

分类号 TP3

密级

UDC

编号

# 中国科学院研究生院

## 博士学位论文

流形学习方法及其在人脸识别中的应用研究

王瑞平

指导教师 高文 教授

中国科学院计算技术研究所

申请学位级别 工学博士 学科专业名称 计算机应用技术

论文提交日期 2010 年 4 月 论文答辩日期 2010 年 5 月

培养单位 中国科学院计算技术研究所

学位授予单位 中国科学院研究生院

答辩委员会主席 查红彬 教授

## 声 明

我声明本论文是我本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，本论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

作者签名: 日期:

## 论文版权使用授权书

本人授权中国科学院计算技术研究所可以保留并向国家有关部门或机构送交本论文的复印件和电子文档，允许本论文被查阅和借阅，可以将本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、汇编本论文。

(保密论文在解密后适用本授权书。)

作者签名: 导师签名: 日期:



## 摘要

信息技术和互联网的飞速发展使得人们可以获取日益丰富的多媒体资源，包括大量的图片、视频、文本以及声音数据等。针对海量数据进行快捷有效的处理，从中提取用户所需要的有价值信息，正在成为机器学习、模式识别和计算机视觉领域的研究者们所共同关注的问题。以具体的人脸识别问题为例，随着网络资源的普及和视频采集设备的发展，研究者可以为每个人收集其不同时期、不同场景、涵盖不同光照模式和姿态变化等各种条件下的多幅图片，并构建规模可观的人脸数据库。如何充分有效的挖掘利用大规模数据库当中的有用信息，以设计性能稳定的高精度人脸识别分类器，则给研究者提出了很大的挑战。从模式识别的角度来看，需要解决如下两个基本问题：第一，如何从人脸图像数据中提取有效、紧致的特征表示；第二，如何针对数据集的分布特性设计合理有效的分类算法。针对人脸图像数据的表示，其核心可以归结为高维数据的降维和特征提取问题。数据降维可以看作是从原始数据中挖掘有效精简信息的过程，这一过程不仅可以去除冗余从而发现数据的低维本质属性，而且可以简化后续数据处理的复杂度。针对人脸图像数据的分类，大规模数据的获取极大推动了近年来基于图像集合进行人脸识别的研究进展。这一问题的关键难点在于，对图像集合中的样本分布进行合理有效的建模，并根据所建的模型综合利用多样本提供的信息进行集合的分类。

本文正是围绕上述两个基本问题，分别在理论研究和应用研究两个层面开展工作，在推进和完善理论研究的同时，将理论研究成果应用于解决模式识别和计算机视觉中的实际问题。在理论研究层面，本文主要从流形学习的角度研究新型高效的非线性降维算法，特别是在流形的局部线性模型表示、解析映射函数学习、内在变化模式刻画等方面开展相关研究。在应用研究层面，本文针对基于图像集合的人脸识别问题，主要从图像集合的流形建模、流形之间的距离定义和计算、多流形判别式分类学习等方面开展相关研究。本文在流形学习算法理论和图像集合人脸识别应用的研究中取得的主要贡献如下：

(1) 提出了一种最大线性嵌入流形学习方法，用以解决非线性降维问题。该方法通过学习一个参数式可逆的解析映射函数，可以将流形数据从原始高维观察空间保距地投影到一个全局低维嵌入空间中。算法从几何直觉的角度，引入一种合理有效的局部线性分块的定义，称为最大线性分块，分块的非线性程度通过块内样本间测地距离与欧氏距离的偏差来度量。首先，原始的流形采样数据集合被分解为一组最大线性分块的组合，每个分块采用一个局部线性模型来建模表示。接下来，算法从每个局部模型内部随机选取一定数量的地标点，利用其测地距离进行多维尺度变换运算，进而得到最终的保距低维嵌入空间并完成局部模型的全局对齐。这一对齐过程称为基于地标点的全局对齐算法。该对齐算法避免了传统方法中存在的局部极值与大规模迭代优化这两个不足，并且只需要解小规模的特征分解问题就能得到有效的闭形式解。针对流形内在结构的描述，本文

算法不仅提供了一种快速有效的本质维数估计方法，而且可以显式建模流形观察数据的内在变化模式，从而可以广泛应用于多种实际问题。

(2) 提出了流形-流形距离的理论框架，将其应用于基于图像集合的人脸识别问题。在该理论框架中，属于每一类别的图像集合采用非线性表观流形来建模，图像集合的分类问题从而形式化为两个流形之间的距离计算问题，即流形-流形距离。注意到实际问题中的图像集合可能表示为三种模式层次，即点、子空间、流形，本文系统研究了这三种模式之间的各种距离，并将它们形式化在一个统一的多层次流形-流形距离框架下。具体地，将流形表示为一组局部线性模型的组合，其中各局部模型采用子空间来刻画。这样，两个流形之间的距离就可以转化为它们各自对应的子空间对之间的距离综合。针对流形-流形距离的计算，本文从理论上和实验上探讨了算法各个模块的多种不同的定义方案，包括局部线性模型构建、局部模型距离度量、局部距离的全局综合等。通过图像集合人脸识别的实验验证表明，本文提出的流形-流形距离作为一种通用的图像集合相似性度量，一致地优于其它对比的非判别式算法，并且达到了与当前领先的基于集合的判别式学习算法可比的性能。

(3) 提出了一种流形判别分析方法，将其应用于基于图像集合的人脸识别问题。该方法从监督学习的角度，将图像集合分类问题形式化为面向分类的多流形学习问题，通过在流形-流形距离框架内引入“最大化流形间隔”的思想来学习判别式的线性投影空间，最终在该空间内对不同类别的流形进行匹配来完成图像集合的分类。具体地，流形判别分析方法采用图嵌入的框架来构建图模型，采用本征图和惩罚图分别刻画流形的类内紧致性与类间可分性。通过设计的最优化目标函数来学习判别式的嵌入空间，使得具有不同类别标号的流形之间能够更好地区分，同时每个流形内部局部区域的数据紧致性得到增强。在分类识别阶段，通过将待测试的流形数据投影到该判别空间中，就可以在流形-流形距离框架下进行更加可靠的分类。通过人脸识别和物体分类两个任务的实验对比表明，本文提出的流形判别分析方法取得了与当前领先的方法相当的性能，同时具有很高的训练和测试效率。

综上所述，本文针对流形学习的理论与应用开展了广泛深入的研究，在传统的单流形学习框架下提出了一种有效的非线性降维算法，该算法在数据降维的功能与效率两方面都展示出良好的特点。本文进一步将该流形学习算法推广到多流形学习问题，在图像集合人脸识别任务中取得了成功的应用，从而在流形学习实用化方面进行了有益的探索。

**关键词：**流形学习；非线性降维；最大线性嵌入；流形-流形距离；流形判别分析；基于图像集合的人脸识别

## **Manifold Learning and Its Application to Face Recognition**

Wang Ruiping (Computer Applications)

Advisor: Prof. Wen Gao

The rapid development of information technology and Internet makes it possible to obtain increasingly rich multimedia resources, including a large number of images, videos, texts and voice data. It has been becoming a difficult problem to process the explosive data in an effective and efficient way. This has also placed a great challenge to researches in the areas of machine learning, pattern recognition and computer vision. Take the task of face recognition for example. Nowadays, for each person it is easy to collect multiple facial images, which are captured in different scenes and times, covering different lighting conditions and pose variations. With the masses of images, we can construct considerably large-scale face databases. In this case, it raises a great challenge to effectively exploit the large databases and design stable face recognition classifiers with high precision. From the perspective of pattern recognition, we need to solve the following two basic problems. The first one is how to extract informative and compact feature representation from face images. The second is how to develop reasonable and effective classification algorithms according to the data distribution in the database. For face image representation, the basic problem can be cast as dimensionality reduction and feature extraction. Dimensionality reduction mainly involves recovering compact, informative and meaningful low-dimensional structures hidden in raw high-dimensional data for subsequent operations, such as classification and visualization. For face images classification, with large-scale databases, the research in recent years has encountered a new application, i.e., face recognition with image set. One of the key problems in this application is to model the image set in a reasonable and effective way and then perform set classification using the information from multiple samples in the set.

This thesis mainly conducts research on the above two basic problems from both theoretical and application views. In the aspect of theoretical research, we focus on manifold learning approach to nonlinear dimensionality reduction. The research mainly addresses the following problems, including local linear model representation of manifold, the learning of analytic mapping function and the characterization of intrinsic modes of variation of the manifold. In the aspect of application research, we focus on the task of face recognition with image set. The research mainly addresses the following problems, including the manifold modeling of image set, the definition and computation of the distance between manifolds and the classification-orientated multi-manifold learning. The main contributions of this thesis are outlined as follows.

We propose a simple but effective nonlinear dimensionality reduction algorithm, named Maximal Linear Embedding (MLE). MLE learns a parametric mapping to recover a single

global low-dimensional coordinate space and yields an isometric embedding for the manifold. Inspired by geometric intuition, we introduce a reasonable definition of locally linear patch, Maximal Linear Patch (MLP), which seeks to maximize the local neighborhood in which linearity holds. The input data are first decomposed into a collection of local linear models, each depicting an MLP. These local models are then aligned into a global coordinate space, which is achieved by applying MDS to some randomly selected landmarks. The proposed alignment method, called Landmarks-based Global Alignment (LGA), can efficiently produce a closed-form solution with no risk of local optima. It just involves some small-scale eigenvalue problems, while most previous aligning techniques employ time-consuming iterative optimization. Compared with traditional methods such as ISOMAP and LLE, our MLE yields an explicit modeling of the intrinsic variation modes of the observation data. Extensive experiments on both synthetic and real data indicate the effectiveness and efficiency of the proposed algorithm.

We address the problem of classifying image sets, each of which contains images belonging to the same class and typically covering large appearance variations. By representing each image set as a manifold, we formulate the problem as the computation of Manifold-Manifold Distance (MMD). Observing that an image set can come in three pattern levels, i.e., point, subspace and manifold, we systematically study the distances among the three levels and innovatively formulate them in a general multi-level MMD framework. Specifically, we express a manifold by a collection of local linear models, each depicted by a subspace. MMD is then converted to integrating the distances between pair of subspaces respectively from one of the involved manifolds. For the computation of MMD, we theoretically and experimentally investigate various different configurations of its ingredients. The proposed method is evaluated on the task of face recognition with image set (FRIS), where identification is achieved by seeking the minimum MMD between two image sets. Our experiments demonstrate that 1) as a general set similarity measure, the proposed MMD consistently outperforms other competing methods without discriminative learning and 2) the MMD is also promisingly comparable to the state-of-the-art discriminant methods over sets.

We propose a novel discriminative learning method, called Manifold Discriminant Analysis (MDA), to solve the problem of image set classification. By modeling each image set as a manifold, we formulate the problem as classification-oriented multi-manifolds learning. Aiming at maximizing “manifold margin”, MDA seeks to learn an embedding space, where manifolds with different class labels are better separated, and local data compactness within each manifold is enhanced. Based on the framework of graph embedding, we design an intrinsic graph to characterize the within-class compactness and another penal graph to reflect the between-class separability. As a result, new testing manifold can be more reliably classified

in the learned embedding space. The proposed method is evaluated on the tasks of object recognition with image sets, including face recognition and object categorization. Comprehensive comparisons and extensive experiments demonstrate the effectiveness of our method.

In summary, this thesis conducts extensive research in both theory and application of manifold learning. Under the traditional single-manifold learning framework, we propose an effective nonlinear dimensionality reduction algorithm, which demonstrates superior characteristics in terms of both effectivity and efficiency. We further generalize the algorithm to the problem of multiple manifolds learning and apply to the task of face recognition with image set. This generalization makes a useful exploration in the practical applications of manifold learning.

