



上海交通大学

博士学位论文

981760

申请上海交通大学博士学位论文

# 图像相似性模型、算法与应用研究

博士生：唐宏 0030322007

指导教师：施鹏飞 教授

专 业：模式识别与智能系统

研究方向：图像检索

上海交通大学

2006 年 1 月

# 图像相似性模型、算法与应用研究

## 摘要

在基于内容的图像检索系统中,两幅图像的相似程度通常被解释为视觉特征之间的某种距离。然而,人们判断两幅图像的相似程度与人类认知过程密切相关,因为相似性是主观认知的结果,取决于人及其所处的环境。为了模拟和分析人类的主观认知过程,心理学研究者进行了广泛的研究,取得了许多研究成果,如几何相似性、特征相似性、匹配相似性和变换相似性。以多维排列为代表的几何相似性模型假设相似性遵循距离公理,可以在低维空间中利用距离来表达对象之间的相似性。然而,许多心理学实验结果表明相似性并不一定遵循距离公理。特征相似性模型将相似性表达成两个对象相同和相异特征之间的线性组合,它可以解释许多违背距离公理的相似性现象。匹配相似性模型将相似性解释为两个对象不同部分的匹配关系和匹配程度。变换相似性模型认为对象之间的相似性是将一个对象转变成另一个对象的难易程度。在分析多维排列和特征相似性模型的基础上,本文提出一些新的相似性分析方法,并将它们用于度量图像之间的相似性:

(1) 提出拉普拉斯线性判别分析算法。该算法的目标函数是极大化类间和类内散度之差,它不仅形式简单而且不需要计算类内散度矩阵的逆,因此避免了传统线性判别分析中的小样本问题。该算法利用全局和局部拉普拉斯矩阵的谱分解获取最优的降维方向。实验结果表明:相对于传统的线性判别分析算法,该算法具有更优的降维效果。

(2) 提出非线性判别排列算法。根据经典多维排列与主成分分析算法在降维上的对偶关系,该算法将拉普拉斯线性判别分析的思想嵌入到等距映射算法的基本框架中,使等距映射算法从非监督学习转变成监督学习算法,从非线性多维排列转变成非线性判别排列。实验结果表明:该算法可以有效地从非线性数据集中获取易于分类的排列结果。

(3) 根据特征相似性理论的特征诊断性原理,分析了模糊特征对比模型中特征强度与相似性之间的强相关性,提出集成语义和视觉特征的图像特征比率模型。该模型不仅在视觉特征相似性度量时削弱了特征强度与相似性之间的强相关性,而且能协调语义和视觉特征在图像整体相似性度量的作用。实验结果表明该算法具有较高的检索精度。

(4) 提出一个集成光谱反射和吸收特征的光谱特征对比相似性模型。基于



地物波谱数据库的实验结果表明：在光谱相似性检索中，相对于传统光谱相似性度量模型，光谱特征对比相似性模型具有更高的检索精度和召回率。

(5) 实现了一个基于内容的图像检索系统。该系统是一个开放的多源影像检索系统，数据源包括 Corel 图像库、遥感图像、多（高）光谱图像和地物波谱数据库。此外，实现了一个基于 Oracle 的遥感图像数据库管理系统，利用它管理和维护检索系统中的高空间分辨率遥感图像。

## 关键词：

相似性度量、图像检索、线性判别分析、多维排列、拉普拉斯矩阵、遥感图像

# Image Similarity Models, Algorithms and Applications

## ABSTRACT

In an image retrieval system, the similarity between two images often is explained as a distance between visual features. However, human similarity judgment of two images is highly correlated with human cognition, because similarity is the result of subjective cognition, which depends on a person and the environment. In order to simulate and analyze the human cognition, many psychologists have done lots of work and achieved many results, for example geometric similarity, feature similarity, alignment-based similarity and transformational similarity. Geometric similarity models exemplified by multidimensional scaling assume that similarity satisfies the metric axiom. However, many experimental results in psychological society show that it is not necessary for similarity to satisfy the metric axiom. Feature similarity model express the similarity between two objects as a combination of common and distinctive features. It can be employed to explain many similarity phenomena, which do not satisfy the metric axiom. In alignment-based similarity models, the similarity concerns alignable and non-alignable features of two objects. In transformational similarity model, one object is very similar the other, when it is very easy to transform one object to the other. Based on the analysis of multidimensional scaling and feature similarity model, some novel approaches are proposed to measure the similarity between two images.

(1) An algorithm of Laplacian Linear Discriminant Analysis (LLDA) is proposed. The objective of the algorithm is to maximize the difference between between-class scatter and within-class scatter. The objective function is simple in style and it is not necessary to calculate the inverse of within-class scatter. Therefore, the small size sample problem does not exist any more. The algorithm captures optimal discriminant vectors using the spectral decomposition of global and local Laplacian matrices. Experimental results



show that the algorithm is promising in terms of dimensionality reduction.

(2) An algorithm is proposed, which is called Nonlinear Discriminant Scaling (NDS). Based on the duality between the classic multidimensional scaling and principal component analysis, the idea of Laplacian discriminant analysis is elegantly embedded into the framework of Isomap. By this way, Isomap is turned from an unsupervised algorithm into a supervised one, nonlinear multidimensional scaling into nonlinear discriminant scaling. Experimental results show that it is easy for NDS to obtain a separable scaling.

(3) According to the diagnostic principle in feature similarity theory, the strong correlation between similarity and feature intensity is pointed out. A novel model is proposed to integrate both semantic and visual features, i.e., image feature ratio model. In this model, the correlation between similarity and feature intensity is weaken and the role of semantic and visual features can be coordinated in similarity measures. In our experiment, the retrieval precision of the proposed model is high than that of other models.

(4) The spectral feature contrast similarity model is proposed, which integrates both spectral reflectance and absorption features. Experimental results, over spectral database, show that the model is better than traditional models in terms of retrieval precision and recall.

(5) An image retrieval system is developed, which integrates multi-source images including general images, remote sensing images, multi- or hyper-spectral images and spectral database. Moreover, a large volume remote sensing image database is developed and employed to manage high-resolution remote sensing images in image retrieval system.

**Key words:** Similarity Measures, Image Retrieval, Linear Discriminant Analysis, Multidimensional Scaling, Laplace Matrix, Remote Sensing Images



## 上海交通大学

### 学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期：06 年 1 月 18 日



## 上海交通大学

### 学位论文版权使用授权书

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

保密☐，在\_\_\_年解密后适用本授权书。

本学位论文属于

不保密☐。

(请在以上方框内打“√”)

学位论文作者签名:

指导教师签名:

日期: 06 年 1 月 8 日

日期: 06 年 1 月 8 日

# 第一章 绪论

## 1.1 引言

图像作为一种直观而形象的信息载体,已被广泛地应用于生活的各个角落。早在远古时代,人们就用简单的图画表达思想、传递感情,如木雕、壁画以及由各种直观的图画简化而来的象形文字,例如汉字。

随着数字电子产品和网络的普及,多媒体数据,特别是数字图像,在获取、存贮、处理和传输等形式上已经取得了飞速发展。然而,在内容上,仍然主要使用文字标注这种传统的图像描述、分组和索引的方式。古人云:一图胜千言。图像内容丰富多彩。然而,对于不同的人或在不同的环境下,同一幅图像的含义可能迥然不同。因此,面对数量庞大的图像,准确、客观和一致地对图像进行文字标注是一件非常困难的事情。

如何有效地根据图像的内容对图像进行描述、分组、索引以及相似性匹配已成为迫切需要解决的问题。基于内容的图像检索(Content-Based Image Retrieval, CBIR)正是为解决上述问题而提出的技术之一,它的目标是在数字图像库及其用户之间搭建一个直观、有效的“人机交互”平台。由于现有技术的限制,检索系统尚不具备与人平等交互的能力,特别是不具备人在图像表达、理解和评价等方面的能力。因此,目前许多研究旨在从不同的角度提升检索系统在人机交互过程中的地位和作用,例如面向对象的图像分割<sup>[143; 144]</sup>、语义特征的获取、描述和使用、相关反馈和心理学相似性度量等。其中,相关反馈技术使图像检索过程更像是人与人之间的“交流”过程。这种交流的最基本问题是如何评价图像之间的相似性。

相似性判别是人们理解图像并进行分类或推理的基础之一,也是图像检索系统所不具备的关键能力之一。长期以来,人们一直致力于理解、揭示和模拟人类如何进行相似性判别。特别是心理学领域的研究人员进行大量的理论和实验研究,取得了大量的研究成果<sup>[30]</sup>,例如几何模型(Geometric Model)<sup>[54; 55; 100-102; 113]</sup>、特征模型(Featural Model)<sup>[26; 114; 115]</sup>、匹配模型(Alignment-based Model)<sup>[29; 71; 72]</sup>、变换模型(Transformational Model)<sup>[25; 34]</sup>等。在图像检索领域内,距离(特



别是欧氏距离)经常会被不加假设地用于表示图像之间的相似性<sup>[94]</sup>。然而大量的心理学实验结果表明,心理学相似性并不一定满足距离公理,即自相似性、最小性、对称性和三角不等式<sup>[26; 52; 53; 114-116]</sup>。

本文拟将心理学领域内若干相似性模型引入到图像检索领域内。

## 1.2 国内外研究现状

基于内容的图像检索技术最早可以追溯到 1992 年 1 月美国国家科学基金委组织的一次研讨会<sup>[103]</sup>。在这次研讨会上,有人指出可视化信息管理不应该被看成是计算机视觉和数据库的一种应用。此后,随着 Mosaic 网络浏览器的诞生,加速了图片在网络上的传输,也使得有效地管理、检索图像日益成为必要。1995 年,由美国国家自然基金委、国防部高等研究计划局和美国航空及太空总署共同出资两千多万美元支持数字图书馆研究计划。此后,许多国家相继在这一领域投入大量资金进行相关研究。本节将简要叙述 CBIR 技术和图像检索中心理学相似性度量的研究现状。

### 1.2.1 基于内容的图像检索

1994 年,IBM 率先推出了第一个商品化的 CBIR 的原型系统—QBIC (Query By Image Content)<sup>[82]</sup>。此后,出现一系列典型的图像检索系统,例如,UIUC (University of Illinois at Urbana-Champaign) 的 MARS<sup>[90]</sup>, MIT (Massachusetts Institute of Technology) 多媒体实验室的 PhotoBook<sup>[84]</sup>, 美国哥伦比亚大学的 VisualSeek<sup>[104]</sup>和加利福尼亚大学 Santa Barbara 分校的 Netra<sup>[66]</sup>等。

近十年来,CBIR 一直是计算机视觉和多媒体数据库领域内的一个研究热点,涌现出大量相关的研究文献。CBIR 技术研究大致可以分为两个阶段<sup>[50]</sup>:第一阶段主要研究 CBIR 系统及其关键技术,如视觉特征提取、高维索引、检索系统设计;第二阶段主要研究如何以相关反馈为依托提高 CBIR 系统的检索效率,如优化查询、诱导学习等。由于几个综述文献<sup>[89; 103]</sup>和书<sup>[145]</sup>已较为完整地综述了 CBIR 的发展,下面拟从检索系统的功能对主要文献进行综述。

图像特征是图像表示的基础,通常包括视觉特征和语义特征。常用的视觉特征有颜色、纹理和形状<sup>[70]</sup>。

颜色特征被广泛地应用到图像检索中,例如颜色集、颜色直方图和颜色矩。

颜色直方图和颜色矩是统计特征,通常独立于图像中地物的空间位置与分布,具有平移、旋转、尺度等不变性。对它们的研究主要集中在颜色空间的选择及其特征表示<sup>[103]</sup>。在颜色空间中,RGB 是最为广泛使用的,然而只有在成像条件和视觉均不变的情况下,RGB 表示才是最佳的选择,例如艺术画、彩色像片和商标等<sup>[16; 39]</sup>。HSV 因为其具有视觉不变性而常被用于图像检索中<sup>[82]</sup>。

在图像检索中,除了传统的统计纹理特征<sup>[87]</sup>(共生矩阵、MRSAR<sup>[108]</sup>)之外,小波纹理特征也获得了极大的关注。Ma和Manjunath利用Gabor小波不同方向和尺度上的子带统计特性(均值和方差)作为图像的纹理特征<sup>[67; 69]</sup>。实验显示该纹理特征具有较好的检索效果,该纹理曾被应用到航空遥感影像的检索中<sup>[68]</sup>。基于人类纹理视觉感知特性,Tamura等人发展了视觉纹理特征的计算方法,其中纹理表示包括粗糙度、对比度、方向性、成线性、规整度和粗糙度<sup>[108]</sup>。Tamura纹理表示和共生矩阵表示的主要区别是:前者表达的纹理特征在视觉上是有意义的,而后的某些纹理特征则不然。由于Tamura纹理的这些优点,它被广泛地应用到许多图像检索系统中,如QBIC、VisualSeek等。Markov随机场也是纹理分析中非常有效的方法。

形状特征在部分检索系统中得到应用,形状特征的不变性<sup>[139]</sup>(平移、旋转、尺度)和实用的分割算法均在一定程度上限制了形状特征的广泛应用。形状特征可以大致分为两类:(1)基于边界的描述子,如傅利叶描述子<sup>[82]</sup>;(2)基于区域的统计特征,如不变矩<sup>[60]</sup>、Zernike 矩、Pseudo-Zernike 矩<sup>[15]</sup>等。

为了提高检索系统的效率,对高维特征进行有效地索引是必然的选择。检索中常用的高维索引技术有 K-d 树、四叉树、R 树及其变体 R+树、R\*树、X 树等<sup>[3; 5]</sup>。此外,还有自组织特征映射、矢量量化等技术也被用于索引高维特征<sup>[58; 59]</sup>。这些索引基本上是基于欧氏距离的,然而检索中的相似性度量可能是非欧氏距离,White 和 Jain 等对相似性索引进行了深入地研究<sup>[120]</sup>。陈学青等人讨论了基于格矢量量化的倒排文件快速索引方法<sup>[137]</sup>。

视觉特征往往不能准确地表达图像语义,不同的人对同一幅图像可能会存在不同的理解,即使同一个人在不同的时候对同一幅图像也可能有不同的理解,这些问题广泛地存在于 CBIR 系统中<sup>[89; 90]</sup>。为了在一定程度上解决这个问题,最早用于文本检索的相关反馈技术被引入到 CBIR 中<sup>[91]</sup>。近几年来,图像检索领域对该问题进行了深入而广泛地研究,它经历了从基于启发式的加权技术到优化学习相似性度量<sup>[135]</sup>。

启发式的反馈技术包括查询点移动和权重调整<sup>[1; 50; 51; 60; 91; 95; 106]</sup>。Rui 和

Huang 等在 MARS 系统中进一步地改进了这些技术<sup>[90, 91]</sup>。Ishikawa 在 MindReader 中将反馈参数的估计看成是一个最小化优化问题<sup>[47]</sup>。Kwak 提出一种区域增长的查询点移动技术<sup>[57]</sup>。Lee 等人探讨了通过累积反馈信息来实现查询点的移动<sup>[60]</sup>。

由于用户提供的反馈信息可以被看成学习问题中的训练样本, 因此, 大量的学习、分类、优化算法被用于相关反馈技术中, 例如贝叶斯学习、支持向量机、人工神经网络、自组织映射、决策树、遗传算法等。其中, 贝叶斯方法获得了较多的关注, 也表现出较高的检索性能, 例如 Cox 等人开发的 PicHunter 系统<sup>[17]</sup>。由于每次用于反馈的样本相对而言是比较少的, 所以适合于小样本学习的支持向量机被广泛地用于交互式图像检索中。Huang 等人提出基于核方法的偏线性判别分析方法和基于支持向量机的一类分类问题; Guo 等人利用支持向量机学习样本的分类边界<sup>[33]</sup>; He 等人利用支持向量机进行短期和长期学习<sup>[40]</sup>。Zhao 等人利用人工神经网络探讨了反馈参数学习和相似性度量<sup>[132]</sup>。Laaksonen 将自组织映射用于图像检索, 开发了 PicSOM 系统<sup>[58, 59]</sup>。

相关反馈是根据反馈信息来推测用户在检索过程中的潜在语义。对于底层视觉特征与高层语义之间的差距被认为是可以通过特征选择和加权来估计。有几种方案被用于解决特征与语义之间的差距和用户的主观性, 如 He 等提出长期学习与短期学习<sup>[40]</sup>、Lu 等提出基于语义和特征的检索框架<sup>[65]</sup>、Yang 等提出两层对等索引技术和用户自适应检索方案等<sup>[126]</sup>。它们的基本思想是通过反馈模拟用户查询意向时, 在数据库内建立一个语义网络, 以标识使用过的图像的语义。

### 1.2.2 图像检索中的相似性度量

正如Smeulders等人所说: 虽然相似性在图像检索中发挥着至关重要的作用, 然而并没有太多的有关图像相似性的系统研究<sup>[103]</sup>。大量的文献不加分析地将几何距离(特别是加权欧氏距离)作为图像相似性度量的工具。Santini和Jain比较分析了心理学相似性中两种主要的相似性理论: 几何相似性和特征相似性, 并通过模糊逻辑将Tversky提出的特征对比模型应用到图像检索中, 提出了模糊特征对比模型<sup>[94]</sup>。此外, Tolias等人利用Santini等人提出的模糊逻辑泛化了Tversky的比率模型(Tversky Index-Min, TIM)<sup>[112]</sup>。Chaira将TIM用于度量颜色直方图之间的相似性<sup>[10, 11]</sup>。Eidenberger等人将Tversky的特征对比模型应用到了基于逻辑判别的检索图像系统中<sup>[21]</sup>。

Santini和Jain认为图像语义不是一成不变的, 是在人机交互过程中即时产生

的 (Emergent Semantics), 是用户的主观意识与图像的客观内容相互交融的结果。为了有效地表达语义的这种主观特征, 他们利用多维排列 (Multidimensional Scaling, MDS) 来表达用户对图像的相似性判别, 并利用这种排列来优化相似性度量模型, 其中包括各种参数条件下的模糊特征对比模型<sup>[92]</sup>。Mojsilovic 等人先利用 MDS 获取图像的语义描述, 然后再通过一系列算法提取这些对象语义<sup>[76]</sup>。

Frigui 等人使用模糊集度量图像之间的相似性<sup>[24]</sup>。He 等人利用相关反馈学习相似性尺度<sup>[40]</sup>。Tang 等人将模糊特征对比模型应用到度量光谱之间的相似性<sup>[109]</sup>。万华林等人分析了基于二部图匹配的图像相似性度量方法<sup>[117]</sup>。戚飞虎等人讨论了协同神经网络的仿射不变性的相似性度量方法<sup>[140]</sup>。秦斌杰等人分析了相关比相似性测度方法<sup>[141]</sup>。

### 1.3 本文的主要工作和创新点

早在 1965 年, Torgerson 就指出了相似性的两种基本类别<sup>[113]</sup>: (1) 客观相似性, 即对象之间的相似性是对象的多维特征之间的某种函数关系, 比如对象之间的欧氏距离; (2) 主观相似性, 即相似性是人 against 研究对象的认知关系, 换句话说, 相似性是主观认知的结果, 它取决于人及其所处的环境。在图像检索系统中, 图像之间的相似性往往被解释为客观相似性, 即两幅图像的特征 (特别是视觉特征) 之间的某种距离, 最常用的距离也是欧氏距离。众所周知, 图像的内容很难用客观且一致的语言来表达, 这是基于内容的图像检索技术得以迅速发展的原因之一。因此, 图像间的相似性是跟人的认知过程密切相关, 属于主观相似性研究的范畴。然而, 由于人类认知本身的复杂性使得用机器去模拟人类认知过程至少在目前尚不现实, 才导致目前这种用客观相似性度量去解释本应该属于主观相似性判别的图像相似性。

以经典多维排列 (Multidimensional Scaling, MDS) 为代表的几何相似性是心理学相似性的基本模型之一。Tenenbaum 等人在借助邻近图上的测地距离将经典 MDS 算法泛化成非线性 MDS 算法, 即 Isoamp 算法<sup>[110]</sup>。由于经典 MDS 算法在结果上与主成分分析 (Principal Component Analysis, PCA) 的结果存在对偶关系。因此, 经典 MDS 算法本质上也是非监督降维技术。本文首先提出了线性拉普拉斯判别分析算法, 并将其应到视觉特征降维中; 然后, 在深入分析拉普拉斯线性判别分析与 PCA 之间内在联系的基础上, 提出了非线性判别排列算法。通过进一步泛化, 它可以被用于半监督的图像相关反馈中, 以直观地展示检索系统中图

像之间的聚类状况。

Tversky 在 1977 年提出了著名的特征对比模型 (Feature Contrast Model, FCM), 它可以被用于解释许多人类相似性判别结果或现象。Santini 等人首先将其引入到图像检索中, 提出了模糊特征对比模型 (Fuzzy Feature Contrast Model, FFCM)。通过深入分析 FCM 与 FFCM 在相似性度量上的差别, 我们进一步提出了诊断性 FFCM, 由于存在大量的文献分析特征间的诊断性, 本文主要集中于特征内的诊断性。“语义鸿沟”是制约图像检索系统实用性的重要因素之一, 也是图像检索技术的热点和难点问题之一。对于人类来说, 图像的内容就是现实世界中的可见对象、现象及其所隐含的抽象概念和相互关系, 表达图像最直观的语言是自然语言。对于机器而言, 图像就是一组相关的数字, 表达图像的“语言”就是各种特征, 例如颜色、纹理或形状描述子。对于两种使用不同“语言”的人, 他们之间的交流难度就可想而知。对于该问题, 目前存在两种不同的解决方案:

(1) 以一种“语言”为基础, 将另一种“语言”显式或隐式地“翻译”成该“语言”, 如以视觉特征为基础的权重调整、相似性尺度选择、以语义为主的语义建模<sup>[62, 118]</sup>、语义提取<sup>[76]</sup>; (2) 将两种语言“翻译”成一种折中的表达形式, 如集成语义和视觉特征的混合索引<sup>[65]</sup>、语义空间学习<sup>[40, 126]</sup>。在分析语义特征与视觉特征在图像相似性度量中作用的基础上, 本文提出一种直接将图像弱语义嵌入到视觉特征向量及其相似性度量的方法。

具体而言, 本文研究主要研究内容和创新点包括:

(1) 提出拉普拉斯线性判别分析算法。该算法的准则函数是极大化类间散度矩阵和类内散度矩阵之差, 它不仅形式简单而且避免了传统线性判别分析中类内散度矩阵求逆运算, 因此也避免了线性判别分析中常见的小样本问题。该算法通过全局和局部拉普拉斯矩阵的谱分解过程获取最优的降维方向。实验结果表明: 相对于传统的线性判别分析算法, 该算法具有更优的降维效果。

(2) 提出非线性判别排列算法。该算法有效地将拉普拉斯线性判别分析的思想嵌入到等距映射算法的基本框架中, 使等距映射算法从非监督学习算法转变成监督学习算法, 从非线性多维排列算法转变成非线性判别排列算法。模拟实验和图像检索实验结果表明: 该算法可以有效地从非线性数据集中获取易于分类的排列结果。

(3) 根据特征相似性理论的特征诊断性原理, 分析了模糊特征对比模型中特征强度与相似性之间的强相关性, 提出集成语义和视觉特征的图像特征比率模型。该模型不仅在视觉特征相似性度量时削弱了特征强度与相似性之间的强相关



性,而且能协调语义和视觉特征在图像整体相似性度量中的作用。实验结果表明该算法具有较高的检索精度。

(4) 改进模糊特征对比模型,提出一个集成光谱反射和吸收特征的光谱特征对比相似性模型。基于地物波谱数据库的实验结果表明:在光谱相似性检索中,相对于传统光谱相似性度量模型,光谱特征对比相似性模型具有更高的检索精度和召回率。

(5) 实现了一个基于内容的图像检索系统。该系统是一个开放的图像检索系统,实现了图像检索的主要功能,如降维、特征提取、高维索引、相似性度量和相关反馈,其中图像包括一般图片(例如 Corel 图像库)、遥感图像、多(高)光谱图像和地物波谱数据库。此外,实现了一个基于 Oracle 的遥感影像数据库管理系统,利用它管理和维护高空间分辨率的遥感图像,它的主要功能包括:海量遥感图像的多分辨率组织、无缝浏览、漫游和缩放等。

## 1.4 本文的章节安排

本文的章节安排如下:

第一章首先分析了心理学相似性度量在图像检索应用研究的目的和意义,其次综述了图像检索技术的发展现状,最后指出本文的主要创新性工作。

第二章介绍了心理学领域中主要的相似性研究成果,主要包括几何相似性、性征相似性、匹配相似性和变换相似性理论,重点讨论了以多维排列为代表的几何相似性和特征相似性在图像检索中的应用。

第三章提出拉普拉斯线性判别分析的概念、模型和算法,并将其应用到视觉特征降维中。通过分析线性拉普拉斯判别分析与经典多维排列算法的关系。

第四章在分析了经典多维排列与主成分分析算法之间关系的基础上,提出了非线性多维排列算法:

第五章深入分析了模糊特征对比模型在图像相似性度量中的相关问题,指出模糊特征对比模型并没有考虑到 Tversky 提出的特征诊断性原理,特别是特征内诊断性,从而提出了诊断性模糊特征对比模型。此外,结合相关关系反馈,提出了可以有效地集成视觉和语义特征的图像特征比率模型。

第六章简要介绍了本研究过程中开发的基于内容的图像检索系统的基本结构和功能,并较为详细地讨论了系统中的两个独特的部分:(1) 基于 Oracle 的海量遥感图像数据库管理系统;(2) 适合于度量光谱相似性的光谱特征对比相似



J-T-U

## 第一章 绪论

---

性模型。

4

最后总结全文工作，并展望心理相似性模型在图像检索中应用的发展前景。

## 第二章 心理学相似性模型

相似性是人类感知、判别、分类和推理等认知活动的基础<sup>[102]</sup>。由于相似性涉及到人的生理、心理、语言、语义、环境以及历史和文化等诸多因素，到目前为止，仍然没有一个通用的理论可以完全地解释各种相似性现象。随着现代科技的发展，在许多应用领域内，迫切需要解决如何解释和模拟人的判别、分类、推理等方面的认知行为。因此，有关相似性的研究仍然层出不穷。本章介绍心理学领域内主要的相似性模型，包括即几何相似性、特征相似性、匹配相似性和变换相似性。

### 2.1 相似性与认知

人类认识世界的过程也是新的事物或现象与人类已知的知识或现象类比的过程。Quine 指出，“相似性是学习、知识和思想的基础，原因是相似性使得人们可以将事物分成不同的类别，从而使得同类的事物具有相似的行为和功能”<sup>[86]</sup>。然而，由于相似性现象的复杂性，很难对相似性本身给出一个统一、明确且广为接纳的定义。目前，对于相似性与认知的关系，存在两种不同的观点<sup>[30]</sup>：（1）相似性是认知的基础。该观点源于一种直观的认识：相似的事物往往具有类似的属性、行为或功能；这也是人类认识事物、对事物进行分类或推理的基础之一。例如，如果说一个动物和鸟很相似，我们通常就会想到，它可能也有翅膀，也会飞<sup>[6]</sup>。再比如，如果当前的问题跟以前某个问题很相似，那么老问题的解决办法可能对解决当前的问题很有帮助<sup>[42]</sup>。（2）相似性具有很强的不确定性，它不适合构成认知（特别是分类）的基础，认知的基础应该是人类的知识，即人类认识世界的理论。一个著名的例子是 Murphy 和 Medin 等提出的场景：在正式的晚宴上，一个穿着整齐的人突然跳入游泳池<sup>[80]</sup>。这个人很可能被认为是喝醉了酒。显然，这种判断不是基于所有喝醉酒的人都会有“和衣跳入游泳池”这个类似的特征，而是基于对人类一般行为解释，即正常人是不会在晚会上和衣跳入游泳池。认为相似性不能很好地解释认知的另一个著名的学者是 Nelson Goodman。他在 1972 年就提出“如果不指定具体的属性，两个对象可以有无数种解释”<sup>[31]</sup>。根据 Goodman 的观点，如果不指定在什么方面两个对象相似，那么此时的相似性解





释是没有意义的；如果仅仅是因为两个对象属于同一类别而说它们是相似的，此时的相似性也是没有意义的概念，因为分类和相似性不能互为解释。然而，相似性与认知一样，是人类尚没有完全揭开的“谜”。因此，有关它们内涵和关系的讨论远没有到结束的时候。

不同的相似性分析方法之间并没有明显的界线，因此出现了一些互为重叠但又不能互为替代的相似性概念和分析方法。Torgerson 将相似性分为主观相似性和客观相似性。主观相似性将相似性看成人类认知的基础和过程，对象间的相似性是人的一种主观认知关系，它并不取决于对象特定的属性，而是人们根据自己对世界的认识、所处的环境和判别的目的所做出的一种主观的、整体的判断。换句话说，主观相似性是因时、因地、因人而异的主观认知过程。客观相似性是指两个对象在多维特征空间中的某种函数关系，一旦指定特征的类别和函数的形式，对象之间的相似性就取决于对象各个特征的测量值，最常用的距离函数就是欧氏距离。

概念相似性 (conceptual similarity) 和感知相似性 (perceptual similarity) 与主观和客观相似性息息相关，却又略有不同。概念相似性主要是指两个对象在抽象的概念或语义特征的类似程度，它又被称为语义相似性 (semantic similarity) 或深层相似性 (deep similarity)。感知相似性是指两个对象视觉特征之间的类似程度，它又被称为视觉相似性 (visual similarity)、表面相似性 (surface similarity) 或物理相似性 (physical similarity)。概念和感知相似性不是孤立和对立的概念，它们互为依存。Medin 和 Ortony 认为概念相似性最初是源于感知相似性，而概念相似性反过来可能会影响乃至主导感知相似性<sup>[74]</sup>。一般而言，主观相似性更注重概念相似性的作用，但它又不排除感知特征在相似性中的作用。虽然客观相似性更多地被用于度量感知特征之间的相似性，但是概念或语义特征同样可以被用于客观相似性。在图像检索中，如何协调语义和视觉特征之间的相互关系是图像相似性度量的一个重要问题。

## 2.2 几何相似性

以多维排列 (Multidimensional Scaling, MDS) 为代表的几何相似性模型是心理学领域内最具代表性，同时也是最具争议的相似性模型之一<sup>[54; 55; 100-102; 113]</sup>。MDS 假设对象  $A$  和  $B$  的相似性判别  $\delta(A, B)$  可以用低维距离空间中点与点之间的距离  $d(A, B)$  来表达，即

$$\delta(A, B) = g[d(A, B)] \quad (2-1)$$

其中,  $g(x)$  是一个单调非增函数。一般情况下, 这里的距离空间就是欧氏空间。MDS 的目标是确定一个低维的距离空间, 并在该空间中寻求一组点, 使得点与点之间的几何距离尽可能地与其所代表的对象之间的相异性值相近。几何相似性的基本假设是对象间的相似性满足距离公理:

(1) 自相似性是一个常量: 所有对象的自相似性是一个常数, 通常为 0, 即  $d(A, A) = d(B, B) = 0$ 。

(2) 极小性: 所有对象的自相似性均大于它与其它对象间的相似性, 换言之, 所有对象与自己的距离小于它与任意别的对象之间的距离, 即  $d(A, B) \geq d(B, B)$  和  $d(A, B) \geq d(A, A)$ 。

(3) 对称性: 两个对象间的相似性是相等的, 即  $d(A, B) = d(B, A)$ 。

(4) 三角不等式: 对象间的距离满足三角不等式, 即  $d(A, B) + d(B, C) \geq d(A, C)$ 。

然而, 许多心理学实验结果表明, 相似性判别并不满足上述的距离公理。例如, Tversky 用识别实验表明对象间的相似性并不一定满足极小性<sup>[114]</sup>。Tversky 曾用“父与子”之间的相似性判别来说明相似性的不对称性。常听说儿子像父亲, 而很少有人会说父亲很像儿子。换言之, 儿子与父亲的相似程度往往被认为比父亲与儿子的相似程度要高。Krumhansl 认为自相似性并不是常量, 而是与对象所处的对象集的密度有关<sup>[53]</sup>。Hahn 认为相似性的非对称性已经是广为接受的观点<sup>[34]</sup>。由于很难用实验直接证实相似性是否满足三角不等式, Tversky 和 Gati 通过单调逼近结构证实了存在相似性不满足三角不等式的现象<sup>[115]</sup>。Hahn 认为 Tversky 和 Gati 并没有指出违背三角不等式的实验有多大程度的代表性<sup>[34]</sup>。

尽管几何相似性所遵循的距离公理存在许多值得质疑的地方, 但是它仍然被广泛地用于心理学领域。1952 年, Torgerson 基于 Richardson 的工作首次提出 MDS 的概念和算法, 即常说的经典多维排列 (Multidimensional Scaling, MDS)<sup>[113]</sup>。从此以后, 出现大量的有关 MDS 的算法, 如非距离多维排列 (Non-metric Multidimensional Scaling, NMDS)<sup>[54; 55; 100; 101]</sup>、加权多维排列 (Weighted Multidimensional Scaling, WMDS)<sup>[8]</sup>、复制多维排列 (Replicated Multidimensional Scaling, RMDS)<sup>[73]</sup>等。MDS 最近一次重要突破是 Tenenbaum 等提出的 Isomap 算法<sup>[110]</sup>, 该算法将 MDS 从线性泛化到非线性的应用领域。现在, MDS 已经成为一种通用的数据降维和可视化方法。根据 MDS 算法输入数据的性质, MDS 算法可大致可以分为: 距离 MDS (metric Multidimensional Scaling, 也被称为经



