

[21] Moore R E. Interval arithmetic and automatic error analysis in digital computing. PhD thesis, Standford University, October 1962

[22] Yao Y Y. Granular computing: basic issues and possible solutions. In: Wang P P, ed. Proceedings of the 5th Joint Conference on Information Sciences, Atlantic, NJ: Association for Intelligent Machinery, 2000, 1:186~189

[23] Yao Y Y, Li X. Comparison of rough-set and interval-srt models for uncertain reasoning. Fundamental Informatics. 1996, 27(1):289~298

[24] Yao Y Y, Wong S K M, Wang L S. A nonnumeric approach to uncertain reasoning. International Journal of General Systems. 1995, 23(2):343~359

[25] Yao Y Y, Ning Z. Granular computing using information table. In: Lin T Y, Yao Y Y, Zadeh L A, eds. Data Miming, Rough Sets and Granular Computing. Heidelberg: Physica-Verlag, 2000, 102~124

[26] Yao Y Y. Information tables with neighborhood semantics.
www2.cs.uregina.ca/~yyao/grc_paper/semantics.ps

[27] Zhang B, Zhang L. Theory and applications of problem solving. North-Holland: Elsevier Science Publishers B.V., 1992

[28] 张钹，张铃.问题求解的理论与应用. 北京：清华大学出版社，1990

[29] Polkowski L, Tsumoto S, Lin T Y. Rough set methods and applications: new developments in knowledge discovery in information systems. New York: Physica-Verlag, 2000

[30] Zhang L, Zhang B. The quotient space theory of problem solving. Proceedings of International Conference on Rough Sets, Fuzzy Set, Data Mining and Granular Computing, 2003, 11-15

[31] Garcia O N, Kreinovich V, Longpré L, Nguyen H T. Complex problems: granularity is necessary, granularity helps. Proceedings of the Vietnam-Japan International Symposium on Fuzzy Systems and Applications, September 30-October 2, Vietnam, 1998, 449-455

[32] Kreinovich V, Bouchon M B. Granularity via non-deterministic computations: what we gain and what we lose. International Journal of Intelligent Systems. 1997, 12: 469-481

[33] Hobbs J R, Vladik K. Optimal choice of granularity in commonsense estimation: why half orders of magnitude. Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference, Vacouver, British Columbia, July 2001, 1343-1348

[34] Hobbs J R. Granularity. Ninth International Joint Conference on Artificial Intelligence, August 1985, 432-435

[35] Allen T F H, Starr T B. Hierarchy: perspectives for ecological complexity. University of Chicago Press, 1982

[36] 张颖，刘艳秋. 软计算方法. 科学出版社，2002

[37] Duba R O. Pattern classification (Second Edition). John&Wiley & Sons, Inc., 2001

[38] Haykin S. Neural networks: A comprehensive foundation. Second edition, Prentice-Hall, 1999

[39] Theodoridis S, Kouthoumbas K. Pattern recognition (Second Edotion). Elsevier Science, 2003

[40] Schurmann J. Pattern classification: a unified view of statistical and neural approaches. John Wiley & Sons, Inc., 1996

[41] Anzai Y. Pattern recognition and machine learning. Academic Press, Inc., 1992

[42] Marques de Sa J P. Pattern recognition: concepts, methods and applications. Springer-Verlag Berlin Heidelberg, 2001

[43] Vapnik V N. Statistical learning theory. Second edition, Springer-Verlag New York, 2000

[44] Vapnik V N. The nature of statistical learning theory. John Wiley, New York, 1998

[45] 钱芳. 基于内容的图像检索中相关反馈方法研究: [博士学位论文]. 北京：清华大学. 2003

[46] 张磊. 基于内容的图像检索中人机协同问题的研究: [博士学位论文]. 北京：清华大学. 2001

[47] 景风. 高效准确的基于内容的图像检索研究: [博士学位论文]. 北京：清华大学. 2004

[48] 梁路宏. 人脸检测与跟踪研究: [博士学位论文]. 北京：清华大学. 2001

[49] Han J W, Kamber M. Data mining:concepts and techniques. San Francisco: Morgan Kaufmann Publishers, 2001