**Keywords:** manifold learning, nonlinear dimensionality reduction, maximal linear embedding, manifold to manifold distance, manifold discriminant analysis, face recognition with image set



# 目 录

<b>摘 要</b>	I
<b>目 录</b>	VII
<b>图目录</b>	XI
<b>表目录</b>	XIII
<b>第一章 绪论</b>	1
1.1 课题研究背景及意义	1
1.2 流形学习概述	3
1.2.1 线性降维方法	3
1.2.2 非线性降维方法	5
1.2.3 综合对比与讨论	11
1.2.4 现有方法的主要问题	13
1.3 基于图像集合的人脸识别概述	14
1.4 本文的研究内容与主要贡献	17
1.5 论文的组织	18
<b>第二章 基于最大线性嵌入的流形学习方法</b>	21
2.1 引言	21
2.2 算法基本思想	22
2.3 最大线性嵌入算法	23
2.3.1 最大线性分块	24
2.3.2 局部线性模型	28
2.3.3 全局对齐算法	28
2.3.4 讨论	34
2.4 实验结果	36
2.4.1 三维人工数据库实验	36
2.4.2 合成人脸数据库实验	41
2.4.3 真实人脸数据库实验	45
2.5 本章小结	47
<b>第三章 基于流形-流形距离的图像集合人脸识别</b>	49

---

3.1 引言 .....	49
3.2 问题形式化框架 .....	50
3.2.1 定义在点和子空间上的距离 .....	50
3.2.2 定义在流形上的距离 .....	52
3.3 流形-流形距离 .....	53
3.3.1 局部线性模型构建 .....	53
3.3.2 局部模型距离度量 .....	54
3.3.3 局部距离的全局综合 .....	58
3.4 讨论 .....	62
3.4.1 关于度量准则的分析 .....	62
3.4.2 算法复杂度分析 .....	63
3.4.3 与前人工作对比 .....	63
3.5 实验结果 .....	64
3.5.1 实验数据库 .....	64
3.5.2 对比方法与参数设置 .....	66
3.5.3 实验对比结果与分析 .....	67
3.5.4 算法各模块评估 .....	69
3.6 本章小结 .....	72
<b>第四章 基于流形判别分析的图像集合人脸识别 .....</b>	<b>73</b>
4.1 引言 .....	73
4.2 算法基本思想 .....	74
4.3 流形判别分析 .....	74
4.3.1 问题形式化 .....	74
4.3.2 局部线性模型构建 .....	76
4.3.3 判别函数的学习 .....	76
4.3.4 基于流形判别分析的分类 .....	78
4.4 讨论 .....	78
4.4.1 与基于单样本的方法对比 .....	78
4.4.2 与基于集合的方法对比 .....	79
4.5 实验结果 .....	79
4.5.1 实验数据库 .....	79
4.5.2 对比方法与参数设置 .....	80
4.5.3 实验对比结果与分析 .....	82
4.6 本章小结 .....	82
<b>第五章 结束语 .....</b>	<b>85</b>

5.1 本文工作总结 .....	85
5.2 下一步研究方向 .....	86
<b>参考文献 .....</b>	<b>89</b>
<b>攻读博士学位期间发表的论文 .....</b>	<b>i</b>
<b>致 谢 .....</b>	<b>iii</b>
<b>作者简介 .....</b>	<b>vii</b>



## 图目录

图 1.1 ISOMAP 算法示意图 .....	6
图 1.2 LLE 算法示意图 .....	6
图 1.3 降维方法分类结构图 .....	11
图 1.4 基于图像集合的人脸识别问题示意图 .....	15
图 2.1 V 型(V-like shape)流形数据示意图 .....	22
图 2.2 最大线性分块(MLP)的直观示意图 .....	23
图 2.3 层次分裂聚类算法(HDC)示意图 .....	25
图 2.4 V 型流形的 HDC 算法聚类结果 .....	27
图 2.5 基于地标点的全局对齐算法(LGA)示意图 .....	30
图 2.6 瑞士卷(swiss-roll)流形的 HDC 算法聚类结果 .....	37
图 2.7 S 曲线(s-curve)流形的 HDC 算法聚类结果 .....	37
图 2.8 不同算法对邻域参数 $k$ 的敏感性对比结果 .....	38
图 2.9 算法 MLE 与 LLC 采用不同数量局部模型的对比结果 .....	39
图 2.10 MLE 算法采用不同数量地标点样本的运行结果 .....	40
图 2.11 验证 MLE 算法的三个理论环节 .....	41
图 2.12 ISOFace 人脸数据库上的低维嵌入结果 .....	42
图 2.13 MLE 算法针对 ISOFace 数据库得到的重构人脸 .....	43
图 2.14 MLE 与 LLC 算法采用不同维数进行重构的结果对比 .....	43
图 2.15 MLE 算法学习得到的 LLEFace 人脸流形低维结构 .....	45
图 2.16 特征脸(Eigenfaces)与语义脸(Latentfaces)对比示意图 .....	46
图 3.1 物体识别任务的三种模式层次（点、子空间、流形）所构成的层级结构 .....	50
图 3.2 定义在点和子空间上的三种距离 .....	51
图 3.3 定义在流形上的三种距离 .....	51

图 3.4 流形-流形距离(MMD)的计算示意图 .....	53
图 3.5 人脸数据库对应的局部模型示意 .....	54
图 3.6 融合数据样例与变化模式度量的 SSD 距离示意 .....	55
图 3.7 人脸数据对应局部模型之间的相似度 .....	57
图 3.8 局部距离的全局综合方案示意 .....	59
图 3.9 MMD 方法对比实验采用的人脸数据库示意 .....	65
图 3.10 人脸图像集合的非线性分值曲线 .....	67
图 3.11 MMD 采用不同聚类方法建立局部模型的对比 .....	69
图 3.12 MMD 采用不同 SSD 距离度量的对比 .....	71
图 3.13 MMD 与 MSM 各自采用不同数量主夹角的性能对比 .....	71
图 3.14 MMD 采用不同的局部距离全局综合方案的对比 .....	72
图 4.1 MDA 方法采用的图嵌入框架示意图 .....	75
图 4.2 物体分类数据库 ETH-80 的样本示意图 .....	80

## 表目录

表 1.1 不同降维方法的参数和计算复杂度对比 .....	12
表 2.1 算法 1：层次分裂聚类算法(HDC)流程 .....	26
表 2.2 算法 2：可视化算法（前向映射）流程 .....	32
表 2.3 算法 3：重构算法（后向映射）流程 .....	32
表 2.4 ISOFace 数据库的重构误差与计算时间（单位：秒）对比 .....	44
表 2.5 LLEFace 数据库的重构误差对比 .....	47
表 3.1 MMD 采用不同全局综合方案对比 .....	62
表 3.2 MMD 对比方法的识别结果 .....	68
表 3.3 不同聚类算法计算时间对比（单位：秒） .....	70
表 4.1 MDA 方法的参数设置 .....	81
表 4.2 MDA 对比方法的识别结果 .....	81



# 第一章 绪论

科学研究的主要任务可以总结为：对未知事物或者现象进行观察，通过对观察数据进行细致的分析，进而探求隐藏在观察表象背后的事物本质。具体到模式识别和机器学习等领域，如何从数据中学习和发现内在的规律性一直是它们的一个主要的研究目标。

## 1.1 课题研究背景及意义

随着信息时代的来临，人们在享受日益丰富的数据资源的同时，也面临着数据爆炸所带来的种种挑战。一个具体的表现就是人们所需要面对和处理的数据较以往有了显著的变化，其主要特点可归纳为：高数据量、高维数、高数据增长率、非结构化、以及不能被人的感知所单独处理。从这些海量数据资源中，如何快捷有效地甄别冗余数据，进而提取人们所需要的有价值的信息，正在成为一个紧迫而困难的问题。

以近年来受到广泛关注的人脸识别问题为例，其研究的基本对象是人脸图像模式。互联网的普及和图像视频采集存储设备的发展，为研究者提供了大量丰富的图像数据集合。然而如何充分有效的挖掘这些数据当中的有用信息，则给研究者提出了很大的挑战。从模式识别的角度来说，我们主要面临两方面的问题：第一，如何从人脸图像数据中提取有效、紧致的特征表示；第二，如何针对数据集的分布特性设计合理有效的分类算法。

针对人脸图像数据的有效表示，其核心可以归结为高维数据的特征提取和降维问题。数据降维可以看作是从原始数据中挖掘有效精简信息的过程，这一过程不仅可以去除冗余从而发现高维数据的低维本质属性以利于分析，而且可以简化后续数据处理阶段的复杂度以利于存储和计算。从数据集分布的统计意义出发，传统多元统计分析方法通常假设数据集的全局分布是线性的，从而采用线性降维方法对数据进行处理。然而，全局线性分布的假设在对问题进行简化处理的同时，却很可能无法正确反映数据分布的内在结构。因此，更一般的假设应该是数据在原始观察空间中呈现为低维非线性的分布特性，从而需要采用非线性降维方法来分析。流形学习作为一种新兴的非线性降维方法，以其对数据本质结构的有效刻画和紧致表示，近年来在模式识别、机器学习和计算机视觉等领域获得了蓬勃发展[5][66][80]。特别是针对人脸识别问题，在传统的线性子空间方法无法奏效的情况下，基于流形学习思想的一系列新型方法如[12][31][32][99][101]提供了很好的解决方案。

针对人脸图像数据的分类，根据分类采用的基本单元的不同，可以将人脸识别问题分为基于单幅静态图像的识别与基于多幅图像构成的集合的识别。基于单幅图像的人脸识别问题经过近二十年的研究已经取得了多方面的进展，包括特征提取与融合算法、分类器设计算法、多分类器组合算法等。然而，随着应用需求的发展，传统基于静态图像的识别算法在很多实际场景中表现不够稳定，比如针对非受控条件下的监控场景，这类

算法的性能很容易受到图像质量、光照和姿态等因素的影响。从另一个方面来看，这类型实际场景虽然提供的图像质量不够好，但是其图像数量通常比较多，因此如何有效利用每类个体的多幅图像中包含的信息就成为新的研究点，这一需求正好促进了基于多幅图像所构成的集合进行人脸识别问题的研究[23][42][44][45][51][58][70][75][87][106]。基于图像集合的人脸识别问题的难点在于，如何有效地刻画与建模集合中多幅图像的分布，并根据所建的模型综合利用多样本提供的信息。近年来，典型的做法主要有两类，分别从参数式与非参数式两个角度对图像集合建模，前者通常利用概率分布函数来表示图像集合的样本分布，后者则根据图像集合中样本的本质变化模式将其建模为线性子空间或者非线性的流形。

本文的研究课题正是从上述两个基本问题出发展开的，具体可以划分为理论研究和应用研究两个层面：

(1) 理论研究层面：针对模式识别中广泛存在的非线性降维问题，本文从流形学习的角度出发，研究新型高效的降维学习算法。流形学习经过近十年发展，研究者们从不同的角度基于不同的几何现象启发提出了为数众多的算法。然而，这些方法在具体的模式识别和计算机视觉应用中却存在很多困境，远没有发挥出人们最初对它们期望的效果，特别是在很多方面其性能都无法与传统的线性降维方法相比。这些方法的局限主要包括：没有提供显式的解析映射函数、参数较多且对参数设置敏感、计算复杂度高等。本文所研究的新型流形学习算法正是从这些方面克服了现有方法的不足，并在实际人脸识别问题中获得了成功的应用。