典 MDS) 和非距离 MDS (non-metric Multidimensional Scaling)。值得注意的是距离和非距离 MDS 的区别并不是假设相似性是否遵循距离公理。正如 Tversky 所说, 这种称呼名不符实, 出于习惯长期沿用<sup>[114]</sup>。距离 MDS 和非距离 MDS 的区别在于输入数据的性质和寻求目标的不同。经典 MDS 算法分析的是定量数据, 它的目标是在低维空间中寻求一组点, 使得点与点之间的距离尽可能地与给定对象的相异性值相同。而非距离 MDS 算法处理的数据可以是定性的, 它的目标除了在低维空间中寻求一组尽可能保持原始空间中对象之间相异性值相近的点, 还要尽可能地保持原始空间中对象之间相似性排序。下面详细介绍这两类典型的 MDS 算法。

### 2.2.1 经典多维排列

假设  $n$  个对象的平方相异性矩阵为

$$\delta = \begin{bmatrix} 0 & \delta_{12} & \cdots & \delta_{1n} \\ \delta_{21} & 0 & \cdots & \delta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1} & \cdots & \delta_{n,n-1} & 0 \end{bmatrix} \quad (2-2)$$

经典 MDS 算法的目标是确定  $n$  个  $k$  维的点  $\mathbf{Y} = \{y_1, y_2, \dots, y_n\}$ , 使得它们之间平方距离矩阵  $\mathbf{D} = \{d_{ij}, i, j \in [1, n]\}$  满足

$$\mathbf{E} = \arg \min (\delta - \mathbf{D}) \quad (2-3)$$

使得相异矩阵与距离之间的差别尽可能少。

表 2-1 城市间的距离 (公里)

Table 2-1 Distances among cities (km)

	上海	南京	杭州	福州	合肥	南昌
上海	0	2710	163	609	405	606
南京	271	0	236	667	142	464
杭州	163	236	0	472	329	449
福州	609	667	472	0	671	445
合肥	405	142	329	671	0	377
南昌	606	464	449	445	377	0

经典 MDS 算法最常用的一个事例就是根据城市之间的直线距离, 利用该算

法获取城市之间分布。表 2-1 列出了华东五省一市的省会城市之间的直线距离(如表 2-1 所示)。图 2-1 (a) 和 (b) 分别是这些城市的实际地理位置和利用经典 MDS 算法获得的分布。比较两幅图, 不难发现, 经典 MDS 获得的排列不仅反映了城市之间的相对距离, 而且也保持了城市之间的方位关系。

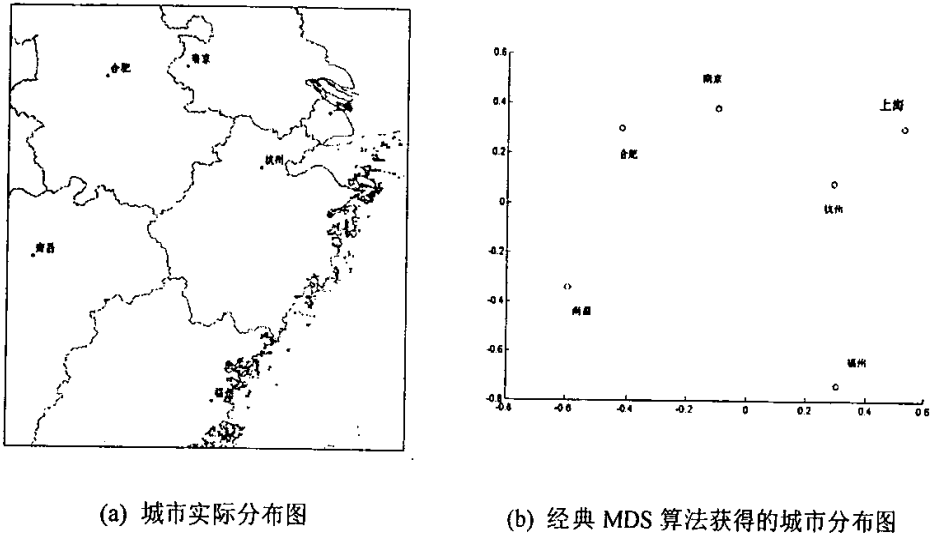


图 2-1 经典 MDS 排列结果与城市之间的实际分布

Figure 2-1 Result of Classic MDS and geological distribution of cities

Torgerson 的算法并没有直接极小化公式 (2-3), 而是通过中心化平方距离矩阵的特征值和特征向量来获取数据点在目标空间中的坐标。该算法的基本思想是首先中心化平方距离矩阵  $\mathbf{D}$ , 设为  $\mathbf{D}'$ 。使得矩阵  $\mathbf{D}'$  中任何行或列的元素之和均为零。然后, 求取矩阵  $\mathbf{D}'$  的特征值和特征向量。其次, 取较大的特征值所对应的特征向量作为目标空间中的坐标。最后, 做适当的平移、旋转和缩放。在城市分布的例子, 选择了两个最大的特征值所对应的特征向量作为城市的地理坐标。此外, 将原始结果旋转了 90 度。与 Torgerson 的算法不同, Young 等人的算法 ALSCAL 是直接通过极小化公式 (2-3) 来获得排列结果<sup>[128]</sup>。

## 2.2.2 非距离多维排列

Shepard 首先提出了非距离多维排列的概念及其算法<sup>[100; 101]</sup>。他假设给定的相异性矩阵  $\delta$  (也被称为逼近数据 (proximity data)) 与待求的欧氏空间中点与点之间的距离  $\mathbf{D}$  满足如下的单调关系:

$$\delta = g(D) \quad (2-4)$$

其中单调函数  $g(x)$  是非距离多维排列最关心的内容之一。非距离 MDS 的目标是：

(1) 确定能保持上述单调性的最小维数；(2) 在给定维数的欧氏空间中求取一组数据点的坐标；(3) 确定单调函数  $g(x)$  的离散形式。基于大量的实验分析，Shepard 认为单调函数应该具有以下形式：

$$g(x) = \exp(-x^\tau) \quad (2-5)$$

其中，参数  $\tau$  需要根据具体的实验来确定<sup>[102]</sup>。

Kruskal 在 Shepard 算法的基础上提出了非距离多维排列的一个度量指标（压力）和一个交互式算法。Kruskal 定义的归一化压力（Stress）为：

$$Stress = \sqrt{\frac{\sum_{i < j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i < j} d_{ij}^2}} \quad (2-6)$$

其中， $d_{ij}$  表示当前排列中点  $i$  和  $j$  之间的距离； $\hat{d}_{ij}$  表示原始数据中对象  $i$  和  $j$  的相异性值。压力的大小表示目标空间中点与点之间距离的排序与给定的相似性排列的吻合程度。压力越大表示越不吻合，此时，一般需要增加目标空间的维数。压力从大到小的突变点可以用于确定目标空间的维数。

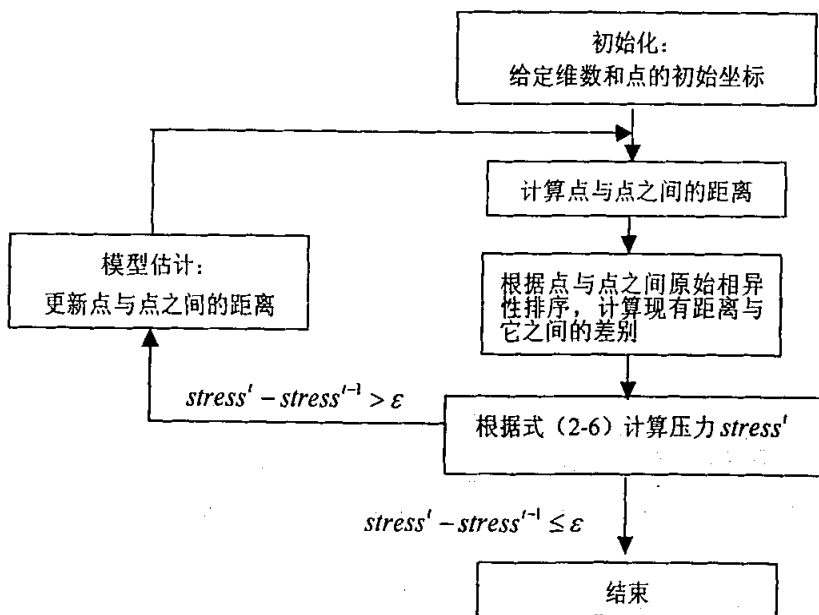


图 2-2 非距离多维排列算法的基本流程  
Figure 2-2 Flow chart of non-metric multidimensional scaling

Kruskal 提出了交互式算法如图 2-2 所示<sup>[54; 55]</sup>。与经典 MDS 算法相比，非距

离 MDS 算法根据较弱的输入获得较多的输出。距离 MDS 要求输入数据是定量的相异性数据(不是直接的相似性判别结果)而且是必须完整的(没有数据缺失)、对称的。非距离多维排列算法的输入数据可以是定量的或定性的、相似性或相异性数据(只需要能对它们进行排序即可)、对称的或非对称的,也可以部分缺失。

## 2.3 特征相似性

在深入分析几何相似性模型所遵循的距离公理的基础上, Tversky 提出了特征相似性理论<sup>[14]</sup>。该理论假设相似性不仅不遵循距离公理, 而且对象的特征不再是按多维结构组织, 而是以集合的形式出现。此外, 他还特别对该模型的特征表示形式做了两点说明: (1) 所研究的对象内容丰富、形式多样; 换言之, 对象的特征是难于穷举的, 而且也是会因人、因地而改变的, 这使得该模型所研究的对象不受具体应用所限制。(2) 特征是二值的(例如, 声音可以表示为有声或无声)或者是可以命名的。因此, Tversky 使用集合而不是几何模型的多维结构来表示对象的特征。

假设  $\Delta = \{a, b, c, \dots\}$  是需要研究的对象集, 其中  $A = \{A_1, A_2, A_3, \dots\}$ ,  $B = \{B_1, B_2, B_3, \dots\}$ ,  $C = \{C_1, C_2, C_3, \dots\}$  分别表示对象  $a, b, c$  的二值特征集。在 Tversky 的相似性实验中, 研究对象在内容和形式上的复杂性, 特征集内的特征是实验者根据自己对实验对象的理解列举出来的。因此, 特征的总数量是可变的。假设  $s(a, b)$  是对象  $a$  和  $b$  之间的相似性。Tversky 的特征相似性理论基于以下几点基本假设:

(1) 匹配性:  $s(a, b) = F(A \cap B, A - B, B - A)$ 。对象  $a$  和  $b$  之间的相似性  $s(a, b)$  可以由一个函数  $F(x)$  表达成, 该函数包含三个如图 2-3 所示的参数, 第一个参数是对象之间的相同特征集, 即  $A \cap B$ ; 第二个参数是存在于对象集  $A$  中却不存在于  $B$  中的相异特征集, 即  $A - B$ ; 最后一个是存在于对象集  $B$  中却不存在于对象  $A$  中的相异特征集, 即  $B - A$ 。

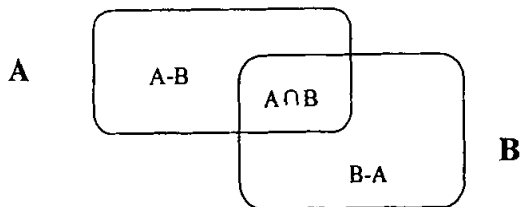


图 2-3 特征集合关系示意图

Figure 2-3 Illustration of the relationship between two feature sets

(2) 单调性: 如果  $A \cap B \supseteq A \cap C$ ,  $A - B \subseteq A - C$ ,  $B - A \subseteq C - A$ , 则不等式  $s(a, b) > s(a, c)$  成立, 即对象间的相似性随着它们之间的相同特征的增加、相异特征的减少而增加。

假设函数  $F(X, Y, Z) = F(A \cap B, A - B, B - A)$ , 如果下式中至少一个等式成立

$$\begin{cases} F(V, Y, Z) = F(W, Y, Z) \\ F(X, V, Z) = F(X, W, Z) \\ F(X, Y, V) = F(X, Y, W) \end{cases} \quad (2-7)$$

则定义  $V \approx W$ 。如果下式中任何一个等式成立, 那么, 两组对象  $(a, b)$  和  $(c, d)$  的相似性中对应地有一组参数相等:

$$\begin{cases} (A \cap B) \approx (C \cap D) \\ (A - B) \approx (C - D) \\ (B - A) \approx (D - C) \end{cases} \quad (2-8)$$

根据上述定义可以得出第三个假设:

(3) 独立性: 假设两组成对对象  $(a, b)$  和  $(c, d)$  与成对对象  $(a', b')$  和  $(c', d')$  有两组相同的参数, 同时, 成对对象  $(a, b)$  和  $(a', b')$  和成对对象  $(c, d)$  和  $(c', d')$  共有另外一组相同的参数, 则下式成立

$$s(a, b) \geq s(a', b') \Leftrightarrow s(c, d) \geq s(c', d') \quad (2-9)$$

此外, Tversky 的相似性理论还假设了独立性、可解性和不变性<sup>[114]</sup>。其中, 满足匹配性和单调性假设的函数被称为匹配函数。在上述基本假设的基础上, Tversky 给出了匹配函数的一种实现形式, 即特征对比模型 (Feature Contrast Model, FCM)。该模型将对象间的相似性表示成三个参数的线性组合:

$$FCM(a, b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad (2-10)$$

其中,  $\theta$ ,  $\alpha$  和  $\beta$  均是非负常量; 特征显著性函数  $f(x)$  通常取集合  $x$  的势。此外, Tversky 还给出了一个比率形式的匹配函数:

$$FCM(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \quad (2-11)$$

上式与许多相似性度量模型在形式上很类似, 如 Bush 和 Mosteller 提出的

$$\frac{f(A \cap B)}{f(A)} \quad [7], \text{ Eisler \& Ekman 提出的 } \frac{f(A \cap B)}{f(A) + f(B)} \quad [22], \text{ Gregson 提出的}$$

$$\frac{f(A \cap B)}{f(A \cup B)} \quad [32]。$$

尽管特征对比模型主要被用于解释心理学领域中的定性相似性判断,但是在其它领域也得到了应用,如语义获取<sup>[76,92]</sup>、图像检索<sup>[21]</sup>、位置相似性<sup>[121]</sup>、光谱相似性度量<sup>[109]</sup>等。

## 2.4 匹配相似性

在几何相似性模型中,相似性取决于特征维数及其各个对象在每个维上的特征值。在特征相似性模型中,相似性取决于对象的相同特征和相异特征。显然,在这两类模型中,相似性并没有考虑到场景中的对象以及对象的属性之间的相互关系。例如,在成对相似性判别时,它们如何对应比较?能找到对应关系的对象或属性与没有对应关系的对象或属性在整体相似性判别时的作用是什么?而这正是匹配相似性模型想要解决的问题。匹配相似性思想最初源于 Genter 等人的类比结构映射理论<sup>[27]</sup>,它的基本观点是:(1)用于相似性判别的元素分为可匹配的(alignable)和不可匹配的(non-alignable);(2)能匹配的元素在相似性的作用大于不可匹配的元素。

Goldstone 等人在 1994 年提出了交互激活和映射相似性模型(Similarity as Interactive Activation and Mapping, SIAM)<sup>[28]</sup>。该模型利用神经网络中节点之间的相互关系解释了能很好匹配的特征和不能很好匹配的特征对整体相似性的影响,前者对相似性的影响要强于后者。该模型利用神经网络中的节点来表示两个场景中的元素,用它们之间的边来表示对应关系及其强度,它的基本步骤如下:

- (1) 为两个场景中的特征添加对应关系,即是否匹配;
- (2) 在特征对应关系的基础上,为两个场景中的对象添加对应关系;
- (3) 特征对应关系与对象对应关系相互影响,即特征和对象的对应关系,如果是一致的,那彼此激活,否则相互抑制。具体实现方式为:如果两个对象被添加了对应关系,那么两个对象的特征对应关系相应地增加,反之减弱;同理,如果两个特征被认为是可以对应的,那么他们所在对象之间的对应也相应地增加,反之亦然。节点之间的相互影响可以用公式表示为:

$$similarity = \frac{\sum_{i=1}^n (match\_value_i * A_i)}{\sum_{i=1}^n A_i} \quad (2-12)$$

其中,  $n$  是节点总数量;  $A_i$  是节点  $i$  被激活的程度;  $match\_value_i$  是第  $i$  节点与其相连接的节点之间的物理相似性。





Markman 和 Gentner 等人利用匹配性差异和非匹配性差异来解释相似性度量<sup>[71]</sup>。所谓匹配性差异是指存在对应关系的两个元素之间的差别,它来源于平行联结过程,即当一个谓项(即属性或关系)与另一个谓项相匹配时,如果这两个谓项各自对应的论点是不等同的表征元素,那么匹配时就出现了匹配性差异。例如,椅子的“功能”是用来“坐”的,而桌子的“功能”是“吃饭用”,这就是匹配性差异,匹配项是“功能”,差异来源于不同的功能。非匹配性差异来源于在一个范畴内的某一表征元素在另一范畴内不存在与之对应的关系。

## 2.5 变换相似性

变换相似性将对象之间的相似性解释为将一个对象转换成另一个对象的难易程度。转换越容易则它们之间越相似,反之亦然。变换相似性的核心是如何定义变换的形式和内容。

Imai 给出了一个变换相似性的例子<sup>[46]</sup>。假设对象的特征可以表示为 XXXXXOO,基本的变换包括:(1)镜像: XXXXXOO  $\rightarrow$  OOOXXXX; (2)平移: XXXXXOO  $\rightarrow$  XXXXOOX (循环平移一位); (3)取反: XXXXXOO  $\rightarrow$  OOOOOXX (设特征 X 和 O 互为相反特征); (4)波长: XXOOXXOOXXOO  $\rightarrow$  XOXOXOXOXO (重复出现的长度为波长,该例是从波长为 4 的序列变为波长为 2 的序列)。

根据 Kolmogorov 的复杂度理论, Li 和 Vitanyi 提出了利用计算机最小程序长度来度量表示对象的复杂程度,而对象之间的相似性被解释为从一种表示形式转换为另一种表示形式的难易程度<sup>[63]</sup>。例如序列 12345678 和 23456789 就很相似,因为两个序列中的元素只相差 1,在计算程序中,它们之间的变换只需要执行加 1 或减 1 操作。然而,相对而言,序列 12345678 和 357911131517 的相似性要低一些,因为他们之间的转换,需要至少两个操作:乘以 2 加 1 或者减 1 以后再除以 2。

## 2.6 图像相似性判断与度量

如果将图像数据库看成一座关于图像的图书馆,那么图像检索系统就像这座图书馆的管理员。用户想要通过检索系统获取图像时的首要任务就是告诉“管理员”什么样的图像是自己想要的。然后,“管理员”根据自己的理解和现有图像

的组织模式去查找用户想要的图像。图 2-4 简要地描绘了这一过程。这与心理学领域中利用相似性模型去解释或模拟人类相似性判断是一致的。略有不同的是心理学领域的研究者更多地关注如何定性地解释人类相似性判别过程中出现有各种现象及其内在机制。图像检索系统则必须定量地度量不同图像之间的相似程度。换言之,心理学相似性可以是定性的或定量的,而图像检索中的相似性度量却必须是定量的。

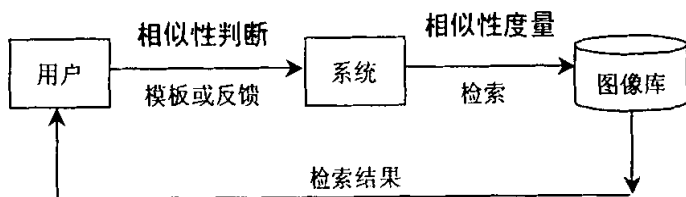


图 2-4 图像检索系统示意图

Figure 2-4 Illustration of image retrieval system

从定量分析的角度而言,几何相似性模型是最适合于度量图像之间的相似性。不过,值得注意的是以 MDS 为代表的几何相似性是一种探索性工具,它与以欧氏距离为代表的距离函数不能完全等同。MDS 的主要目标是在低维空间中以距离这个广为接受的形式可视化人的相似性判断结果,而不是直接度量两幅图像之间的相似性。但是它可以为用户在反馈过程中提供参考,即用户可以通过 MDS 直观地了解目前系统对图像库中某些用户感兴趣的图像之间相似关系或程度。本文第四章在非线性 MDS 的基础上提出了非线性判别排列的概念和算法,将 MDS 从非监督的探索性工具泛化成监督的非线性判别分析工具,从而为用户提供了一种新的分析工具,即可视化检索系统中图像之间的聚类状况。

特征相似性主要是为定性分析对象间相似性而设计的。Santini 等人利用模糊逻辑将其引入图像检索中,用以度量图像之间的相似性。本文第五章根据 Tversky 诊断性原理分析了 Santini 等人的模型特征对比模型中特征强度与诊断性之间的关系,提出顾及特征诊断性的模糊特征对比模型。其次,根据语义与视觉特征之间的关系,提出了图像特征比率模型,该模型集成了图像检索中的弱语义与视觉特征。

匹配和变换相似性需要先创建特征之间的对应关系,然后度量对象或场景之间的相似性。创建特征之间的对应关系需要人为参与,这是一个主观的过程。因此,本文并没有讨论匹配和变换相似性在图像检索中的应用。



## 2.7 本章小结

本章介绍了心理学领域内有关相似性研究的主要模型：几何相似性、特征相似性、匹配相似性和变换相似性。以 MDS 为代表的几何相似性模型是心理学领域应用最为广泛的模型，它假设相似性判断可以在低维的欧氏空间中利用距离来表达。特征相似性不再遵循距离公理，而是对象相同和相异特征之间的线性组合。匹配相似性理论将相似性解释为两个对象不同部分对应关系和对应程度。变换相似性理论认为对象之间的相似性就是将一个对象转变成另一个对象的难易程度。最后结合图像检索的需要讨论了它们在图像检索中的应用。

## 第三章 拉普拉斯线性判别分析

图像视觉特征是图像内容的一种表现形式，是基于内容的图像检索的基础。一方面，为了使得图像视觉特征具有更强的代表性和表现力，图像检索系统往往采用若干种视觉特征，如颜色、纹理和形状特征等。另一方面，随着特征类别的增加，特征维数也急剧增加，从而使得系统的效率降低。因此，降维是提高系统效率关键的问题之一。本章首先简要介绍 MPEG-7 视觉描述子，包括颜色描述子、纹理描述子和形状描述，然后重点阐述一种新的线性判别分析算法——拉普拉斯线性判别分析。最后利用拉普拉斯判别分析算法降低视觉特征的维数。

### 3.1 视觉特征

MPEG 委员会在成功制订 MPEG-1, MPEG-2 和 MPEG-4 的基础上，进一步制订了“多媒体对象描述接口”，即 MPEG-7。MPEG-7 的目标是标准化多媒体对象描述形式，使得一般的数字设备可以在一定程度上理解多媒体的内容。该标准的主要内容包括：系统、描述定义语义、音频、视频、多媒体描述结构和一致性等<sup>[77]</sup>。对于静态图像而言，MPEG-7 采用了一组视觉描述子，即颜色、纹理和形状描述子。在本研究过程中开发的应用系统采用了 MPEG-7 的一些颜色和纹理描述子<sup>[70]</sup>。

#### 3.1.1 颜色描述子

颜色是最直观、最常用和最具表现力的图像特征之一。MPEG-7 采用了四个颜色描述子：主色 (Dominant Color)、可分级颜色 (Scalable Color)、颜色结构 (Color Structure) 和颜色布局 (Color Layout)。

由于颜色描述子往往与颜色空间紧密相关，在详细阐述颜色描述子之前，有必要先简要地介绍一下 MPEG-7 中所使用的颜色空间，其中包括单色空间 (Monochrome Space)、RGB (Red Green Blue)、HSV (Hue Saturation bright Value)、YCrCb 和一个新定义的颜色空间：HMMD (Hue Max Min Difference)。单色空间只定义了颜色中的亮度因子，相当于 YCrCb 空间中的 Y (luminance)，

它可以定义在 RGB 的任意一种颜色上。RGB 即常用的红绿蓝三色空间。HSV 被认为是更接近人类视觉特征的颜色空间。它与 RGB 之间的转换公式如下：

$$H = \begin{cases} (0 + \frac{G-B}{MAX-MIN}) \times 60, & \text{如果 } R = MAX \\ (2 + \frac{B-R}{MAX-MIN}) \times 60, & \text{如果 } G = MAX \\ (4 + \frac{R-G}{MAX-MIN}) \times 60, & \text{如果 } B = MAX \end{cases} \quad (3-1)$$

$$S = \frac{MAX - MIN}{MAX}$$

$$V = MAX$$

其中,  $MAX = \max(R, G, B)$   $MIN = \min(R, G, B)$ 。

YCrCb 空间中 Y 是指颜色的亮度 (luminance), 而 Cr 和 Cb 均是指色度 (Chromance), 它与 RGB 颜色空间的转换关系为:

$$\begin{cases} Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \\ Cb = -0.169 \times R - 0.331 \times G + 0.500 \times B \\ Cr = 0.500 \times R - 0.419 \times G - 0.081 \times B \end{cases} \quad (3-2)$$

HMMD 是一种新定义的颜色空间, H 是指色调, 两个 M 分别表示 RGB 三色中的最大和最小值, D 是最大和最小值之间的差。

#### (一) 主色描述子

主色是指图像中或图像特定区域内占主导地位的颜色。主色描述子描述了图像的主导色彩及其在图像中的百分比、空间一致性和颜色变化的方差, 即

$$F = \{(c_i, p_i, v_i), s\} \quad (3-3)$$

其中,  $c_i, p_i, v_i$  分别表示第  $i$  个主色、及其所占的百分比和方差,  $s$  表示主色在整个图像上的一致性 or 同质性 (homogeneity)。

#### (二) 可分级颜色描述子

可分级颜色描述子是一般的颜色直方图的另一种表现形式。该描述子在 HSV 空间中, 将 RGB 均匀地量化成 256 个级, 并非线性地表示成一个 11 位的值, 然后对原始直方图进行 Haar 变换, 使得直方图的表示具有可分级属性 (多分辨率表示)。因此, 直方图表示的精度从 16 位到 1000 位之间。位数越多, 直方图描述的精度越高, 但是当位数达到一定程度后, 精度的增加会急剧减缓。MPEG-7 核心实验表明, 64 位的直方图就可以获得较好的检索效果, 尽管更高位或 256 位可以获得更高的检索效果。表 3-1 列出了几种典型的 Haar 系数与 HSV 空间量化系数之间的关系。

表 3-1 Haar 系数的个数与 HSV 空间量化

Table 3-1 The number of Haar coefficient and the quantification of HSV space

Haar 系数的个数	H	S	V
16	4	2	2
32	8	2	2
64	8	2	4
128	8	4	4
256	16	4	4

### （三）颜色结构描述子

颜色直方图描述图像中颜色的全局分布，它并不能刻画局部颜色的分布。为了弥补直方图的这一缺点，MPEG-7 采用了颜色结构描述子。首先将图像划分为  $8 \times 8$  的结构元；然后计算颜色直方图，其中直方图中每个颜色出现的次数不再是整个图像中该颜色的数量，而是它出现在不同结构元中的数量。换言之，同一结构元中，同一种颜色被当成一个整体，在统计出现次数时，它们整体被记为出现一次。该描述子采用 HMMD 颜色空间。

### （四）颜色布局描述子

颜色布局描述子旨在刻画整幅图像或图像中任意区域内颜色的空间分布。具体生成方式可以分为两个阶段：（1）图像分块和选择代表性颜色；（2）离散余弦变换（Discrete Cosine Transformation, DCT）和量化。首先，将图像划分为  $8 \times 8$  的块，并提取每个块的代表性颜色，通常是区域内的颜色平均值。然后，对  $8 \times 8$  块的代表颜色进行 DCT 变换，并根据 Zigzag 描述方式选择一些低频系数作为描述子。

## 3.1.2 纹理描述子

MPEG-7 采用了三种纹理描述子，即纹理浏览描述子（texture browsing descriptor）、同质纹理描述子（homogeneous texture descriptor）和局部边界直方图描述子（local edge histogram descriptor）。

### （一）纹理浏览描述子

纹理浏览描述子刻画了纹理的三个主要特征：方向性（directionality）、粗糙度（coarseness）和规则性（regularity）。这些特征也是人类感知图像纹理整体特



性的重要方面。因此,该描述子适合于图像浏览。纹理方向性被定义为纹理的主导方向,从 0 到 150 度,每隔 30 度作为一个方向。因此,可以用 7 个数值来表示,其中,0 表示该纹理没有主导方向,1-6 分别表示 0 到 150 度中 6 个方向。纹理粗糙度与纹理的基元尺寸、分辨率等密切相关。MPEG-7 采用四个数(0-3)表示纹理的粗糙度,其中,0 表示纹理元非常精细,而 3 表示非常粗糙。由于一幅图像可能多于 1 个主要方向,可以设置最多两个主要方向和粗糙度。规则性在一定程度上取决于纹理的方向性和粗糙度,其中 0 表示没有明显的规律性,3 表示图像中的纹理具有很强的方向和一致的粗糙度。

### (二) 同质纹理描述子

该描述子提供了一种定量描述图像纹理的方法,它是频率域中不同方向和尺度上的统计特性。首先使用二维 Gabor 函数过滤图像,在频率域内获得不同方向和尺度上的信息;然后,计算它们的均值和方差。一种典型的频率域划分方法是以 30 度为单位划分为 6 个方向,每个方向上有 5 个尺度,因此,共获得 30 个通道。对每个通道计算其均值和方差,形成 60 维的特征空间。

### (三) 局部边界直方图描述子

与颜色分布描述子相类似,局部边界直方图描述子刻画了边界的空间分布,其中定义了五种边界类型:水平、垂直、45 度对角、135 度对角和没有特定方向的边界。该描述子的计算方法是,首先将图像划分为  $4 \times 4$  的块,然后计算各种边界类型出现的频率。

## 3.1.3 形状描述子

MPEG-7 采用了两种形状描述子:基于区域和基于轮廓,它们分别类似于常用的基于矩和傅利叶描述子的形状特征描述子。尽管形状描述子具有很强的表现力,但是对于一般的图像检索而言,提取有意义的对象形状仍然是一件具有挑战性的工作。

## 3.2 拉普拉斯线性判别分析

如上所述,许多 MPEG-7 视觉描述子的维数都很高,特别是与颜色直方图相关的描述子,如可分级颜色描述子和颜色结构描述子等。在图像检索中,为了使图像的内容更为丰富,往往需要组合使用多种视觉特征。因此,特征的维数一般



会很高。为了提高系统效率,降维就成为必然的选择。已经有大量文献讨论有关降维的问题。降维技术大致可以分为四类<sup>[127]</sup>: (1) 以主成分分析 (Principle Component Analysis, PCA) 为代表的基于重构误差分析的降维方法; (2) 以线性判别分析 (Linear Discriminant Analysis, LDA) 为代表的基于类别信息获取最优分类方向的降维方法; (3) 以独立主元分析 (Independent Component Analysis, ICA) 为代表的降维方法; (4) 以局部线性嵌入 (Locally Linear Embedding, LLE) 和等距映射 (Isometric Mapping, Isomap) 为代表的基于流形学习的降维方法。其中, PCA 和 LDA 是最常用的两种降维技术。

PCA 属于非监督学习算法,它的目标是尽可能地减少数据的重构误差,它并没有考虑不同类之间的可分性。监督学习算法 LDA 的目标是使得降维后,类间样本的散度最大,而类内样本的散度最小。假设存在  $M$  个训练样本  $\mathbf{X} = \{\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_M\}$ , 每个训练样本是一个  $N$  维的向量。样本存在  $L$  个类别,每个类别中的样本数为  $l_i$  ( $i=1, 2, \dots, L$ )。传统 LDA 算法的目标是寻求一个投影方向  $\bar{\mathbf{v}}$ , 极大化投影后样本的类间散度与类内散度之间的比。Fisher 线性判别分析的准则函数为

$$J(\bar{\mathbf{v}}^*) = \arg \max \frac{\bar{\mathbf{v}}^T \mathbf{S}_b \bar{\mathbf{v}}}{\bar{\mathbf{v}}^T \mathbf{S}_w \bar{\mathbf{v}}} \quad (3-4)$$

其中,  $\mathbf{S}_b$  和  $\mathbf{S}_w$  分别为样本的类间散度和类内散度矩阵, 它们可以表示为:

$$\mathbf{S}_b = \frac{1}{M} \sum_i l_i (\bar{\mathbf{m}}_i - \bar{\mathbf{m}}_0)(\bar{\mathbf{m}}_i - \bar{\mathbf{m}}_0)^T \quad (3-5)$$

$$\mathbf{S}_w = \frac{1}{M} \sum_{i=1}^L \sum_{j=1}^{l_i} (\bar{\mathbf{x}}_j^i - \bar{\mathbf{m}}_i)(\bar{\mathbf{x}}_j^i - \bar{\mathbf{m}}_i)^T \quad (3-6)$$

其中,  $\bar{\mathbf{x}}_j^i$  表示第  $i$  类第  $j$  个样本;  $\bar{\mathbf{m}}_i$  和  $\bar{\mathbf{m}}_0$  分别为第  $i$  类样本和所有样本的均值向量。此外, 样本的总体散度  $\mathbf{S}_t$  与类间和类内散度之间存在如下的关系

$$\mathbf{S}_t = \mathbf{S}_b + \mathbf{S}_w \quad (3-7)$$

然而, 在实际应用中, 训练样本的数量总是有限的。当训练样本的数量小于特征的维数时, 就会出现所谓的“小样本问题”。此时, 传统 LDA 算法将不能获得最优的降维方向。为了解决这个问题, 人们提出了许多方法。Tian 等人提出使用类内散度矩阵的伪逆  $\mathbf{S}_w^+$  来代替逆矩阵  $\mathbf{S}_w^{-1}$ <sup>[111]</sup>。向矩阵  $\mathbf{S}_w$  中添加一个小的扰动矩阵, 使之成为满秩矩阵<sup>[44; 133]</sup>, 类似的方法还有正则化 LDA<sup>[23]</sup>和惩罚性 LDA<sup>[38]</sup>, 这类方法的缺点是不稳定。Cheng 等人提出了行列式分解法<sup>[14]</sup>, 该方法依次分解矩阵  $\mathbf{S}_t$  和  $\mathbf{S}_b$ , 该方法不仅计算量大, 而且并没有求出  $\mathbf{S}_w$  的空空间 (Null Space)