[50] Westhead D R, Parish J H, Twyman R M. Bioinformatics. Oxford, 2002

[51] UCI Machine Learning Repository. www.ics.uci.edu/~mlearn/MLRepository.html

[52] Smola J, Scholkopf B, Muller K R. The connection between regularization operators and support vector kernels. Neural Networks, 1998, 11:637-649

[53] Andras P. The equivalence of support vector machine and regularization neural networks. Neural Processing Letters. 2002, 65:97-104

[54] Chapelle O, Vapnik V. Model selection for Support Vector Machines. NIPS. 1999, 230-236

[55] Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural networks. 1990, 3:551-560

[56] 段晓军. 神经网络的函数逼近能力分析. 模糊数学与系统. 1998, 12(4):79-84

[57] Wang J Q, Wu X D, Zhang C Q. Support vector machines based on K-means clustering for real-time business intelligence systems. Int. J. Business Intelligence and Data Mining, 2005, 1(1):54-64

[58] Boley D, Cao D W. Training Support Vector Machine using Adaptive Clustering. SIAM International Conference on data Mining. 2004, 126-137

[59] Pavlov D, Chudova D, Smyth P. Towards scalable support vector machines using squashing. In Knowledge Discovery and Data Mining.2000, 295-299.

[60] Shih L, Rennie J D M, Chang Y H, Karger D R. Text bundling: Statistics-based data reduction. In Twentieth International Conference on Machine Learning. 2003, 696-703

[61] 边肇祺，张学工. 模式识别. 清华大学出版社. 2000

[62] Platt J. Fast training of support vector machines using sequential minimal optimization. In:Scholkopf B, Burges C, Smola A, eds. Advances in Kernel Methods:Support Vector Learning. Cambridge, MA:MIT Press, 1999. 185-208

[63] Chang C C, Lin C J. Libsvm-a library for support vector machines. http://www.csie.ntu.edu.tw/ cjlin/libsvm/index.html, Tech. Rep.

[64] Scholkopf B, Smola A, Williamson R C, and Bartlett P L. New support vector algorithms. Neural Computation. 2000, 12:1207 – 1245

[65] Kishore J K, Patnaik L M, Mani V, Agrawal V. K. Genetic programming based pattern classification with feature space partitioning. Information Sciences. 2001, 131:65-86

[66] Simpson P K. Fuzzy Min-Max neural netwoks-part 1: classification. IEEE transaction on neural netwoks. 1992, 3(5):776-786

[67] Garcke J, Griebel M. Data mining with sparse grids using simplicial basis functions. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, 87 - 96

[68] Zhang Q, Benveniste. Wavelet networks. IEEE Transactions on Neural Networks, 1992, 3:889-898

[69] Blayvas I, Kimmel R. Multiresolution approximation for classification. www.mgnet.org/mgnet/Conferences/ CMCIM02/students/blayvas3.pdf

[70] Blayvas I, Kimmel R. Efficient classification via multiresolution training set approximation. CS Dept. Technical Report CIS-2002-01.

[71] Blayvas I, Kimmel R. Machine learning via multi-resolution approximation. Special issue on multi-resolution analysis, IEICE Trans. Inf.&Syst., July, 2003, E86-D(7):1172-1180

[72] Singh S. Multiresolution estimates of classification complexity. IEEE transaction on pattern recognition and machine intelligence. 2003, 25(12):1534-1539

[73] Singh S. PRISM - A novel framework for pattern recognition. Pattern Anal Applic. 2003, 6: 134-149

[74] Yu H, Yang J, Han J W. Classifying large data sets using SVMs with hierarchical clusters. SIGKDD, 2003, 306-315