(2) 应用研究层面。以基于图像集合的人脸识别为应用范例，本文将上述新型流形学习方法进行了推广。如前所述，当前的图像集合识别方法通常采用线性子空间或者非线性流形来对集合进行建模。本文从集合数据分布的广泛性和一般性出发，利用流形建模图像集合，并将图像集合的分类问题形式化为流形之间的匹配或者说距离计算问题。具体地，本文系统地研究了点、子空间和流形三种模式层次之间的距离关系，并将它们形式化在统一的多层次流形-流形距离框架下。针对流形-流形距离的计算问题，本文深入研究了该框架的各环节并提出了一套完整的计算流程。为了有效利用离线训练数据的信息，本文进一步从判别学习角度出发，将图像集合分类问题形式化为面向分类的多流形学习问题，通过在流形-流形距离框架内引入“最大化流形间隔”的思想来学习判别式的特征空间，通过在该空间内对不同类别的流形进行匹配来完成图像集合的可靠分类。

综上所述，流形学习方法的理论研究与其在人脸识别问题中的应用研究，对于推进模式识别、机器学习和计算机视觉等多个领域的研究进展都具有广泛的意义和价值。本章接下来首先简要回顾流形学习与图像集合人脸识别两个问题的相关研究进展，通过分析当前方法存在的主要问题引出本文的主要研究内容，并简要介绍本文在理论研究和应用研究两方面所作出的相应贡献。

## 1.2 流形学习概述

如前所述，计算机视觉和模式识别领域的很多应用问题都需要处理复杂高维的数据集，特别是需要从原始高维观察数据中寻找有意义的、紧致的内在低维数据分布结构，用于后续的数据处理分析，比如特征提取、分类、可视化等[48][49][55][56][71][99][101][102]。以人脸图像数据为例，假定待处理的数据是一个人在不同条件下采集的多幅人脸图像，这些图像涵盖了若干种不同的变化模式如姿态、光照和表情等。那么，如何从给定的这些观察图像中发现数据的内在变化模式，比如姿态的角度、光照的方向、表情的类别等参数？很显然，给定的原始数据来自于高维的图像空间，而期望的结果则是反映或者控制数据变化模式与自由度的低维参数。解决这类问题的有效工具就是各种降维方法。通常来说，假定观察数据的内在结构是输入样本空间的一个非线性的低维流形，该流形的自由度即对应数据本质的变化模式，这种情形下的降维问题则被称为流形学习。

所谓流形就是一般的几何对象的总称，比如各种维数的曲线曲面等；在数学上，流形被定义为一个满足局部欧几里德属性的拓扑空间。与传统降维方法一样，流形学习通过某种显式或者隐式的映射将高维空间中的数据在低维空间中重新表示；和以往方法不同的是，流形学习中对数据分布的假设，即上面所说的观察数据采样于一个潜在的流形上或者说对于这组数据存在一个潜在的流形。目前的流形学习方法，通常都是将原始数据投影到一个低维嵌入空间，即所谓的 *embedding space*，在这个空间内保持原始数据的一些几何属性。对于不同的方法，对于流形性质的要求也各不相同，从而也就产生了在流形假设下的各种不同性质的假设，比如保持流形测地距离关系[80]、保持流形的局部重构表示关系[66]、保持流形的局部近邻关系[5]等等。

形式上，流形学习可以看作是从一组观察数据中推导其产生式模型的过程。给定一组观察数据集合  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ， $\mathbf{x}_i \in \mathbb{R}^D$ ，其中  $N$  为样本数量， $D$  为观察数据的维数。假设这些观察数据采样自一个本质维数  $d$  ( $d < D$ ) 的平滑流形，那么流形学习的目标可以看作是寻找一个从观察空间到低维嵌入空间的映射： $F(\mathbf{X}) \rightarrow \mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ ， $\mathbf{y}_i \in \mathbb{R}^d$ ，以及其一对一的逆映射： $F^{-1}(\mathbf{Y}) \rightarrow \mathbf{X}$ ，这一映射函数应该同时保持流形数据的全局结构与局部几何关系。在很多算法中，映射函数  $F$  往往都不具有显式的表达形式，而仅仅通过隐含映射的方式给出训练数据集合的降维结果。在本文后续讨论中，假设采样数据集  $\mathbf{X}$  已经做过零均值处理。本节接下来简要回顾经典的线性降维和非线性降维方法，在非线性降维的框架内重点介绍若干种得到广泛应用的流形学习方法，最后从降维问题的目标和形式化以及性能评测等方面对这些方法进行深入讨论。

### 1.2.1 线性降维方法

线性降维方法通常假设处理的数据具有全局线性的分布，具有代表性的两个方法分别是主成分分析(principal component analysis, PCA)[38]和多维尺度变换(multidimensional scaling, MDS)[14]。

(1) 主成分分析(principal component analysis, PCA)[38]。PCA 是统计学中被广泛采用的数据降维工具，主要用于去除数据观察维数之间的统计相关性。在流形学习的框架下，PCA 的目标可以表述为寻找一组线性投影  $\mathbf{W}$ ，使之最小化原始观察数据经过降维处理之后再重构回原始空间时的重构误差，目标函数的数学形式如下：

$$\min J(\mathbf{W}) = \sum_{i=1}^N \left\| \mathbf{x}_i - \mathbf{W}\mathbf{W}^T \mathbf{x}_i \right\|^2, \text{ s.t. } \mathbf{W}^T \mathbf{W} = \mathbf{I}. \quad (1-1)$$

该目标函数通过对数据  $\mathbf{X}$  的协方差矩阵  $\mathbf{X}\mathbf{X}^T$  进行特征分解来求解，即  $\mathbf{X}\mathbf{X}^T \mathbf{W} = \lambda \mathbf{W}$ 。降维后的低维表示为：  $\mathbf{Y} = \mathbf{W}^T \mathbf{X}$ 。

(2) 多维尺度变换(multidimensional scaling, MDS)[14]。MDS 最早是在心理学领域发展起来的多元分析方法，主要用于做数据的可视化，其基本出发点是要在低维嵌入空间中保持任意点对之间对应于原始高维空间的距离关系。具体地，经典 MDS(classical scaling)方法的输入为点对之间的欧氏距离矩阵  $\mathbf{D}$ ，矩阵的元素  $d_{ij}$  表示原始空间中  $\mathbf{x}_i$  与  $\mathbf{x}_j$  之间的欧氏距离。算法优化的目标函数如下：

$$\min \phi(\mathbf{Y}) = \sum_{i,j} (d_{ij}^2 - \|\mathbf{y}_i - \mathbf{y}_j\|^2). \quad (1-2)$$

其中  $\|\mathbf{y}_i - \mathbf{y}_j\|$  为低维空间中样本  $\mathbf{y}_i$  与  $\mathbf{y}_j$  之间的欧氏距离。上述目标函数的最优解可以通过对原始高维数据的 Gram 矩阵  $\mathbf{K} = \mathbf{X}^T \mathbf{X}$  进行特征分解得到，其中 Gram 矩阵的元素可以通过对距离矩阵的元素进行双中心化(double-centering)操作来得到：

$$k_{ij} = -\frac{1}{2} (d_{ij}^2 - \frac{1}{n} \sum_l d_{il}^2 - \frac{1}{n} \sum_l d_{jl}^2 + \frac{1}{n^2} \sum_m d_{lm}^2). \quad (1-3)$$

由于在 classical scaling 方法中，高维空间数据点之间的距离度量通常使用欧氏距离，此时 MDS 与 PCA 互为对偶问题，二者的降维结果仅相差一个线性变换的关系[71]。事实上，MDS 就是理论上保持欧氏距离的一个经典方法，由于 MDS 得到的低维表示中心在原点，所以又可以说保持内积，即利用低维空间中的内积近似高维空间中的内积。

除了上述经典的 PCA 和 MDS 方法，近年来从流形学习的视角出发，研究者们提出一系列全新的线性降维算法，典型的方法包括局部保持投影(locality preserving projections, LPP)[32]，邻域保持嵌入(neighborhood preserving embedding, NPE)[31]，局部判别嵌入(local discriminant embedding, LDE)[12]，无监督判别投影(unsupervised discriminant projection, UDP)[101] 以及正交邻域保持投影(orthogonal neighborhood preserving projections, ONPP)[48]等。这些算法在保持流形数据的某些局部或者全局几何属性约束下，可以在一定程度上发现隐含在数据集当中的流形结构，并且在人脸识别和掌纹识别等实际问题应用中得到了验证。

总的来说，线性降维方法普遍具有模型简洁、对数据依赖性低以及计算高效等优点。然而，由于这类方法中固有的对数据全局线性分布的假设，在复杂数据分布模式条件下很难取得理想的效果。因此，非线性降维方法的研究就逐渐引起研究者们更多的重视。

## 1.2.2 非线性降维方法

非线性降维(nonlinear dimensionality reduction, NLDR)的历史最早可以追溯到经典算法 Sammon's mapping [69]。在此之后，很多非线性算法不断涌现出来，包括自组织映射(self-organizing maps, SOM)[46]，主曲线(principal curve)及其扩展[30][81]，自动编码神经网络(auto-encoder neural networks)[3][15]和产生式拓扑映射(generative topographic maps, GTM)[8]等。随着支持向量机(support vector machine, SVM)的成功广泛应用，基于核学习的降维方法逐渐兴起，比如核主成分分析(kernel principal component analysis, Kernel PCA)[62][72]，它们为非线性降维问题提供了一种不同以往的解决思路。这类算法的核心思想是，将在原始空间中非线性分布的数据，通过某种核函数投影变换到某个隐式的高维特征空间中，使得数据在该高维空间呈现全局线性分布，并进而可以采用线性降维算法的思路对特征空间内的投影数据进行处理。总的来说，与上述线性降维算法相比，非线性算法在降维功能方面取得了一定的改进，但是这些非线性算法同时失去了线性方法的很多优点，它们往往具有计算复杂度高、代价函数难以设计以及众多的参数需要调节等特点，这些不足在一定程度上限制了此类方法的应用范围。

最近十年来，针对非线性降维问题，研究者们受认知科学领域人脑感知信息方式的启发，开创了一条思路新颖、充满活力且具有深刻理论基础和广泛应用前景的研究路线，这就是建立在数据非线性流形分布假设基础上的流形学习框架[73]。大体来说，这个领域的经典算法都可以看作是将流形学习形式化为一个优化目标函数的问题，通过该目标函数来指定流形数据的哪些几何特性需要在降维过程中保持[86]。接下来，按照算法发表的时间顺序，回顾在模式识别和机器学习领域应用最为广泛的几种流形学习算法。

(1) ISOMAP(isometric feature mapping)算法[80]。该算法的目标是在其低维嵌入空间中保持流形上样本点对之间的测地距离(geodesic distance)，其理论基础为线性降维方法 MDS。不同的是，ISOMAP 在流形学习的理论框架内将经典 MDS 方法中的欧氏距离替换为刻画非线性流形上样本本质差异的测地距离。所谓的测地线，就是流形上加速度为零的曲线，等同于欧氏空间中的直线。ISOMAP 实际上就是把任意两点的测地距离作为流形的几何描述，采用 MDS 的框架来在理论上保持这些点对之间的测地距离。在实际计算中，由于测地距离无法精确获得，通常是采用样本点之间的最短距离来近似，相关的算法一般采用图论中的最短路径算法如 Dijkstra 算法，Floyd 算法等。ISOMAP 算法主要有三点不足：第一，算法的拓扑稳定性容易受到邻域图参数设置的影响，当邻域个数设置不合理时可能出现所谓的短路(short-circuiting)现象[2]；第二，算法对流形数据采样中出现的空洞区域(holes)比较敏感，这同样是由于空洞区域的存在会导致测地距离的计算偏差；第三，由于算法要求任意两点之间都存在测地线，因此需要保证流形的连通区域为凸区域。从上述分析可以看出，ISOMAP 算法的核心思想和成功关键在于利用测地距离来度量流形上的样本距离，而其降维性能的优劣也同样取决于测地距离的估算是否正确。图示 1.1 给出了该算法的示意图。