中的最优判别向量。Swets 等人提出一种两阶段方法, 即 PCA+LDA<sup>[107]</sup>。该方法在执行 LDA 之前, 先执行 PCA, 并选取部分较大特征值所对应的特征向量, 并将样本投影到这些向量上, 从而使得投影后的类内散度矩阵  $S_w$  不再是奇异矩阵。由于该方法丢弃了部分特征值较小的特征向量, 因此, 可能会失去部分最优的判别向量。此外, Chen 等人提出了另一种两阶段方法, 即 LDA+PCA<sup>[11, 13]</sup>和公共判别向量 (common discriminant vectors)<sup>[11]</sup>。对于小样本问题, Chen 等人的算法首先将样本投影到类内散度矩阵  $S_w$  的空空间中, 然后极大化该空间中样本的类间散度矩阵。当空空间的维数较大时, 这类算法非常有效。当  $S_w$  的空空间维数很低时, 该方法并不能提取  $S_w$  的非空空间中最优判别向量。另一类著名的空空间分析方法是直接 LDA (Direct LDA)<sup>[129]</sup>, 该方法先对角化类间散度矩阵  $S_b$ , 然后再对角化类内散度矩阵  $S_w$ , 并选取零特征值和较小特征值所对应特征向量。该算法实际上并没有极大化  $S_w$  为空而  $S_b$  非空时的最优判别向量。

假设训练样本  $X = \{\bar{x}_1, \dots, \bar{x}_M\}$  的  $L$  个类别标志为  $\{c_1, \dots, c_L\}$ 。Li 等人定义了极大边界准则函数<sup>[61]</sup>:

$$J = \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L l_i l_j d(c_i, c_j) \quad (3-8)$$

如果用各个类的均值向量来表示, 上式可以重写为

$$J = \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L l_i l_j d(\bar{m}_i, \bar{m}_j) \quad (3-9)$$

Li 等人进一步指出上式并没有考虑到各个类本身的散度, 并将上式改写为

$$J = \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L l_i l_j (d(\bar{m}_i, \bar{m}_j) - \text{tr}(S_i) - \text{tr}(S_j)) \quad (3-10)$$

$$= \text{tr}(S_b - S_w)$$

此外, Li 等人提出了一个快速两步离异值分解算法。它的目标是获取变换矩阵  $P$ , 使之同时对角化矩阵  $S_b$  和  $S_i$ :

$$P^T S_b P = \Lambda \quad (3-11)$$

$$P^T S_i P = I \quad (3-12)$$

其中,  $P = \Phi \Theta^{-1/2} \Psi$ , 而  $\Theta$  和  $\Psi$  分别为矩阵  $S_i$  的特征值和特征向量矩阵,  $\Phi$  为矩阵  $\Theta^{-1/2} \Psi^T S_b \Psi \Theta^{-1/2}$  的特征向量矩阵。快速两步离异值分解算法的基本步骤为:

(1) 假设  $S_i = YY^T$ , 则矩阵  $Y$  的奇异值分解获得  $\Theta$  和  $\Psi$ ; (2) 假设矩阵  $\Theta^{-1/2} \Psi^T S_b \Psi \Theta^{-1/2} = ZZ^T$ , 则通过矩阵  $Z$  的奇异值分解获得  $\Phi$ 。该方法存在两个主要的问题: (1) 从公式 (3-12) 可以看出, 变换矩阵  $P$  将所有样本的协方差矩



阵对角化成一个单位矩阵,即变换后训练样本的协方差矩阵是一个单位阵。因此,降维后任意方向上的协方差矩阵均接近相等,从而使得原来方差小的方向在相似性度量时(如欧氏距离)被隐式地分配了更大的权重,相反,原来方差大的方向的权重被减轻。这种情况会随着降维后维数的增加而加剧。(2)当小样本问题出现时(即  $\bar{p}_i^T \mathbf{S}_w \bar{p}_i = 0$  且  $\bar{p}_i^T \mathbf{S}_b \bar{p}_i \neq 0 (\bar{p}_i \in \mathbf{P})$ ),对于任何向量  $\bar{p}_i (\bar{p}_i \in \mathbf{P})$ ,均有  $\bar{p}_i^T \mathbf{S}_b \bar{p}_i = \bar{p}_i^T \mathbf{S}_i \bar{p}_i = 1$ 。因此,该算法并不能区分哪一个方向是最优的判别方向。

### 3.2.1 准则函数

定义以下的准则函数

$$\tilde{J}(\tilde{\mathbf{v}}^*) = \arg \max_{\tilde{\mathbf{v}}} \frac{\tilde{\mathbf{v}}^T \mathbf{S} \tilde{\mathbf{v}}}{\tilde{\mathbf{v}}^T \tilde{\mathbf{v}}} \quad (3-13)$$

其中,矩阵  $\mathbf{S} = \mathbf{S}_b - \mathbf{S}_w$ 。从获取判别向量能力的角度,下面证明了上式与经典的 LDA 准则函数是等同的。此外,在深入分析获取各种情况下最优判别向量的基础上,本节提出两个有效获取判别向量的算法,一个是基于 Li 的两阶段离异值分解的改进算法;另一个是全新的基于拉普拉斯谱分解算法。为区别起见,公式(3-13)被称为拉普拉斯准则函数,本章提出的算法被统称为拉普拉斯线性判别分析(Laplacian Linear Discriminant Analysis, LLDA)算法,并分别记为 LLDA1 和 LLDA2。

定理 3-1: 假设  $\mathbf{T}$  是  $N \times N$  单位正交矩阵,  $\mathbf{S}_b$  和  $\mathbf{S}_w$  分别为类间和类内散度矩阵,

则函数  $f(x) = \frac{\bar{\mathbf{x}}^T \mathbf{S}_b \bar{\mathbf{x}}}{\bar{\mathbf{x}}^T \mathbf{S}_w \bar{\mathbf{x}}}$  和  $g(x) = \frac{\bar{\mathbf{x}}^T \mathbf{T}^T \mathbf{S}_b \mathbf{T} \bar{\mathbf{x}}}{\bar{\mathbf{x}}^T \mathbf{T}^T \mathbf{S}_w \mathbf{T} \bar{\mathbf{x}}}$  取得相同的最大值。

证明: 因为类内散度矩阵  $\mathbf{S}_w$  是半正定矩阵,因此有  $\bar{\mathbf{x}}^T \mathbf{S}_w \bar{\mathbf{x}} \geq 0$ 。下面分两种情况证明该定理:

(1) 当对于任意  $\bar{\mathbf{x}}$ ,  $\bar{\mathbf{x}}^T \mathbf{S}_w \bar{\mathbf{x}} = 0$  成立时(即在  $\mathbf{S}_w$  的空空间内),显然  $\bar{\mathbf{x}}^T \mathbf{T}^T \mathbf{S}_w \mathbf{T} \bar{\mathbf{x}} = 0$ , 此时  $f(x) = g(x) = +\infty$ 。

(2) 当  $\bar{\mathbf{x}}^T \mathbf{S}_w \bar{\mathbf{x}} > 0$  时,  $\bar{\mathbf{x}}^T \mathbf{T}^T \mathbf{S}_w \mathbf{T} \bar{\mathbf{x}} > 0$ , 因此有



$$\begin{aligned}
 g(\bar{x}) &= \frac{\bar{x}^T \mathbf{T}^T \mathbf{S}_b \mathbf{T} \bar{x}}{\bar{x}^T \mathbf{T}^T \mathbf{S}_w \mathbf{T} \bar{x}} \\
 &= \frac{\bar{x}^T (\mathbf{T}^T \mathbf{S}_w \mathbf{T})^{-1} \mathbf{T}^T \mathbf{S}_b \mathbf{T} \bar{x}}{\bar{x}^T \bar{x}} \\
 &= \frac{\bar{x}^T (\mathbf{S}_w \mathbf{T})^{-1} \mathbf{S}_b \mathbf{T} \bar{x}}{\bar{x}^T \bar{x}} \\
 &= \frac{\bar{x}^T \mathbf{S}_b \bar{x}}{\bar{x}^T \mathbf{S}_w \bar{x}} \\
 &= f(\bar{x})
 \end{aligned} \tag{3-14}$$

根据上述两种情况，定理得证。□

根据定理 3-1 可知，将样本投影到任意一组单位正交基上，LDA 的最优判别方向不变。不失一般性，假设单位正交矩阵  $\mathbf{T}$  满足

$$\mathbf{T}^T \mathbf{S}_i \mathbf{T} = \mathbf{I} \tag{3-15}$$

其中， $\mathbf{I}$  为单位矩阵。

定理 3-2: 假设对于  $x \in R$ ，存在  $b(x) + w(x) = 1$ ，且满足  $1 \geq b(x) \geq 0$  和  $1 \geq w(x) \geq 0$ ，

则函数  $J_1(x) = \frac{b(x)}{w(x)}$  和  $J_2(x) = b(x) - w(x)$  之间存在如下的关系：它们在同一点取得极大值。

证明：分两种情况证明该定理：

$$\because 1 \geq b(x) \geq 0 \text{ 且 } 1 \geq w(x) \geq 0$$

$$\therefore 0 \leq J_1(x) \leq +\infty \text{ 和 } -1 \leq J_2(x) \leq 1$$

(1) 如果  $w(x) = 0$ ，则

$$\begin{aligned}
 J_1(x) &= \frac{b(x)}{w(x)} \\
 &= +\infty
 \end{aligned} \tag{3-16}$$

$$\because b(x) + w(x) = 1$$

$$\therefore b(x) = 1$$

$$\begin{aligned}
 J_2(x) &= b(x) - w(x) \\
 &= b(x) \\
 &= 1
 \end{aligned} \tag{3-17}$$

因此，此时  $J_1(x)$  和  $J_2(x)$  均达到其极大值。

(2)  $1 \geq w(x) > 0$

$$\begin{aligned} J_1(x) &= (1 - w(x)) / w(x) \\ &= 1/w(x) - 1 \end{aligned} \quad (3-18)$$

从 (3-8) 式可以看到:  $J_1(x)$  是关于  $w(x)$  的单调递减函数。

$$\begin{aligned} J_2(x) &= b(x) - w(x) \\ &= 1 - 2w(x) \end{aligned} \quad (3-19)$$

同理, 从 (3-9) 式可以看到:  $J_2(x)$  也是关于  $w(x)$  的单调递减函数。此时  $J_1(x)$  和  $J_2(x)$  均是关于  $w(x)$  的单调递减函数。因此, 它们同时取得极大值。

根据上述两种情况, 定理得证。□

根据定理 3-1 和 3-2 可知, 就获取判别向量而言, LLDA 准则函数与 LDA 的准则函数是等同的。

### 3.2.2 谱分解

根据公式 (3-5) 和 (3-6), 矩阵  $S$  可以展开为

$$\begin{aligned} S &= \frac{1}{M} \sum_i l_i (\bar{m}_i - \bar{m}_0)(\bar{m}_i - \bar{m}_0)^T - \frac{1}{M} \sum_{i=1}^L \sum_{j=1}^{l_i} (\bar{x}_j^i - \bar{m}_i)(\bar{x}_j^i - \bar{m}_i)^T \\ &= \frac{1}{M} (2 \sum_{i=1}^L l_i \bar{m}_i \bar{m}_i^T - \sum_{i=1}^M \bar{x}_i \bar{x}_i^T - M \bar{m}_0 \bar{m}_0^T) \\ &= \frac{1}{M} (2XWX^T - XX^T - \frac{1}{M} X\bar{e}(X\bar{e})^T) \\ &= \frac{1}{M} X((I - \frac{1}{M} \bar{e}\bar{e}^T) - 2(I - W))X^T \end{aligned} \quad (3-20)$$

其中,  $I$  是  $M \times M$  的单位矩阵,  $\bar{e}$  是  $M \times 1$  的单位向量 (其元素均为 1); 不失一般性, 假设对训练样本矩阵  $X$  中的样本进行了重新排列, 使得同一类样本排列在一起。因此,  $W$  可以表示成一个分块对角矩阵

$$W = \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_L \end{bmatrix} \quad (3-21)$$

其中,  $W_i$  是一个  $l_i \times l_i$  的矩阵, 即

$$\mathbf{W}_i = \begin{bmatrix} 1/l_i & 1/l_i & \cdots & 1/l_i \\ 1/l_i & 1/l_i & \cdots & 1/l_i \\ \vdots & \vdots & \ddots & \vdots \\ 1/l_i & 1/l_i & \cdots & 1/l_i \end{bmatrix} \quad (3-22)$$

在进一步讨论矩阵  $\mathbf{S}$  的谱分解之前, 先给出两个定义。

**定义 3-1:** 对于一个加权图  $G$ , 其权重矩阵为  $\mathbf{W}_G = \{w_{ij}, i, j \in [1, M]\}$ , 图中所有顶点的度所形成的对角矩阵为  $\mathbf{D}_G = \text{diag}\{d_1, \cdots, d_M\}$ , 其中  $d_i$  是第  $i$  个顶点的度, 且

满足  $d_i = \sum_{j=1, j \neq i}^M w_{ij}$ 。图  $G$  的拉普拉斯矩阵定义为

$$\begin{aligned} \mathbf{L}_G &= \mathbf{D}_G - \mathbf{W}_G \\ &= \begin{cases} -w_{ij}, & i \neq j \\ d_i = \sum_{j=1, j \neq i}^M w_{ij}, & i \in [1, M] \end{cases} \end{aligned} \quad (3-23)$$

**定义 3-2:** 全局拉普拉斯矩阵  $\mathbf{L}_g^{(c)}$  是指图  $G$  上所有节点均存在边相连, 且其权重为  $c$ 。相反, 局部拉普拉斯矩阵  $\mathbf{L}_l^P$  是指图  $G$  上所有节点被分割成  $P$  个互不连通的组, 组间节点互不相通, 组内节点均有边相连, 且其权重为  $1/p_i$  ( $i=1, \cdots, P$ ), 其中,  $p_i$  为第  $i$  组内节点的个数。

根据定义 3-1 和 3-2, 公式 (3-20) 可以改写为

$$\mathbf{S} = \frac{1}{M} \mathbf{X}(\mathbf{L}_g^{(1/M)} - 2\mathbf{L}_l^t) \mathbf{X}^T \quad (3-24)$$

其中,  $\mathbf{L}_g^{(1/M)}$  和  $\mathbf{L}_l^t$  分别为全局和局部拉普拉斯矩阵。为了简化符号, 如不特别

标明,  $\mathbf{L}_g^{(1/M)}$  和  $\mathbf{L}_l^t$  将分别简写为  $\mathbf{L}_g$  和  $\mathbf{L}_l$ 。此外, 定义  $\mathbf{L}_t = \mathbf{L}_g - 2\mathbf{L}_l$ 。因此,

矩阵  $\mathbf{S}$  可以进一步简写为

$$\begin{aligned} \mathbf{S} &= \frac{1}{M} \mathbf{X}(\mathbf{L}_g - 2\mathbf{L}_l) \mathbf{X}^T \\ &= \frac{1}{M} \mathbf{X} \mathbf{L}_t \mathbf{X}^T \end{aligned} \quad (3-25)$$

根据瑞利商原理, 公式 (3-13) 的最优向量是公式中的矩阵  $\mathbf{S}$  的最大的特征值所对应的特征向量。因此, 拉普拉斯线性判别分析就转化成矩阵  $\mathbf{S}$  的谱分析。根据定义 3-1 和 3-2, 矩阵  $\mathbf{S}$  的核心是两个拉普拉斯矩阵。因此, 将在拉普拉斯矩阵谱分析的基础上分析矩阵  $\mathbf{S}$  的谱。首先, 给出有关拉普拉斯矩阵谱的定理。  
**定理 3-3:**  $M \times M$  全局拉普拉斯矩阵  $\mathbf{L}_g$  的特征值为 0 和 1, 而且它们的重数分别

为 1 和  $M-1$ 。

证明：由 (3-23) 式容易验证 0 是全局拉普拉斯矩阵  $\mathbf{L}_g$  的一个特征值，其对应的特征向量为  $M \times 1$  的向量  $\bar{\mathbf{e}}$ 。

假设矩阵  $\mathbf{L}_g$  的特征值 1 对应的特征向量为  $\bar{\mathbf{x}}^T = \{x_1, x_2, \dots, x_M\}$ ，则有

$$\mathbf{L}_g \bar{\mathbf{x}} = 1 \times \bar{\mathbf{x}} \quad (3-26)$$

对于矩阵  $\mathbf{L}_g$  的每一行，均有（设为第  $i$  行）

$$\frac{M-1}{M} x_i - \frac{1}{M} \sum_{j=1, j \neq i}^M x_j = x_i \quad (3-27)$$

进而有

$$\sum_{j=1}^M x_j = 0 \quad (3-28)$$

从而上式可以看到，特征值 1 的对应的特征向量为  $\bar{\mathbf{x}}^T = \{x_1, x_2, \dots, x_M\}$ ，它的最大线性无关组的组数为  $M-1$ 。因此，特征值 1 的重数为  $M-1$ 。定理得证。

□

定理 3-4:  $M \times M$  局部拉普拉斯矩阵  $\mathbf{L}_l^L$  的特征值为 0 和 1，而且它们的重数分别为  $L$  和  $M-L$ 。

证明：从 (3-21) 可以看出，局部拉普拉斯矩阵  $\mathbf{L}_l^L$  所在图的权重矩阵  $\mathbf{W}$  是由  $L$  个  $l_i \times l_i$  分块矩阵  $\mathbf{W}_i$  组成。然而，每个块矩阵又可以被看成一个  $l_i \times l_i$  全局拉普拉斯矩阵  $\mathbf{L}_g^{(l_i)}$  所在图的权重矩阵。局部拉普拉斯矩阵  $\mathbf{L}_l^L$  可以被看成是  $L$  个  $l_i \times l_i$  全局

拉普拉斯矩阵  $\mathbf{L}_g^{(l_i)}$  的直和。因此，局部拉普拉斯矩阵  $\mathbf{L}_l^L$  的特征值也是  $L$  个  $l_i \times l_i$

全局拉普拉斯矩阵  $\mathbf{L}_g^{(l_i)}$  的特征值的集合。故定理得证。

□

根据定理 3-3 和 3-4，存在正交矩阵  $\mathbf{U}$  和  $\mathbf{V}$ ，它们可以同时对角化矩阵  $\mathbf{L}_g$  和  $\mathbf{L}_l$ ，即

$$\begin{cases} \mathbf{U}^T \mathbf{L}_g \mathbf{U} = \Lambda_g \\ \mathbf{V}^T \mathbf{L}_l \mathbf{V} = \Lambda_l \end{cases} \quad (3-29)$$

其中， $\Lambda_g = \text{diag}(\underbrace{0, 1, \dots, 1}_{M-1})$ ， $\Lambda_l = \text{diag}(\underbrace{0, \dots, 0}_L, \underbrace{1, \dots, 1}_{M-L})$ 。

定理 3-5: 存在正交矩阵  $\mathbf{Y} = \{\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_M\}$ ，满足  $\mathbf{Y}^T \mathbf{L}_l \mathbf{Y} = \Lambda_l$ ，其中对角矩阵

$$\Lambda_i = \Lambda_g - 2\Lambda_i = \text{diag}(0, \underbrace{1, \dots, 1}_{L-1}, \underbrace{-1, \dots, -1}_{M-L})。$$

证明：为了便于说明，将全局和局部拉普拉斯矩阵分别展开为

$$\mathbf{L}_g = \begin{bmatrix} \bar{L}_g^1 \\ \bar{L}_g^2 \\ \vdots \\ \bar{L}_g^M \end{bmatrix} = \begin{bmatrix} \frac{M-1}{M} & -\frac{1}{M} & \cdots & -\frac{1}{M} \\ -\frac{1}{M} & \frac{M-1}{M} & \cdots & -\frac{1}{M} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{M} & -\frac{1}{M} & \cdots & \frac{M-1}{M} \end{bmatrix} \quad (3-30)$$

$$\mathbf{L}_l = \begin{bmatrix} L_{11} & & & \\ & L_{22} & & \\ & & \ddots & \\ & & & L_{LL} \end{bmatrix} \quad (3-31)$$

其中， $L_{ii}$  ( $i \in [1, L]$ ) 也是一个  $l_i \times l_i$  拉普拉斯矩阵，即

$$\mathbf{L}_{ii} = \begin{bmatrix} \frac{l_i-1}{l_i} & -\frac{1}{l_i} & \cdots & -\frac{1}{l_i} \\ -\frac{1}{l_i} & \frac{l_i-1}{l_i} & \cdots & -\frac{1}{l_i} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{l_i} & -\frac{1}{l_i} & \cdots & \frac{l_i-1}{l_i} \end{bmatrix} \quad (3-32)$$

假设  $\bar{c}_i$  是一个  $M \times 1$  的列向量，当其元素的位置对应训练矩阵  $\mathbf{X}$  中第  $i$  类样本时，它的值为 1，否则为 0。进一步定义  $M \times 1$  的向量  $\bar{y}$  为

$$\bar{y} = \sum_{i=1}^L \sum_{j=1}^{l_i} \alpha_{ij} \bar{c}_i \quad (3-33)$$

其中， $\alpha_{ij}$  是实数。

下面从两个方面证明该定理。

(1) 假设  $\alpha_{ij} = a_i$  ( $j \in [1, l_i]$ ) 且所有的  $a_i$  ( $i \in [1, L]$ ) 不全为 0。

此时，向量  $\bar{y}$  可以写成

$$\bar{y} = \begin{bmatrix} a_1 \bar{e}_1 \\ a_2 \bar{e}_2 \\ \vdots \\ a_L \bar{e}_L \end{bmatrix} \quad (3-34)$$

其中， $\bar{e}_i$  是一个元素均为 1 的  $l_i \times 1$  的列向量。因此，对于任意的  $i \in [1, L]$ ，下式成立

$$\begin{aligned} L_u \bar{y} &= L_u a_i \bar{e}_i \\ &= a_i L_u \bar{e}_i \\ &= 0 \end{aligned} \quad (3-35)$$

因此, 此时的  $\bar{y}$  是矩阵  $L_l$  的特征值 0 所对应的特征向量, 即

$$L_l \bar{y} = 0 \bar{y} \quad (3-36)$$

又由 (3-34) 式可知,  $\bar{y}$  是由  $L$  个不全为 0 的数所决定 (即  $L$  个线性无关的数组成)。因此, 它处于一个维数为  $L$  的线性子空间中。

为了讨论此时  $\bar{y}$  与全局拉普拉斯矩阵  $L_g$  之间的关系, 在上述条件的基础上, 进一步地假设

(a) 对于所有  $i$  ( $i \in [1, L]$ ),  $a_i = a_0$  ( $a_0 \neq 0$ )

此时,  $\bar{y}_1 = a_0 \bar{e}$  满足

$$\begin{aligned} L_g \bar{y}_1 &= a_0 L_g \bar{e} \\ &= 0 \times \bar{y}_1 \end{aligned} \quad (3-37)$$

(b)  $\sum_{i=1}^L l_i a_i = 0$

对于矩阵  $L_g$  的第  $j$  行  $\bar{L}_g^j$  ( $j \in [1, M]$ ), 此时  $\bar{y}$  满足

$$\begin{aligned} \bar{L}_g^j \bar{y} &= \frac{M-1}{M} a_j - (l_j - 1) \frac{1}{M} - \frac{1}{M} \sum_{i=1, i \neq j}^L l_i a_i \\ &= a_j - \frac{1}{M} \sum_{i=1}^L l_i a_i \\ &= a_j \end{aligned} \quad (3-38)$$

因此, 存在如下关系

$$L_g \bar{y} = 1 \times \bar{y}, \quad (3-39)$$

其中,  $\bar{y}$  由  $L$  个线性相关的实数组成, 因此, 它处于  $L-1$  维的线性子空间中, 也存在  $L-1$  正交向量  $\{\bar{y}_2, \dots, \bar{y}_L\}$ , 使得下式成立

$$L_g \bar{y}_i = \bar{y}_i \quad (i \in [2, L]) \quad (3-40)$$

根据假设 (a) 和 (b), 显然向量  $\bar{y}_1$  与向量组  $\{\bar{y}_2, \dots, \bar{y}_L\}$  是正交的。

根据公式 (3-36)、(3-37) 和 (3-40), 正交向量组  $Y_1 = \{\bar{y}_1, \dots, \bar{y}_L\}$  满足

$$\begin{cases} L_l Y_1 = Y_1 \text{diag}(\underbrace{0, \dots, 0}_L) \\ L_g Y_1 = Y_1 \text{diag}(0, \underbrace{1, \dots, 1}_{L-1}) \end{cases} \quad (3-41)$$

(2) 假设  $\sum_{j=1}^L \alpha_{ij} = 0$  and  $\alpha_{kj} = 0$  ( $k \in [1, L]$  and  $k \neq i$ )。





对于任意的  $i$  ( $i \in [1, L]$ ), 向量  $\bar{y}$  满足

$$\begin{aligned} L_i \bar{y} &= \begin{bmatrix} \frac{l_i-1}{l_i} \alpha_{i1} - \frac{1}{l_i} \sum_{j=2}^{l_i} \alpha_{ij} \\ \frac{l_i-1}{l_i} \alpha_{i2} - \frac{1}{l_i} \sum_{j=1, j \neq 2}^{l_i} \alpha_{ij} \\ \vdots \\ \frac{l_i-1}{l_i} \alpha_{il_i} - \frac{1}{l_i} \sum_{j=1, j \neq l_i}^{l_i} \alpha_{ij} \end{bmatrix} \\ &= \begin{bmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{il_i} \end{bmatrix} \end{aligned} \quad (3-42)$$

因为  $\alpha_{kj} = 0$  ( $k \in [1, L]$  且  $k \neq i$ ), 上式可以写成

$$\mathbf{L}_i \bar{y} = [\underbrace{0, \dots, 0}_{\sum_{i=1}^{i-1} l_i}, \alpha_{i1}, \alpha_{i2}, \dots, \alpha_{il_i}, \underbrace{0, \dots, 0}_{M - \sum_{i=1}^i l_i}]^T = \bar{y}. \quad (3-43)$$

此外, 对于任意行向量  $\bar{L}_g^j$  ( $j \in [1, M]$ , 假设  $j = \sum_{i=1}^{i-1} l_i + k$ ), 向量  $\bar{y}$  满足

$$\begin{aligned} \bar{L}_g^j \bar{y} &= \frac{M-1}{M} \alpha_{ik} - \frac{1}{M} \sum_{j=1, j \neq k}^{l_i} \alpha_{ij} \\ &= \alpha_{ik} - \frac{1}{M} \sum_{j=1}^{l_i} \alpha_{ij} \\ &= \alpha_{ik} \end{aligned} \quad (3-44)$$

即

$$\mathbf{L}_g \bar{y} = \bar{y} \quad (3-45)$$

因为  $\sum_j \alpha_{ij} = 0$  且  $\alpha_{kj} = 0$  ( $k \in [1, L]$ ,  $k \neq i$ ),  $\bar{y}$  处于一个  $M-L$  维的线性子空间中。因此, 必存在  $M-L$  正交的向量组 (设为  $\mathbf{Y}_2 = \{\bar{y}_{L+1}, \dots, \bar{y}_M\}$ ), 使得下式成立

$$\begin{cases} \mathbf{L}_i \mathbf{Y}_2 = \mathbf{Y}_2 \text{diag}(\underbrace{1, \dots, 1}_{M-L}) \\ \mathbf{L}_g \mathbf{Y}_2 = \mathbf{Y}_2 \text{diag}(\underbrace{1, \dots, 1}_{M-L}) \end{cases} \quad (3-46)$$

根据假设 (1) 和 (2), 正交向量组  $\mathbf{Y}_1$  和  $\mathbf{Y}_2$  分别满足条件  $\alpha_{ij} = a_i$  ( $a_i$  不全

为 0) 和  $\sum_j \alpha_{ij} = 0$  且  $\alpha_{kj} = 0$ , 故下式成立

$$\begin{aligned} \mathbf{Y}_1^T \mathbf{Y}_2 &= a_i \sum_j^L \alpha_{ij} \\ &= 0 \end{aligned} \quad (3-47)$$

即正交向量组  $\mathbf{Y}_1$  和  $\mathbf{Y}_2$  是彼此正交的。

不失一般性, 假设  $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2\} = \{\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_M\}$ , 根据公式 (3-41)、(3-46) 和 (3-47), 它可以同时对角化全局和局部拉普拉斯矩阵, 且满足

$$\mathbf{Y}^T \mathbf{L}_i \mathbf{Y} = \Lambda_i, \quad (3-48)$$

其中对角矩阵  $\Lambda_i = \Lambda_g - 2\Lambda_l = \text{diag}(0, \underbrace{1, \dots, 1}_{L-1}, \underbrace{-1, \dots, -1}_{M-L})$ 。因此, 定理得证。

□

根据定理 3-5, 在拉普拉斯判别分析中, 矩阵  $\mathbf{S}$  的两个拉普拉斯矩阵, 可以同时被对角化, 而且其大于零的特征值个数为  $L-1$ 。接下来的问题是, 矩阵  $\mathbf{S}$  存在多少个非零特征值。在回答这个问题之前, 先给下面的定理及其证明。

定理 3-6: 假设  $\mathbf{B}$  是一个  $N \times M$  的矩阵, 对角矩阵  $\Lambda = \text{diag}(\Lambda_1, \dots, \Lambda_M)$  中存在  $K$  个大于零的元素, 则矩阵  $\mathbf{H} = \mathbf{B}\mathbf{A}\mathbf{B}^T$  最多存在  $K$  个大于零的特征值。

证明: 不失一般性, 假设对角矩阵  $\Lambda$  中的正、零和负的对角元素个数分别为  $k$ ,  $j$  和  $i$ , 其中  $M = k + j + i$ 。不妨将对角矩阵  $\Lambda$  写成分块矩阵的形式

$$\Lambda = \begin{bmatrix} \Lambda_{k \times k}^p & & \\ & \Lambda_{j \times j}^z & \\ & & \Lambda_{i \times i}^n \end{bmatrix} \quad (3-49)$$

其中,  $\Lambda_{k \times k}^p$ ,  $\Lambda_{j \times j}^z$  和  $\Lambda_{i \times i}^n$  分别为矩阵  $\Lambda$  中正、零和负的对角元素所组成的对角矩阵。由于矩阵  $\mathbf{B}$  是一个  $N \times M$  的矩阵, 它也可以如下形式的分块矩阵

$$\mathbf{B} = \{\mathbf{B}_{N \times k}^1, \mathbf{B}_{N \times j}^2, \mathbf{B}_{N \times i}^3\} \quad (3-50)$$

利用前面的块表示, 矩阵  $\mathbf{H}$  可以进一步展开为

$$\begin{aligned} \mathbf{H} &= \mathbf{B}\mathbf{A}\mathbf{B}^T \\ &= [\mathbf{B}_{N \times k}^1, \mathbf{B}_{N \times j}^2, \mathbf{B}_{N \times i}^3] \begin{bmatrix} \Lambda_{k \times k}^p & & \\ & \Lambda_{j \times j}^z & \\ & & \Lambda_{i \times i}^n \end{bmatrix} [\mathbf{B}_{N \times k}^1, \mathbf{B}_{N \times j}^2, \mathbf{B}_{N \times i}^3]^T \\ &= \mathbf{B}_{N \times k}^1 \Lambda_{k \times k}^p (\mathbf{B}_{N \times k}^1)^T + \mathbf{B}_{N \times j}^2 \Lambda_{j \times j}^z (\mathbf{B}_{N \times j}^2)^T + \mathbf{B}_{N \times i}^3 \Lambda_{i \times i}^n (\mathbf{B}_{N \times i}^3)^T \\ &= \mathbf{B}_{N \times k}^1 \Lambda_{k \times k}^p (\mathbf{B}_{N \times k}^1)^T + 0 - \mathbf{B}_{N \times i}^3 (-\Lambda_{i \times i}^n) (\mathbf{B}_{N \times i}^3)^T \\ &= \mathbf{B}_{N \times k}^1 (\Lambda_{k \times k}^p)^{1/2} (\mathbf{B}_{N \times k}^1 (\Lambda_{k \times k}^p)^{1/2})^T - \mathbf{B}_{N \times i}^3 (-\Lambda_{i \times i}^n)^{1/2} (\mathbf{B}_{N \times i}^3 (-\Lambda_{i \times i}^n)^{1/2})^T \end{aligned} \quad (3-51)$$

为了简化上式, 定义  $\mathbf{H}_1 = \mathbf{B}_{N \times k}^1 (\Lambda_{k \times k}^p)^{1/2}$ 、 $\mathbf{H}_3 = \mathbf{B}_{N \times i}^3 (-\Lambda_{i \times i}^n)^{1/2}$ 、 $\mathbf{H}_p = \mathbf{H}_1 \mathbf{H}_1^T$  和



$\mathbf{H}_n = \mathbf{H}_3 \mathbf{H}_3^T$ 。则上式可以简化为

$$\begin{aligned}\mathbf{H} &= \mathbf{H}_1 \mathbf{H}_1^T - \mathbf{H}_3 \mathbf{H}_3^T \\ &= \mathbf{H}_p - \mathbf{H}_n\end{aligned}\quad (3-52)$$