[75] Liang Y, Page E W. Multiresolution learning paradigm and signal prediction. IEEE transactions on signal processing. 1997, 45(11): 2858-2864

[76] Liang Y. Adaptive neural network activation functions in multiresolution learning. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Nashville, Tennessee, USA, Oct. 2000, 4: 2601-2606

[77] Chan W C, Chan L W. Transformation of Back-propagation networks in multiresolution learning. IEEE International Conference on Neural Networks, 1994, 290-294

[78] Iske A, Levesley J. Multiresolution methods in scattered data modeling. New York, Springer, 2004

[79] Bhatt R, Gaw D, Meystel A. Learning in a multiresolutional conceptual framework. in Proceedings of IEEE International Symposium on Intelligent Control, 1988, 564-568

[80] Dumais S, Chen H. Hierarchical classification of web content. Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval, 2000, 256-263

[81] Ediriwickrema J, Khorram S. Hierarchical maximum-likelihood classification for improved accuracies. IEEE transactions on geoscience and remote sensing. 1997, 35(4): 810-816

[82] Ojala T, PietikaÈinen M, MaÈenpaÈa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence. 2002, 24(7): 971-987

[83] Greenspan H, Goodman R, Chellappa R, Anderson C H. Learning texture discrimination rules in a multiresolution system. IEEE Transactions on pattern analysis and machine intelligence. 1994, 16(9): 894-901

[84] Das N, Bhattacharya B B, Dattagupta J. Hierarchical classification of permutation classes in multistage interconnection networks. IEEE transaction on computers. 1994, 43(12): 1439-1444

[85] Foresti G L, Gentili S. A hierarchical classification system for object recognition in underwater environments. IEEE Journal of oceanic engineering. 2002, 27(1): 66-78

[86] Guo N R, Li T H S, Kuo C L. Hierarchical fuzzy model for classification problem. Proc. IEEE IECON 2002, 2096-2101

[87] Kil D H, Shin F B. A unified approach to hierarchical classification. IEEE Proc. ICASSP, 1549–1552

[88] Ricard M R, Crawford M M. Multiscale hierarchical classification of wetland environments using SAR data. IEEE International Geoscience and Remote Sensing Symposium Proceedings, 1998, 354-356

[89] Chou Y Y, Shapiro L G. A hierarchical multiple classifier learning algorithm. IEEE International Conference on Pattern Recognition, 2000, 152-155

[90] Brun M, Barrera J, etc. Multi-resolution classification trees in OCR design. IEEE Proceedings of XIV Brazilian Symposium on Computer Graphics and Image Processing, 2001, 59-66

[91] Bakshi B R, Stephanopoulos G. Wavelets as basis functions for localized learning in multi-resolution hierarchy. IEEE International Joint Conference on Neural Networks, vol. II, Baltimore, MD, 1992, 140-145

[92] Shao X H, Cherkassky V. Multi-resolution support vector machine. IEEE International Joint Conference on Neural Networks, 1999, 1065 – 1070

[93] Garcke J, Griebel M, Thess M. Data mining with sparse grids. Computing. 2001, 67:225-253

[94] Schölkopf B, Mika S, Burges C J C, Knirsch P, Müller K R, Ratsch G, and Smola A J. Input space versus feature space in kernel-based methods. IEEE Trans. on Neural Network, 1999, 10 (5) : 1000-1017

[95] Burges C J C. Simplified support vector decision rules. Proc. 13th International Conference on Machine Learning, Morgan Kaufmann, 1996, 71-77

[96] Scholkopf B, Smola A J. Learning with kernels. Cambridge, MA: The MIT Press, 2002

[97] Lin K M, Lin C J. A study on reduced support vector machines. IEEE Transactions on Neural Networks, 2003, 14:1449–1459

[98] 张铃，张钹. M-P 神经元模型的几何意义及其应用. 软件学报. 1998, 9(5):334-338

[99] 朱志刚，林学闾，石定机. 数字图像处理. 电子工业出版社. 1998

[100] Bertozzi M, Broggi A, Fascioli A. Vision-based intelligent vehicles: State of the art and perspectives. Robotics and Autonomous Systems. June 2000, 32: 1-16

[101] Dufourd D, Dalgalarrondo A. Quantitative evaluation of image processing algorithms for unstructured road edge detection. in Proc. of SPIE Unmanned Ground Vehicle Technology V, Orlando, USA, Apr. 2003, 5083: 440-451.