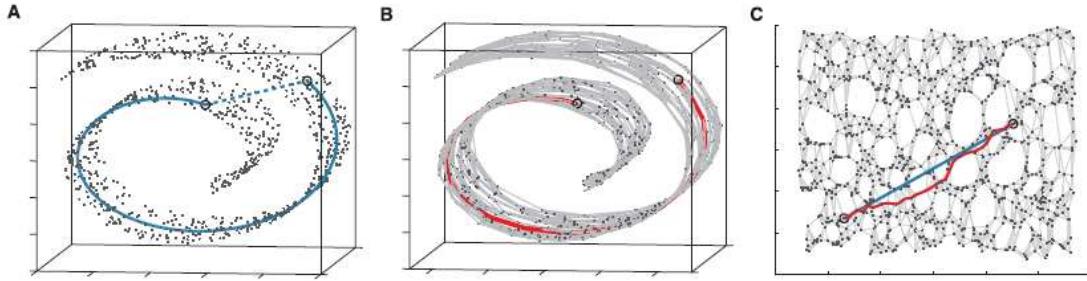


图 1.1 ISOMAP 算法示意图。A 图中流形上两点之间的蓝色虚线为欧氏距离，蓝色实线表示刻画样本本质差异的“测地距离”；B 图中的红色曲线为利用训练样本计算得到的样本间最短路径距离，用以近似其真实的测地距离；C 图为算法最终得到的降维嵌入空间，降维过程的目标就是保持样本间的最短路径距离，从而近似保持流形的测地距离[80]。

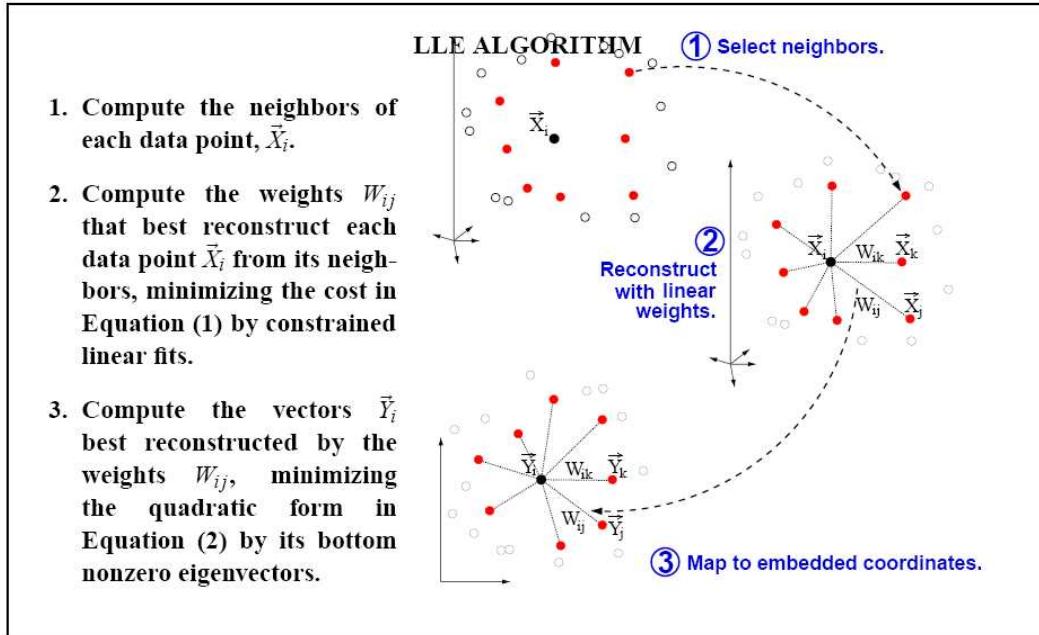


图 1.2 LLE 算法示意图。算法主要包括三个步骤：(1) 计算样本点的局部近邻；(2) 计算每个样本由其近邻线性重构的系数表示；(3) 计算保持了局部几何属性的低维嵌入空间[66]。

(2) LLE (locally linear embedding) 算法[66]。该算法的基本思想是非线性流形在其较小的局部区域上可以近似看成欧氏空间，即满足局部线性。因此，在每个样本所处的局部邻域内，该样本可以用其近邻点在最小二乘意义下进行最优的线性重构表示。LLE 就是将这种样本之间的近邻重构表示关系作为流形上局部几何属性的刻画。以此为动机，LLE 算法降维的目标空间就应该保持流形的这种局部几何，即在低维嵌入空间中的样本之间应该保持与原始空间一致的线性重构表示关系。具体地，针对流形上的每个样本  $x_i$ ，LLE 利用其  $k$  个最近邻  $x_{ij}$  进行线性重构来表示  $x_i$ ，各近邻样本对应的重构权重为  $w_{ij}$ ，最优权重的优化目标函数如下：

$$\min \epsilon(\mathbf{W}) = \sum_i \left\| \mathbf{x}_i - \sum_j w_{ij} \mathbf{x}_{ij} \right\|^2. \quad (1-4)$$

公式(1-4)的优化问题最终转化为线性方程组的求解。利用原始高维空间的重构权重作为约束, LLE 的降维目标即为寻求样本的低维表示, 以保持样本之间的重构关系不变, 其优化目标如下:

$$\min \phi(\mathbf{Y}) = \sum_i \left\| \mathbf{y}_i - \sum_j w_{ij} \mathbf{y}_{ij} \right\|^2. \quad (1-5)$$

该优化问题通过对稀疏矩阵  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$  进行特征分解来求解, 其中  $\mathbf{I}$  为单位矩阵,  $\mathbf{W}$  为以  $w_{ij}$  作为元素的权重矩阵。样本的降维坐标依次采用矩阵  $\mathbf{M}$  对应于最小若干个特征值的特征向量, 其中对应于零特征值的特征向量作为平凡解(trivial solution)被排除。LLE 算法的不足主要包括对流形空洞区域的不稳定, 对低维坐标进行归一化协方差约束所带来的全局结构变形等。图 1.2 给出了 LLE 算法的流程图。

(3) Laplacian Eigenmaps 算法[5]。该算法的基本思想是利用一个无向带权图来描述流形上样本点之间的近邻距离关系, 图的顶点对应于流形的样本, 顶点之间的边对应于样本之间的近邻关系, 边的权重对应于近邻样本之间的某种距离或者相似性度量。算法降维的目标就是通过图嵌入(graph embedding)的方式寻找样本的低维表示, 在嵌入空间中保持样本之间的局部近邻关系。算法在数学上将上述目标形式化为一个以权重作为惩罚系数的最小二乘代价函数, 形式如下:

$$\min \phi(\mathbf{Y}) = \sum_{i,j} \left\| \mathbf{y}_i - \mathbf{y}_j \right\|^2 w_{ij} = 2 \mathbf{Y} \mathbf{L} \mathbf{Y}^T, \text{ 其中 } w_{ij} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right). \quad (1-6)$$

上式中, 权重  $w_{ij}$  的定义采用了 Gaussian kernel 的形式, 且仅在各样本与其  $k$  近邻之间有非零取值, 即只有当  $\mathbf{x}_i$  与  $\mathbf{x}_j$  为对方的  $k$  近邻时, 权重  $w_{ij}$  才有上述定义, 否则  $w_{ij}$  取值为 0。因此, 算法的权重矩阵  $\mathbf{W}$  是一个稀疏矩阵。上式中的矩阵  $\mathbf{L}$  为图拉普拉斯矩阵, 定义为  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ , 其中  $\mathbf{D}$  为图上的度数矩阵, 其元素定义为  $d_{ii} = \sum_j w_{ij}$ , 这是一个对角矩阵。从公式(1-6)可以看出, 样本  $\mathbf{x}_i$  与  $\mathbf{x}_j$  之间的距离越小, 它们之间的权重越大, 从而对目标函数的贡献也越大。上述目标函数的求解最终转化为拉普拉斯矩阵的特征分解问题, 具体形式如下:

$$\mathbf{L} \mathbf{y} = \lambda \mathbf{D} \mathbf{y} . \quad (1-7)$$

样本的降维坐标即为上述广义特征分解问题中对应于最小特征值的特征向量。与 LLE 类似, Laplacian Eigenmaps 同样需要对降维空间的坐标进行单位协方差约束以排除平凡解。

(4) Hessian LLE (HLLE)算法[16]。该算法的基本思想是通过局部切空间估计流形上的 Hessian 矩阵, 将其作为流形局部几何属性的刻画, 并利用该矩阵的零空间完成降维。在数学上,  $d$  维流形的每个局部区域都与  $d$  维欧氏空间的一个开子集拓扑同胚, 从而满足局部保距(isometric)属性, 因此其 Hessian 矩阵具有  $d+1$  维的零空间。具体地, 由于流形局部等距于欧氏空间中的一个开子集, 因此流形上每点的局部区域到其开子集的

映射函数为线性函数, 该函数的混合二阶偏导数为零, 从而流形上的局部 Hessian 系数构成的二次型也为零。将这些局部区域扩展到流形上的全局 Hessian 矩阵, 就可以得到该矩阵具有  $d+1$  维的零空间。与 LLE 算法相比, 该算法可以得到更加可靠的降维嵌入结果, 但相应的计算代价也变得很高。

(5) LTSA (Local tangent space alignment) 算法[103]。该算法与 Hessian LLE 类似, 采用流形上局部区域的切空间坐标来作为其局部几何表示, 这里的切空间通过局部区域内样本的 PCA 主子空间来建立。算法的核心思想为: 假设流形上局部线性属性的假设成立, 那么原始观察空间的高维数据点与其局部切空间之间存在线性映射关系; 同样地, 相应的嵌入空间的低维数据点与该局部切空间之间也存在线性映射关系。LTSA 算法通过建立这些切空间与目标嵌入空间之间的线性变换关系, 就可以将这些局部切空间对齐(即所谓的 alignment)到统一的全局低维线性空间。具体地, 针对每个样本点  $\mathbf{x}_i$  与其  $k$  个最近邻  $\mathbf{x}_{ij}$ , 算法首先采用 PCA 计算该点的  $d$  维局部切空间  $\Theta_i$ 。根据上述假设, 在切空间  $\Theta_i$  的样本坐标  $\theta_{ij}$  与低维嵌入空间的样本坐标  $\mathbf{y}_{ij}$  之间存在一个线性变换  $\mathbf{L}_i$ 。因此, 算法的优化目标可以表述如下:

$$\min_{\mathbf{Y}_i, \mathbf{L}_i} \sum_i \|\mathbf{Y}_i \mathbf{J}_k - \mathbf{L}_i \Theta_i\|^2 \quad (1-8)$$

其中,  $\mathbf{J}_k$  为规模为  $k$  的中心化操作矩阵, 即完成类似于公式(1-3)的相应变换。该目标函数的最优解可以通过对一个对齐矩阵  $\mathbf{B}$  进行特征分解来获得, 降维坐标  $\mathbf{Y}$  即为对应于最小非零特征值的特征向量。矩阵  $\mathbf{B}$  的元素通过如下迭代求和得到:

$$\mathbf{B}_{N_i N_i} = \mathbf{B}_{N_{i-1} N_{i-1}} + \mathbf{J}_k (\mathbf{I} - \mathbf{V}_i \mathbf{V}_i^T) \mathbf{J}_k \quad (1-9)$$