显然,  $\mathbf{H}_p$  和  $\mathbf{H}_n$  均是半正定矩阵。不妨假设,  $\mathbf{H}_p$  的秩为  $r \leq \min(N, k) \leq K$ 。则存在正交矩阵  $\mathbf{U}$  和  $\mathbf{V}$  可以分别对角化矩阵  $\mathbf{H}_p$  和  $\mathbf{H}_n$ , 并满足

$$\begin{cases} \mathbf{H}_p = \mathbf{U} \Lambda_p \mathbf{U}^T, \\ \mathbf{H}_n = \mathbf{V} \Lambda_n \mathbf{Z}^T \mathbf{V}^T \end{cases}\quad (3-53)$$

其中,  $\mathbf{Z} = \mathbf{U} \Lambda_p^{-1/2}$ , 对角矩阵  $\Lambda_p$  和  $\Lambda_n$  中的对角元素均为非负数。

令  $\mathbf{Q} = \mathbf{V}^T \mathbf{Z}^T$ , 则矩阵  $\mathbf{S}$  即为矩阵  $\mathbf{H}$  的特征向量

$$\mathbf{Q} \mathbf{H} \mathbf{Q}^T = \begin{bmatrix} \mathbf{I}_{r \times r} & & \\ & 0 & \\ & & 0 \end{bmatrix} - \Lambda_n\quad (3-54)$$

从上式可以看出, 矩阵  $\mathbf{H}$  最多具有  $r$  个正的特征值。由于  $r \leq K$ , 定理得证。

□

由于公式 (3-25) 中的  $\mathbf{S} = \frac{1}{M} \mathbf{X} \mathbf{L}_t \mathbf{X}^T$ , 因此, 舍去系数  $1/M$  后,  $\mathbf{S}$  可以写成

$$\mathbf{S} = \mathbf{X} \mathbf{Y} \Lambda_t (\mathbf{X} \mathbf{Y})^T\quad (3-55)$$

其中,  $\mathbf{L}_t = \mathbf{Y} \Lambda_t \mathbf{Y}^T$ 。根据定理 3-6, 对角矩阵  $\Lambda_t$  大于零的对角元素的个数为  $L-1$ , 矩阵  $\mathbf{S}$  最多具有  $L-1$  个大于零的特征值。这一结论与传统的 Fisher 线性判别分析中有关  $\mathbf{S}_b$  大于零的特征值个数的结论是一致的。这些非零的特征值所对应的特征向量可以用于组建线性降维的变换矩阵。

### 3.2.3 算法

根据瑞利商原理, 公式 (3-4) 和 (3-13) 可以分别转化为一般特征值和特征向量的求解问题

$$\mathbf{S}_w^{-1} \mathbf{S}_b \bar{\mathbf{x}} = \lambda \bar{\mathbf{x}}\quad (3-56)$$

$$\mathbf{S} \bar{\mathbf{x}} = \lambda \bar{\mathbf{x}}\quad (3-57)$$

比较上述两式, 不难发现一个显著的差别: 公式 (3-56) 中需要求取类内散度矩阵的逆矩阵  $\mathbf{S}_w^{-1}$ , 当矩阵  $\mathbf{S}_w$  不可逆时, 公式 (3-56) 就不能求解, 这也是所谓的“小样本问题” (Small Sample Size problem, SSS)。然而, 公式 (3-57) 不存在这个问题。一般而言, 当训练样本的数量小于样本的维数时, 就会出现小样本问



题。然而,当训练样本数量大于样本的维数时,也有可能出现类内散度矩阵  $S_w$  不可逆的情况。为了便于分析,根据散度矩阵的秩,将小样本问题分为以下四种情况:

- (1) SSS0:  $rank(S_w) = rank(S_t) = N$ ;
- (2) SSS1:  $rank(S_w) = rank(S_t) = r$ , 且  $N > r \geq L-1$ ;
- (3) SSS2:  $rank(S_w) < rank(S_t)$ , 且  $rank(S_t) < rank(S_w) + rank(S_b)$ ;
- (4) SSS3:  $rank(S_w) < rank(S_t)$ , 且  $rank(S_t) = rank(S_w) + rank(S_b)$ ;

其中,  $rank(X)$  表示矩阵  $X$  的秩,  $N$  和  $L$  分别表示样本的维数和训练样本的类别数。

SSS0 表示类内散度矩阵  $S_w$  可逆,即不存在小样本问题; SSS1、SSS2 和 SSS3 表示  $S_w$  不可逆,但最优判别向量的分布和求法不尽相同。

众所周知, LDA 的目标是寻求最优的判别向量  $\bar{x}$ , 使得样本投影到该向量上, 类间散度 ( $\bar{x}^T S_b \bar{x}$ ) 最大, 类内散度 ( $\bar{x}^T S_w \bar{x}$ ) 最小。当类内散度  $\bar{x}^T S_w \bar{x}$  的最小值为 0 时, 该类向量所张成的空间被称为  $S_w$  的空空间 (null space), 记为  $S_w(0)$ ,

它的补空间记为  $\overline{S_w(0)}$ 。因此, 空空间  $S_w(0)$  中可能包含很强的判别向量, 此时, 由于存在空空间, 矩阵  $S_w$  不可逆, 即存在小样本问题。另外, 由于  $S_t = S_b + S_w$ , 当  $\bar{x}^T S_t \bar{x} = 0$  时, 必然会有  $\bar{x}^T S_b \bar{x} = 0$  和  $\bar{x}^T S_w \bar{x} = 0$ , 这就是 SSS1 所示的情况。由于类间散度  $\bar{x}^T S_b \bar{x}$  也为零, 显然, 该方向对于分类而言是没有意义的 (因为类与类之间几乎完全不可分)。换句话说, 最优判别向量不可能存在于类间散度矩阵

的空空间 ( $S_b(0)$ ) 内。值得注意的是, 类间散度矩阵的非空空间  $\overline{S_b(0)}$  的最大维数为  $L-1$ 。根据上述分析可知, SSS1 中的空空间  $S_t(0)$  和  $S_w(0)$  内不可能存在最优的判别向量, 为了使矩阵  $S_w$  可逆, 可以事先将这些空空间去除掉。

当  $rank(S_w) < rank(S_t)$  时, 存在一些包含强判别向量的空间, 即  $\{\bar{x} | \bar{x}^T S_w \bar{x} = 0 \text{ 且 } \bar{x}^T S_b \bar{x} \neq 0\}$ 。

当  $rank(S_t) < rank(S_w) + rank(S_b)$  时 (即 SSS2), 存在类内散度矩阵  $S_w$  的空空间  $S_w(0)$ , 但该空间的维数小于类间散度矩阵的非空空间的维数, 即  $rank(S_w(0)) < rank(\overline{S_b(0)})$ , 该空间包含最优的判别向量, 但并没有包含所有最优判别向量。因此, 最优的判别向量既可能存在于类内散度矩阵  $S_w$  的空空间  $S_w(0)$



中,也可能存在于类间散度矩阵的非空空间 $\overline{S_b(0)}$ 中。

当 $\text{rank}(S_t) = \text{rank}(S_w) + \text{rank}(S_b)$ (即 SSS3),类内散度矩阵 $S_w$ 的空空间 $S_w(0)$ 的维数等于非空空间 $\overline{S_b(0)}$ 的维数。因此,最优判别向量只可能存在于类内散度矩阵 $S_w$ 的空空间 $S_w(0)$ 中,即类间散度矩阵的非空空间 $\overline{S_b(0)}$ 与类内散度矩阵 $S_w$ 的空空间 $S_w(0)$ 的交集就张成了最优判别向量所在的空间。

为了求取上述四种情况下的最优判别向量,本章提出了两个算法,分别称之为 LLDA1 和 LLDA2。

#### (一) 算法 1: LLDA1

为了弥补 Li 等人提出的快速两步奇异值分解算法的缺陷,提出了一个如图 3-1 所示的算法,即 LLDA1。该算法是基于上面对“小样本问题”的分析,适用于各种情况下求取最优判别向量。它的基本步骤如下:

步骤 1: 计算矩阵 $S_t, S_b$ 和 $S_w$ 的秩,并根据三个秩之间的关系分为四种情况,即 SSS0、SSS1、SSS2 和 SSS3;

步骤 2: 如果是 SSS3,则根据式(3-57)直接计算矩阵 $S$ 的特征值和特征向量,并取若干大于零的特征值所对应的特征向量作为变换矩阵;

步骤 3: 如果是 SSS0、SSS1 和 SSS2,则先对角化散度矩阵 $S_t$ ,并取舍弃特征等于零的特征值及其对应的特征向量,即 $U^T S_t U = \Lambda_1 > 0$ (其中 $\Lambda_1 > 0$ 表示对角矩阵中对角元素大于零,下同)。

步骤 4: 将训练样本投影到正交矩阵 $U\Lambda_1^{-1/2}$ 上,并计算投影后类间散度矩阵与类内散度矩阵之差,即 $S' = \Lambda_1^{-1/2} U^T (S_b - S_w) U \Lambda_1^{-1/2}$ 。

步骤 5: 此时,再细分为两种情况,一种情况是类内散度矩阵的空空间 $S_w(0)$ 中不包含最优判别向量,即 SSS0 和 SSS1;另一种情况是,最优判别向量既可能包含在空空间 $S_w(0)$ 中,也可能包含在非空空间 $\overline{S_b(0)}$ 中,即 SSS2。

步骤 6: 如果是 SSS0 和 SSS1,求取矩阵 $S'$ 的特征值和特征向量。并可以选取大于-1的特征值所对应的特征向量,即 $V^T S' V = \Lambda_2 > -1$ (如果大于零的特征值数量可以满足预期降维的需要,就选择大于零的特征值对应的特征向量,因为特征值越大,它所对应的特征向量越优)。最后,确定的变换矩阵是它与总样本散度矩阵的非零特征向量 $U$ 的乘积,即 $T = UV$ 。值得注意的是,这与 Li 等人提出的快速

两阶段奇异值分解算法不同的是：Li 等人的算法中变换矩阵是  $T' = \Lambda_1^{-1/2} U V$ ，该变换矩阵会使得样本变换后所有维上的方差均为 1，从而会改变数据本身在不同维上分布的差异。

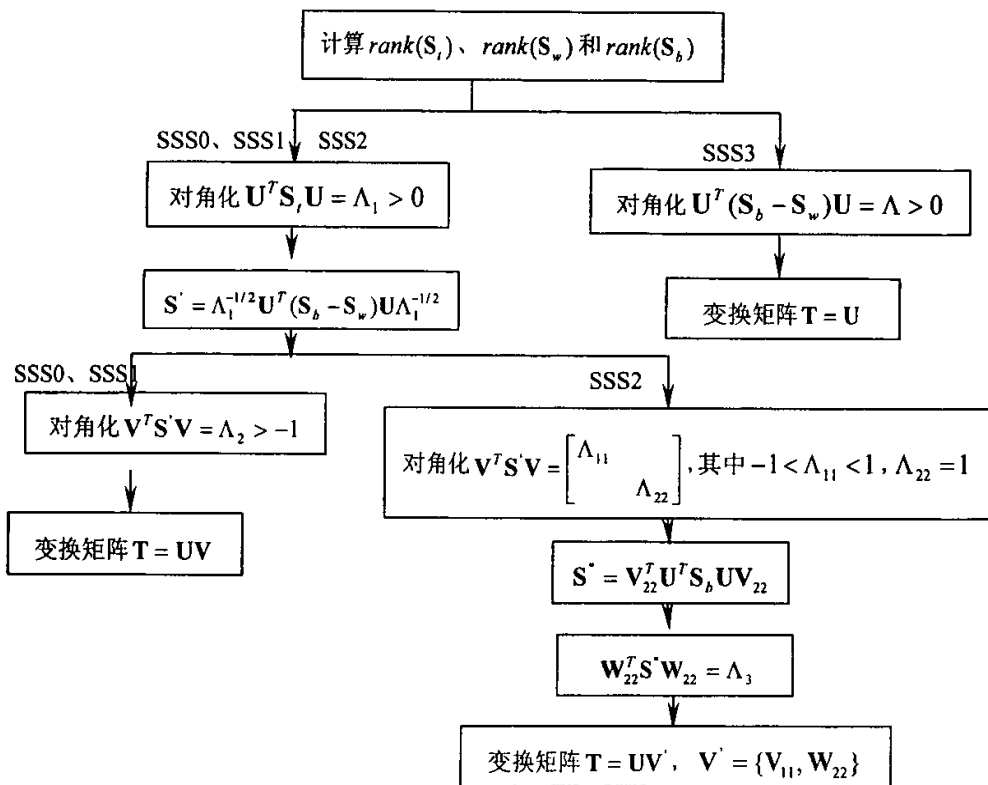


图 3-1 LLDA1 算法示意图

Figure 3-1 Flow chart of the algorithm LLDA1

步骤 7: 如果是 SSS2, 将矩阵  $S'$  的特征值分为三组, 一组是等于 -1 的特征值, 舍弃它们所对应的特征向量; 第二组是处于 -1 和 1 之间的特征值, 这是处于非空空间  $\overline{S_w(0)}$  中的判别向量; 最后一组是等于 1 的特征值, 这些向量一起张成了包含判别向量的空空间  $S_w(0)$ 。由于第三组空间中的特征值均相等, 并不能确定哪个向量更优, 因此, 需要进一步处理。即, 将样本投影到正交矩阵  $UV_{22}$  后, 求取类间散度矩阵的特征值和特征向量, 并选取较大的特征值所对应的特征向量。因此, 最后的变换矩阵为  $T = UV'$ , 其中矩阵  $V'$  是由非空空间  $\overline{S_w(0)}$  中的判别向量



$V_{11}$  和空空间中的  $W_{22}$  组成, 即  $V' = \{V_{11}, W_{22}\}$ 。

## (二) 算法 2: LLDA2

LLDA1 虽然很全面, 但却很复杂, 计算量也很大。LLDA2 是根据 LLDA 准则函数中拉普拉斯矩阵的谱分解过程来求取最优的投影向量。它的步骤如下:

步骤 1: 根据类别重新组织训练样本矩阵中各个样本的位置, 使得同一类样本彼此相邻;

步骤 2: 根据样本总数  $M$ 、每个类别内的样本个数  $l_i$  ( $i \in [1, L]$ ) 及其样本的排列顺序, 计算全局和局部拉普拉斯矩阵  $L_g$  和  $L_l$ , 以及  $L_i$ ;

步骤 3: 计算矩阵  $L_i$  的特征值和特征向量;

步骤 4: 选择矩阵  $L_i$  的大于零的特征值所对应的特征向量, 即式 (3-40) 所示的  $Y_p = \{\bar{y}_i, i \in [2, L]\}$ ;

步骤 5: 将样本投影到正交矩阵  $Y_p$ , 并计算矩阵  $S = XY_p(XY_p)^T$ ;

步骤 6: 根据式 (3-48) 计算矩阵  $S$  的特征值和特征向量;

步骤 7: 取所有的特征向量组成变换矩阵  $T$ 。

比较上述两个算法, 不难发现, 从计算的步骤来看, LLDA2 要比 LLDA1 简单得多。现在的问题是 LLDA2 是否跟 LLDA1 一样能够获取所有的判别向量吗?

根据定理 3-5 可知, 存在单位正交矩阵  $Y$  满足

$$L_i = Y \Lambda_i Y^T \quad (58)$$

其中,  $L_i = L_g - 2L_l$ , 其中对角矩阵  $\Lambda_i = \Lambda_g - 2\Lambda_l = \text{diag}(\underbrace{0, 1, \dots, 1}_{L-1}, \underbrace{-1, \dots, -1}_{M-L})$ 。因

此, 拉普拉斯准则函数中的矩阵  $S$  可以重写为

$$\begin{aligned} S &= X L_i X^T \\ &= X Y \Lambda_i Y^T X^T \end{aligned} \quad (59)$$

根据对角形  $\Lambda_i$  中对角元素为正、零和负将矩阵  $Y$  进行分块, 即  $Y = [Y^p \ Y^z \ Y^n]$ , 进一步假设  $B = [B^p \ B^z \ B^n] = [XY^p \ XY^z \ XY^n]$ , 根据定理 3-6 及其证明, 上式可以重写为

$$\begin{aligned} S &= B_{N \times (L-1)}^p \Lambda_{(L-1) \times (L-1)}^p (B_{N \times (L-1)}^p)^T + B_{N \times 1}^z \Lambda_{1 \times 1}^z (B_{N \times 1}^z)^T + B_{N \times (N-L)}^n \Lambda_{(N-L) \times (N-L)}^n (B_{N \times (N-L)}^n)^T \\ &= S^p + S^z + S^n \end{aligned}$$

(60)

其中,  $N$  和  $L$  分别为训练样本的维数和类别数;  $S^p = B_{N \times (L-1)}^p \Lambda_{(L-1) \times (L-1)}^p (B_{N \times (L-1)}^p)^T$ ,  $S^z = B_{N \times 1}^z \Lambda_{1 \times 1}^z (B_{N \times 1}^z)^T$  和  $S^n = B_{N \times (N-L)}^n \Lambda_{(N-L) \times (N-L)}^n (B_{N \times (N-L)}^n)^T$ 。由定理 3-6 及其证明可知, 矩阵  $S^p$ 、 $S^z$  和  $S^n$  分别为正定、零和负定矩阵。因此, 最优判别向量应该在

正定矩阵  $S''$  的非空空间  $\overline{S''(0)}$  中。因此, LLDA2 跟 LLDA1 一样能够获取所有的判别向量。

此外, LLDA2 具有两个 LLDA1 所不具备的特点:

(1) LLDA2 更适合于小样本情况, 即样本数量  $M$  较少, 且明显小于特征维数  $N$ 。

在 LLDA1 中, 因为  $S_b$ 、 $S_w$  和  $S$  均是  $N \times N$  的矩阵, 从计算和存贮量大, 特别是求解矩阵  $S$  的特征值和特征向量, 它的运算量随着  $N$  的增加而迅速增加。然而, LLDA2 中的矩阵  $L_g$ 、 $L_l$  和  $L_t$  均是  $M \times M$  的矩阵, 因此, 存贮量要求少。另外, 由于  $L_t$  是非常简单而有规律的矩阵, 其特征值求解相对较快。尽量 LLDA2 需要第二次计算矩阵  $S$  的特征值和特征向量, 由于此时矩阵的  $S$  秩的已经非常小 (小于或等于类别数减 1), 因此, 计算量并不大。

(2) 在相同训练模式下, LLDA2 可以极大地节约运算量

所谓的相同训练模式是指训练样本的类别数相同, 每个类别内的训练样本数也相同, 所不同的是训练样本本身, 即维数和在每个维上的值。从 LLDA2 的步骤可以看到, 如果两个训练任务具有相同的训练模式, 它们的第 2 步到 4 步是相同的。因此, 在这样的多任务下, 第 2 到第 4 步只需运行一次, 从而可以极大地减少运算量。

### 3.2.4 非监督拉普拉斯线性判别分析

尽管从形式上看, 拉普拉斯判别分析只是修改了传统的判别函数。实际上, 它从另一角度揭示了线性判别分析的本质: 线性判别分析是在全局和局部拉普拉斯图之间取得一个折中。因此, 拉普拉斯线性判别分析可以被看成是一个与 Fisher 线性判别分析一样的通用分析框架。下面探讨在非监督情况下, 如何实现或逼近拉普拉斯线性判别分析。

不难验证, 矩阵  $\frac{1}{M} \mathbf{X} \mathbf{L}_g \mathbf{X}^T$  实际上就是  $M$  个训练样本矩阵  $\mathbf{X}$  的协方差矩阵。

而  $\frac{1}{M} \mathbf{X} \mathbf{L}_l \mathbf{X}^T$  就是训练样本中, 每个类的类内协方差矩阵的直和。因此, 拉普拉斯判别分析的目标是在极大化所有样本的协方差矩阵的同时, 极小化每个类的类内协方差矩阵。换句话说, 它的目标就是寻找一些方向, 使得所有样本在这此方



向的投影越分散越好，而每个类的样本在这些方向上越聚集越好。然而，如果将上述目标解读成，寻找一些方向，使得所有样本在这此方向的投影越分散越好，而原来临近的样本在这些方向上越聚集越好。这样就可以得到非监督情况下的拉普拉斯判别分析的准则函数：

$$\tilde{J}_U(\tilde{\mathbf{v}}) = \arg \max \frac{\tilde{\mathbf{v}}^T \mathbf{X}(\mathbf{L}_g - \alpha \times \mathbf{L}_k) \mathbf{X}^T \tilde{\mathbf{v}}}{\tilde{\mathbf{v}}^T \tilde{\mathbf{v}}} \quad (3-61)$$

其中， $\alpha$  是一个大于等于 2 的常数； $\mathbf{L}_k$  被称之为  $k$  近邻拉普拉斯矩阵，它所在的图是所有样本的  $k$  近邻图，用公式表示为：

$$\mathbf{L}_k = \mathbf{I} - \mathbf{W}_k \quad (3-62)$$

其中， $\mathbf{W}_k$  是  $k$  近邻图的权重矩阵，用公式表示如下：

$$\mathbf{W}_k = \{w_{ij} \mid w_{ij} = \begin{cases} 1/k, & i \text{ 和 } j \text{ 是邻近点} \\ 0, & i \text{ 和 } j \text{ 不是邻近点} \end{cases}, i \in [1, M]\} \quad (3-63)$$

一般而言，在提取判别向量的能力方面，非监督拉普拉斯线性判别分析要比拉普拉斯判别分析要弱。然而，通过适当地选择参数  $k$  和  $\alpha$ ，非监督分辨分析能力可能达到一般线性分辨分析的水平。参数的选择需要具体问题大量的实验分析。尽量还没有探索出如何选取这些参数的方法，因此，它仍然不是可供实际使用的成熟方法。然而，在某些没有类别信息可供使用的情况下，非监督拉普拉斯线性判别分析仍然是一种可供选择的方法。

### 3.4 实验

本次实验所用的图像数据库是由 100 000 多幅图像组成的 Corel 图像集，随机从中选择 95 个主题的图像，每个主题包含 100 幅图像，其中主题涉及动物、植物、自然和人造景观等。同一个主题的图像被当成一个类别。为了使视觉特征更具有代表性和表现力，实验选用了 MPEG-7 中的 6 种视觉描述子，共有 674 维视觉特征，这此描述子包括可分级颜色描述子、颜色结构描述子、主色描述子、颜色布局描述子、同质纹理描述子和边界直方图描述子<sup>[70]</sup>。提取这些描述子的代码来自于 MPEG-7 实验模型软件<sup>[77]</sup>。表 3-2 列出了实验中所使用的各种描述子的基本信息。在本文后续章节中，如不做特别说明，所有基于 Corel 图像库的检索实验采用的图像及其视觉描述子均与本实验相同。

表 3-2 实验使用的 MPEG-7 视觉描述子

Table 3-2 MPEG-7 visual descriptors in experiments



名称	颜色空间	维数	简述
可分级颜色	HSV	256	$(H, S, V) = (16, 4, 4)$
颜色结构	HMMD	256	标准的非线性量化表
主色	LUV	8	2 个主色以及它们的百分比
颜色布局	YCrCb	12	$(Y, Cr, Cb) = (6, 3, 3)$
同质纹理	RGB	62	6 个方向 和 5 个尺度
边界直方图	RGB	80	划分为 16 子图像、5 滤波器

为了从不同角度测试算法的性能,本章做了两组实验。在第一组实验中,从每个图像类中随机选择 5 幅图像作为训练样本(共 475 幅图像),其余每类图像中的 95 图像均被当作测试图像(即查询图像)。然后,相对于任何给定的查询图像,计算它与图象库中所有图像之间的距离;最后,根据返回的最相似的若干幅图像,计算检索精度;这又分为两种情况:一种是每次检索出与查询图像最相似的 20 幅图像,另一种是检索出 40 幅与查询图像最相似的图像。第二组实验设置跟第一组基本相同,不同之处是训练和测试图像的数量。该实验中,随机从每个类中选择 10 幅图像作为训练样本,其余 90 幅图像作为测试样本。尽管从表面上看,这两组实验的设置极为相似,但是实际上它们分别属于两种不同类型的小样本问题。在第一组实验中,训练样本的总体、类内和类间散度矩阵的秩分别为 474、380 和 94。因此,所有最优判别向量应处于类内散度矩阵的零特征值所对应的特征向量所张成的空间内,它属于四类小样本问题中的第中类,即 SSS3 (因为  $rank(S_t) = rank(S_w) + rank(S_b)$ )。在第二组实验中,训练样本的总体、类内和类间样本矩阵的秩分别为 648、648 和 94,它属于四类小样本问题中的第二类,即 SSS1 (因为  $rank(S_t) = rank(S_w) = 648 < 674$ )。因此,可以舍弃总体散度矩阵或类内散度矩阵的零特征值所对应的特征向量。

表 3-2 和 3-3 列出了第一组实验中各个算法在不同降维情况下的图像检索精度。从中可以看到,LLDA1 和 LDLA2 的检索精度明显高于其它算法,其中 LLDA1 要略优于 LLDA2。根据检索精度的高低,其它算法依次为 DLDA、LDA 和 MMC。与其它算法相比,MMC 不仅检索精度最低,而且随着维数的增加,它的检索精度与其它算法的差距越来越大。这是因为(1)降维后任意方向上的协方差矩阵均接近相等,从而使得原来方差小的方向在相似性度量时被隐式地分配了更大的



权重,相反,原来方差大的方向的权重被减轻;(2)对于任何向量  $\bar{p}_i$  ( $\bar{p}_i \in \mathbf{P}$ ),均有  $\bar{p}_i^T \mathbf{S}_b \bar{p}_i = \bar{p}_i^T \mathbf{S}_t \bar{p}_i = 1$ 。因此,该算法并不能区分哪一个方向是最优的判别方向。

表 3-3 反馈 20 幅图像和不同维数下的检索精度

Table 3-3 Retrieval precisions with 20 feedback images and various dimensionalities

维 数	10	20	30	40	50	60	70	80	90
LLDA1	26.052	30.508	31.913	32.626	32.981	33.181	33.363	33.502	33.552
LLDA2	24.712	29.774	31.346	32.225	32.606	32.798	32.943	33.027	33.065
MMC	8.3195	9.3368	10.1	10.199	10.953	11.65	12.217	12.674	12.841
DLDA	18.131	23.654	26.554	28.204	29.005	29.718	30.193	30.528	30.838
LDA	13.745	17.452	19.648	21.05	22.23	23.075	23.752	24.204	24.586

表 3-4 反馈 40 幅图像和不同维数下的检索精度

Table 3-4 Retrieval precisions with 40 feedback images and various dimensionalities

维 数	10	20	30	40	50	60	70	80	90
LLDA1	20.571	24.053	25.022	25.563	25.838	25.989	26.11	26.181	26.227
LLDA2	19.339	23.236	24.364	25.038	25.297	25.439	25.533	25.601	25.639
MMC	5.1179	5.7882	6.2634	6.4468	6.9961	7.4784	7.8574	8.0745	8.1253
DLDA	14.199	18.316	20.498	21.656	22.223	22.713	23.097	23.273	23.482
LDA	10.076	12.721	14.201	15.163	15.987	16.606	17.052	17.365	17.669

表 3-4 和 3-5 列出了第二组实验结果。如果从检索精度高低排序而言,这组结果跟第一组实验结果完全吻合。不同之处是:在第二组实验中,所有算法的检索精度均有所提高。这主要是因为第二组实验中增加了一倍的训练样本,使得基于训练样的统计特性更接近所有样本的分布特征。另一个值得注意的情况是:在该组实验中,MMC 跟其它算法在检索精度上的差距要明显少于第一组实验中它们之间的差距。这是因为该实验中,所有判别向量不再处于训练样本的类内散度矩阵的空空间内,判别向量之间的优劣可以通过目标矩阵的特征值大小来加以区别。而在第一组实验中,所有最大的特征值增为 1,不能区分这些向量中哪些向量更优。

表 3-5 反馈 20 幅图像和不同维数下的检索精度



Table 3-5 Retrieval precisions with 20 feedback images and various dimensionalities

维数	10	20	30	40	50	60	70	80	90
LLDA1	23.703	30.023	32.514	33.782	34.696	35.164	35.489	35.648	35.762
LLDA2	25.438	30.214	31.978	32.775	33.179	33.393	33.566	33.704	33.752
MMC	15.485	18.798	19.983	20.726	21.195	21.445	21.399	21.468	21.359
DLDA	23.314	28.87	30.942	31.959	32.223	32.478	32.749	32.874	32.881
LDA	14.239	17.912	20.156	21.711	22.765	23.553	24.121	24.377	24.616

表 3-6 反馈 40 幅图像和不同维数下的检索精度

Table 3-6 Retrieval precisions with 40 feedback images and various dimensionalities

维数	10	20	30	40	50	60	70	80	90
LLDA1	18.762	23.92	25.867	26.82	27.573	27.819	28.014	28.137	28.197
LLDA2	19.895	23.696	25.012	25.553	25.846	26.02	26.113	26.176	26.199
MMC	10.489	12.789	13.42	13.781	13.952	13.953	13.867	13.866	13.699
DLDA	18.847	23.133	24.563	25.212	25.294	25.4	25.475	25.471	25.38
LDA	9.8563	12.454	14.092	15.207	15.891	16.467	16.812	16.847	16.974

上述的实验结果表明，本章提出的拉普拉斯线性判别分析算法可以适应不同类型的小样本问题。此外，从表 3-3 到 3-6 中可以看到，当维数大于 30 以后，检索的精度并不再有明显地提高。因此，在本文后续实验中，（如果不做特别说明）所有视觉特征向量都是利用 LLDA1 算法降到 30 维时的结果，其中原始的视觉特征向量是表 3-2 所示的六种视觉描述子。

### 3.5 本章小结

本章简要介绍 MPEG-7 颜色、纹理和形状描述子，提出拉普拉斯线性判别分析算法，即 LLDA1 和 LLDA2。LLDA1 改进了 Li 等人提出的两阶段奇异值分解算法，弥补了该算法中的两大缺点：（1）变换矩阵将所有样本的协方差矩阵对角化成一个单位矩阵，即变换后训练样本的协方差矩阵是一个单位阵。因此，降维后任意方向上的协方差矩阵均接近相等，从而使得原来方差小的方向在相似性度量时（如欧氏距离）被隐式地分配了更大的权重，相反，原来方差大的方向的权重被减轻。这种情况会随着降维后维数的增加而加剧。（2）当小样本问题出现时



(即  $\bar{p}_i^T \mathbf{S}_w \bar{p}_i = 0$  且  $\bar{p}_i^T \mathbf{S}_i \bar{p}_i \neq 0 (\bar{p}_i \in \mathbf{P})$ ), 对于任何向量  $\bar{p}_i (\bar{p}_i \in \mathbf{P})$ , 均有  $\bar{p}_i^T \mathbf{S}_b \bar{p}_i = \bar{p}_i^T \mathbf{S}_i \bar{p}_i = 1$ 。因此, 该算法并不能区分哪一个方向是最优的判别方向。

LLDA2 将线性判别分析转化成准则函数中全局和局部拉普拉斯矩阵的谱分解。

该算法在相同训练模式下, LLDA2 可以极大地节约运算量, 因此, 它更适合于具有相同训练模式的多任务情况。实验结果表明本章提出的算法不仅可以适用于不同类型的小样本问题, 而且降维效果优于传统的线性判别分析算法。

## 第四章 非线性判别排列

在图像检索中,图像既是需要描述的对象,也可以被看成人机交互过程中系统所能使用的、用于描述其它图像内容的“语言”。因此,如何让系统尽可能像人类一样理解、表达和运用这种“语言”是检索系统成功与否的关键。一种解决办法是模拟人类理解、表达和运用这种“语言”的行为,这包括两个部分:(1)人类理解图像的行为,特别是人的图像相似性判别行为;(2)图像本身的内涵,特别是图像语义。然而,由于这两个问题都非常复杂,很难对它们进行有效地建模。另一种解决方法是根据人类理解、表达和判别图像相似性的结果,探索这些结果内在的主导结构或模式。以多维排列(Multidimensional Scaling, MDS)为代表的几何相似性模型<sup>[54; 55; 100-102; 113]</sup>已经成为心理学领域用于探索相似性判别结果中内在结构或模式的主要方法之一<sup>[110]</sup>。近年来, MDS 也被应用到图像检索中。例如, Santini 等人以 MDS 方法为核心根据用户反馈来优化相似性度量<sup>[92]</sup>; Mojsilovic 等人利用 MDS 方法获取不同的人对图像内容相似性判别之间的内在关系,并利用这种关系去描述图像语义<sup>[76]</sup>。

MDS 的目标是确定一个低维的距离空间,并在该空间中寻求一组点,得使这些点与点之间的几何距离尽可能地与其所代表的对象之间的相异性值相近。与线性判别分析(Linear Discriminant Analysis, LDA)不同的是, MDS 是一种非监督学习算法,没有利用不同类样本之间的可分性信息。它们都是多元线性分析算法,对于非线性数据,它们都无能为力。近年来,随着非线性分析方法的引入,以 MDS、LDA 和 PCA (Principle Component Analysis) 为代表的多元线性分析方法又成为许多以统计分析为主要手段的研究领域的热点问题之一,例如人工智能、模式识别等。最典型的算法有:(1)流形学习算法,如等距映射(Isometric mapping, Isomap)<sup>[110]</sup>、局部线性嵌入(Locally Linear Embedding, LLE)<sup>[88; 96]</sup>、拉普拉斯特征映射(Laplacian Eigenmaps, LE)<sup>[4]</sup>等;(2)核方法,如核主成分分析(Kernel PCA)<sup>[97]</sup>等;(3)主曲线和主曲面<sup>[36; 37; 97]</sup>。这些算法的一个共同特点是它们都属于非监督学习算法。因此,它们与经典的 MDS 和 PCA 一样可以用于探索和理解数据内在低维的流形结构。然而,它们并没有考虑到如何利用已知的观察结果去辅助探索和理解数据的内在结构。在这方面已经存在一些研究成果,如局部保持投影(Locality Preserving Project, LPP)<sup>[41]</sup>、LPE<sup>[75]</sup>和监督等距