[102] Bertozzi M, Broggi A. GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. IEEE Transactions on Image Processing. Jan. 1998, 7(1): 62-81

[103] Broggi A. Robust real-time lane and road detection in critical shadow conditions. In Proc. IEEE International Symposium on Computer Vision, Coral Gables, Florida, Nov. 1995, 353-359

[104] Broggi A, Bertè S. Vision-based road detection in automotive systems: a real-time expectation-driven approach. Journal of Artificial Intelligence Research. December 1995, 3:325-348

[105] Pomerleau D. RALPH: Rapidly adapting lateral position handler. In Proc. IEEE Intelligent Vehicles Symposium, Detroit, Michigan, Sep. 1995, 506-511

[106] Southall B, Taylor C J. Stochastic road shape estimation. In Proc. IEEE International Conference on Computer Vision, ICCV, Vancouver, BC, Canada, July 2001, 1:205 -212.

[107] Tarel J P, Guichard F. Combined dynamic tracking and recognition of curves with application to road detection. In Proc. IEEE International Conference on Image Processing, Vancouver, BC, Canada, Sep. 2000, 1:216 -219

[108] Kanatani K, Watanabe K. Reconstruction of 3-D road geometry from images for autonomous land vehicles. IEEE Transactions on Robotics and Automation. Feb. 1990, 6(1):127 -132

[109] Kluge K, Thorpe C. Representation and recovery of road geometry in YARF. in Proc. IEEE Symposium on Intelligent Vehicles, July 1992, 114-119.

[110] Katsande P, Liatsis P. Adaptive order explicit polynomials for road edge tracking. In Proc. IEEE International Conference on Advanced Driver Assistance Systems, ADAS, Sept. 2001, 63 -67.

[111] Crisman J D, Thorpe C E. SCARF: A color vision system that tracks roads and intersections. IEEE Trans. on Robotics and Automation. Feb. 1993, 9(1): 49-58

[112] Crisman J, Thorpe C E. UNSCARF: A color vision system for the detection of unstructure roads. In Proc. IEEE Int. Conf. Robotics&Automation, Apr. 1991, 2496-2501

[113] Serge B, Michel B. Road segmentation and obstacle detection by a fast watershed transform. In Proc. IEEE Intelligent Vehicles Symposium, October 1994, 296-301.

[114] Rasmussen C. Grouping dominant orientations for ill-structured road following. In proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, July 2004, 1: 470 – 477

[115] Rasmussen C. Combining laser range, color, and texture cues for autonomous road following. In proc. IEEE International Conference on Robotics and Automation, May 2002, 4: 4320 – 4325.

[116] Onoguchi K, Takeda N, Watanabe M. Planar projection stereopsis method for road extraction. IEICE Trans. Inf.&Syst. Sept. 1998, E81-D(9):1006-1018

[117] Chapuis R, Aufrere R, Chausse F. Accurate road following and reconstruction by computer vision. IEEE Transactions on Intelligent Transportation Systems. Dec. 2002, 3:261 – 270.

[118] Oh S Y, Lee J H, Choi D H. A new reinforcement learning vehicle control architecture for vision-based road following. IEEE Transactions on Vehicular Technology. May 2000, 49:997 – 1005.

[119] Paezold F, Franke U. Road recognition in urban environment. In proc. IEEE International Conference on Intelligent Vehicles, 1998, 87-91

[120] Gibbs F W J, Thomas B T. The fusion of multiple image analysis algorithms for robot road following. In proc. IEEE Fifth International Conference on Image Processing and its Applications, July 1995, 394 – 398.