其中,  $N_i$  为样本  $\mathbf{x}_i$  的邻域点标号集合。为了保证特征分解的矩阵为对称矩阵, 需要对迭代后的矩阵  $\mathbf{B}$  进行对称化操作, 即最终对矩阵  $\frac{1}{2}(\mathbf{B} + \mathbf{B}^T)$  进行特征分解。LTSA 算法的一大优点是降维效果稳定且计算效率很高。

(6) Semidefinite embedding (SDE) 算法[93], 也被称为 Maximum Variance Unfolding (MVU)[108]。该算法来源于 Kernel PCA 的理论框架。由于 Kernel PCA 主要在 PCA 中引入核学习来进行非线性扩展, 其不足在于如何选择合适的核函数。SDE 算法则试图采用优化算法在训练数据上学习出相应的最优核内积矩阵即 Gram 矩阵。算法的基本思想是首先在训练数据上构建邻域图  $G$ , 然后在降维的过程中最大化样本点之间的欧氏距离, 即所谓的 Maximum Variance, 同时保持邻域点之间的距离不变从而保持流形局部结构不被破坏。这一优化目标最终形式化为一个半正定规划(semidefinite programming problem, SDP)问题。具体地, 算法构建邻域图的过程与 LLE 和 Laplacian Eigenmaps 等算法完全相同, 通过计算样本间的欧氏距离为每个样本  $\mathbf{x}_i$  寻找其  $k$  个最近邻  $\mathbf{x}_{ij}$  ( $j=1, 2, \dots, k$ )。接下来, 算法在保持邻域点对距离不变的约束下, 最大化所有样本之间的距离平方和, 目标函数形式化为:

$$\max \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2, \quad s.t. \quad \|\mathbf{y}_i - \mathbf{y}_j\|^2 = \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad \text{for } \forall (i, j) \in G \quad (1-10)$$

通过定义低维坐标  $\mathbf{Y}$  上的核内积矩阵  $\mathbf{K}$ ，算法将上述目标函数转化为如下的 SDP 问题：

$$\begin{aligned} & \max \text{Trace}(\mathbf{K}), \quad s.t. \\ (1) \quad & k_{ii} + k_{jj} - 2k_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad \text{for } \forall (i, j) \in G \\ (2) \quad & \sum_{i,j} k_{ij} = 0 \\ (3) \quad & \mathbf{K} \geq 0 \end{aligned} \quad (1-11)$$

该 SDP 问题的最优解内积矩阵  $\mathbf{K}$  即作为 Kernel PCA 的核矩阵输入进行后续的降维运算。

(7) Conformal eigenmaps (CE) 算法[74]。该算法主要解决局部保持算法如 LLE 和 Laplacian Eigenmaps 等仅能保持流形上特定的近邻关系而不能显式地保持流形上更一般的局部特征如距离和角度这一不足。算法采用 LLE 或 Laplacian Eigenmaps 的降维结果作为初始输入，通过施加一个额外的约束条件以最大程度地保持邻域点之间的夹角，进而得到保角的(conformal, i.e. angle-preserving)低维嵌入空间。该约束条件在数学上形式化为最大化降维前后流形局部邻域内三角形的相似度。具体地，考虑样本  $\mathbf{x}_i$  与其近邻点  $\mathbf{x}_j$  和  $\mathbf{x}_k$  构成的三角形  $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ ，在降维后的低维嵌入空间中对应的三角形为  $(\mathbf{y}_i, \mathbf{y}_j, \mathbf{y}_k)$ ，三角形之间的相似性关系如下：

$$\frac{\|\mathbf{y}_i - \mathbf{y}_j\|^2}{\|\mathbf{x}_i - \mathbf{x}_j\|^2} = \frac{\|\mathbf{y}_j - \mathbf{y}_k\|^2}{\|\mathbf{x}_j - \mathbf{x}_k\|^2} = \frac{\|\mathbf{y}_k - \mathbf{y}_i\|^2}{\|\mathbf{x}_k - \mathbf{x}_i\|^2} . \quad (1-12)$$

利用上式的三角形相似度构造代价函数，算法的优化目标最终形式化为求解一个小规模的半正定规划(SDP)问题。算法在求解该 SDP 问题的过程中，同时可以得到对流形内在本质维数的一个估计。

(8) Diffusion maps 算法[49]。该算法的基本思想与 ISOMAP 中保持距离的思想比较类似，不同的是 diffusion maps 所保持的距离称为 diffusion distance，而不是 ISOMAP 中所采用的测地距离。正如前面介绍 ISOMAP 方法的不足时所提到的，通过最短路径距离估算得到的近似测地距离是否正确，在很大程度上影响算法降维结果的正确性。为了避免测地距离计算不准确这一问题，diffusion maps 方法从动态系统的角度度量流形上两个样本点之间的相似度或者距离，利用马尔可夫随机游走(Markov random walk)的思想来定义样本邻接图上任意两点之间的路径距离，即所谓的 diffusion distance。由于这种距离综合了邻接图上所有路径的信息，因此比通过样本估算的测地距离更加可靠，特别是针对短路问题(short-circuiting)更加鲁棒。具体地，与 ISOMAP 和 LLE 等算法类似，diffusion maps 首先建立样本邻接图，采用公式(1-6)中的高斯核对邻接图中的边进行加权，从而得到权重矩阵  $\mathbf{W}$ 。对矩阵  $\mathbf{W}$  的每一行进行归一化处理得到马尔可夫随机游走初始步骤的相似度矩阵  $\mathbf{P}^{(1)}$ ，其元素定义如下：

$$p_{ij}^{(1)} = \frac{w_{ij}}{\sum_k w_{ik}} . \quad (1-13)$$

从动态系统理论的角度将矩阵  $\mathbf{P}^{(1)}$  看作一个马尔可夫矩阵，该矩阵即给出了动态过程的前向转移概率矩阵。因此， $\mathbf{P}^{(1)}$  中的元素反映了在单步状态转移中从一个样本点到另一个样本点的转移概率。经过  $t$  步转移之后，得到的前向转移概率矩阵  $\mathbf{P}^{(t)}$  即为  $(\mathbf{P}^{(1)})^t$ 。通过该矩阵中两点之间的转移概率就可以计算它们的 diffusion distance 如下：

$$D^{(t)}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_k \frac{(p_{ik}^{(t)} - p_{jk}^{(t)})^2}{\psi(\mathbf{x}_k)^{(0)}}} , \text{ 其中 } \psi(\mathbf{x}_i)^{(0)} = \frac{m_i}{\sum_j m_j}, m_i = \sum_j p_{ij} . \quad (1-14)$$

公式中的分母项  $\psi(\mathbf{x}_i)^{(0)}$ ，其核心思想是为邻接图中密度大的区域分配较大的权重。从上式定义可以看出，针对转移概率较大的两个样本点，它们之间的 diffusion distance 则相对较小，而且这种距离定义方式综合考虑了邻接图上所有路径的信息。diffusion maps 算法降维的优化目标为保持样本间的 diffusion distance 即  $D^{(t)}(\mathbf{x}_i, \mathbf{x}_j)$ ，根据随机游走框架中的谱图理论，优化目标转化为对转移概率矩阵  $\mathbf{P}^{(t)}$  进行特征分解：

$$\mathbf{P}^{(t)} \mathbf{v} = \lambda \mathbf{v} . \quad (1-15)$$

最优的降维坐标  $\mathbf{Y}$  即取对应于最大特征值的前若干个特征向量并且以相应的特征值进行加权。其中对应于最大特征值  $\lambda_1$  的第一个特征向量是平凡解，从而被排除。 $\mathbf{Y}$  的最终表达式如下：

$$\mathbf{Y}^T = \{\lambda_2 \mathbf{v}_2, \lambda_3 \mathbf{v}_3, \dots, \lambda_{d+1} \mathbf{v}_{d+1}\} . \quad (1-16)$$

上面介绍的八种流形学习算法均可以归类为非参数式的、基于谱分解低维嵌入思想的方法。可以看出，为数众多的流形学习方法有着各自不同的目标函数和优化过程，因此研究者们尝试将它们形式化在统一的框架下面，相关的代表性工作包括核主成分分析 (Kernel PCA) 框架[28]，图嵌入(graph embedding)框架[99]，以及黎曼流形学习框架 (Riemannian manifold learning, RML)[56]等。此外，与传统的批处理(batch mode)训练模式不同，一些研究者致力于解决增量式的流形学习问题，代表性工作如[50][105]，这些算法可以处理流形数据无法一次性获取而只能在线序贯采集，并不断利用现有数据更新学习模型的情形。

除了上述非参数式算法之外，流形学习的另一条研究主线是参数式的基于局部模型拟合进而完成全局对齐思想的算法，这类型算法的代表工作包括全局协调算法(global coordination)[67]，流形坐标图册算法(manifold charting)[10]，局部线性协调算法(locally linear coordination, LLC) [79]，以及协调因子分析算法(coordinated factor analysis, CFA) [85][86]。此类方法的通常做法是首先在流形的局部区域利用一些子空间学习方法（如主成分分析[38]，因子分析[19][25]）建立一组局部线性模型，然后通过最优化某个目标函数，将这些来自不同局部模型的低维表示结果综合起来得到流形数据的全局统一坐标表

示, 从而完成非线性降维。在模型训练阶段完成之后, 算法可以学习得到参数式的映射函数, 从而用于对未知测试数据从高维观察空间到低维嵌入空间的投影或者反向的虚拟样本重构等任务。

### 1.2.3 综合对比与讨论

本小节首先从降维问题的目标和形式化方面对上面两个小节介绍的方法进行分类对比分析, 之后针对流形学习算法的性能评测和一些开放性问题展开讨论。

根据优化目标函数的性质, 降维方法可以分为基于凸目标函数(convex)和基于非凸目标函数(nonconvex)两大类[59], 如下图 1.3 所示。其中, 凸目标函数的优化过程中不存在局部极值的问题, 上节介绍的八种非参数式基于谱分解低维嵌入的算法都属于这一类别; 非凸目标函数的优化则存在局部极值, 上节介绍的参数式基于局部模型拟合后全局对齐的算法如 LLC 和 manifold charting 则属于这一类别。进一步地, 根据谱分解的目标矩阵是否为稀疏矩阵, 可以将基于凸目标函数的方法划分为基于满矩阵谱分解(full spectral)和基于稀疏矩阵谱分解(sparse spectral)。其中, 基于 full spectral 的方法包括 PCA, classical scaling, ISOMAP, Kernel PCA, SDE(MVU), diffusion maps 等。这些方法的优化目标可以归结为保持观察数据各维之间的方差或者数据点对之间的距离与相似度。基于 sparse spectral 的方法包括 LLE, Laplacian Eigenmaps, Hessian LLE, LTSA 等。可以看出, 这些方法都侧重于保持流形上的局部几何结构。