映射 (Supervised Isometric Mapping, SIsomap)<sup>[19]</sup>等。

本章在分析经典流形学习算法的基础上,提出基于监督学习的非线性判别排列 (Nonlinear Discriminant Scaling, NDS) 算法。基于模拟数据和 Corel 图像数据库的实验结果表明,本章提出的算法可以获取非线性数据集内的可分信息。

## 4.1 经典流形学习算法

等距映射 (Isomap) 和局部线性嵌入 (LLE) 是两个相关而又不同的流形学习算法。它们都是学习和探索高维数据内在结构的非监督学习算法,前者的目标是在低维空间中保持高维数据的全局非线性结构,后者的目标是在低维空间中保持数据之间的局部线性结构。本节简要地介绍这两个经典流形学习算法,并将在下节讨论本章提出的一个基于 Isomap 基本框架的监督学习算法,即非线性判别排列。

### 4.1.1 等距映射算法

经典MDS的目标函数为中心化的平方距离矩阵,即

$$J(\bar{y}^*)_{MDS} = \arg \max_{\bar{y}} \frac{\bar{y}^T (-\frac{1}{2} \mathbf{L}_g \mathbf{D} \mathbf{L}_g) \bar{y}}{\bar{y}^T \bar{y}} \quad (4-1)$$

Isomap的目标函数为中心化的平方测地距离矩阵,即

$$J(\bar{y}^*)_{Isomap} = \arg \max_{\bar{y}} \frac{\bar{y}^T (-\frac{1}{2} \mathbf{L}_g \mathbf{G} \mathbf{L}_g) \bar{y}}{\bar{y}^T \bar{y}} \quad (4-2)$$

其中,  $\bar{y}$  是排列后获得的点的坐标矩阵;  $\mathbf{G}$  和  $\mathbf{D}$  分别为原始数据的测地距离和欧氏距离;  $\mathbf{L}_g$  是第三章中所定义的全局拉普拉斯矩阵。

比较经典 MDS 和 Isomap 的目标函数,不难发现:本质上, Isomap 的目标与经典 MDS 一样,是在低维欧氏空间中寻求一组点,使得这些点与点之间的欧氏距离能够反映输入空间中对象之间的相异性(即距离);不同的是 Isomap 输入空间中对象之间的距离不再是欧氏距离,而是基于数据内在邻近关系的测地距离,其中测地距离是指输入空间中由数据集所构筑的邻近图上点与点之间的最短路径。该距离不再仅仅取决于两个对象的特征之间的关系,而是取决于整个特征集的

分布状态。

Isomap 算法的基本步骤如下：

(1) 计算数据之间的欧氏距离，并根据欧氏距离构筑邻近图。有两种构筑邻近图的办法，一个是选择一个最短的距离  $\varepsilon$ ，为每对距离小于  $\varepsilon$  的点连接一条边，边的权重为它们之间的距离；另一种方法，对于任意给定的点，连接与它距离最近的  $K$  个点，其中， $K$  是事先选定的一个常数。

(2) 估算测地距离，即邻近图上两点之间的最短距离。

(3) 根据测地距离，利用经典 MDS 算法获得降维后每个点的坐标。

### 4.2.1 局部线性嵌入

LLE 的目标是通过数据的局部拓扑关系发现数据的局部线性结构，该算法的基本流程是：

(1) 构筑邻近图，比如  $k$  邻近图：

(2) 对于任意一个点  $x_i$  与其  $k$  个临近点  $x_j$  ( $j \in [1, k]$ )，计算权重  $w_{ij}$ ，使得利用  $k$  个临近点  $x_j$  重构点  $x_i$  的误差最少，即极小化函数

$$\varepsilon(\mathbf{w}) = \sum_i \left| x_i - \sum_{j=1}^k w_{ij} x_j \right|^2 \quad (4-3)$$

其中  $\sum_{j=1}^k w_{ij} = 1$ 。

(3) 在低维欧氏空间中计算新的点  $\tilde{y} = \{y_i\}$ ，极小化下式

$$J(\tilde{y}) = \sum_i \left| y_i - \sum_{j=1}^k w_{ij} y_j \right|^2 \quad (4-4)$$

其中，权重  $w_{ij}$  是通过公式 (4-3) 计算的结果。

从公式 (4-3) 和 (4-4) 可以看出，它们之间的联系纽带就是每个样本利用邻近点重构其自身时的  $k$  个权重。正是通过这些局部的重构权重来实现变换前后数据之间的局部线性关系。

## 4.2 非线性判别排列

He 等人从拉普拉斯算子的角度分析了 PCA、LDA 和 LPP 之间的区别和联系<sup>[41]</sup>。本节在前章内容的基础上，通过分析 Isomap 和 LLE，进一步探讨在非线性





状态下, 判别分析与全局和局部拉普拉斯算子之间的联系。在分析之前, 先引用一个结论: 经典 MDS 的结果实际上等同于 PCA 的变换结果<sup>[18]</sup>。根据前节的分析可知, Isomap 是 MDS 的非线性泛化的结果。因此, Isomap 实际上也可以看成是非线性的 PCA。正是受这个观点的启发, 本章提出了非线性判别排列算法。该算法有效地将拉普拉斯线性判别分析 (Laplacian Linear Discriminant Analysis, LLDA) 的思想嵌入到 Isomap 算法的框架中, 是 LLDA 非线性泛化的结果。下面从 PCA 与经典 MDS 算法之间的关系来理解 Isomap 的目标函数, 进而导出非线性判别排列算法。

### 4.2.1 多维排列与主成分分析之间的对偶关系

广义地说, PCA 与 MDS 都是一种变换, 不失一般性, 设为

$$\phi: \mathfrak{R}^N \rightarrow \mathfrak{R}^n, \bar{x} \mapsto \bar{y} \quad (4-5)$$

其中, PCA 的核心任务是确定变换函数  $\phi(\bar{x})$ ; 然而, MDS 最关心的是变换后的结果  $\bar{y}$ , 并不需要直接求取变换函数。

#### (一) 主成分分析

众所周知, 为了确定 PCA 所需要的变换函数, 首先要求取样本的协方差矩阵。假设存在一组中心化样本集  $\mathbf{X} = \{\bar{x}_1, \dots, \bar{x}_M\}$ , 其中, 包括  $M$  个  $N$  维的样本, 样本集的均值向量需足

$$\begin{aligned} \bar{m}_0 &= \frac{1}{M} \sum_{i=1}^M \bar{x}_i \\ &= 0 \end{aligned} \quad (4-6)$$

假设协方差矩阵  $\mathbf{C} = \frac{1}{M} \mathbf{X} \mathbf{X}^T$  的特征值和特征向量分别为  $\Lambda = \text{diag}(\lambda_i)$  和  $\mathbf{V} = \{v_i\}$

( $i \in [1, N]$ ), 则有

$$\mathbf{C} \mathbf{V} = \mathbf{V} \Lambda \quad (4-7)$$

PCA 变换后的结果为

$$\mathbf{Y} = \mathbf{X}^T \mathbf{V} \quad (4-8)$$

假设矩阵  $\mathbf{C}_g = \frac{1}{M} \mathbf{X}^T \mathbf{X}$  的特征值和特征向量分别为  $\Lambda_g$  和  $\mathbf{U}$ , 则公式 (4-7) 左乘

矩阵  $\mathbf{X}^T$  得:

$$\begin{aligned}
 \mathbf{X}^T \mathbf{C} \mathbf{V} &= \mathbf{X}^T \mathbf{V} \Lambda \\
 \Rightarrow \left( \frac{1}{M} \mathbf{X}^T \mathbf{X} \right) (\mathbf{X}^T \mathbf{V}) &= (\mathbf{X}^T \mathbf{V}) \Lambda \\
 \Rightarrow \mathbf{C}_g (\mathbf{X}^T \mathbf{V}) &= (\mathbf{X}^T \mathbf{V}) \Lambda
 \end{aligned} \tag{4-9}$$

因为矩阵  $\mathbf{C}$  和  $\mathbf{C}_g$  具有相同的特征值，因此有  $\Lambda_g = \Lambda$  和  $\mathbf{U} = \mathbf{X}^T \mathbf{V}$ 。从上式可以看到，如果不降维的话，PCA 变换后的矩阵  $\mathbf{Y} = \mathbf{U}$ 。

## (二) 经典多维排列

给定一组中心化样本  $\mathbf{X} = \{\bar{x}_1, \dots, \bar{x}_M\}$ ，假设它们之间平方欧氏距离矩阵为  $\mathbf{D} = \{d_{i,j}\}$ ，则任意两点间的距离平方为

$$\begin{aligned}
 d_{i,j} &= \sum_{k=1}^n (\bar{x}_{i,k} - \bar{x}_{j,k})^2 \\
 &= \bar{x}_i^T \bar{x}_i - \bar{x}_i^T \bar{x}_j - \bar{x}_j^T \bar{x}_i + \bar{x}_j^T \bar{x}_j
 \end{aligned} \tag{4-10}$$

结合公式 (4-6)，对上式所有的  $i$  求和得：

$$\begin{aligned}
 \sum_{i=1}^M d_{i,j} &= \sum_{i=1}^M (\bar{x}_i^T \bar{x}_i) - \left( \sum_{i=1}^M \bar{x}_i^T \right) \bar{x}_j - \bar{x}_j^T \left( \sum_{i=1}^M \bar{x}_i \right) + M \bar{x}_j^T \bar{x}_j \\
 &= \sum_{i=1}^M (\bar{x}_i^T \bar{x}_i) + \bar{m}_0^T \bar{x}_j + \bar{x}_j^T \bar{m}_0 + M \bar{x}_j^T \bar{x}_j \\
 &= \sum_{i=1}^M (\bar{x}_i^T \bar{x}_i) + 0 + 0 + M \bar{x}_j^T \bar{x}_j \\
 &= \sum_{i=1}^M (\bar{x}_i^T \bar{x}_i) + M \bar{x}_j^T \bar{x}_j
 \end{aligned} \tag{4-11}$$

同理，可得

$$\begin{aligned}
 \sum_{j=1}^M d_{i,j} &= M \bar{x}_i^T \bar{x}_i - \bar{x}_i^T \sum_{j=1}^M \bar{x}_j - \left( \sum_{j=1}^M \bar{x}_j^T \right) \bar{x}_i + \sum_{j=1}^M (\bar{x}_j^T \bar{x}_j) \\
 &= M \bar{x}_i^T \bar{x}_i + \sum_{j=1}^M (\bar{x}_j^T \bar{x}_j)
 \end{aligned} \tag{4-12}$$

$$\begin{aligned}
 \sum_{i=1}^M \sum_{j=1}^M d_{i,j} &= M \sum_{i=1}^M \bar{x}_i^T \bar{x}_i - 2 \sum_{i=1}^M \bar{x}_i^T \sum_{j=1}^M \bar{x}_j + M \sum_{j=1}^M \bar{x}_j^T \bar{x}_j \\
 &= 2M \sum_{i=1}^M \bar{x}_i^T \bar{x}_i
 \end{aligned} \tag{4-13}$$

借助中心化算子  $\mathbf{L}_g$ （它的定义与第三章的全局拉普拉斯矩阵相同），可使距离平方矩阵  $\mathbf{D}$  中任意行或列上元素的和为零（double-centering），即：

$$\begin{aligned}
 -\frac{1}{2} \mathbf{L}_g \mathbf{D} \mathbf{L}_g^T &= -\frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M (d_{ij} - \frac{1}{M} \sum_{i=1}^M d_{ij} - \frac{1}{M} \sum_{j=1}^M d_{ij} + \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M d_{ij}) \\
 &= \sum_{i=1}^M \sum_{j=1}^M \tilde{x}_i^T \tilde{x}_j \\
 &= \mathbf{X}^T \mathbf{X}
 \end{aligned} \tag{4-14}$$

根据公式 (4-9) 和 (4-14), 在变换前后没有误差的假设下, 下式成立:

$$\begin{aligned}
 -\frac{1}{2} \mathbf{L}_g \mathbf{D} \mathbf{L}_g^T &= \mathbf{X}^T \mathbf{X} \\
 &= \mathbf{M} \mathbf{C}_g
 \end{aligned} \tag{4-15}$$

矩阵  $\mathbf{C}_g$  的特征向量为 PCA 变换后的结果, 即  $\mathbf{U} = \mathbf{X}^T \mathbf{V}$ 。换言之, 经典 MDS 与 PCA 的变换结果在不降维 (或相同的重构精度) 的假设下是一致的。

### (三) 结论及其推广

比较式 (4-9) 和 (4-16), 可以得到下面的结论: 在变换前后没有误差 (或相同误差) 的假设下, 经典 MDS 排列结果与 PCA 变换后的结果是完全相同的。在实际应用中, 无论是 PCA 还是 MDS, 变换后空间的维数都要低于或远远低于原始空间的维数。对于 PCA 而言, 维数的选择一般都取决于数据协方差矩阵的特征值。而 MDS 的维数常常取决于变换前后数据间距离矩阵之间的冗差, 而这种冗差的一个明显标志就是中心化后的距离矩阵 (即  $-\frac{1}{2} \mathbf{L}_g \mathbf{D} \mathbf{L}_g^T$ ) 的特征值大小。

从式 (4-1) 和 (4-2) 可以看出: Isomap 不过是用数据间的平方测地距离矩阵  $\mathbf{G}$  替代平方欧氏距离矩阵  $\mathbf{D}$ , 从而将经典的 MDS 算法从线性应用领域泛化到非线性应用领域。根据上述有关 PCA 与经典 MDS 算法在变换结果上的同一性, Isomap 也可以被看成是 PCA 算法的非线性泛化。正是基于这一认识, 提出基于 Isomap 框架的非线性 LLDA 算法, 即非线性判别排列算法。

## 4.2.2 非线性判别排列算法的准则函数

假设存在  $M$  个  $N$  维的训练样本  $\mathbf{X} = \{\bar{x}_1, \dots, \bar{x}_M\}$ , 样本中存在  $L$  个类别, 每个类别中的样本数为  $l_i$  ( $i=1, 2, \dots, L$ )。第三章中定义的拉普拉斯线性判别分析的准则函数为

$$J(\bar{\mathbf{v}}^*)_{LLDA} = \arg \max \frac{\bar{\mathbf{v}}^T \mathbf{S} \bar{\mathbf{v}}}{\bar{\mathbf{v}}^T \bar{\mathbf{v}}} \tag{4-16}$$

其中,

$$S = \frac{1}{M} X(L_g^{1/M} - 2L_l)X^T \quad (4-17)$$

且  $L_l$  满足

$$L_l = \begin{bmatrix} L_g^{1/l_1} & & & \\ & L_g^{1/l_2} & & \\ & & \ddots & \\ & & & L_g^{1/l_L} \end{bmatrix} \quad (4-18)$$

样本集  $X$  的协方差矩阵  $C$  可以表示为:

$$\begin{aligned} C &= \frac{1}{M} \sum_{i=1}^M (\bar{X}_i - \bar{m}_0)(\bar{X}_i - \bar{m}_0)^T \\ &= \frac{1}{M} XX^T - \frac{1}{M^2} (Xe)(Xe)^T \\ &= \frac{1}{M} X(I - \frac{1}{M} ee^T)X^T \\ &= \frac{1}{M} XL_gX^T \end{aligned} \quad (4-19)$$

其中,  $\bar{m}_0$  是所有样本的均值向量。

同理, 对于  $M$  个训练样本  $X$  中所有第  $i$  类  $l_i$  个样本集  $X_i$  的协方差矩阵  $C_i$  可以表示为

$$C_i = \frac{1}{M} X_i L_g^{1/l_i} X_i^T \quad (4-20)$$

设矩阵  $C^L$  为训练样本  $X$  中  $L$  个类的类内协方差的直和, 即

$$C^L = \begin{bmatrix} C_1 & & & \\ & C_2 & & \\ & & \ddots & \\ & & & C_L \end{bmatrix} \quad (4-21)$$

则公式 (4-17) 可以重写成

$$S = C - 2 \times C^L \quad (4-22)$$

上式表明, 拉普拉斯线性判别分析的目标函数是训练样本的总体协方差与 2 倍的类内协方差矩阵之间的差。此外, 由上述分析可知, 协方差矩阵与经典 MDS 在结果上存在对偶关系。因此, 非线性判别排列(Nonlinear Discriminant Scaling, NDS)的准则函数定义为

$$J(\tilde{\mathbf{v}}^*)_{NDS} = \arg \max \frac{\tilde{\mathbf{v}}^T \tilde{\mathbf{S}} \tilde{\mathbf{v}}}{\tilde{\mathbf{v}}^T \tilde{\mathbf{v}}} \quad (4-23)$$

其中, 矩阵  $\tilde{\mathbf{S}}$  被定义为

$$\begin{aligned} \tilde{\mathbf{S}}_g &= \frac{1}{M} \left\{ -\frac{1}{2} \mathbf{L}_g \mathbf{G} \mathbf{L}_g^T - 2 \left( -\frac{1}{2} \mathbf{L}_l \mathbf{G} \mathbf{L}_l^T \right) \right\} \\ &= -\frac{1}{2M} (\mathbf{L}_g \mathbf{G} \mathbf{L}_g^T - 2 \mathbf{L}_l \mathbf{G} \mathbf{L}_l^T) \end{aligned} \quad (4-24)$$

式中矩阵  $\mathbf{G}$  是数据集的平方测地距离矩阵。

### 4.2.3 非线性判别排列算法

根据上面的分析可知, 非线性判别排列沿用了 Isomap 泛化经典 MDS 算法的思路。因此, 非线性判别排列算法是基于 Isomap 的基本框架, 它的基本步骤如下:

- (1) 计算数据之间的欧氏距离, 并根据欧氏距离构筑邻近图。有两种构筑邻近图的办法, 一种方法是选择一个距离  $\varepsilon$ , 为每对距离小于  $\varepsilon$  的点连接一条边, 边的权重为它们之间的距离; 另一种方法, 对于任意给定的点, 连接与它距离最近的  $K$  个点, 其中,  $K$  是事先选定的一个常数。在我们的实验中  $k=8$ 。
- (2) 计算测地距离, 即邻近图上两点之间的最短路径;
- (3) 根据训练数据的类别信息, 计算全局和局部拉普拉斯矩阵, 即  $\mathbf{L}_g$  和  $\mathbf{L}_l$ 。
- (4) 根据公式 (4-24) 计算目标矩阵  $\tilde{\mathbf{S}}_g$ ;
- (5) 计算目标矩阵  $\tilde{\mathbf{S}}_g$  的特征值和特征向量;
- (6) 选择较大的特征值对应的特征向量作为原始数据在低维空间中的坐标;

## 4.3 实验

为了从不同角度分析 NDS 算法的特点, 本章做了三组实验, 实验数据包括:

- (1) 两层 Swiss-Roll 数据;
- (3) Corel 图像库中两类动物图像 (狐狸和狼);
- (4) 交互式图像检索中的正负反馈图像。

### 4.3.1 两层 Swiss Roll 数据

两层 Swiss-Roll 数据是一个典型的两类非线性数据集, 每个 Roll 被看成一个

类，由 500 个样本点组成（如图 4-1 所示）。

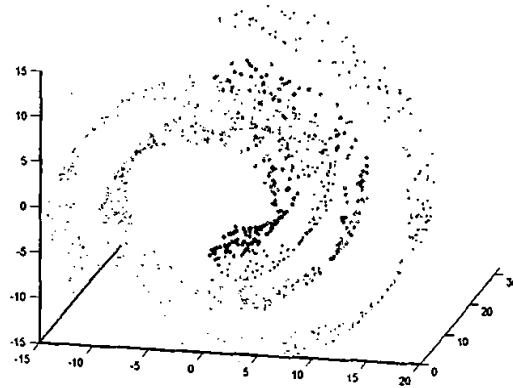
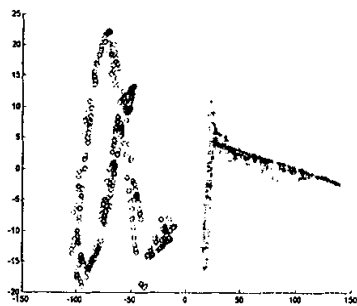
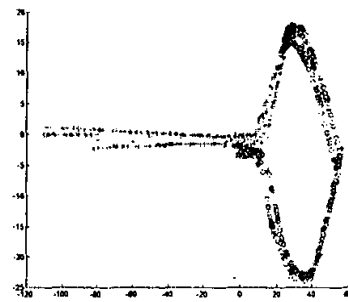


图 4-1 两层 Swiss-Roll 数据

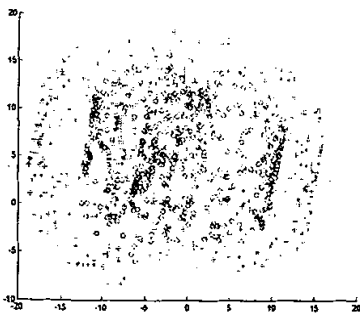
Figure 4-4 Two-layer Swiss-Roll data set



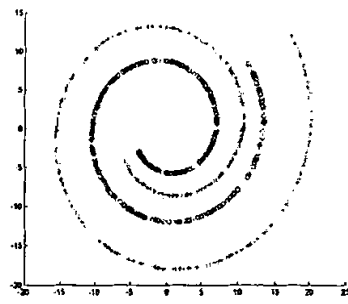
(a) NDS



(b) Isomap



(c) LDA



(d) MDS

图 4-2 两层 Swiss-Roll 数据二维排列结果

Figure 4-2 Two dimensional scaling of two-layer swiss roll data set

图 4-2 是 Swiss-Roll 数据降维后的结果。从中可以看出，只有 NDS 的结果是完全线性可分的。对于线性分类器，Isomap 结果将不可避免地存在一定的误识别率。然而，PCA 和 LDA 的结果是线性不可分的。该实验结果表明：对于非线性数据集，NDS 算法的降维结果具有较强的可分性。

### 4.3.2 两类图像

为了测试 NDS 对 Corel 图像之间的非线性可分性，从 Corel 图像库中选择了如图 4-3 所示的两类动物图像（狐狸和狼，每类 100 幅图像），从视觉特征而言，它们是非常相似，但却属于不同类别的图像。

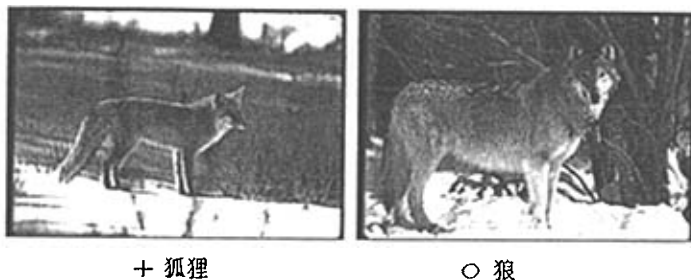


图 4-3 实验的两类动物图像

Figure 4-3 Two kinds of animal Images in experiments

各种算法对这两类图像的排列结果如图 4-4 所示，其中十字架表示狐狸，圆圈表示狼。从图 4-4 (b) 和 (d) 可以看到，无论是欧氏距离还是测地距离，这两类图像的排列高度重叠，完全不可分。原因在于：(1) 它们是非监督学习算法，类别信息并没有被嵌入到这些算法中；(2) 它们的目标不是获取易于分类的结果，而是尽可能地保持数据之间的距离（相异性），因此，从结果可以看到，这两类图像本身非常相似，任何线性分类器都难于区分他们。

LDA 的排列结果具有一定的可分性，但是两类图像之间的交叉仍然非常严重。尽管 NDS 的排列结果也存在一些重叠，但是类与类之间的可分性非常明显，很容易找到一个线性分类器以很小的错误率就可以将它们分开。

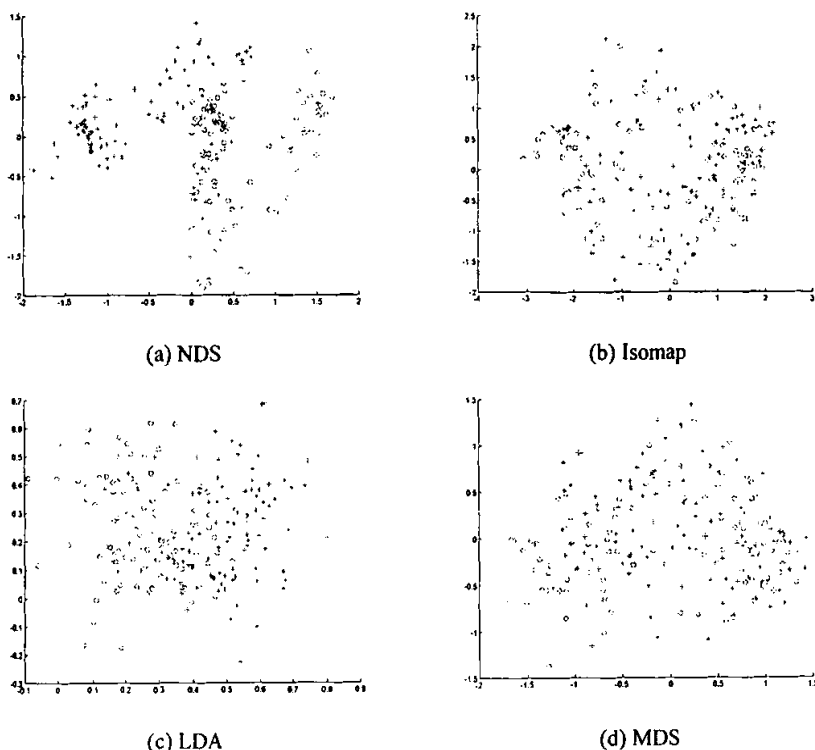


图 4-4 两类动物图像的排列结果

Figure 4-4 Scaling of two kinds of animal images

### 4.3.3 交互式图像检索中的非线性判别排列

NDS 和 MDS 都是探索性 (exploratory) 数据分析工具, 它们的目标是在低维的线性空间中可视化数据集的某些分布特征。然而, 这些分布特征并不能用类似于 PCA 或 LDA 的变换矩阵来表达。对数据分布特征的解释需要人为地参与。尽管 Santini 等人提出利用 MDS 优化相似性度量模型的方法<sup>[92]</sup>。该方法假设用户可以根据反馈图像之间的相似程度在平面上排列图像, 使得图像之间相异性与它们在平面上的距离相吻合。然后, 优化相似性度量模型, 使得 MDS 的排列结果与用户的排列结果相吻合。尽管这是很具有新意的想法, 但是却一件很难实现的事情。一方面, 人为按上述方式地排列几十幅图像本身是一件非常困难的事情; 另一方面, MDS 本身并不是建模工具, 而是一种探索性分析工具, 任何两幅图像之间细微的相异性变化可能会导致完全不同的排列结果。因此, 根据两种 MDS 排列结果的吻合程度优化得出的相似性度量模型也一样不稳定。相反, NDS 和 MDS 更适合于作为交互式图像检索中辅助的可视化分析工具, 例如, 用户可以



将它们作为探索图像库中某些图像之间的相互关系一种可视化工具。下面, 将图 4-3 所示的两幅图像作为查询图像, 分别用 NDS 和 MDS 来可视化查询过程中反馈图像之间的分布。其中, 反馈中的加权方法为 Rui 等人的逆方差加权算法; 图中圆圈表示与查询图像同类的图像 (正样本), 方框表示不同于查询图像的图像 (负样本), 十字架表示是该轮反馈中最新出现的图像。

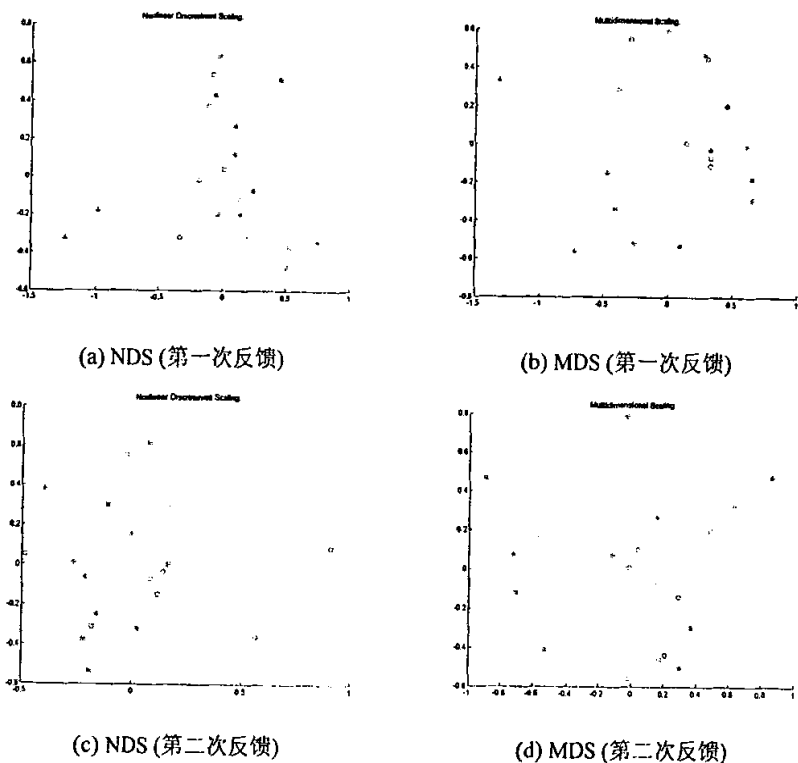


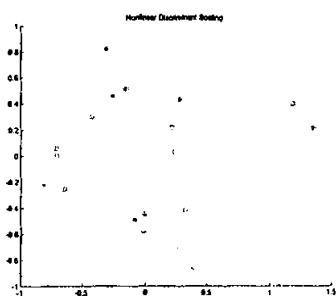
图 4-5 狐狸图像作为查询模板时的排列结果

Figure 4-5 Scaling of images when the fox image acts as a querying image

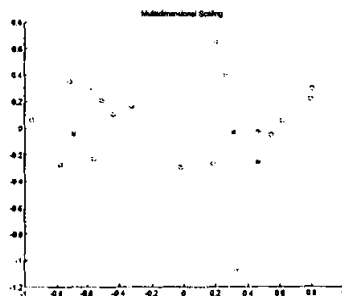
图 4-5 (a) 表明尽管负样很少, 但正负样本之间的分界线较明显, 即易于分类; 同时, 图 4-5 (b) 显示负样散落在正样本的周围。这两点表明通过加权较难获得新的同类图像。图 4-5 (c) 和 (d) 验证了这一点。

图 4-6 展示了完全不同于图 4-5 的情况。第一次反馈中的正样本较多, 虽然 NDS 所示的两类分布较清楚, 但是没有线性分类器可以被它们完全分开, 此外, MDS 所示的正样并不是聚集在一起, 而是成长条状分布, 这表明该图像可能处于某种低维的非线性流形上, 可能通过选择合适的加权方法来逼近这种分布。第二次反馈中出现不少新的正样本。此时, NDS 排列结果表明, 正负样分类易于

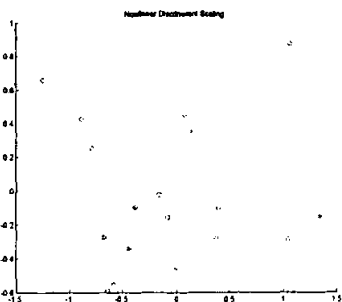
进行, 但 MDS 排列没有明显的规律可循。第三次反馈虽然仍有新的正样本出现, 但数量已经非常少。因此, 再继续反馈的意义已经不大。



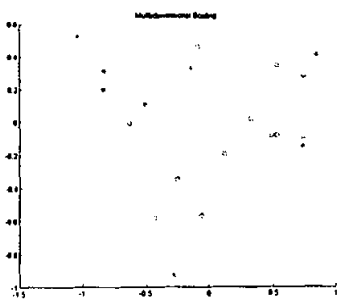
(a) NDS (第一次反馈)



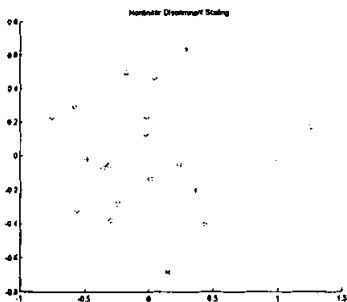
(b) MDS (第一次反馈)



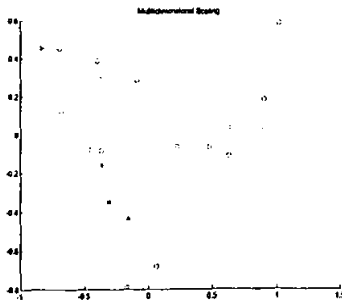
(c) NDS (第二次反馈)



(d) MDS (第二次反馈)



(e) NDS (第三次反馈)



(f) MDS (第三次反馈)

图 4-6 狼图像作为查询模板时的排列结果

Figure Scaling of images when the wolf image acts as a querying image

## 4.4 本章小结

在分析经典 MDS 与 PCA 之间对偶关系的基础上, 本章提出了非线性判别排列算法。该算法将拉普拉斯线性判别思想嵌入到 Isomap 的基本框架之下, 将



LLDA 从线性泛化到非线性应用领域, 将 Isomap 从非监督学习转化为监督学习算法。基于模拟数据和图像数据库的实验结果表明, 本章提出的算法可以从非线性数据集中获得易于分类的排列结果。