[121] Apostoloff N, Zelinsky A. Robust vision based lane tracking using multiple cues and particle filtering. In proc. IEEE Intelligent Vehicles Symposium, June 2003, 558 – 563.

[122] Lutzeler M, Baten S. Road recognition for a tracked vehicle. In Proc. SPIE Enhanced and Synthetic Vision, 2000, 171–180.

[123] Xiuqing Y, Jilin L, Weikang G. An integrated vision system for alv navigation. International Journal of Pattern Recognition and Artificial Intelligence. 2000, 14:929–940

[124] Golland P, Bruckstein A M. Motion from color. Computer Vision and Image Understanding, December 1997, 68(3):346-362

[125] Birchfield S. Elliptical head tracking using intensity gradients and color histograms. In proc. IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, California, June 1998, 232-237

[126] Baluja S. Evolution of an artificial neural network based autonomous land vehicle controller. IEEE Transaction on System, Man and Cybernetics-Part B:Cybernetics. June 1996, 26(3): 450-463

[127] Conrad P, Foedisch M. Performance evaluation of color based road detection using neural nets and support vector machines. In proc. IEEE Applied Imagery Pattern Recognition Workshop, Oct. 2003, 157 – 160

致　谢

衷心感谢导师张钹教授对本人的精心指导，他的言传身教将使我终生受益。

感谢实验室的林福宗教授，王宏副教授，来春红老师，以及实验室全体老师和同学们的热情帮助和支持！

感谢李建民老师以及讲习教授组的 Alfred M. Bruckstein 教授，他们分别在机器学习和计算机视觉领域和我进行了许多讨论，使我受益匪浅。

最后要特别感谢我的父母以及我的先生张坤龙对我工作上的支持和鼓励，不管我遇到什么困难，他们总是给予我很大的鼓励和帮助。

本课题承蒙国家自然科学基金以及 863 高技术项目资助，特此致谢。

声　明

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

签　名：_____日　期：_____

个人简历、在学期间发表的学术论文与研究成果

个人简历

1975 年 6 月 1 日出生于河南省郑州市。1993 年 9 月考入华中理工大学计算机科学与工程系计算机软件专业，1997 年 7 月本科毕业并获得工学学士学位。1997 年 9 月免试进入华中科技大学计算机学院计算机软件与理论专业，2000 年 7 月研究生毕业并获得工学硕士学位。2001 年 3 月考入清华大学计算机科学与技术系攻读计算机科学与技术博士至今。

发表的学术论文

[1] Yinghua He, Bo Zhang and Jianmin Li, A New Multiresolution Classification Model Based on Partition of Feature Space, IEEE International Conference on Granular Computing, Beijing, China, July 25-27, 2005, pp.462-467

[2] Yinghua He, Hong Wang and Bo Zhang, Color based road detection in urban traffic scenes, IEEE Transactions on Intelligent Transportation Systems, Vol. 5 , Issue: 4 , Dec. 2004 pp. 309-318

[3] Yinghua He, Hong Wang and Bo Zhang, Color based road detection in urban traffic scenes, IEEE International Conference on Intelligent Transportation Systems, Shanghai, China, October 12-15, 2003, pp. 515-519

[4] Yinghua He, Hong Wang and Bo Zhang, Background updating in illumination-variant scenes, IEEE International Conference on Intelligent Transportation Systems, Shanghai, China, October 12-15, 2003, pp. 730-735

[5] Yinghua He, Hong Wang and Bo Zhang, Moving area detection based on color background in outdoor scenes, 7th Joint Conference on Information Sciences (JCIS) , North Carolina, USA, September 26-20, 2003 pp. 623-626

[6] Yinghua He, Bo Zhang and Jianmin Li, A New Multiresolution Classification Model Based on Partition of Feature Space, accepted by International Journal of Intelligent Systems