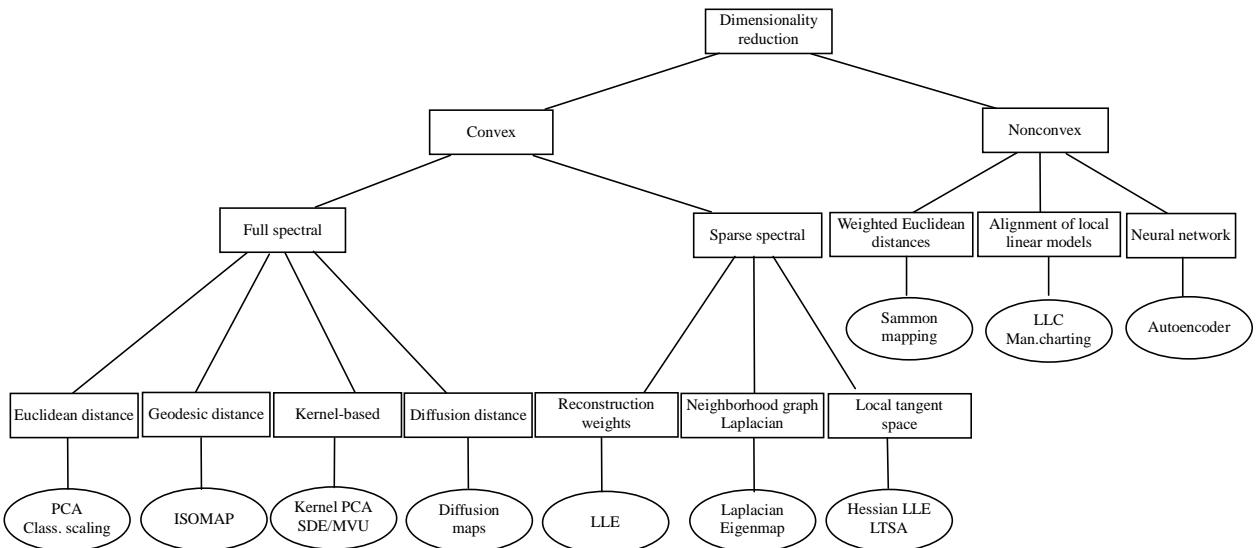


图 1.3 降维方法分类结构图[59]

接下来, 针对降维算法的一些主要属性进行对比分析。这里主要考虑四方面的属性:

- (1) 高低维空间之间的参数式映射, (2) 算法优化的自由参数, (3) 算法时间复杂度, (4) 算法空间复杂度。不同降维算法的属性对比如表 1.1 所示。

考虑算法属性 (1), 可以看出大部分非线性降维算法都是非参数式的, 即它们不能

提供从原始高维空间到低维嵌入空间的显式映射函数，从而不能解析处理测试数据的 out-of-sample 问题。反过来，这些非参数式算法也不能从低维空间的数据点来重构高维空间中的样本并进而度量重构过程中的误差，因此也就无法有效获知原始高维空间的信息在多大程度上在低维空间中得到了保持。

考虑算法属性 (2)，可以看到绝大多数非线性降维算法的目标函数中都包含自由参数，这些参数在算法应用到不同数据集合时需要做不同的优化设置。针对目标函数非凸的算法 LLC 和 manifold charting，还需要一些额外的参数，比如目标函数优化的学习率 (learning rate)，最大迭代次数等。针对算法包含自由参数这一属性，其优点是降维算法具有更大的灵活性，缺点则是参数的优化调节往往对算法的降维性能影响很大。

考虑算法属性 (3) (4)，表 1.1 中给出了各算法计算代价最高部分的时空复杂度，这些指标在很大程度上决定了算法的实用性。可以看出，算法复杂度主要受两方面因素决定：1) 数据集的属性，比如样本数量  $N$ ，样本维数  $D$ ；2) 算法的自由参数，比如降维的目标维数  $d$ ，近邻图中邻域个数  $k$ ，迭代次数  $i$ 。此外，表中的参数  $p$  为稀疏矩阵中非零元的比例， $m$  为 LLC 和 manifold charting 采用的混合因子分析中局部模型的个数。

表 1.1 不同降维方法的参数和计算复杂度对比

Technique	Parametric	Parameters	Computational	Memory
PCA	yes	none	$O(D^3)$	$O(D^2)$
Classical scaling	no	none	$O(N^3)$	$O(N^2)$
ISOMAP	no	$k$	$O(N^3)$	$O(N^2)$
Kernel PCA	no	$\kappa(\cdot, \cdot)$	$O(N^3)$	$O(N^2)$
MVU	no	$k$	$O((Nk)^3)$	$O((Nk)^3)$
Diffusion maps	no	$\sigma, t$	$O(N^3)$	$O(N^2)$
<hr/>				
LLE	no	$k$	$O(pN^2)$	$O(pN^2)$
Laplacian Eigenmaps	no	$k, \sigma$	$O(pN^2)$	$O(pN^2)$
Hessian LLE	no	$k$	$O(pN^2)$	$O(pN^2)$
LTSA	no	$k$	$O(pN^2)$	$O(pN^2)$
LLC	yes	$m, k$	$O(imd^3)$	$O(Nmd)$
Manifold charting	yes	$m$	$O(imd^3)$	$O(Nmd)$

针对流形学习算法的实验验证，目前的研究工作采用的数据集主要仍是一些人工生成的数据和真实的流形结构已知的数据。典型的人工数据集包括瑞士卷(swiss-roll)、S 曲线(s-curve)、helix 曲线、twinpeaks 曲面、punctured sphere 曲面等。典型的真实数据集包括 ISOMAP 和 LLE 算法文献中分别采用的两个人脸数据库、ORL 人脸库、MNIST 手写体数字库、COIL20 物体数据库等。具体的数据集产生方法和介绍详见参考文献[56][59]。

在算法性能评测方面，早期起步阶段的研究主要采用定性评估的方式，通过将已知流形结构的数据集进行降维投影到二维或者三维的坐标空间中进行可视化分析。近几年来，随着大量的新算法不断涌现，基于标准数据库采用一些公共的定量方式来评测不同算法逐渐成为研究者们的共识。

针对算法降维结果的可靠性，文献[109]提出两种量化指标来度量算法在保持流形局部结构方面的性能，这两种指标分别称为可信度(trustworthiness)与连续性(continuity)。可信度指标主要度量在降维后的低维空间中离得太近的样本的比例，其具体定义如下：

$$T(k) = 1 - \frac{2}{Nk(2N-3k-1)} \sum_{i=1}^N \sum_{j \in U_i^{(k)}} (r(i, j) - k) . \quad (1-17)$$

其中， $r(i, j)$ 表示在低维空间中样本  $j$  在样本  $i$  的  $k$  近邻中的排序序号， $U_i^{(k)}$  表示在低维空间中是样本  $i$  的  $k$  近邻但是在高维空间中不是  $i$  的  $k$  近邻的样本标号集。形象地理解，这个指标用来惩罚算法将高维空间中原本不邻近的样本经过降维后变得紧致的程度。类似地，连续性指标定义如下：

$$C(k) = 1 - \frac{2}{Nk(2N-3k-1)} \sum_{i=1}^N \sum_{j \in V_i^{(k)}} (\hat{r}(i, j) - k) . \quad (1-18)$$

其中， $\hat{r}(i, j)$  表示在高维空间中样本  $j$  在样本  $i$  的  $k$  近邻中的排序序号， $V_i^{(k)}$  表示在高维空间中是样本  $i$  的  $k$  近邻但是在低维空间中不是  $i$  的  $k$  近邻的样本标号集。形象地理解，这个指标用来惩罚算法将高维空间中原本近邻的样本经过降维后变得分散的程度。

针对参数式降维算法的解析映射功能，文献[86]采用样本重构的方式，将原始样本从高维观察空间投影到低维嵌入空间，之后利用逆映射再将低维样本反投影回高维空间，计算原始样本与重构样本之间的差别。这种重构误差指标在一定程度上可以度量算法的泛化推广能力。

流形学习经过近十年来的蓬勃发展，在机器学习的理论研究领域和计算机视觉的应用研究领域均引起了研究者们广泛而持续的关注。然而，在实际问题应用中，目前的算法还很难产生出令人们满意的效果。特别是在最近几年来，流形学习的发展遭遇了一定的瓶颈期，这主要是由于制约流形学习算法实用化的若干开放性问题始终没有得到完美的解决。这些问题包括算法对样本稠密采样的要求、对噪声数据的敏感性、算法中多个自由参数的组合调节、以及对数据流形分布过于理想的假设等。因此，从长远看来，流形学习方法距离真正成为传统线性降维工具的替代框架还有很长的路要走。

#### 1.2.4 现有方法的主要问题

综合分析前人工作，很多方法无法有效解决所谓的 out-of-sample 问题，即它们只能为训练数据提供嵌入空间的坐标，而无法处理未知的测试数据降维。为此，文献[7]针对 ISOMAP, LLE 和 Laplacian Eigenmaps 等算法提出了一种通用的解决框架。然而，该方法同时也具有明显的不足之处，作为非参数方法，它需要存储并访问整个训练数据集合。

如果待处理的训练数据规模庞大且维数较高时, 算法的这一需求的代价就愈发显得高昂, 特别是需要将训练模型得到的流形结构推广到测试数据的时候。显而易见, 更好的解决方案应该是学习从高维数据空间到低维嵌入空间的一个显式的参数式映射函数。

流形学习研究的一个核心问题是寻找流形的低维嵌入空间, 另一个重要的研究问题则是挖掘隐含在观察数据中的流形本质结构, 从而有助于发现流形上的重点感兴趣区域并进行后续特别处理[35][39]。为此, 前人工作大多侧重于估计流形的内在维数[24][53][65]。然而, 仅仅估计流形维数还不足以深入分析流形的内在结构。为了推测流形数据的内在变化模式, 已知的方法通常都是将训练数据投影到低维嵌入空间后, 通过可视化观察这些数据的分布走势来进行间接分析[66][80][93]。这种推测手段的基本假设是, 低维嵌入空间的每一维坐标轴分别对应于原始流形观察数据的一种变化模式。更有效的方式应该是对流形数据的这些本质变化模式进行显式的建模刻画。

此外, 与对应的线性降维算法相比, 目前的绝大部分非线性流形学习算法在计算效率方面都不够优越, 它们往往需要求解与训练样本集合同等规模的特征分解问题[5][10][66][80]或者需要采用诸如期望最大化(expectation maximization, EM)此类的迭代式算法来进行模型优化[67][85]。

### 1.3 基于图像集合的人脸识别概述

在传统的物体识别问题中, 分类器的训练和测试通常都是基于物体的单个或者很少量的样本进行的。随着摄像机和大容量存储设备等硬件技术的普及发展, 在很多新的应用场景如视频监控、视频检索等任务中, 可以获取到物体的大量图像数据, 从而为后端分类问题的训练和测试阶段提供大量的样本。以具体的人脸识别问题为例, 目前从视频序列或者个人相册中可以收集到每个人的多幅不同情况下采集的照片, 这些照片自然地构成对应于每个注册个体的一个 *gallery* 图像集合。在测试阶段, 同样可以收集每个测试个体的多幅图像作为 *probe* 图像集合, 从而取代传统方法中的单个 *probe* 图像。基于不同人的这些 *gallery* 和 *probe* 图像集合, 就可以设计基于图像集合的人脸识别系统, 如图 1.4 所示。本文中将这类识别问题称之为基于图像集合的物体识别(Object Recognition with Image Set, ORIS)。在这类识别问题中, 每个集合通常包含属于同一个物体类别的多个图像样本, 这些图像涵盖了物体广泛的表观变化模式, 比如姿态视角的变化、非刚性变形、光照角度的变化等等。目前, 在场景可控的条件下, 基于单幅图像的传统物体识别方法在一定程度上已经取得了成功应用。针对情况更为复杂的实际场景, 如果能够获取待识别物体多种变化模式下的图像样本集合, 将集合当中多个样本的信息综合有效利用起来, 则无疑能够在很大程度上改进分类算法的性能[1][21][43]。