## 第五章 图像特征比率模型

特征相似性理论是心理学领域中重要的相似性理论之一。在深入分析和批判几何相似性理论的基础上, Tversky 提出了不同于距离公理的相似性原则, 并在此基础上提出特征对比模型 (Feature Contrast Model, FCM)<sup>[114]</sup>。该理论可以解释许多几何相似性模型所不能解释的人类相似性判别现象, 如相似性不对称性和非常量的自相似性等。由于人类相似性判别时所使用的特征往往是定性的, 例如, 关于两个人的身高, 一般说 A 比 B 高 (或矮、差不多一样高)。因此, 特征对比模型采用的特征是二值的, 它更适用于度量两个对象关于定性特征的相似程度。尽管定量特征通过二值化处理也可以被应用到特征对比模型中, 但是这种处理是极其复杂的<sup>[115]</sup>。图像特征包括语义特征和视觉特征。然而, 由于语义特征本身的复杂性, 图像之间的相似性通常使用视觉特征来度量。Santini 等人利用模糊逻辑扩展了 FCM 的应用范围, 提出了模糊特征对比模型 (Fuzzy Feature Contrast Model, FFCM)<sup>[94]</sup>。此外, Tolas 等人提出了另一个类似的模型, 即 Tversky Index-Min (TIM)<sup>[112]</sup>, 并被应用到图像检索中<sup>[110, 111]</sup>。根据 Tversky 提出的特征论断性原理, 特征在相似性度量中的重要性取决于两个因素: 特征强度和特征诊断性。然而, 在图像视觉特征相似性度量时, FFCM 和 TIM 均没有考虑特征论断性在相似性度量的作用, 从而导致图像的特征强度和相似度量值成强相关性。此外, 这两类模型均只被应用于图像的视觉特征, 并没有考虑到如何集成视觉特征与语义特征。为了解决上述两个问题, 本章提出了图像特征比率模型 (Image Feature Ratio Model, IFRM)。

### 5.1 特征对比模型在图像检索中的应用

本节简述两个特征对比模型的改进模型, 并分析它们在图像检索中存在的几个问题。

#### 5.1.1 模糊特征对比模型

特征相似性理论中的特征是二值的, 表示研究对象是否拥有该属性。这种方式非常适合表达定性的语义特征。然而, 图像视觉特征通常是定量的, 一般用连

续区间的某个数来表示图像该属性的测量值,例如颜色,一般是用 0 到 255 之间的数值来表示,只有简单的黑白图像才使用 0 或 1 这种二值模式来表示。因此,彩色图像的视觉特征一般不适宜用二值方式来表达。在特征对比模型中,为了能利用对象的连续定量特征,需要将这些定量特征转化为一系列的二值特征。例如,在 FCM 中,长度  $l$  将被划分为一个长度序列,如  $l_1 < l_2 < \dots < l_n$ ,从而将其表示成一组特征集  $A_1 \subset A_2 \subset \dots \subset A_n$ 。很显然,在图像相似性度量中,这种特征表示模式既不实用也会引入一些新的问题。为了解决这些问题, Santini 等人将模糊逻辑引入到 FCM 中,从而将 FCM 扩展到更宽的应用领域,即对象特征用二值特征形式表示可能存在问题或者根本不可能的领域<sup>[94]</sup>,例如描述图像的视觉特征。

定义 5-1: 假设  $\Omega$  是一个集合,  $\phi: \Omega \rightarrow R^m$  是集合  $\Omega$  中元素的测量值,  $P\omega$  是集合  $\Omega$  中元素  $\omega$  的一个判别值,则  $P\omega$  为真值 (或隶属度)

$$T(P\omega) = \mu(\phi(\omega)) \quad (5-1)$$

其中,  $\mu: R^m \rightarrow [0, 1]$ 。

在 FCM 中,一句话“那人的嘴很宽”只能用两个值来表示,即“是”或“不是”。在引入定义 5-1 中的模糊逻辑以后,那句话的判别结果将取决于一些物理量的测量结果,如嘴的宽度或整个脸的宽度等。所有测量结果可以表示成下面的一个向量:

$$\mu(\phi) = \{\mu_1(\phi), \mu_2(\phi), \dots, \mu_N(\phi)\}, \quad (5-2)$$

其中,  $\mu_i(\phi)$  ( $0 < i < N$ ) 是第  $i$  个物理量测量  $\phi_i$  的模糊预测结果,它表示就这个判别而言,对象属于模糊集  $\mu(\phi)$  的程度。

在模糊特征对比模型中, Santini 等人采用了与特征对比模型相同的特征显著性函数  $f(x)$ :

$$f(x) = \sum_{i=1}^N \mu_i(x) \quad (5-3)$$

即对象之间的相似性是对象间每个特征之间相互关系的总和。

模糊集合  $A$  和  $B$  之差的传统定义为

$$A - B = \min\{\mu_A, 1 - \mu_B\} \quad (5-4)$$

根据上式的定义,模糊集自身之差为

$$\begin{aligned} A - A &= \min\{\mu_A, 1 - \mu_A\} \\ &= \begin{cases} \mu_A, & \text{if } \mu_A < 0.5 \\ 0, & \text{if } \mu_A = 0.5 \\ 1 - \mu_A, & \text{if } \mu_A > 0.5 \end{cases} \end{aligned} \quad (5-5)$$

从上式可以看到, 传统模糊集之差的定义不能保证模糊集与其自身的差总为零 (即空集)。为了解决这个问题, Santini 等人给出了模糊集之差的另一种定义:

$$A - B = \max\{\mu_A - \mu_B, 0\} \quad (5-6)$$

假设图像数据库中第  $i$  幅图像和查询图像的视觉特征集分别为  $V_i = \{v_1^i, v_2^i, \dots, v_N^i\}$  和  $V_q = \{v_1^q, v_2^q, \dots, v_N^q\}$ , 则模糊特征对比模型 (Fuzzy Feature Contrast Model, FFCM) 可以定义为:

$$\begin{aligned} FFCM(V_i, V_q) &= f(\theta(u(V_i) \cap u(V_q)) - \alpha(u(V_i) - u(V_q)) - \beta(u(V_q) - u(V_i))) \\ &= \theta \sum_{k=1}^N \min\{\mu_k(v_k^i), \mu_k(v_k^q)\} \\ &\quad - \alpha \sum_{k=1}^N \max\{\mu_k(v_k^i) - \mu_k(v_k^q), 0\} \\ &\quad - \beta \sum_{k=1}^N \max\{\mu_k(v_k^q) - \mu_k(v_k^i), 0\} \end{aligned} \quad (5-7)$$

其中,  $\theta$ 、 $\alpha$  和  $\beta$  是正常数。

### 5.1.2 广义 Tversky 系数

Tolias 等人改进了 Tversky 的另一个特征匹配函数:

$$GTI(A, B) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \quad (5-8)$$

使用公式 (5-3) 作为特征显著性函数和传统的集合定义, 从而得到一个新的模型, 即 Tversky Index-Min (TIM) <sup>[112]</sup>:

$$\begin{aligned} TIM(A, B) &= \\ &= \frac{\sum_{k=1}^N \min\{\mu_k(v_k^a), \mu_k(v_k^b)\}}{\sum_{k=1}^N \min\{\mu_k(v_k^a), \mu_k(v_k^b)\} + \alpha \sum_{k=1}^N \min\{\mu_k(v_k^a), 1 - \mu_k(v_k^b)\} + \beta \sum_{k=1}^N \min\{1 - \mu_k(v_k^a), \mu_k(v_k^b)\}} \end{aligned} \quad (5-9)$$

Tolias 等人提出的 TIM 沿用了 FFCM 中的模糊逻辑, 不同之处包括: (1) 它是 Tversky 比率模型的泛化。(2) 使用了不同的集合差定义 (即公式 5-4)。FFCM 集合差的定义如公式 (5-6) 所示。在后面的章节将结合模拟实验进一步分析这两种定义对相似性度量的影响。一方面, TIM 沿用了模糊特征对比模型主要思想。另一方面, 无论是本章的模拟实验还是图像检索实验均表明: 该模型并不具备模糊特征对比模型的许多优点。因此本章主要讨论模糊特征对比模型。

### 5.1.3 模糊特征对比模型在图像检索中存在的主要问题

图像相似性度量包括两个重要的元素：(1) 特征空间；(2) 相似性度量模型<sup>[94]</sup>。其中特征包括语义特征和视觉特征，它们之间的最大差别在于：语义特征是抽象的、定性的，而且可能随着环境的改变而改变；然而，视觉特征是相对客观的，可以定量地描述。狭义的图像特征一般是指相对客观的、定量的视觉特征。模糊特征对比模型利用模糊逻辑将特征对比模型扩展到可以度量图像视觉特征之间的相似性。这是否意味着模糊特征对比模型自然而然地继承特征对比模型适合度量二值的、定性的特征（特别是语义特征）相似性的优势呢？答案是否定的。因为图像的视觉特征是按多维结构组织的，特征向量是事先设定的多维空间中的点；而二值的语义特征是按集合结构组织的，集合中的元素是可变的，换言之，不可能事先设定图像语义特征的类型或数量，因为语义特征是随着环境的改变而改变的。如果将这两类特征不加区别地组合在一起，并利用模糊特征对比模型去度量图像之间的相似性，那么必将导致两类特征因为其数量的多少而主导整个相似性度量的结果。例如，当视觉特征的维数很高时，语义特征在整个相似性度量中的作用会非常小。在交互式图像检索中，为了保证图像特征具有更大的表现力，一般会选择多种类型的视觉特征（如主色描述子、颜色结构描述子、同质纹理描述子等）。因此，视觉特征的维数一般会很高。此外，对于检索返回的相似图像，从视觉特征而言，它们已经很相似。此时，区别这些图像更为有效特征是语义特征，也就是说应该让语义特征在整个相似性度量起更大的作用。模糊特征对比模型并没有提供这样一个机制，以适应图像检索的需要。

根据 Tversky 提出的特征诊断性原理 (Diagnostic Principle)<sup>[114]</sup>，特征在相似性度量中的作用由两个因子所决定：(1) 特征强度，即对象的物理测量值；(2) 特征诊断性，即对象分类时的重要性。在 FCM 中，特征诊断性是通过改变对象的特征集合的大小来实现。然而，在多维特征结构中，特征的类型是预先设定好的，可变只是每个特征的强度。因此，在 FFCM 中，由于多维结构的存在，FCM 的特征诊断性思想不可能再存在。从而出现这样一种情况：相对于同一个参考对象，两个对象的特征强度与参考对象具有相同的差别，但是具有较大特征强度的对象将总是会被判别为更相似于参考对象。

为了克服上述两个问题，本章提出了图像特征比率模型。

## 5.2 图像特征比率模型

图像内容是三维现实世界的二维映射。因此,图像之间关于其内容的相似性度量也是人类对现实世界中各种对象、事物、场景或现象相似性判别的自然延伸。然而,任何一张图像都已经脱离了其内容所处的时间和空间。对于不同的人或同一个人不同的时间和空间内,图像的内容都可能不同,甚至可能完全相反。不变的(或变化细微的)是图像的颜色、纹理、形状和场景等。前者是主观、不容易被捕捉的语义特征,同时也是判别图像相似性的重要因素之一。后者是图像上相对客观的视觉特征,较容易被捕捉并用于度量图像之间的相似性。任意给定两幅图像,人们在判别它们的相似性时,一般并不会单独使用语义特征或视觉特征,而是会根据自己对图像理解做出一个整体判别。然而,由于技术水平的限制,目前的图像检索系统更多地利用图像的视觉特征度量两幅图像之间的相似性。这也使得有必要将原本更适合于度量语义特征相似性的特征对比模型改造成适合于度量视觉特征的相似性。

假设图像数据库  $\Delta = \{I_1, I_2, I_3, \dots\}$  中第  $i$  幅图像  $I_i$  的视觉特征向量为  $V_i = \{v_1^i, v_2^i, v_3^i, \dots, v_N^i\}$ 。在数据库的使用过程中,图像之间的相似性度量不再仅仅基于事先选定的视觉特征向量,而是不断地嵌入一些语义特征,从而丰富图像的特征集  $F_i = \{V_i, S_i\}$ 。在图像检索过程中,用户会从返回的图像中指定相似的图像。因此,从用户反馈中获取图像语义是一个渐进的过程。语义特征集合可以定义成一个动态变化的集合,即  $S_i = \{s_1^i, s_2^i, s_3^i, \dots\}$ 。

虽然模糊特征对比模型继承了特征对比模型的核心思想,即对象之间的相似性被解释成相同和相异特征的线性组合。但是,在实际的图像相似性度量中,特征对比模型中特征的可扩展性被固定的多维特征结构所替代。在综合 FCM 和 DFFCM 各自优点的基础上,提出一个新的模型,即图像特征比率模型(Image Feature Ratio Model, IFRM):

$$m(I_i, I_q) = \frac{1 + FCM(S_i, S_q)}{1 + \sum_{k=1}^N \mu_k(v_k^q) - DFFCM(V_i, V_q)} \quad (5-10)$$

其中,  $FCM(S_i, S_q)$  是特征对比模型;  $DFFCM(V_i, V_q)$  是诊断性模糊特征对比模型,它弥补了模糊特征对比模型中没有考虑特征诊断性的缺陷,下一小节将详细



讨论该模型； $\mu_k(v_k^q)$  表示定义在第  $k$  个视觉特征中的模糊成员函数； $\sum_{k=1}^N \mu_k(v_k^q)$  是查询图像的视觉特征自相似性值。

## 5.2.1 特征内诊断性

尽管从形式上看，FFCM 秉承了 FCM 的主要思想，但是从模糊逻辑的使用和该模型在图像相似性度量中的具体应用来说，该模型隐式地引入了 FCM 所抛弃的一个重要假设：特征是按多维结构组织的。为了更清楚地表明这一点，将公式 (5-7) 写成：

$$\begin{aligned} FFCM(V_i, V_q) &= \sum_{k=1}^N FFCM_k(V_i, V_q) \\ &= \sum_{k=1}^N (\theta \min\{\mu_k(v_k^i), \mu_k(v_k^q)\} \\ &\quad - \alpha \max\{\mu_k(v_k^i) - \mu_k(v_k^q), 0\} \\ &\quad - \beta \max\{\mu_k(v_k^q) - \mu_k(v_k^i), 0\}) \end{aligned} \quad (5-11)$$

从上式可以看出：模糊特征对比模型的相似性度量可以表示为  $N$  维特征空间中每维特征上的模糊特征对比之和，即  $\sum_{k=1}^N FFCM_k(V_i, V_q)$ 。这一点非常类似于按维数结构组织的几何距离，如海明距离。

对于 FCM 而言，由于特征是以集合形式存在，没有维数的概念。因此，集合之间的关系不会受制于预定的维数结构。图 5-1 表示了两种集合之间的关系，其中，图 5-1(a)表示的是两个二值特征集合之间的关系，这也是 FCM 所基于的特征集之间的关系；图 5-1(b)表示的是模糊特征对比模型所使用的两个模糊特征集在一维特征空间上的相互关系。比较图 5-1(a)和 5-1(b)，不难发现两种集合之间存在的差别：(1) 模糊集中的元素值都是从 0 到 1，也就是说，模糊集中的元素有一个共同的起点、终点和增长方向；(2) 模糊集之差可以定义为  $D1$ ，也可以定义为  $D2$ ，然而，二值特征集之差为  $A-B$  和  $B-A$ ；(3) 模糊集中的全集可以定义为具有更大真值的集合如  $A \cup B = A$ ，也可以定义了固定的值 1，然而，二值特征集的全集为  $A \cup B = (A-B) \cup (A \cap B) \cup (B-A)$ 。所有上述的差别都源自同一个问题：模糊集是按维数结构组织的。从图像相似性度量的应用需求出发，将视觉特征组织成多维结构是自然而然的选择，但是对于 FCM，这将影响该模型的基本假设之一（即非维数结构）及其所决定的特征在相似性度量中的作用。

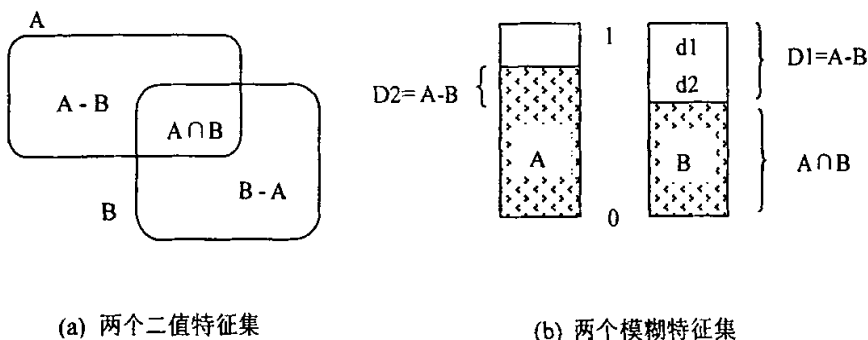


图 5-1 特征集合之间的关系

Figure 5-1 The relationship between two feature sets

如上所述,在 FCM 中,特征在相似性度量中的作用由两个因子决定:(1)特征强度,即对象属性的物理测量值;(2)特征诊断性,即对象属性在分类时的重要性。在 FCM 中,特征的诊断性是通过对象的特征集合的大小来实现的<sup>[53]</sup>。然而,在多维特征结构中,特征的类型是预先设定的,可变只是每个特征的强度。因此,在 FFCM 中,由于多维结构的存在,FCM 中的特征论断性思想不再存在。然而,在多维结构下,仍然可以模拟 FCM 中的特征诊断性思想。

为了简化说明,以一种简单对称的情况来说明:在模糊特征对比模型中,特征强度与相似性度量值之间的关系。假设  $\theta$ ,  $\alpha$  和  $\beta$  为 1, 只有一维特征,公式 (5-7) 可以重写为:

$$FFCM(V_i, V_q) = \min\{\mu(v^i), \mu(v^q)\} - \max\{\mu(v^i) - \mu(v^q), 0\} - \max\{\mu(v^q) - \mu(v^i), 0\} \quad (5-12)$$

由于没有考虑到特征诊断性,FFCM 在相似性度量时,特征强度与相似性度量值成强相关:

(1) 自相似性值与特征强度成正比

证明: 如果  $\mu(v^i) = \mu(v^q)$ , 则

$$\begin{aligned} FFCM(V_i, V_q) &= \min\{\mu(v^i), \mu(v^q)\} - \max\{\mu(v^i) - \mu(v^q), 0\} - \max\{\mu(v^q) - \mu(v^i), 0\} \\ &= \mu(v^q) - \max(0, 0) - \max(0, 0) \\ &= \mu(v^q) \end{aligned} \quad (5-13)$$

从上式可以看到:对象的自相似性值与其特征强度成正比:特征强度越大的对象,其自相似性值越大。

(2) 假设两个对象的特征  $V_i$  和  $V_j$  与同一个参考对象的特征  $V_q$  相距相同的距

离时, 强度较大的对象将被判别为与参考对象更为相似。

证明: 不失一般, 假设  $|v^q - v^i| = |v^j - v^q| = a > 0$  且  $v^i < v^q < v^j$ 。

因为,  $v^i < v^q$ , 则有  $v^i - v^q = -a$  和  $v^q - v^i = a$ 。

$$\begin{aligned}
 &FFCM(V_i, V_q) \\
 &= \min\{\mu(v^i), \mu(v^q)\} - \max\{\mu(v^i) - \mu(v^q), 0\} - \max\{\mu(v^q) - \mu(v^i), 0\} \\
 &= \min\{\mu(v^q) - a, \mu(v^q)\} - \max\{-a, 0\} - \max\{a, 0\} \\
 &= \mu(v^q) - a - 0 - a \\
 &= \mu(v^q) - 2a
 \end{aligned} \tag{5-14}$$

因为,  $v^j > v^q$ , 则有  $v^j - v^q = a$  和  $v^q - v^j = -a$ 。

$$\begin{aligned}
 &FFCM(V_j, V_q) \\
 &= \min\{\mu(v^j), \mu(v^q)\} - \max\{\mu(v^j) - \mu(v^q), 0\} - \max\{\mu(v^q) - \mu(v^j), 0\} \\
 &= \min\{\mu(v^q) + a, \mu(v^q)\} - \max\{a, 0\} - \max\{-a, 0\} \\
 &= \mu(v^q) - a - 0 \\
 &= \mu(v^q) - a
 \end{aligned} \tag{5-15}$$

显然,  $FFCM(V_i, V_q) < FFCM(V_j, V_q)$ 。换句话说, 两个对象的特征  $V_i$  和  $V_j$  与同一个参考对象的特征  $V_q$  相距相同的距离时, 强度较大的对象将被判别为与参考对象更为相似。

为了削弱上述两种特征强度对相似性度量的影响, 提出了诊断性模糊特征对比模型 (Diagnostic Fuzzy Feature Contrast Model, **DFFCM**):

$$\begin{aligned}
 &DFFCM(V_i, V_q) \\
 &= \theta \sum_{k=1}^N \eta(v_k^R) \times \varpi(\mu_k(v_k^i), \mu_k(v_k^q)) \times \min\{\mu_k(v_k^i), \mu_k(v_k^q)\} \\
 &\quad - \alpha \sum_{k=1}^N \eta(v_k^R) \times \rho(\mu_k(v_k^i), \mu_k(v_k^q)) \times \max\{\mu_k(v_k^i) - \mu_k(v_k^q), 0\} \\
 &\quad - \beta \sum_{k=1}^N \eta(v_k^R) \times \rho(\mu_k(v_k^i), \mu_k(v_k^q)) \times \max\{\mu_k(v_k^q) - \mu_k(v_k^i), 0\}
 \end{aligned} \tag{5-16}$$

其中,  $\eta(v_k^R)$  是由检索中反馈图像集  $V^R$  所决定的特征间的诊断性因子;  $\varpi(\mu_k(v_k^i), \mu_k(v_k^q))$  和  $\rho(\mu_k(v_k^i), \mu_k(v_k^q))$  是特征内相同特征和相异特征的诊断性因子。

相异性特征的诊断性函数表示为

$$\rho(x, y) = 1 + d(x) + d(y) \tag{5-17}$$

其中  $d(x)$  为特征  $x$  的诊断性, 它可以表示为

$$d(x) = \exp\left(-\frac{(x-0.5)^2}{a}\right) \quad (5-18)$$

其中  $a$  是一个非零常量。

共同特征的诊断性函数可以表示为

$$\varpi(x, y) = \frac{1}{1 + \exp(-b \times |\min(x, y) - 0.5|)} \quad (5-19)$$

其中  $b$  是一个非零常量, 在本章的实验中, 常数  $a$  和  $b$  设为 2。

在交互式图像检索中, 已经有大量的文献讨论了特征间诊断性, 例如[60; 83; 90; 91; 117; 132; 135; 144]。因此, 本文并没有讨论特征间诊断性加权方法。

### 5.2.2 弱语义及其嵌入方法

在基于内容的图像检索系统中, 图像之间的相似性度量是对人类相似性判别的模拟。然而, 由于现有技术的限制, 系统还不可能被训练成人一样, 这就出现了所谓的“语义鸿沟”。如果将图像检索比喻成两个人之间的交流的话, 不难发现, 语义鸿沟的直接来源就是彼此之间的“语言”不通。对于人来说, 图像的内容就是现实世界中可见对象及其相互关系, 表达图像最直观的语言就是自然语言。对于机器而言, 图像是一组相关的数字, 表达图像的“语言”可能是图像本身, 也可能是图像的各种特征, 例如颜色、纹理或语义特征。对于两种使用不同“语言”的人, 他们之间的交流必然非常困难。对于该问题, 目前存在两种不同的解决方案: (1) 以一种“语言”为基础, 将另一种“语言”显式或隐式地“翻译”成该“语言”, 如以视觉特征为基础的权重调整<sup>[91]</sup>、相似性尺度的选择<sup>[92]</sup>、以语义为主的语义建模<sup>[76]</sup>、语义提取<sup>[118]</sup>; (2) 将两种语言“翻译”成一种折中的表达形式, 如集成语义和视觉特征的混合索引<sup>[126]</sup>、语义空间学习<sup>[40]</sup>。Zhang 等人提出了将知识直接嵌入特征向量的方法, 该文中的知识是根据现有特征诱导出的<sup>[131]</sup>。

关于图像的语义, 有两种截然相反的观点: (1) 语义是图像固有的属性, 可以根据视觉特征提取语义或建立语义模型<sup>[76]</sup>; (2) 图像本身并没有任何语义, 语义是即时产生的, 即只有人们看到图像时, 图像的语义才会在人们脑海里形成<sup>[92]</sup>。这两种观点反映了图像语义的一个重要的特性: 多层次性。根据从具体到抽象的顺序, 图像语义大致可以分为三个层次: 对象、场景、抽象概念(如自然现象、社会事件或人类活动等)。认为图像语义可以提取或模拟的观点通常是指图像的

对象或场景层次的语义。这种模式最大的问题是如何分割出图像中的自然对象以及如何使得基于对象的模型具有广泛的适应性。这两个问题严重地限制了这种语义表达模式的应用范围。在语义是即时产生的观点中,语义更多地是指图像的抽象语义。抽象语义本身的不确定性使得对图像的解释因人而异。Santini 等人认为图像的抽象语义可以通过探索图像数据库中图像之间的低维空间排列来获得<sup>[92]</sup>。在图像检索的实际应用中,语义更多地被表达成前述两种观点的一个折中:以图像的视觉特征为基础,将用户反馈的图像及其相互关系通过某种途径传递到图像的相似性度量模型中,如权重调整或者距离模型的选择。这种方法假设图像的语义是可以通过改变相似性度量模型来表达的。按照这种假设,如果使用图像的直方图作为图像特征,那么仍然可以通过改变相似性度量方式来检索某些形状、纹理或其它含有高层次语义的图像。显然,这种假设过于牵强。本章提出另一种表达语义的方法,即在与用户的交互过程中不断地向特征向量中嵌入一些弱语义,通过扩充特征来增强图像的语义表现力,其中弱语义是指用户指定的跟查询图像相似的图像集合。语义集合中的元素之所以被称为弱语义是因为它既不是直接从图像视觉特征中导出的语义,也不是对图像语义的直接建模。在本章的实验中利用 Yang 等人提出的双层对等索引技术来组织反馈样本<sup>[126]</sup>。

### 5.2.3 图像特征比率模型的基本特点

该模型具有以下几个基本特点。首先,图像的语义和视觉特征之间的相似性度量都源于同一个相似性理论。因此,图像特征比率模型(IFRM)也是特征匹配函数的表示形式之一。其次,语义和视觉特征被无缝集成在同一个模型中。尽管这也是 FCM 的最初的想法,然而,由于 FCM 对特征表示的一些限制条件使得 FCM 很难集成语义和视觉特征去度量图像之间的相似性。在 FFCM 中,视觉特征的复杂转换机制被模糊逻辑所替换,遗憾的是在图像检索中,图像的语义和视觉特征不可能跟 Tversky 的实验一样直接由实验者所指定,只能事先选择视觉特征、间接地从用户反馈中获取一些相对较弱的语义特征。此外,为了尽可能地充分表达图像的各类视觉特征,视觉特征的维数往往是很高的,然而考虑到系统的实用性,每次反馈给用户的图像不可能太多。因此,如果利用 FFCM 去度量图像的可变语义和固定的视觉特征之间的相似性,则出现两种特征之间相互干扰,典型的情况是弱语义的作用会被忽略掉。第三,图像特征比率模型为语义和视觉特征之间的协调工作提供了一个非常灵活的机制。例如,一方面,该模型在



缺少语义特征时仍然可以正常使用。另一方面,在图像间的视觉特征变得越来越相似时,语义特征就会在整体相似性度量中发挥越来越重要的作用。从公式(5-10)可以看到,当图像库中图像跟查询图像在视觉特征上越来越相似时,公式(5-10)的分母将趋向于1。此时,图像的整体相似将主要由图像之间的语义特征所主导。因此,该模型可以区分视觉特征相似而语义特征存在差别的图像。假设相对于查询图像 $I_q$ ,就视觉特征而言,图像 $I_i$ 比图像 $I_j$ 更相似,即

$$FFCM(V_i, V_q) > FFCM(V_j, V_q) \quad (5-20)$$

进而有

$$0 \leq \sum_{k=1}^N \mu_k(v_k^q) - FFCM(V_i, V_q) < \sum_{k=1}^N \mu_k(v_k^q) - FFCM(V_j, V_q) \quad (5-21)$$

从公式(5-10)可以看到随着图像 $I_i$ 的视觉特征与查询图像 $I_q$ 的视觉特征越来越相似时,图像特征比率模型中的分母将趋向于1,即

$$1 + \sum_{k=1}^N \mu_k(v_k^q) - FFCM(V_i, V_q) \rightarrow 1 \quad (5-22)$$

此时,图像特征比率模型的值将主要取决于图像的弱语义特征之间的相似性。换句话说,根据图像特征比率模型,视觉上相似而语义有差别的图像将更容易被区分开。

## 5.3 实验

实验包括两部分:(1)通过模拟数据直观地分析 DFFCM 和 FFCM、TIM 在相似性度量上的差别,特别是两种不同的集合差定义对相似性度量的影响;(2)根据 Corel 图像库的检索实验,比较 IFRM 与其它模型的检索精度。

### 5.3.1 模拟分析

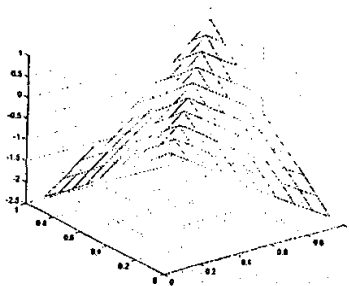
为了直观地展示 DFFCM 与 FFCM 和 TIM 在相似性度量上的差异,假设两个简单对象,它们只有一维定量特征(模糊真值从0到1),且所有模型中的 $\theta$ , $\alpha$ 和 $\beta$ 均取1。DFFCM 和 FFCM 的相似性度量如图5-2(a)和(c)所示,它们的四个剖面如图5-2(b)和(d)所示。图5-2(b)和(d)中最上面的线为两个对象模糊真值相等时的剖面,它被称为自相似性剖面,其它三条剖面线自左至

右分别为其中一个对象的模糊真值分别为 0.3, 0.5 和 0.9 时, 另一个对象和它之间的相似性度量值。

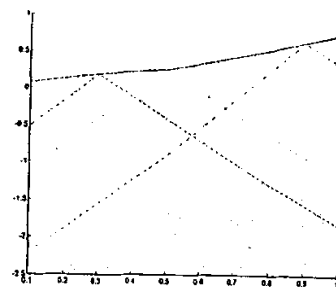
表 5-1 相似性非对称性

$ \mu(\phi) - \mu(\varphi) $	0.1	0.2	0.3	0.4
DFFCM	0.0301	0.0704	0.1209	0.1810
FFCM	0.1000	0.2000	0.3000	0.4000

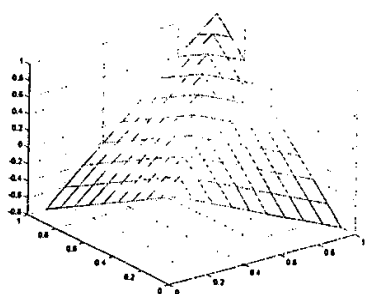
从图 5-2 (a) 和 (c) 中, 较难发现它们之间的差别。比较图 5-2 (b) 和 (d), 可以发现: (1) 对象间的相似性度量与对象本身的特征强度存在强相关性, 例如 5-2 (d) 中自相似性跟对象的模糊真值成正比; (2) 假设两个对象距离同一个参考对象相同的距离时, 强度较大的对象将被判别为与参考对象更为相似 (由特征强度引起的不对称性)。这些正是 FFCM 在相似性度量时的缺陷。同时, DFFCM 在很大程度上克服了这些缺点, 特别是由特征强度引起的相似性不对称性。表 5-1 列出了当参考对象的模糊真值为 0.5 时, 两个测试对象的模糊真值与参考对象的距离为 0.1、0.2、0.3 和 0.4 时, 它们与参考对象的相似性度量之差。这种差值越大, 表明相似性度量的非对称性越严重。从表 5-1 可以看出, 随着距离的增大, FFCM 的相似性度量的不对称性越明显, 且均是模糊真值越大的对象相似性值更大。虽然 DFFCM 也存在相似性度量的不对称性, 但是相对于 FFCM 而言, 这种不对称性已经得到极大地削弱。



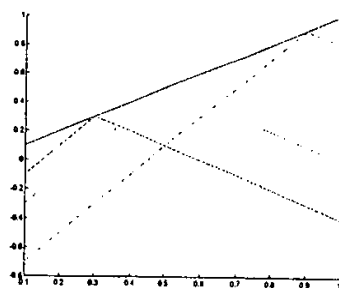
(a) DFFCM



(b) DFFCM 相似性表面的四个剖面



(c) FFCM



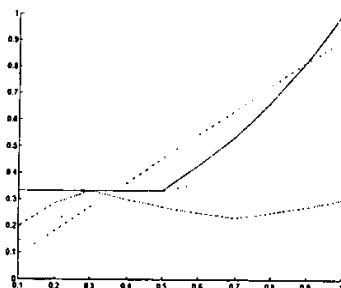
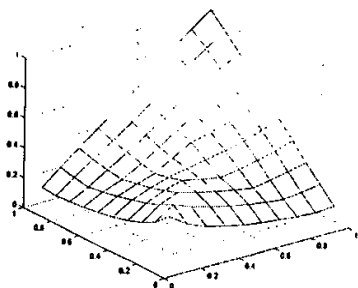
(d) FFCM 相似性表面的四个剖面

—— 自相似性    ----  $\mu(\phi) = 0.3$     .....  $\mu(\phi) = 0.5$     - - -  $\mu(\phi) = 0.9$

图 5-2 DFFCM 和 FFCM 的相似性表面及其四个剖面

Figure 5-2 The similarity surface and its four profiles of DFFCM and FFCM

TIM 采用了不同于 FFCM 和 DFFCM 的集合差的定义(如公式(5-4)和(5-6)所示)。为了直观地比较这两种集合差对相似性度量的影响,将公式(5-6)替代 TIM 中公式(5-4)的定义,并称之为 TIM2。TIM 和 TIM2 的相似性度量及其四个剖面如图 5-3 所示。TIM 和 TIM2 的相似性表面明显不同于 DFFCM 和 FFCM。从 TIM 的相似性剖面可以看出, TIM 在相似性度量时存在以下几个难于解释的相似性度量现象:(1) 自相似性在模糊真值小于 0.5 时为一个常量,而大于 0.5 时,它随着模糊真值的增加而急速增加;(2) 当一个对象的模糊真值为 0.9 时,另一个对象模糊真值在 0.35 到 0.85 之间时,相似性度量大于其自相似性,然而,其它情况下,相似性均小于自相似性值;(3) 当一个对象的模糊真值为 0.3 时,另一个对象模糊真值大于 0.3 时,相似性度量值先降后升。从图 5-3 (d) 可以看到,不再存在如图 5-2 (b) 所示的不合常理的相似性度量现象,原因是 TIM2 采用了公式(5-6)所示的集合差定义。由此可以看出,采用公式(5-6)所示集合差定义更为合适。





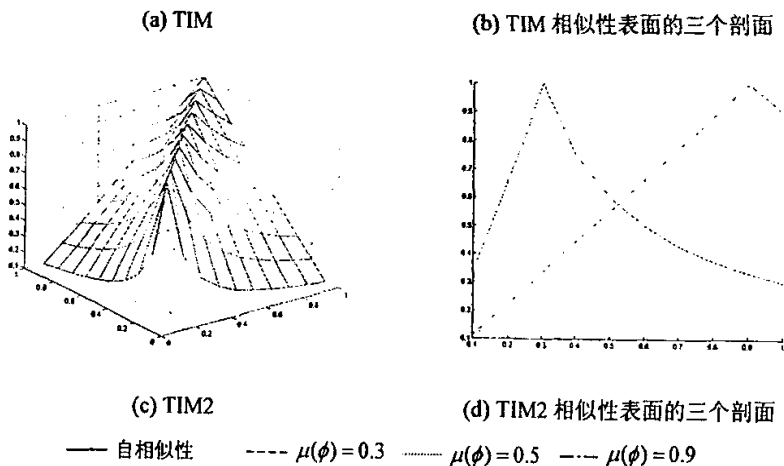


图 5-3 TIM 和 TIM2 的相似性表面及其四个剖面

Figure 5-3 The similarity surface and its four profiles of TIM and TIM2

### 5.3.2 检索精度

本实验分为两部分：(1)比较 DFFCM 与 FFCM、TIM、L1 norm(L1)、Euclidean Distance (ED) 的检索性能。实验中，所有选择的 Corel 图像均将作为一次查询图像，然后统计平均的检索精度；(2)测试 IFRM 检索性能。实验中，Corel 图像库中的所有图像均作为一次查询图像，每次返回最相似的 20 幅图像，然后根据 Corel 图像库已有的类别标识，计算检索精度。

图 5-4 显示了五种相似性模型的检索精度。DFFCM 的检索精度最高，TIM 的检索精度最低。这一方面说明 DFFCM 中引入的特征内论断性因子较好地适应了图像相似性度量的需要；另一方面，TIM 极低的检索精度与模拟实验中 TIM 所展示的难于解释的相似度量行为相吻合。FFCM 和 L1 的检索精度相当，它们均略高于欧氏距离。

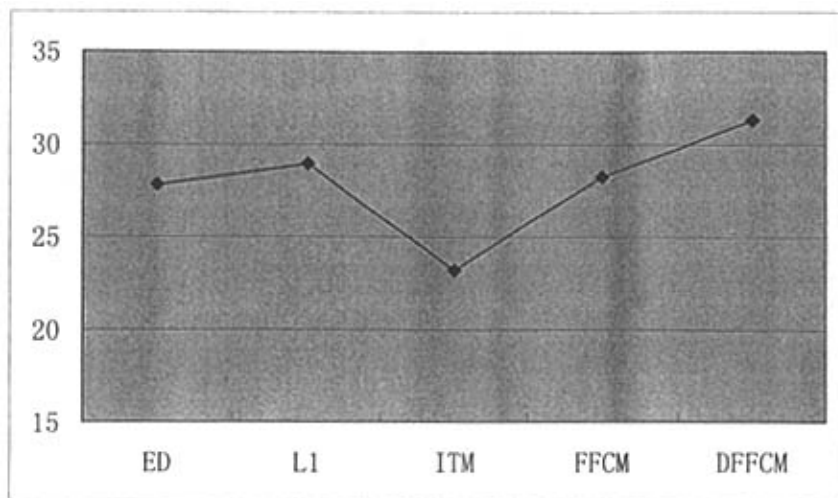


图 5-4 几个相似性模型的检索精度

Figure 5-4 The retrieval precisions of several similarity models

在反馈实验中，将正反馈样本作为查询的弱语义，从而构筑类似于 Tversky 提出的语义特征集合。其中，每个查询图像执行 5 次的反馈。为了比较，实现了文献<sup>[91]</sup>和<sup>[126]</sup>中相关反馈算法及其相似性度量方法。为区别起见，分别用 Rui 和 Yang 来代表文献[91]和[126]中的反馈算法和其相似性度量方法。实验中， $FCM(S_i, S_q)$  的函数  $f(x)$  取集合  $x$  的势，参数  $\alpha$ ， $\theta$  和  $\beta$  均取 1；模糊特征对比模型  $FFCM(V_i, V_q)$  中的模糊隶属函数采用了与文献[94]相同的函数，即：

$$\mu(x) = (1 + \exp(-\frac{x - \bar{x}}{\sigma}))^{-1} \quad (5-23)$$

其中， $x$ ， $\bar{x}$  和  $\sigma$  分别为视觉特征及其均值和方差。

实验结果如图 5-5 所示。结果表明，IFRM 的检索精度明显高于其它两个算法，且随着反馈的进行，检索精度有进一步提升的潜力。Rui 算法的检索精度最低，这是由于该算法没有考虑语义特征在图像相似性度量中的作用。Yang 算法随着反馈的进行，检索精度也得到了明显的提高，但它仍然低于 IFRM 的增长速度。

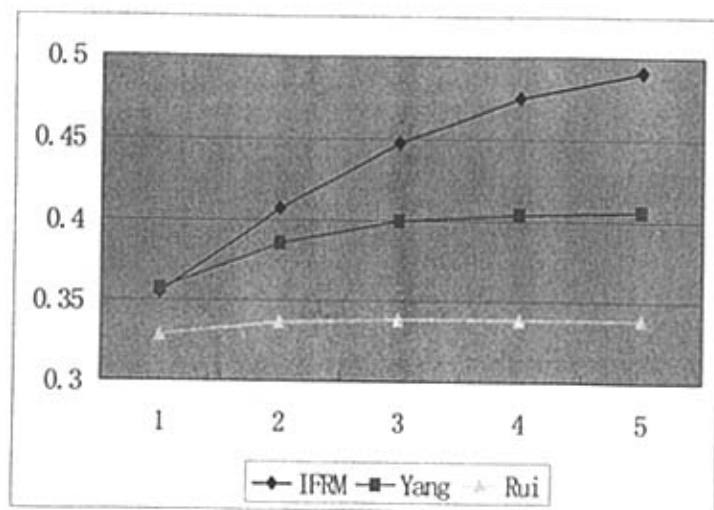


图 5-5 检索精度图

Figure 5-5 The graph of retrieval precision

## 5.4 本章小结

本章首先介绍了特征对比模型的两类典型的扩展模型，即 FFCM 和 TIM，然后指出模糊特征对比模型在图像视觉特征相似性度量中的缺点，提出了改进模型，即诊断性模糊特征对比模型。然后，提出了弱语义嵌入方法及其相似性度量模型，即图像特征比率模型。最后，模拟和图像检索实验表明，本章提出的图像特征比率模型在图像检索中是有效的。

## 第六章 图像检索原型系统

在本文的研究过程中开发了一个图像检索的原型系统。该系统不仅实现了一般图像的检索功能,还实现了基于海量遥感图像数据库管理系统的遥感图像和光谱的检索功能。在简要介绍该系统的主要结构和功能之后,本章将重点介绍该系统中的海量遥感图像数据库管理系统和适用于度量地物光谱相似性的光谱特征对比相似性模型。

### 6.1 系统概述

该系统是一个开放的图像检索系统,它支持多种图像数据源。目前已经实现的影像数据源有:(1)一般的图像,如 Corel 图像库;(2)高空间分辨率遥感图像,即覆盖全上海市的 1:50 000 万的遥感图像;(3)多光谱遥感图像,如 TM 图像;(4)高光谱数据,如美国地质调查局的地物光谱库。其中,高空间分辨率遥感图像采用 Oracle 数据库存储,其它均以文件形式存储。该系统的主要结构和功能如下图所示:

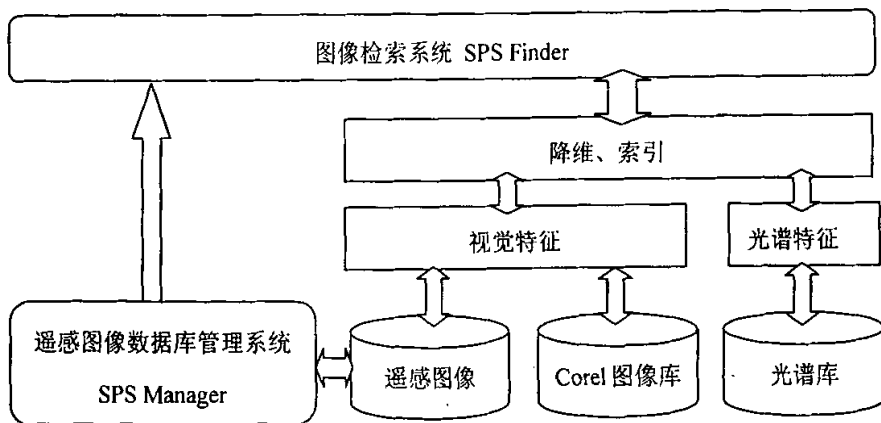


图 6-1 系统构架

Figure 6-1 The framework of system

从上图可以看到,系统实现了检索系统的基本功能,例如特征提取、降维、索引以及相关反馈。原型系统的主界面如图 6-2 所示。由于上述功能与一般的图

像检索系统并没有本质的区别,因此不再赘述。下面只对系统中两个独特的部分进行详细地介绍:(1)海量遥感图像数据库管理系统;(2)光谱特征对比相似性模型。

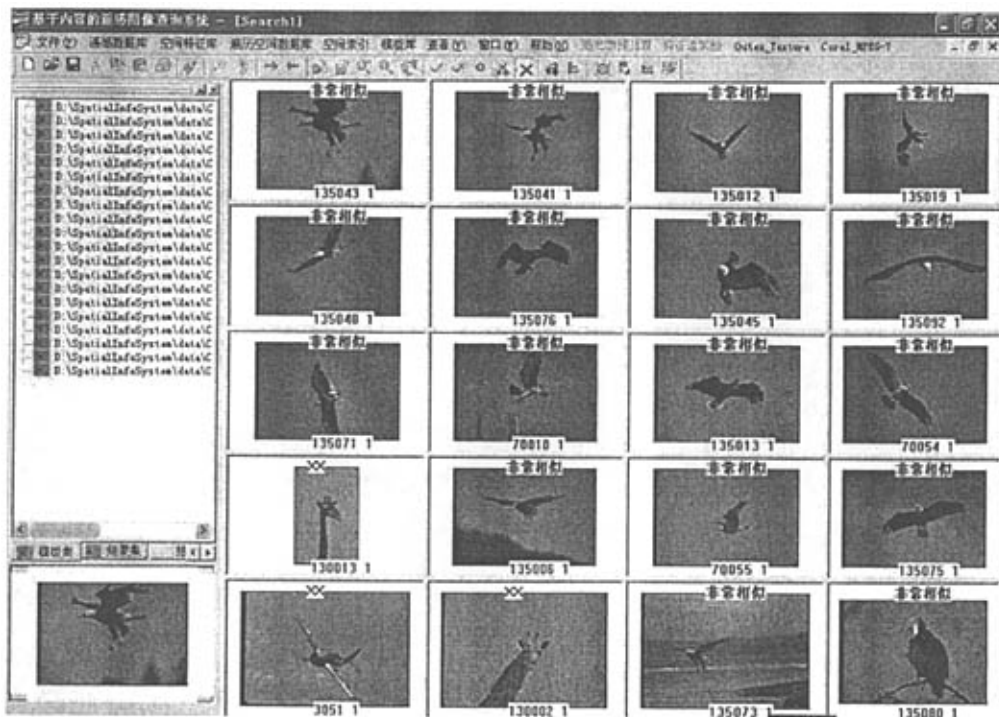


图 6-2 多源图像检索系统界面

Figure 6-2 The interface of multi-source image retrieval system

## 6.2 海量遥感图像数据库管理系统

与一般图像不同,遥感图像是国家空间数据基础设施和数字地球建设中最重要空间数据源之一,它除了在内容上具有很强的地域性以外,还具有多传感器、多分辨率、多波段、多时相等特点。一方面,随着航空航天技术的飞速发展,遥感图像数据在快速获取、存储、压缩、传输、管理与检索等方面面临着巨大的挑战<sup>[138]</sup>。另一方面,随着计算机技术的发展,高性能服务器、磁盘阵列、高速宽带网接入等技术为海量空间数据的存储、管理、实时显示、漫游、传输、检索和共享提供了一定的技术基础。

为了实时显示、无缝缩放、漫游和快速获取多尺度遥感图像,需要对各种尺度(或空间分辨率)的图像信息进行有效的组织和管理。遥感图像数据的多尺度组织和分块组织是解决上述问题的两个关键技术<sup>[142]</sup>。图像数据的多尺度组织可

以用图像金字塔结构来实现。简单地说,图像数据库是由多层同一地理范围内、不同尺度的遥感图像组成,如下图所示。

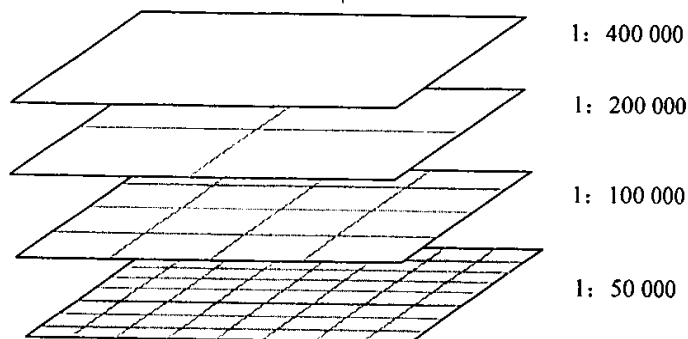


图 6-3 多尺度遥感图像金字塔示意图

Figure 6-3 Illustration of multi-scale remote sensing image pyramid

为方便说明,假设上图中每一个长方形方格代表一幅完整的遥感图像,且每幅图像的行列数大致相同。从上图可以看出,金字塔上不同尺度的遥感图像反映了地表上同一区域内地物的不同详尽程度,越往下层图像文件越多、单位地表上像素数量越多、空间分辨率越高、地物越清晰。金字塔上不同尺度的遥感图像可能来自不同的传感器或不同时相的数据,也可能从高分辨图像重采样而得。因此,在浏览图像时,可以根据用户所需的分辨率,确定应该从金字塔上哪一层提取数据,并进一步根据地理坐标确定具体的图像文件。

一般而言,单幅遥感图像的行列数会超出计算机屏幕的显示范围。因此,在浏览图像时,屏幕可见区域上显示的只是图像中的部分区域。因此,没有必要在每次浏览时,均将索引到的整幅图像读到内存中。为了加快图像的浏览速度,采用图像数据的分块技术。同时,这也为压缩图像提供了可能,从而极大地节约存储空间和减少图像读取时间。分块后的图像可以分开存贮,也可以仍然存贮在数据库的同一条记录内。由于在读写二进制大对象时,可以跳跃地读取图像数据,将所有的分块图像存贮在一起可以减少I/O次数,效率更高。图像块的大小,即图像块的行数和列数,通常取2的幂次方。图像分块需要考虑以下几个因素:图像的局部相关程度、压缩算法、图像数据类型、图像块缓冲区的管理算法、用户感兴趣区域的大小、网络的传输单元等,例如每幅图像被分成若干个 $512 \times 512$ 的块,不足 $512 \times 512$ 的边缘部分将并入其最邻近的一块。

从上面的讨论可知:在图像金字塔上浏览图像数据时,有两个关键算法:(1)空间定位;(2)重采样。这里“空间定位”是指如何根据用户当前缩放或漫游的

状态,确定地理位置、显示范围、显示分辨率以及应该读取的图像和图像在金字塔上的层数。空间空位的输入参数是显示窗的尺寸和对应的地理范围,根据这两个范围之比获得大概的显示分辨率,从而得到应该从金字塔的哪一层读取数据。确定金字塔的层数以后,需要确定应该读取该层中的哪些图像。为了加速这个过程,可以对所有图像块进行四叉树编码。四叉树编码的主要思想是根据每幅图像块在图层中的行列号生成一个数,该数通常被称之为Mortan码。结合图像所在的层,根据Mortan码可以生成一个在整个金字塔上唯一的ID号。从而,在图像的ID号与其所在的层数、行列数之间建立一个映射关系。借助这种映射关系可以实现快速空间定位。在获取图像以后,采用最近邻法对图像进行重采样。

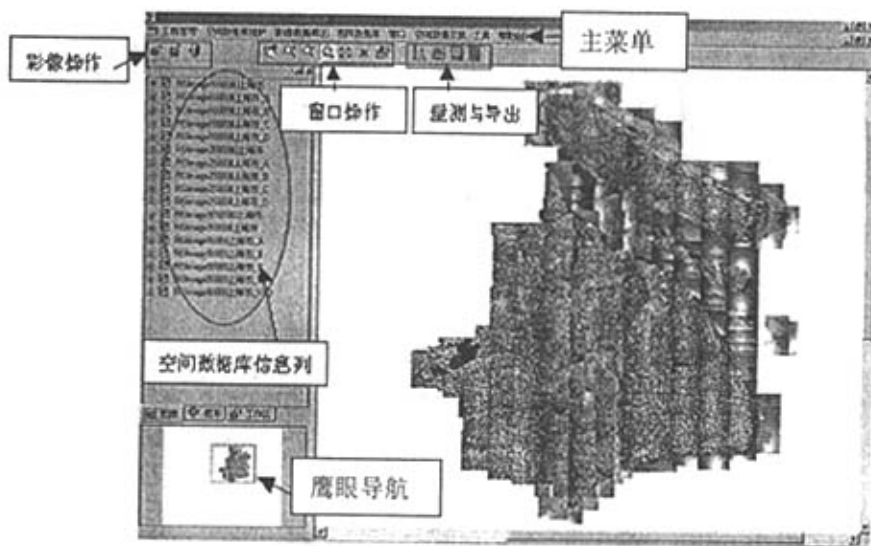


图 6-4 遥感图像管理系统界面

Figure 6-4 Interface of remote sensing image management system

上图是遥感图像管理系统中上海市遥感图像全景图。图 6-5 是放大后的陆家嘴区域。遥感图像数据管理系统是基于内容的图像检索系统的支持系统,它在后台组织、管理和维护所有的遥感图像,以便于检索系统可以从中检索和浏览遥感图像。图 6-6 是在图像数据库管理系统提供遥感影像中浏览图像检索结果。

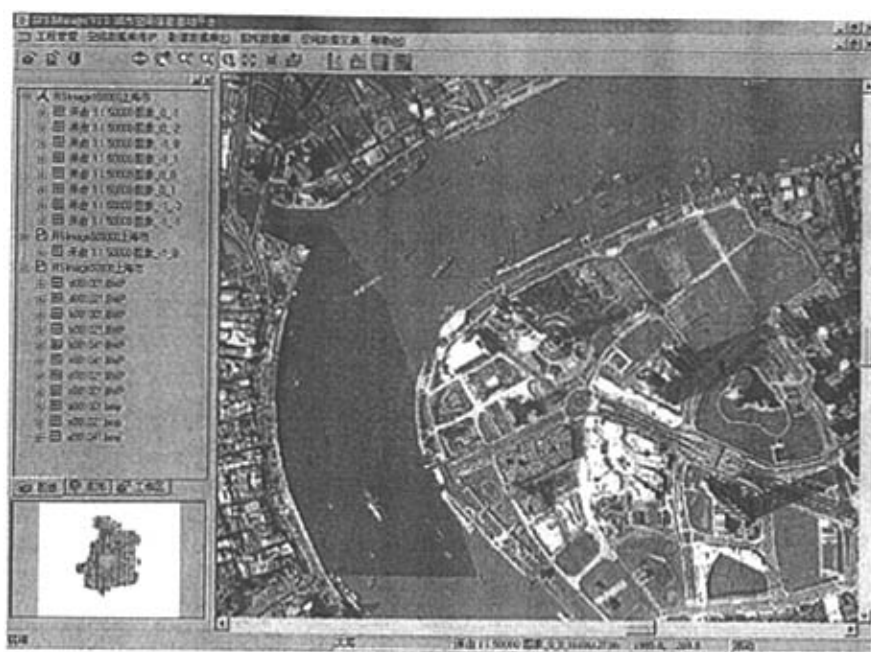
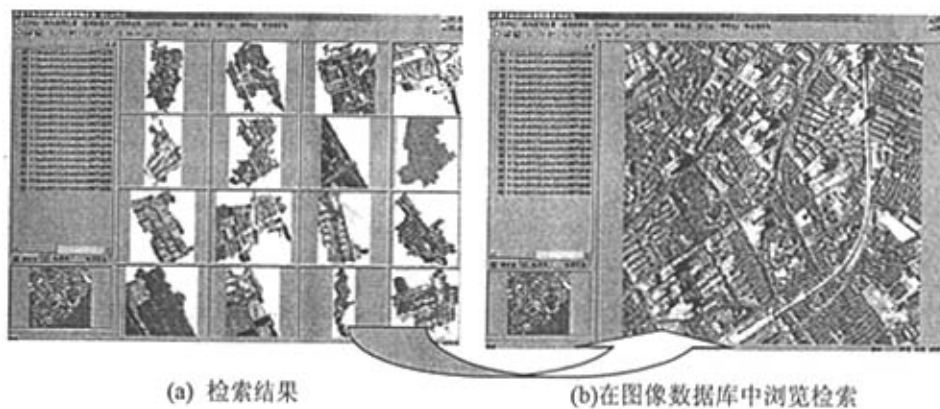


图 6-5 放大后图像

Figure 6-5 Enlarged image



(a) 检索结果

(b)在图像数据库中浏览检索

图 6-6 在整个图像库内浏览图像检索结果

Figure 6-6 Browsing retrieved images in whole image database

### 6.3 光谱特征对比相似性模型

现有的光谱相似性量度方法要么考虑光谱波形之间的相似程度，如光谱角制



图 (Spectral Angle Mapper, SAM), 要么考虑光谱特征波段之间的相似性, 如光谱吸收指数 (Spectral Absorption Index, SAI)。特征相似性理论将相似性解释成对象间的相同特征和相异特征的线性组合。为此, 本章提出了一种基于模糊特征对比模型的光谱特征对比相似性模型, 其中利用核主成分分析 (kernel Principal Component Analysis, PCA) 去除波段间的强相关性。

### (1) 核 PCA 去除波段间的强相关性

核 PCA 是基于核方法的 PCA。假设存在一个非线性映射  $f$  将输入空间  $\mathbf{R}^N$  中的样本变换到特征空间  $\mathbf{F}$  中<sup>[97]</sup>:

$$f: \mathbf{R}^N \rightarrow \mathbf{F}, \quad r \rightarrow R \quad (6-1)$$

其中, 映射函数  $f$  隐含在满足 Mercer 定理的核函数中, 即

$$k(x, y) = f(x) \bullet f(y) \quad (6-2)$$

在输入空间中, 核函数  $k(x, y)$  对应于特征空间中的内积。

假设特征空间  $\mathbf{F}$  中存在  $M$  个零均值样本  $f(r_i)$  ( $\sum_{i=1}^M f(r_i) = 0, i=1, \dots, M$ ), 它们的协方差矩阵为

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^M f(r_i) f(r_i)^T \quad (6-3)$$

假设  $\lambda_j$  ( $\lambda_j \neq 0$ ) 和  $\bar{v}_j$  分别为协方差矩阵  $\mathbf{C}$  的特征值及其对应的特征向量, 即:

$$\lambda_j \bar{v}_j = \mathbf{C} \bar{v}_j \quad (6-4)$$

由于向量  $\bar{v}_j$  处于特征空间中所有向量  $f(r_1), \dots, f(r_M)$  张成的空间, 因此, 存在一组系数  $\bar{w} = (w_1, w_2, \dots, w_M)$  满足下面的表达式:

$$\bar{v}_j = \sum_{i=1}^M w_i f(r_i) \quad (6-5)$$

将式 (6-3) 和 (6-5) 代入式 (6-4) 得:

$$\mathbf{K} \bar{w} = M \lambda_j \bar{w} \quad (6-6)$$

其中, 核矩阵  $\mathbf{K} = \{k_{ij}\}$ , ( $k_{ij} = k(f(r_i), f(r_j))$ )。

显然, 公式 (6-6) 就是求取核矩阵的特征值和特征向量。可以求出特征空间中的所有样本的协方差矩阵的特征值和特征向量, 因此, 从理论上可以求取特征空间中的正交向量基, 即  $\mathbf{V} = \{\bar{v}_j\}$  ( $j=1, \dots, N(N \leq M)$ )。

根据表达式 (6-2) 和 (6-5), 第  $j$  个变换后的光谱向量  $f(r_j)$  投影到正交基  $\mathbf{V}$  上, 第  $k$  维的坐标值 ( $r^*$ ) 可以如下表示:

$$\begin{aligned}
 r^k &= \vec{V}^k \cdot f(r_j) \\
 &= \sum_{i=1}^M w_i (f(r_i) \cdot f(r_j)) \\
 &= \sum_{i=1}^M w_i k_{ij}
 \end{aligned} \tag{6-7}$$

从上式可以看到，只要获得系数向量  $\vec{w}$  和核矩阵  $\mathbf{K}$ ，就可以求取特征空间中的样本在正交基  $\mathbf{V}$  上的投影。

## (2) 光谱的集合表示

一条光谱曲线可以表示成一个光谱集合  $I$ ，它包括两个子集  $A$  和  $A^-$ ，如下式所示（见图 6-7）：

$$\begin{cases} I = A \cup A^- \\ A \cap A^- = \Phi \end{cases} \tag{6-8}$$

其中， $\Phi$  为空集。

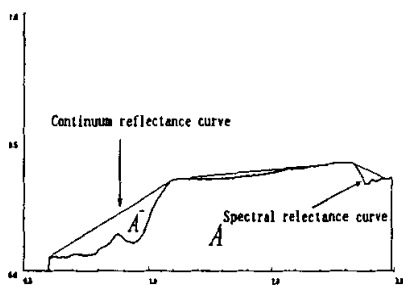


图 6-7 光谱集合示意图

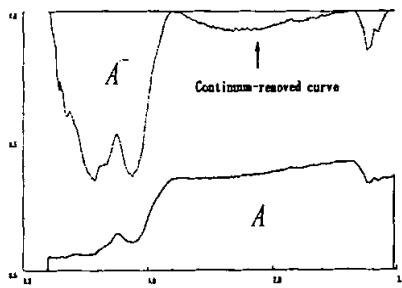


图 6-8 归一化以后的光谱集合示意图

Figure 6-7 Illustration of spectral set

Figure 6-8 Illustration of normalized spectra set

为了把子集  $A^-$  放在一个可以比较的框架下，将子集  $A^-$  用连续统曲线归一化，如下式所示：

$$A^- = \{a_{wr}^- \mid a_{wr}^- = a_{wr}^- / l_w, a_{wr}^- \in A^-\}, \tag{6-9}$$

归一化以后的光谱集合如图 6-8 所示。为方便起见，将  $A^-$  仍然写成  $A^-$ （下同）。

至此，可以将光谱集合的元素  $x_w$  用下式表示：

$$x_w = \begin{cases} r_w, x_w \in A \\ c_w, x_w \in A^- \end{cases} \tag{6-10}$$

## (3) 光谱相似性度量模型

基于上述的光谱集合表示，两个光谱集的并集  $I_1 \cup I_2$  可以细分为以下四个子

$$\begin{cases} A_1 \cap A_2 \\ A_1 \cap A_2^- \\ A_1^- \cap A_2 \\ A_1^- \cap A_2^- \end{cases} \quad (6-11)$$

从特征对比模型中的特征集可以发现,只有三个子集 ( $A_1 \cap A_2$ ,  $A_1 \cap A_2^-$  和  $A_1^- \cap A_2$ ) 出现在特征对比模型中,而子集  $A_1^- \cap A_2^-$  没有被考虑,这是因为在特征对比模型中,集合中的元素  $x$  满足以下面的条件:

$$x = \begin{cases} 1, & x \in A \\ 0, & x \in A^- \end{cases} \quad (6-12)$$

然而,光谱集合中的元素却并不满足公式 (6-12),即光谱集合的补集是非空的。因此,子集  $A_1^- \cap A_2^-$  在光谱相似性度量时应该给予应有的考虑。此外,两个光谱集合中的相异子集之间的交集 (即  $A_1 \cap A_2^-$  和  $A_1^- \cap A_2$ ) 并没不能体现光谱特征之间的重叠,可以不用考虑。相反,同一光谱内的两个子集之间  $A \cap A^-$  可以表现光谱相似性的强相关性,不失一般性,用函数  $\varphi(A, A^-)$  来刻画这一特征。基于以上的分析,提出一个适合光谱的相似性度量模型<sup>[109]</sup>:

$$\begin{aligned} S(I_1, I_2) &= F(A_1, A_1^-, A_2, A_2^-) \\ &= \alpha f(A_1 \cap A_2) + \beta f(A_1^- \cap A_2^-) + \gamma (\varphi(A_1, A_1^-) \cap \varphi(A_2, A_2^-)) \end{aligned} \quad (6-13)$$

其中,显著性函数  $f(x)$  与特征对比模型中的定义一样;而  $\varphi(x, y) = xy$ , 它融合了光谱的吸收和反射特征。

#### (4) 实验分析

实验数据来自美国地质调查局 (United States Geological Survey, USGS) 的光谱库 (spc\_lib04a)。光谱库中共有共 498 条光谱,光谱的波段数为 480,波长在 0.2 到 3.0 微米之间,光谱分辨率小于 10 纳米。实验中选择了几种典型的相似性度量方法,即欧氏距离 (ED)、光谱角制图 (SAM) 和光谱信息散度 (SID)<sup>[12]</sup>,模糊特征对比模型 (FFCM)<sup>[94]</sup>,统计它们的平均检索精度和召回率,如图 6-9 所示。

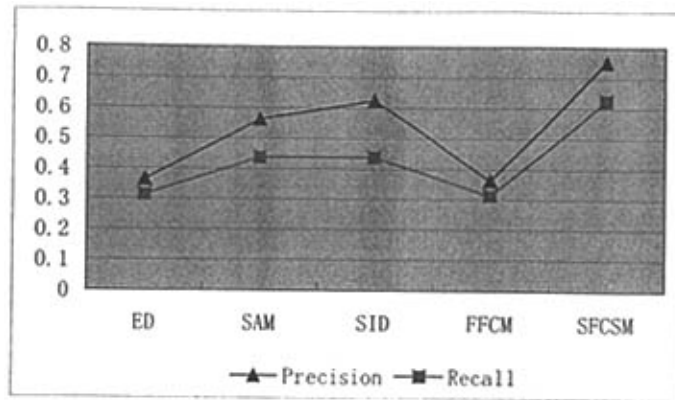


图 6-9 光谱检索精度

Figure 6-9 Spectral retrieval precisions

从上图可以看到，无论是检索精度还是召回率，光谱特征对比相似性模型均比传统的相似性模型都要好。在原型系统中，光谱检索界面如下图所示。

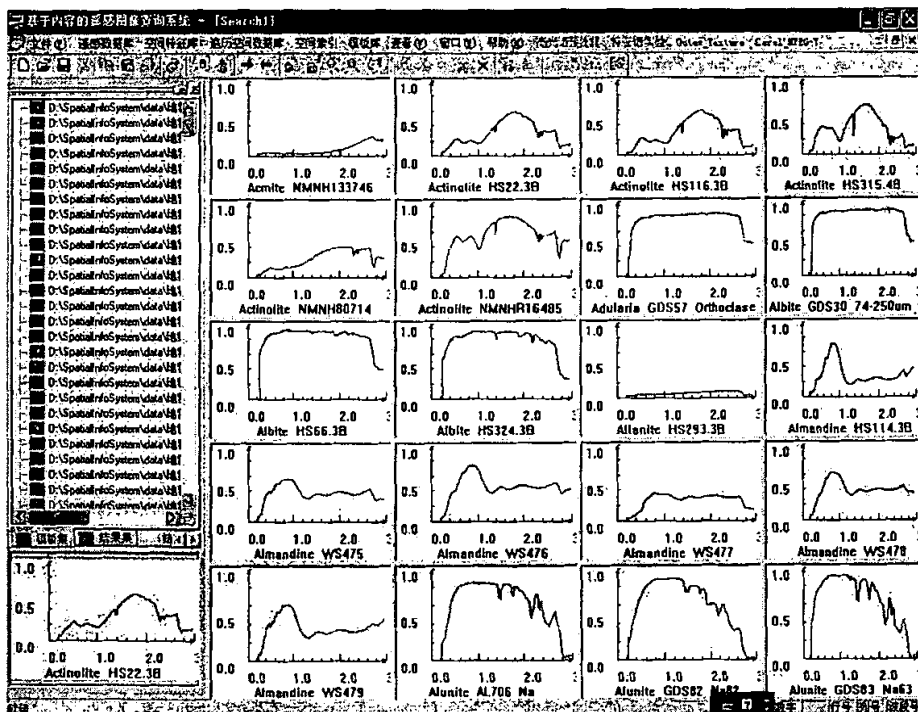


图 6-10 光谱相似性检索界面

Figure 6-10 The interface of spectral similarity



## 6.4 本章小结

在简要介绍图像检索原型系统的基础上，本章重点描述了两方面内容：（1）海量遥感图像数据库管理系统的主要实现技术，例如按金字塔组织、索引多分辨率遥感图像；（2）适用于度量光谱相似性的光谱特征对比相似性模型，其中包括利用核主成分分析去除波段之间的强相关性、光谱的集合表形式和新的光谱相似性度量模型。