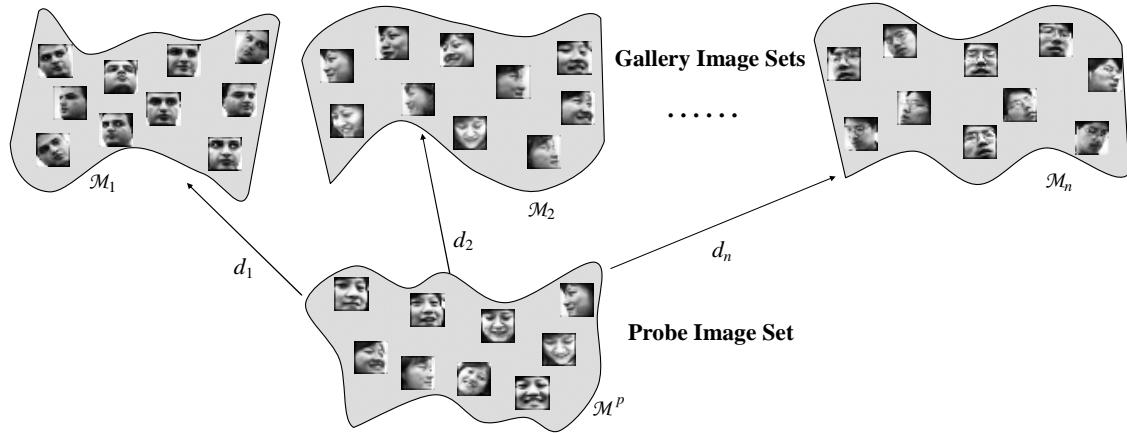


图 1.4 基于图像集合的人脸识别问题示意图。在身份注册阶段，所有人均采用一组图像来注册得到 gallery 图像集合；在测试阶段，待测试个体同样采用一组图像注册得到 probe 图像集合，用来进行识别。图示中的人脸样本来自于 Honda/UCSD 数据库[51]。

针对基于图像集合的物体识别问题，计算机视觉研究者们在过去的十多年中投入了越来越多的关注，并且提出了一系列解决方法[23][42][44][45][51][58][70][75][87][106]。其中，有一类工作主要集中于基于视频序列的识别问题，它们主要依靠挖掘利用视频帧中的时间和空间一致性和相关性信息来辅助进行物体识别。这类工作可以看作基于图像集识别问题的特例。一般而言，在基于图像集合的物体识别问题中，gallery 或者 probe 集合中的图像样本并不要求必须来自于视频序列的连续帧图像。它们可以是待识别个体的多幅无序采样得到的样本，比如说来自于个人相册中的多张照片或者视频监控序列中的若干帧不连续的图像。事实上，视频序列中包含待识别物体的不连续帧的出现非常普遍，比如在一些人脸监控场景中，镜头前面来回移动的人通常不会一直处于面对摄像机的状态[43]。此外，即使视频的每一帧都捕捉到了被监控的个体，后续的物体检测定位算法也很难保证将每帧的物体位置都准确标记出来。因此，上述基于帧间时序连续性假设的视频识别算法如[45][51][58][106]，在具体应用中往往受到很多实际情况的制约。由于本文主要研究广义条件下的图像集物体识别问题，这些基于视频帧间信息假设的算法在下述章节中将较少提及。

从图像集合的建模角度来看，用于物体识别的集合匹配方法大致分为两类：基于模型的参数式方法、与模型无关的非参数式方法。其中，基于模型的参数式方法主要包括概率建模法[75]，流形密度散度方法[1]。这类方法一般采用参数式的概率分布函数来建模图像集合，进而采用诸如 K-L 散度(Kullback-Leibler Divergence, KLD)之类的度量来计算两个概率分布函数之间的相似度。在概率建模法[75]中，人脸的各种变化模式被建模为人脸图像空间中的单一高斯分布函数。显然，这种简单的概率建模无法刻画复杂的模式分布情形；因此，流形密度散度方法[1]采用了更加符合真实复杂条件的高斯混合模型(Gaussian mixture models, GMM)来取代单高斯模型，这种建模方式由于其适用性更好从

而也更加可靠。不难看出，基于模型的参数式方法有一些明显的不足之处，一方面概率模型的参数估计比较困难，这个过程通常需要庞大的数据量作为支撑；另一方面，基于概率分布的集合建模方式背后有一个基本的假设，即要求训练集合与测试集合之间具有比较强的统计相关性，如果实际问题中的训练与测试集数据的概率分布相差较大时，这些方法的性能就很可能会受到很大影响[42][43]。

由于参数式建模方法的上述局限，研究者们将更多的注意力投向了非参数式方法，这类方法一般对数据的分布不做任何先验假设，因此相对来说具有更加广泛的应用范围。其中，一项早期的研究工作是文献[70]针对视频人脸匹配提出的相关方法。该方法将集合中的图像样本分别单独对待，通过匹配两个集合中最近的图像对来获得集合的相似度，以此完成分类。可以看出，这种基于单样本孤立匹配的方法没有从整体上考虑集合样本之间的相关性，很容易受到噪声数据的干扰[70]。此外，由于需要比较两个集合中的任意一对样本，这类方法的计算复杂度也非常高。

近些年来，随着对子空间学习方法的深入研究，研究者们开始尝试利用子空间方法来对图像集合中的数据变化模式进行全局建模。这一思路正逐渐成为图像集合分类问题的主流研究路线，其开创性工作是文献[98]针对人脸识别提出的互子空间方法(Mutual Subspace Method, MSM)。这类方法通常将图像集合建模为线性子空间[23][29][43][96]或者更为一般的非线性流形[20][21][27][42][89][104]。基于这种建模方式，集合匹配识别的问题就可以转化为子空间或者流形匹配的问题，从而采用子空间或者流形上的各种可能的相似性度量函数与匹配策略进行分类。

从分类目标的角度来看，上述非参数式方法根据其侧重点不同大致可以划分为如下两类。第一类方法主要集中于如何定义相似性度量，如[20][27][42][98]。在文献[20][27]中，图像集合中的一些代表性样本，称之为样例(exemplars)，首先被提取出来作为集合的一组局部模型表示。然后，采用单个图像之间的一些距离函数来度量这些样例点之间的相似度，进而综合得到集合之间的相似度，以此进行集合匹配。可以看出，这种做法与上述非参数方法[70]比较类似，不同之处在于文献[70]需要进行两个集合中所有样本对之间的匹配，而基于样例的方法只需要对两个集合的样例点对进行匹配。在文献[42][98]中，主夹角(principal angles)[9][34]被用来度量两个线性子空间或者流形上的两个局部线性模型之间的距离。作为一种子空间相似性度量，主夹角方法可以有效地捕捉两个子空间之间的公共的数据变化模式[42]。第二类方法则主要集中于根据某种特定的相似性度量函数，学习判别式的分类器[21][23][29][43][96][104]。文献[23][43]采用主夹角作为子空间相似性度量，以最大化不同类别的区分性为目标来学习判别式的投影空间，然后在该投影空间中进行分类。将主夹角嵌入到核函数中，文献[29][96]就将线性的分类方法扩展得到了非线性方法，从而适应于更复杂的数据分布情形。为了显式建模图像集合数据的非线性分布模式，文献[21][104]采用了类似[20][27][42][51]的做法，利用非线性流形来建模图像集合，从流形上建立一组局部线性模型来提取局部的判别信息。

总体而言，上述两类方法（参数式与非参数式方法）各自具有不同的优缺点，从而

适用于不同的问题。在实际应用中,如果可以获得较强的先验知识,确认训练数据和测试数据具有相似的统计属性,那么基于模型的参数式方法就有可能获得更好的结果[1]。然而,对于物体识别问题的一般情形而言,由于采集条件的不同,训练集和测试集数据的统计分布在多数情况下可能很不相同。此时,不需要对数据分布做任何先验模型假设的非参数式方法就更为适用这种问题了[42]。正如诸多文献[20][21][27][42][104]中所推崇的,由于不同图像集合之间以及同一集合内部的数据分布均可能非常复杂,呈现出一定程度上的非线性变化(如图1.4所示),因此采用足够灵活的图像集合建模表示方式来刻画这种非线性分布就显得尤为重要。假设每个集合中的图像样本都各自分布在一个非线性流形上,那么相应的物体识别问题就可以转化为基于一定的相似性度量函数进行非线性流形的匹配问题。本文的研究课题正是受此启发,以非线性流形之间的距离计算为出发点,分别从产生式(无监督地计算流形-流形距离)和判别式(监督地学习判别空间以最大化流形之间的间隔)两个角度来研究图像集合的匹配分类问题。

## 1.4 本文的研究内容与主要贡献

综合前面两个小节的回顾分析,针对流形学习算法和基于图像集合的人脸识别应用,本文分别在理论研究和应用研究两个层面开展工作,在推进和完善理论研究的同时,将理论研究成果应用于解决模式识别和计算机视觉中的实际问题。在理论研究层面,本文主要从流形学习的角度研究新型高效的非线性降维算法,特别是在流形的局部线性模型表示、解析映射函数学习、内在变化模式刻画等方面开展相关研究。在应用研究层面,本文针对基于图像集合的人脸识别问题,主要从图像集合的流形建模、流形之间的距离定义和计算、多流形判别式分类学习等几个角度开展研究。

本文旨在推动流形学习算法理论与应用研究,特别是在扩展流形学习在人脸识别问题中的应用前景方面做了一些有益的尝试。本文的研究内容与主要贡献总结如下:

- 提出了基于最大线性嵌入的流形学习方法

该方法通过学习一个参数式可逆的解析映射函数,可以将流形数据从原始高维观察空间保距地投影到一个全局低维嵌入空间中。算法从几何直觉的角度,引入一种合理有效的局部线性分块的定义,称为最大线性分块。最大线性分块定义为非线性流形上满足一定非线性程度约束的样本规模最大的局部区域,分块的非线性程度通过块内样本间测地距离与欧氏距离的差异来度量。首先,原始的流形采样数据集合被分解为一组最大线性分块的组合,每个分块采用一个局部线性模型来建模,从而将流形的全局非线性结构通过这组局部线性模型来组合表示。接下来,算法从每个局部模型内部随机选取一定数量的地地标点,利用其测地距离进行多维尺度变换运算,进而得到最终的保距低维嵌入空间并完成局部模型的全局对齐。这一对齐过程称为基于地地标点的全局对齐算法。该对齐算法可以同时保持流形数据的局部几何特性和全局结构,避免了传统算法中常见的局部

极值与大规模迭代优化这两个缺点，并且只需要求解小规模的特征分解问题就能得到有效的闭形式解。针对流形内在结构的描述，本文算法不仅提供了一种快速有效的本质维数估计方法，而且可以显式建模流形观察数据的内在变化模式，从而广泛应用于多种实际问题。

### ● 提出了基于流形-流形距离框架的图像集合人脸识别方法

针对基于图像集合的物体识别问题，本文采用非线性表观流形来建模属于每一类别的图像集合，并将集合分类问题形式化为流形-流形距离的计算问题，即计算训练所采用的 *gallery* 图像集对应的 *gallery* 流形和测试所采用的 *probe* 图像集对应的 *probe* 流形之间的距离。注意到实际问题中的图像集合可能表示为三种模式层次，即点、子空间、流形，本文系统研究了这三种模式之间的各种距离，并将它们形式化在一个统一的多层次流形-流形距离框架下。具体地，将流形表示为一组局部线性模型的组合，其中各局部模型采用子空间来刻画。这样，两个流形之间的距离就可以转化为它们各自对应的子空间对之间的距离综合。针对流形-流形距离的计算，本文从理论上和实验上探讨了算法各个模块的多种不同的定义方案，包括局部线性模型构建、局部模型距离度量、局部距离的全局综合等。通过图像集合人脸识别的实验验证表明，本文提出的流形-流形距离作为一种通用的图像集合相似性度量，一致地优于其它对比的非判别式算法，并且达到了与当前领先的基于集合的判别式学习算法可比的性能。