## 第七章 结论与展望

本文深入分析了心理学相似性理论中的几何和特征相似性,结合图像检索的应用需求,提出了一些新的相似性模型和分析方法。本文的创新点包括以下几个方面:

(1) 提出拉普拉斯线性判别分析算法。该算法的准则函数是极大化类间与类内散度矩阵的差,它不仅形式简单而且不要求取类内散度矩阵的逆。该算法将获取最优判别向量的过程转化成全局和局部拉普拉斯矩阵的谱分解。

(2) 提出非线性判别排列算法。该算法有效地将拉普拉斯线性判别分析的思想嵌入到等距映射算法的基本框架中,使等距映射算法从非监督学习转变成监督学习算法,从非线性多维排列转变成非线性判别排列算法。该算法可以有效地从非线性数据集中获取易于分类的排列结果。

(3) 根据特征相似性理论的特征诊断性原理,分析了模糊特征对比模型中特征强度与相似性之间的强相关性,提出集成语义和视觉特征的图像特征比率模型。该模型不仅在视觉特征相似性度量时削弱了特征强度与相似性之间的强相关性,而且能协调语义和视觉特征在图像整体相似性度量中的作用。

(4) 提出一个集成光谱反射和吸收特征的光谱特征对比相似性模型。

图像相似性度量是模式识别和人工智能领域的基础问题之一,仍有许多问题值得深入研究,例如:

(1) 图像相似性判别实验研究。在心理学相似性实验中,相似性判别的对象通常是现实世界中的对象、场景或事物,而不是图像。因此,对于特定应用领域的图像,为了有效地描述图像的内容,有必要实验分析用户对图像的相似性判别的共性和差异。

(2) 进一步探索以 MDS 为代表的几何相似性模型在图像语义获取、分析和相似性度量中的应用。图像相似性判别是人为解释图像内容的过程,以 MDS 为代表的几何模型可以为用户直观地展示系统内图像之间的相互关系,如相似性、聚类状况等。

(3) 其它心理学相似性理论在图像检索中的应用,如匹配相似性和变换相似性。匹配相似性可以有效地解释图像中不同场景之间相似程度,前提是有效地分割图像中的对象或场景。



- (4) 研究图像语义及其相似性度量方法。

## 参考文献

- [1] Aggarwal, G., Ashwin, T. V., Ghosal, S., An image retrieval system with automatic query modification [J], IEEE Transactions On Multimedia, 2002,4(2): 201-214.
- [2] Aksoy, S., Haralick, R. M., Feature normalization and likelihood-based similarity measures for image retrieval [J], Pattern Recognition Letters, 2001,22(5): 563-582.
- [3] Beckmann, N., Kriegel, H. P., Schneider, R., R\*-tree. An efficient and robust access method for points and rectangles, Proceedings of the ACM SIGMOD International Conference on Management of Data, 1990, 322-331.
- [4] Belkin, M., Niyogi, P., Laplacian eigenmaps for dimensionality reduction and data representation [J], Neural Computation, 2003,15(6): 1373-1396.
- [5] Berchtold, S., Keim, D. A., Kriegel, H. P., The X-tree: An index structure for high-dimensional data, Proc. 22nd VLDB Conference, 1996.
- [6] Bradley, C. L., Similarity and Categorization: a review, Oxford University Press, 2002.
- [7] Bush, R. R., Mosteller, F., A model for stimulus generalization and discriminant [J], Psychological Review, 1951,58: 413-423.
- [8] Carroll, J. D., Chang, J. J., Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition [J], Psychometrika, 1970,35(3): 283-319.
- [9] Cevikalp, H., Neamtu, M., Wilkes, M. et al., Discriminative common vectors for face recognition [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2005,27(1): 4-13.
- [10] Chaira, T., Ray, A. K., Fuzzy approach for color region extraction [J], Pattern Recognition Letters, 2003,24(12): 1943-1950.
- [11] Chaira, T., Ray, A. K., Fuzzy measures for color image retrieval [J], Fuzzy Sets And Systems, 2005,150(3): 545-560.
- [12] Chang, C. I., An information-theoretic approach to spectral variability,



similarity, and discrimination for hyperspectral image analysis [J], IEEE Transactions On Information Theory, 2000,46(5): 1927-1932.

[13] Chen, L. F., Liao, H. Y. M., Ko, M. T. et al., A new LDA-based face recognition system which can solve the small sample size problem [J], Pattern Recognition, 2000,33(10): 1713-1726.

[14] Cheng, Y. Q., Zhuang, Y. M., Yang, J. Y., Optimal Fisher Discriminate Analysis Using the Rank Decomposition [J], Pattern Recognition, 1992,25(1): 101-111.

[15] Chong, C. W., Raveendran, P., Mukundan, R., A comparative analysis of algorithms for fast computation of Zernike moments [J], Pattern Recognition, 2003,36(3): 731-742.

[16] Ciocca, G., Schettini, R., Content-based similarity retrieval of trademarks using relevance feedback [J], Pattern Recognition, 2001,34(8): 1639-1655.

[17] Cox, I. J., Miller, M. L., Minka, T. P. et al., The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments [J], IEEE Transactions On Image Processing, 2000,9(1): 20-37.

[18] Cox, T. F., Cox, M. A. C., Multidimensional Scaling (2nd ed.), Chapman & Hall: London, 2001.

[19] Daschiel, H., Datcu, M., Information mining in remote sensing image archives: System evaluation [J], IEEE Transactions On Geoscience And Remote Sensing, 2005,43(1): 188-199.

[20] Datcu, M., Daschiel, H., Pelizzari, A. et al., Information mining in remote sensing image archives: System concepts [J], IEEE Transactions On Geoscience And Remote Sensing, 2003,41(12): 2923-2936.

[21] Eidenberger, H., Breiteneder, C., Visual Similarity Measurement with the Feature Contrast Model, SPIE IS&T Electronic Imaging Conference (Storage and Retrieval for Media Databases), Santa Clara, USA, 2003.

[22] Eisler, H., Ekman, G., A mechanism of subjective similarity [J], Acta Psychologica, 1959,16(2): 1-10.

[23] Friedman, J. H., Regularized discriminant analysis [J], Journal of the American Statistical Association, 1989,84(405): 165--175.

[24] Frigui, H., Interactive image retrieval using fuzzy sets [J], Pattern



Recognition Letters, 2001,22(9): 1021-1031.

- [25] Garner, W. R., The processing of information and structure, Wiley, 1974.
- [26] Gati, I., Tversky, A., Weighting common and distinctive features in perceptual and conceptual judgements [J], Cognitive Psychology, 1984,16(3): 341-370.
- [27] Gentner, D., Structure-mapping: A theoretical framework for analogy [J], Cognitive Science, 1983,7(1): 155-170.
- [28] Goldstone, R., Medin, D., Similarity, interactive. activation, and mapping [J], Journal of Experimental Psychology: Memory and Cognition, 1994,20: 3-28.
- [29] Goldstone, R. L., Alignment-based nonmonotonicities in similarity [J], Journal of Experimental Psychology: Learning, Memory, and Cognition, 1996,22(4): 988-1001.
- [30] Goldstone, R. L., Similarity, MIT Press, 1999, 763-765.
- [31] Goodman, N., Seven structures on similarity, The Bobbs-Merrill Co., 1972.
- [32] Gregson, R. A. M., Psychometrics of similarity, Academic Press, 1975.
- [33] Guo, G. D., Jain, A. K., Ma, W. Y. et al., Learning similarity measure for natural image retrieval with relevance feedback [J], IEEE Transactions On Neural Networks, 2002,13(4): 811-820.
- [34] Hahn, U., Chater, N., Richardson, L. B., Similarity as transformation [J], Cognition, 2003,87(1): 1-23.
- [35] Hahn, U., Ramscar, M., Similarity and categorization, Oxford University Press, 2001.
- [36] Hastie, T., Principal Curves and Surfaces [D], Stanford University, PhD, 1984.
- [37] Hastie, T., Stuetzle, W., Principal Curves [J], Journal of the American Statistical Association, 1989,84(406): 502-516.
- [38] Hastie, T., Tibshirani, R., Penalized discriminant analysis [J], The Annals of Statistics, 1995,23(1): 73-102.
- [39] Hastings, S., Query Categories in a Study of Intellectual Access to Digitized Art Images, Proceedings of the 58th ASIS Annual Meeting, Chicago, 1995, 3-8.
- [40] He, X. F., King, O., Ma, W. Y. et al., Learning a semantic space from user's relevance feedback for image retrieval [J], IEEE Transactions On Circuits And

Systems For Video Technology, 2003,13(1): 39-48.

[41] He, X. F., Yan, S. C., Hu, Y. X. et al., Face recognition using Laplacianfaces [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2005,27(3): 328-340.

[42] Heit, E.,Rubinstein, J., Similarity and property effects in inductive reasoning [J], Journal of Experimental Psychology: Learning, Memory & Cognition, 1994,20(2): 411-422.

[43] Holowczak, R. D., Artigas, F. J., Chun, S. A. et al., An experimental study on content-based image classification for satellite image databases [J], IEEE Transactions On Geoscience And Remote Sensing, 2002,40(6): 1338-1347.

[44] Hong, Z. Q.,Yang, J.-Y., Optimal Discriminant Plane for a Small Number of Samples and Design Method of Classifier on the Plane [J], Pattern Recognition, 1991,24(4): 317-324.

[45] Hsieh, J. W.,Grimson, W. E. L., Spatial template extraction for image retrieval by region matching [J], IEEE Transactions On Image Processing, 2003,12(11): 1404-1415.

[46] Imai, S., Pattern similarity and cognitive transformations. [J], Acta Psychologica, 1977,41(6): 433-447.

[47] Ishikawa, Y. R.,Suramanya, F. C., MindReader: Query databases through multiple example, Proc. Of the 24th VLDB Conf, NY, 1998.

[48] Jia, L. H.,Kitchen, L., Object-based image similarity computation using inductive learning of contour-segment relations [J], IEEE Transactions On Image Processing, 2000,9(1): 80-87.

[49] Kegl, B., Krzyzak, A., Linder, T. et al., Learning and design of principal curves [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2000,22(3): 281-297.

[50] Kherfi, M. L., Ziou, D., Bernardi, A., Combining positive and negative examples in relevance feedback for content-based image retrieval [J], Journal Of Visual Communication And Image Representation, 2003,14(4): 428-457.

[51] King, I.,Jin, Z., Integrated probability function and its application to content-based image retrieval by relevance feedback [J], Pattern Recognition, 2003,36(9): 2177-2186.



- [52] Krantz, D. H., Tversky, A., Similarity of rectangles: an analysis of subjectivity dimensions [J], *Journal of Mathematical Psychology*, 1975, 12: 4-34.
- [53] Krumhansl, C. L., Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship between Similarity and Spatial Density [J], *Psychological Review*, 1978, 85(5): 445-463.
- [54] Kruskal, J. B., Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis [J], *Psychometrika*, 1964a, 29(2): 1-27.
- [55] Kruskal, J. B., Nonmetric multidimensional scaling: a numerical method [J], *Psychometrika*, 1964b, 29(2): 115-129.
- [56] Kumar, S., Ghosh, J., Crawford, M. M., Best-bases feature extraction algorithms for classification of hyperspectral data [J], *IEEE Transactions On Geoscience And Remote Sensing*, 2001, 39(7): 1368-1379.
- [57] Kwak, J. W., Cho, N. I., Relevance feedback in content-based image retrieval system by selective region growing in the feature space [J], *Signal Processing-Image Communication*, 2003, 18(9): 787-799.
- [58] Laaksonen, J., Koskela, M., Laakso, S. et al., PicSOM - content-based image retrieval with self-organizing maps [J], *Pattern Recognition Letters*, 2000, 21(13-14): 1199-1207.
- [59] Laaksonen, J., Koskela, M., Oja, E., PicSOM - Self-organizing image retrieval with MPEG-7 content descriptors [J], *IEEE Transactions On Neural Networks*, 2002, 13(4): 841-853.
- [60] Lee, K. M., Street, W. N., Incremental feature weight learning and its application to a shape-based query system [J], *Pattern Recognition Letters*, 2002, 23(7): 865-874.
- [61] Li, H. F., T., J., Zhang, K. S., Efficient and Robust Feature Extraction by Maximum Margin Criterion, *Advances in Neural Information Processing Systems*, Vancouver, Canada, 2003, 97 - 104.
- [62] Li, J., Wang, J. Z., Automatic linguistic indexing of pictures by a statistical modeling approach [J], *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 2003, 25(9): 1075-1088.
- [63] Li, M., Vitanyi, P., *An Introduction to Kolmogorov Complexity and its Applications* (2nd ed.), Springer-Verlag, 1997.

- [64] Li, S. Z., Chan, K. L., Wang, C. L., Performance evaluation of the nearest feature line method in image classification and retrieval [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2000,22(11): 1335-1339.
- [65] Lu, Y., Zhang, H. J., Liu, W. Y. et al., Joint semantics and feature based image retrieval using relevance feedback [J], IEEE Transactions On Multimedia, 2003,5(3): 339-347.
- [66] Ma, W. Y., Manjunath, B. S., NeTra: A toolbox for navigating large image databases [J], Multimedia Systems, 1999,7(3): 184-198.
- [67] Ma, W. Y., Manjunath, B.S., Texture features and learning similarity, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996, 425-430.
- [68] Manjunath, B. S., Ma, W. Y., Browsing large satellite and aerial photographs, IEEE International Conference on Image Processing, 1996a.
- [69] Manjunath, B. S., Ma, W. Y., Texture features for browsing and retrieval of image data [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996b,18(8): 837-842.
- [70] Manjunath, B. S., Ohm, J. R., Vasudevan, V. V. et al., Color and texture descriptors [J], IEEE Transactions On Circuits And Systems For Video Technology, 2001,11(6): 703-715.
- [71] Markman, A. B., Gentner, D., Splitting the differences: A structural alignment view of similarity [J], Journal of Memory and Language, 1993a,32(4): 517-535.
- [72] Markman, A. B., Gentner, D., Structural alignment during similarity comparisons [J], Cognitive Psychology, 1993b,25(4): 431-467.
- [73] McGee, V. C., Multidimensionnal scaling of n sets of similarity measures: A nonmetric individual differences approach [J], Multivariate Behavioral Research, 1968,3: 233-248.
- [74] Medin, D. L., Ortony, A., Psychological essentialism, Cambridge University Press, 1989.
- [75] Min, W. L., Lu, K., He, X. F., Locality pursuit embedding [J], Pattern Recognition, 2004,37(4): 781-788.
- [76] Mojsilovic, A., Gomes, J., Rogowitz, B., Semantic-friendly indexing and



quering of images based on the extraction of the objective semantic cues [J], International Journal Of Computer Vision, 2004,56(1-2): 79-107.

[77] MPEG. MPEG-7 Visual Part of the Experimentation Model (Version 9.0) ISO/IEC JTC1/SC29/WG11 N3914, 2001, pp.

[78] Muller, H., Pun, T., Squire, D., Learning from user behavior in image retrieval: Application of market basket analysis [J], International Journal Of Computer Vision, 2004,56(1-2): 65-77.

[79] Muneesawang, P., Guan, L., Automatic machine interactions for content-based image retrieval using a self-organizing tree map architecture [J], IEEE Transactions On Neural Networks, 2002,13(4): 821-834.

[80] Murphy, G. L., Medin, D. L., The role of theories in conceptual coherence [J], Psychological Review, 1985,92(3): 289-316.

[81] Nakazato, M., Manola, L., Huang, T. S., ImageGrouper: a group-oriented user interface for content-based, image retrieval and digital image arrangement [J], Journal Of Visual Languages And Computing, 2003,14(4): 363-386.

[82] Niblack, W., Barber, R., W.Equitiz et al., The QBIC project: Querying images by content using color, texture and shape, Proc. SPIE Storage and Retrieval for Image and Video Databases, 1994, 173-187.

[83] Peng, J., Multi-class relevance feedback content-based image retrieval [J], Computer Vision And Image Understanding, 2003,90(1): 42-67.

[84] Pentland, A., Picard, R. W., Sclaroff, S., Photobook: content-based manipulation of image databases, International Journal of Computer Vision, 1996, 233-254.

[85] Qiu, G. P., Lam, K. M., Frequency layered color indexing for content-based image retrieval [J], IEEE Transactions On Image Processing, 2003,12(1): 102-113.

[86] Quine, W. V., Ontological Relativity and Other Essays, Columbia University Press, 1969, 165.

[87] Randen, T., Husoy, J. H., Filtering for texture classification: A comparative study [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 1999,21(4): 291-310.

[88] Roweis, S. T., Saul, L. K., Nonlinear dimensionality reduction by locally linear embedding [J], Science, 2000,290(5500): 2323-2326.

[89] Rui, Y., Huang, T. S., Chang, S. F., Image retrieval: Current techniques, promising directions, and open issues [J], Journal Of Visual Communication And Image Representation, 1999,10(1): 39-62.

[90] Rui, Y., Huang, T. S., Mehrotra, S., Content-based image retrieval with relevance feedback in MARS, IEEE International Conference on Image Processing, 1997, 815-818.

[91] Rui, Y., Huang, T. S., Ortega, M. et al., Relevance feedback: A power tool for interactive content-based image retrieval [J], IEEE Transactions On Circuits And Systems For Video Technology, 1998,8(5): 644-655.

[92] Santini, S., Gupta, A., Jain, R., Emergent semantics through interaction in image databases [J], IEEE Transactions On Knowledge And Data Engineering, 2001,13(3): 337-351.

[93] Santini, S., Jain, R., Similarity is a geometer [J], Multimedia Tools And Applications, 1997,5(3): 277-306.

[94] Santini, S., Jain, R., Similarity measures [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 1999,21(9): 871-883.

[95] Santini, S., Jain, R., Integrated browsing and querying for image databases [J], IEEE Multimedia, 2000,7(3): 26-39.

[96] Saul, L. K., Roweis, S. T., Think globally, fit locally: Unsupervised learning of low dimensional manifolds [J], Journal Of Machine Learning Research, 2004,4(2): 119-155.

[97] Scholkopf, B., Smola, A., Muller, K. R., Nonlinear component analysis as a kernel eigenvalue problem [J], Neural Computation, 1998,10(5): 1299-1319.

[98] Schroder, M., Rehrauer, H., Seidel, K. et al., Interactive learning and probabilistic retrieval in remote sensing image archives [J], IEEE Transactions On Geoscience And Remote Sensing, 2000,38(5): 2288-2298.

[99] Schwarz, J., Staenz, K., Adaptive threshold for spectral matching of hyperspectral data [J], Canadian Journal Of Remote Sensing, 2001,27(3): 216-224.

[100] Shepard, R. N., The analysis of proximities: Multidimensional scaling with an unknown distance function. I. [J], Psychometrika, 1962a,27(2): 125-140.

[101] Shepard, R. N., The analysis of proximities: Multidimensional scaling with an unknown distance function. II. [J], Psychometrika, 1962b,27(3): 219-246.



[102] Shepard, R. N., Toward a universal law of generalization for psychological science [J], Science, 1987,237(4820): 1317-1323.

[103] Smeulders, A. W. M., Worring, M., Santini, S. et al., Content-based image retrieval at the end of the early years [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2000,22(12): 1349-1380.

[104] Smith, J. R., Querying by color regions using VisualSEEK content-based visual query system, Intelligent multimedia information retrieval, 1997, 23 - 41.

[105] Squire, D. M., Muller, W., Muller, H. et al., Content-based query of image databases: inspirations from text retrieval [J], Pattern Recognition Letters, 2000,21(13-14): 1193-1198.

[106] Stejic, Z., Takama, Y., Hirota, K., Relevance feedback-based image retrieval interface incorporating, region and feature saliency patterns as visualizable image similarity criteria [J], IEEE Transactions On Industrial Electronics, 2003,50(5): 839-852.

[107] Swets, D. L., Weng, J., Using discriminant eigenfeatures for image retrieval [J], IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996,18(8): 831-836.

[108] Tamura, H., Mori, S., Yamawaki, T., Texture features corresponding to visual perception [J], IEEE Transactions on Systems, Man, and Cybernetics, 1978,8(6).

[109] Tang, H., Fang, T., Shi, P. F., Spectral similarity measure based on fuzzy feature contrast model [J], Optics Communications, 2004, 238(1-3): 123-137.

[110] Tenenbaum, J. B., de Silva, V., Langford, J. C., A global geometric framework for nonlinear dimensionality reduction [J], Science, 2000,290(5500): 2319-2322.

[111] Tian, Q., Barbero, M., Gu, Z.-H. et al., Image classification by the foley-sammon transform [J], Optical Engineering, 1986,25(7): 834-840.

[112] Toliás, Y. A., Panas, S. M., Tsoukalas, L. H., Generalized fuzzy indices for similarity matching [J], Fuzzy Sets And Systems, 2001,120(2): 255-270.

[113] Torgerson, W. S., Multidimensional scaling of similarity [J], Psychometrika, 1965,30(4): 379-393.

[114] Tversky, A., Features of similarity [J], Psychological Review, 1977,84(4):





327-352.

[115] Tversky, A., Gati, I., similarity, separability, and the triangle inequality [J], Psychological Review, 1982,89(2): 123-154.

[116] Tversky, A., Krantz, D. H., The dimensional representation and metric structure of similarity data [J], Journal of Mathematical Psychology, 1970,7(3): 4-34.

[117] Wan, H. L., Hu, H., Shi, Z. Z., Image similarity measurement using max weighted bipartite matching, IEEE International Conferences on Info-tech & Info-net, Beijing, 2001.

[118] Wang, J. Z., Li, J., Wiederhold, G., SIMPLcity: Semantics-sensitive integrated matching for picture libraries [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2001,23(9): 947-963.

[119] Weng, S. F., Zhang, C. S., Lin, Z. G., Exploring the structure of supervised data by Discriminant Isometric Mapping [J], Pattern Recognition, 2005,38(4): 599-601.

[120] White, D., Jain, R., Similarity indexing: Algorithms and performance, Proc. SPIE Storage and Retrieval for Image and Video Databases, 1996.

[121] Winter, S., Location similarity of regions [J], Isprs Journal Of Photogrammetry And Remote Sensing, 2000,55(3): 189-200.

[122] Wu, P., Manjunath, B. S., Newsam, S. et al., A texture descriptor for browsing and similarity retrieval [J], Signal Processing-Image Communication, 2000,16(1-2): 33-43.

[123] Wu, Y., Huang, T. S., Towards self-exploring discriminating features for visual learning [J], Engineering Applications Of Artificial Intelligence, 2002,15(2): 139-150.

[124] Yang, J., Frangi, A. F., Yang, J. Y. et al., KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition [J], IEEE Transactions On Pattern Analysis And Machine Intelligence, 2005,27(2): 230-244.

[125] Yang, J., Jin, Z., Yang, J. Y. et al., Essence of kernel Fisher discriminant: KPCA plus LDA [J], Pattern Recognition, 2004,37(10): 2097-2100.

[126] Yang, J., Li, Q., Zhuang, Y. T., Towards data-adaptive and user-adaptive image retrieval by peer indexing [J], International Journal Of Computer Vision, 2004,56(1-2): 47-63.



- [127] Yang, Q., Tang, X., Recent Advances in Subspace Analysis for Face Recognition, SINOBIOMETRICS, 2004, 275-287.
- [128] Young, F. W., Quantitative Analysis of Qualitative Data [J], Psychometrika, 1981, 46(4): 357-388.
- [129] Yu, H., Yang, H., A direct LDA algorithm for high-dimensional data - with application to face recognition [J], Pattern Recognition, 2001, 34(10): 2067-2070.
- [130] Zhang, C., Chen, T. S., An active learning framework for content-based information retrieval [J], IEEE Transactions On Multimedia, 2002, 4(2): 260-268.
- [131] Zhang, Y. G., Zhang, C. S., Zhang, D., Distance metric learning by knowledge embedding [J], Pattern Recognition, 2004, 37(1): 161-163.
- [132] Zhao, T., Tang, L. H., Ip, H. H. S. et al., On relevance feedback and similarity measure for image retrieval with synergetic neural nets [J], Neurocomputing, 2003, 51: 105-124.
- [133] Zhao, W., Chellappa, R., Krishnaswamy, A., Discriminant Analysis of Principal Components for Face Recognition, Proc. Third IEEE International Conference Automatic Face and Gesture Recognition, 1998, 336-341.
- [134] Zhou, X. S., Huang, T. S., Edge-based structural features for content-based image retrieval [J], Pattern Recognition Letters, 2001, 22(5): 457-468.
- [135] Zhou, X. S., Huang, T. S., Relevance feedback in image retrieval: A comprehensive review [J], Multimedia Systems, 2003, 8(6): 536-544.
- [136] 常建芳, Markman, A. B., 莫雷, 归类的相似性对比模型研究终述 [J], 心理科学, 2004, 27(2): 423-425.
- [137] 陈学青, 罗航哉, 薛向阳 et al., 基于格矢量量化和倒排文件的快速图像检索方法 [J], 电子学报, 2000, 28(8): 94-96.
- [138] 方涛, 李德仁, 龚建雅, GeoImageDB 多分辨率无缝影像数据库系统的开发与实现 [J], 武汉测绘科技大学学报, 1999, 24(3): 189-193.
- [139] 李向阳, 潘云鹤, 基于椭圆的形状描述与检索方法 [J], 中国科学 E 辑, 2002, 32(5): 674-684.
- [140] 戚飞虎, 赵同, Horace, H., 基于协同神经网络的仿射不变性检索相似性度量方法 [J], 红外与毫米波学报, 2002, 21(5): 327-331.
- [141] 秦斌杰, 庄天戈, 利用相关比相似性测度多分辨率配准 MR 和 PET 医学图像的方法 [J], 中国生物医学工程学报, 2003, 22(1): 4-5.



[142] 吴信才, 郭玲玲, 李军, RDBMS 和 COM 的海量遥感影像数据的管理和 Web 发布 [J], 中国图象图形学报, 2002, 7(4): 384-387.

[143] 张永平, 赵荣椿, 郑南宁, 基于变分的图像分割算法 [J], 中国科学 E 辑, 2002, 32(1): 133-144.

[144] 章毓晋, 图象分割, 科学出版社, 2001.

[145] 章毓晋, 基于内容的视觉信息检索, 科学出版社, 2003.

## 攻读博士学位期间发表或录用的学术论文

- (1) Tang Hong, et al., Laplacian Linear Discriminant Analysis, **Pattern Recognition**, Vol.39, No.1, pp.136-139, 2006 [SCI]
- (2) Tang Hong, et al., Nonlinear Discriminant Mapping Using the Laplacian of a graph, **Pattern Recognition**, Vol.39, No.1, pp.156-159, 2006 [SCI]
- (3) Tang Hong, et al., Spectral Similarity Measure based on Fuzzy Feature Contrast Model, **Optics Communications**, Vol.238, No.1-3, pp.123-137, 2004 [SCI]
- (4) Tang Hong, et al., Intra-dimensional Feature Diagnosticity in Fuzzy Feature Contrast Model, **Information Science**, (Accept With Minor Revision) [SCI]
- (5) Tang Hong, et al., Unsupervised Linear Discriminant Analysis, **Journal of Shanghai Jiao Tong University** (English Edition), (Accepted) [EI]
- (6) 唐宏等, 基于完全波形的反射和吸收特征的非线性光谱相似性度量与检索, **光谱学与光谱分析**, 25 (2): 307-310 FEB 2005
- (7) 唐宏等, 光谱角制图模型的误差源分析与改进算法, **光谱学与光谱分析**, 2005, 25 (8): 1180-1183
- (8) 唐宏等, 基于弱语义嵌入的图像相似性度量, **高技术通讯** (录用) [EI]
- (9) Tang Hong et al. Nonlinear Spectral Similarity Measure, **IEEE International Geosciences and Remote Sensing Symposium (IGARSS)**, 20-24 Sept. 2004, Anchorage Alaska USA [EI]
- (10) Tang Hong, et al., Image Feature Ratio Model: Similarity Measures with Both Semantic and Visual Features in Interactive Image Retrieval, **The Fourth International Symposium on Multispectral Image Processing and Pattern Recognition**, WuHan [EI]



## 致 谢

时光如白驹过隙,似乎漫长而短暂的博士求学生活已接近尾声。回首之间,无数亲人、师长、同学和朋友的支持、鼓励、关怀和帮助仍历历在目。没有他们无私的帮助,很难想象我是否能如期完成这一论文。

首先要衷心感谢我的导师施鹏飞教授。在我博士学习期间,施教授不仅给我提供了一个良好宽松的研究环境,在平时生活中也给予的关怀和帮助。从他身上,我不但学到治学方法,而且学到很多做人的道理,施教授的孜孜以求的治学态度和高尚的做人风范是我毕生的楷模。

需要特别感谢方涛教授。本文研究得到了方涛教授主持的国家自然科学基金项目“基空间数据挖掘的若干理论与技术问题的研究”、国家 863 项目“基于内容的遥感影像相似性检索理论与关键技术研究”和上海市科委项目“面向 Tb 量级海量空间信息的存储管理系统”的支持。方教授刻苦钻研、执着追求的精神不断地激励我克服科研中各种困难,在此对他表示衷心的感谢和崇高的敬意。

感谢南京师范大学盛业华教授在我学习期间给与的指导和帮助。此外,还要感谢本所的各位老师:刘重庆教授、刘允才教授、杨杰教授、杨新教授、胡福乔副教授、赵宇明副教授、严京旗老师、霍宏老师、朱莉莉老师、吴妍青老师和吴沂军老师等,他们在我的博士生学习期间都曾从不同方面给予过很大的帮助。

衷心感谢曾经在实验室朝夕相处的同学,他们是杜培军、吴颖谦、允疆、何彬彬、孙庆先、石琴琴、刘春婷、柳稼航、许刚、李楠、余杰千、田明杰、游文建、谭睿、柴可夫、钱晋、王晓东、吴海涛、韩博慧等。我将终身铭记他们曾给予我的关心和帮助。还要感谢我的同班和同宿舍同学,同窗情意终生难忘,一起度过的时光将成为我无法忘记的回忆。

最后,我要衷心感谢我的父母,在我长达二十多年求学生涯中,他们无私支持始终是激发我不懈前进的动力和源泉。特别要感谢我的妻子,是她与我共度这一段艰苦而清贫的生活,是她热情的鼓励、无私的支持和无微不至的照顾支持着我圆满完成本文的工作。

唐 宏

2006 年 1 月 15 日

# 上海交通大学学位论文答辩决议书

申请者	唐 宏	所在学科(专业)	模式识别与智能系统
论文题目	图像相似性模型、算法与应用研究		
答辩日期	2006-01-19	答辩地点	上海交通大学闵行校区电信大楼2号楼412室

## 答辩委员会成员

担任职务	姓名	职称	所在工作单位	备注	签名
主席	刘允才	教授	上海交通大学	答辩主席	刘允才
委员	邵世煌	教授	东华大学	无	邵世煌
委员	席裕庚	教授	上海交通大学	无	席裕庚
委员	方涛	教授	上海交通大学	无	方涛
委员	徐孝涵	教授	国家自然科学基金委	无	徐孝涵
委员	陈雨良	研究员	上海技术物理研究所	无	陈雨良
委员	施鹏飞	教授	上海交通大学	导师	施鹏飞

## 评语和决议:

图像相似性模型、算法与应用的研究是计算机视觉领域中经典而至今尚未得到很好解决的研究课题。作者就图像相似性、特征相似性、匹配相似性和变换相似性等问题进行了研究。论文具有重要的理论意义和应用价值。论文的研究工作取得了如下结果。

提出了拉普拉斯线性判别分析算法和非线性判别排列算法,将线性判别的思想嵌入到算法中,有效地从非线性数据集中获取易于分类的排列结果;分析了模糊特征对比模型中特征强度与相似性之间的强相关性,提出集成语义和视觉特征的图像特征比率模型,不仅在视觉特征相似性度量时削弱了特征强度与相似性之间的强相关性,而且能协调语义和视觉特征在图像整体相似性度量中的作用;提出一个集成反射和吸收特征的光谱特征对比相似性模型,部分成果在遥感图像检索得到了应用。

论文立论正确,条理清楚,层次分明,实验数据可靠,取得了创新性的成果。论文工作体现了作者已经掌握本专业的坚实宽广的基础理论和系统深入的专门知识,独立从事科研能力强。

## 表决结果:

该同学在答辩过程中表述清楚,能正确回答所提问题,经答辩委员会无记名投票,一致同意通过论文答辩。建议校学位委员会授予工学博士学位。

答辩委员会主席 刘允才 (签名)

2006年1月19日