### ● 提出了基于流形判别分析方法的图像集合人脸识别方法

针对基于图像集合的物体识别问题，本文进一步从监督学习的角度对流形-流形距离框架进行扩展并提出流形判别分析方法。该方法将图像集合分类问题形式化为面向分类的多流形学习问题，通过在流形-流形距离框架内引入“最大化流形间隔”的思想来学习判别式的特征空间，最终在该空间内对不同类别的流形进行匹配来完成图像集合的分类。具体地，流形判别分析方法采用图嵌入的框架来构建图模型，采用本征图和惩罚图分别刻画流形的类内紧致性与类间可分性。通过设计的最优化目标函数来学习判别式的嵌入空间，使得具有不同类别标号的流形之间能够更好地区分，同时每个流形内部局部区域的数据紧致性得到增强。在分类识别阶段，通过将待测试的流形数据投影到该判别式空间中，就可以进行更加可靠的分类。通过人脸识别和物体分类两个任务的实验对比表明，本文提出的流形判别分析方法取得了与当前领先的方法相当的性能，同时具有很高的训练和测试效率。

## 1.5 论文的组织

本文各章节安排如下：

第1章为绪论部分，主要介绍了流形学习理论及其应用于人脸识别问题的研究背景、

研究现状以及目前存在的问题和挑战。论文全面回顾了流形学习方法的研究发展，从线性降维和非线性降维的角度分别对相关算法进行介绍；接下来，论文总结了基于图像集合的人脸识别问题的研究进展，从图像集合的建模方式与分类目标两个角度分别介绍了相关算法的基本思想。最后，给出本文研究的课题内容与主要贡献。

第2章介绍了基于最大线性嵌入的流形学习方法。通过分析现有方法的不足，首先提出算法的研究动机与基本思想。接下来详细给出算法的核心模块构成，包括局部线性模型的建立与全局对齐算法，并针对流形内在本质结构的描述进行深入讨论。算法的有效性通过人工三维数据与真实人脸数据上的一系列实验来进行验证，并与经典的流形学习方法进行定性与定量对比。本章最后详细分析了最大线性嵌入算法与经典的ISOMAP和LLE算法的联系与区别，并就算法可能的改进方案进行了讨论。

第3章提出了流形-流形距离的理论框架，并将其应用于图像集合人脸识别问题。首先系统调研了点、子空间、流形三种模式层次之间的各种距离，并将它们形式化在一个统一的多层次流形-流形距离框架下。接下来，本章详细介绍流形-流形距离的三个组成模块，深入探讨了各模块的多种可能定义方案；并针对算法框架的两个理论计算问题进行了详细讨论。最后，本章通过在三个公开人脸数据库上的对比实验来评测流形-流形距离方法的有效性，并与相关的识别算法进行实验对比和理论分析。

第4章介绍了流形判别分析方法，并将其应用于图像集合人脸识别问题。基于上一章的流形-流形距离框架，本章首先从判别分析的角度，对流形之间的匹配和分类问题进行表述，并简要给出算法的基本思想。接下来，具体介绍算法的训练与测试阶段的各个环节，包括问题的形式化、局部模型的构建、判别函数的学习和测试算法流程。最后，本章通过在人脸识别和物体分类两个具体任务上的对比实验来验证算法的性能，并进一步给出理论分析和讨论。

第5章结论部分，在对本文工作进行概要总结的基础上，针对下一步的研究方向和相应的扩展进行了深入讨论。



## 第二章 基于最大线性嵌入的流形学习方法

### 2.1 引言

流形学习是近十年来模式识别和机器学习领域的一大研究热点和难点，并且在计算机视觉领域的很多应用问题中得到了推广，比如物体识别、运动轨迹分析、人脸表情姿态以及年龄分析等等。针对当前流形学习研究中存在的若干问题，本章提出一种新型的非线性降维算法，称为最大线性嵌入(Maximal Linear Embedding, MLE)。与现有算法相比，MLE 具有如下几个特点：

1. 针对流形局部线性属性的刻画，MLE 算法提出了一种紧致而高效的局部线性分块概念，称为最大线性分块(Maximal Linear Patch, MLP)。MLP 定义为非线性流形上满足一定线性尺度约束的分块规模最大的局部区域。利用流形上划分的 MLP 可以建模得到一组局部线性模型，从而将流形的全局非线性结构通过这组局部线性模型来组合表示。
2. 为了完成流形全局数据的统一降维，MLE 采用基于地标点的全局对齐方法(Landmarks-based Global Alignment, LGA)将上述局部线性模型进行一致的对齐。LGA 方法的对齐过程可以同时保持流形数据的局部几何特性和全局结构，并最终得到一个全局统一的低维坐标空间为原始流形提供保距的(isometric)低维嵌入。与之前算法采用的对齐方法相比，LGA 方法的对齐过程还有效地避免了经常面临的局部极值问题。
3. MLE 算法在完成训练数据降维的同时，最终学习得到一个非线性可逆的参数式映射函数，从而有效解决降维结果推广的 out-of-sample 问题。该映射函数具有闭形式的解析解，可以很容易地对未知测试数据进行正向(forward)到低维空间的降维投影和反向(backward)到高维空间的虚拟样本重构。
4. 针对流形内在本质结构的描述，MLE 算法不仅为流形本质维数的估计提供了一种简单高效的解决方法，而且首次成功尝试了对隐含在流形观察数据中的内在变化模式进行显式的建模表示。
5. 在计算效率方面，MLE 算法表现出良好的性能。算法非迭代式的学习过程只需要求解小规模的特征分解问题，其分解的矩阵规模与局部线性模型的个数成比例，而与训练样本的数量无直接关系。

本章接下来的安排如下：第 2 节简要提出 MLE 算法的研究动机与基本思想；第 3 节详细介绍算法的核心环节，包括局部线性模型的建立与全局对齐算法，并针对流形内在本质结构的描述进行深入讨论；第 4 节通过在人工三维数据与真实人脸数据上的一系列实验验证算法的有效性，并与经典的流形学习方法进行定性与定量对比；第 5 节对算法进行总结，特别是深入讨论了 MLE 与经典算法 ISOMAP 和 LLE 的联系与区别，并就算法可能的改进方案进行了分析。

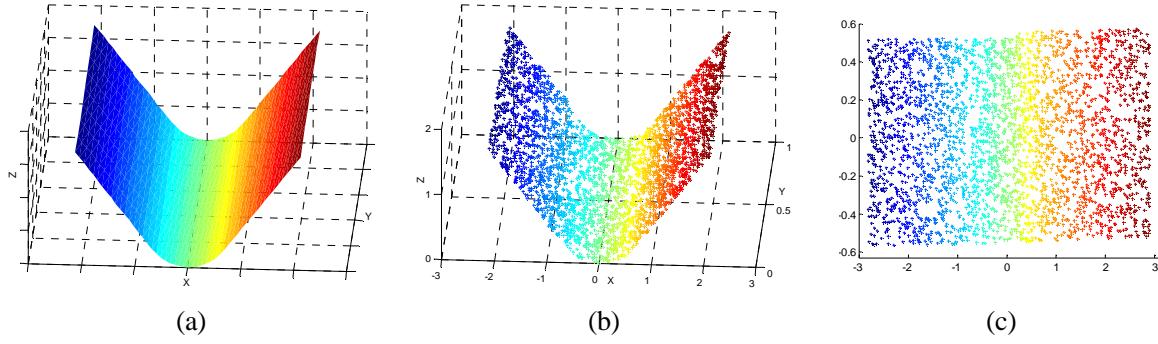


图 2.1 V 型(V-like shape)流形数据示意图。(a) 原始的三维流形曲面; (b) 流形上采样得到的 3000 个样本点; (c) 最大线性嵌入方法(MLE)学习得到的具有保距特性的低维嵌入。

## 2.2 算法基本思想

基于可信集(trusted-set)[10]的流形学习方法,如 ISOMAP 和 LLE,均通过  $k$  近邻( $k$ -NN)或者  $\epsilon$ -ball 邻域的方式来定义其局部线性分块,这种局部分块的样本规模通常都是固定的并且数量也比较少。由于这种局部分块的定义方式没有自适应地考虑样本点邻域的真实结构,因此可能将非线性流形上的一些大规模线性分块拆分为很多琐碎的细小分块。可以看出,这种做法很不经济,这里所谓的“经济”是指应该尽量避免 LLE 算法中的极度稠密的局部线性分块构造方式。同时,有研究表明[2],针对基于 trusted-set 方式定义的局部分块,邻域大小的一些轻微改变就可能导致此类算法的结果出现不稳定。为了克服固定邻域规模所带来的这一问题,研究者们尝试提出了一些解决方法,相关的工作包括[63][92][94]。然而,这些改进方法中所定义的局部分块本质上仍是基于样本最近邻的方式,而没有显式的考虑样本更大邻域范围内的真实的线性/非线性结构。

MLE 算法正是根据流形数据的真实线性/非线性结构来定义一种紧致而高效的局部线性分块。算法动机来自于直觉的几何观察:如图 2.1 所示的“V 型”数据,这是一个分段线性的流形,它由两个二维平面平滑地连接而成,分布在每个平面上的样本点天然地张成一个“全局”的二维线性子空间。很显然,基于 trusted-set 方式定义的局部分块方式并不能利用数据的这一“全局”信息。因此,MLE 算法考虑以更一般的方式来定义流形上的局部线性分块,从而引入最大线性分块(Maximal Linear Patch, MLP)的概念。具体地,MLP 尝试利用尽可能多的“全局”信息并进而张成一个最大化的线性子空间,其非线性程度通过分块内欧氏距离与测地距离的偏差来度量。图 2.2 给出了 MLP 线性分块的示意图。直觉看来,MLP 所张成的线性子空间的维数应该就是流形的内在本质维数。

基于上述 MLP 的几何直观定义,本文创新性地提出一种有效的层次聚类算法,将流形数据划分为一组 MLP 的集合。然后,针对每个 MLP,通过子空间学习方法可以很容易构建一个局部线性模型作为其低维表示。由于 PCA 方法具有简单高效且解析求解等优点,本文采用 PCA 来完成这一局部模型的建模过程。

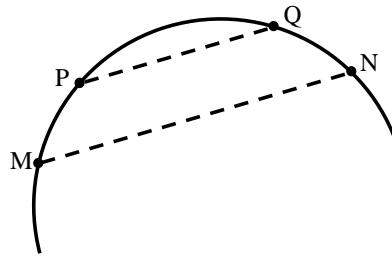


图 2.2 最大线性分块(MLP)的直观示意图。图中实线半圆弧表示一维非线性流形，其中点 P 到 Q 的一段圆弧可以看作是符合期望的 MLP 分块，因为其对应的欧氏距离（即 PQ 之间的虚线段）可以较好地近似其测地距离（即 PQ 之间的实线圆弧）；而点 M 到 N 的一段圆弧则由于弯曲程度太大而无法作为符合期望的 MLP 分块，其欧氏距离（即 MN 之间的虚线段）与测地距离（即 MN 之间的实线圆弧）之间的偏差太大。

局部模型构造完成之后，接下来就需要将它们按照一定的规则一致地对齐到一个统一的全局坐标空间中，同时进行显式的参数式投影函数的学习。针对局部模型的全局对齐这一问题，本文第一章中提到的参数式方法如[10][67][79][85][86]提供了一些可能的候选解决方案。然而，文献[10][67][85][86]中的几种对齐方法有些需要利用 LLE 与 ISOMAP 算法的降维结果作为其迭代初始化输入，有些则由于其庞大的局部模型数量需要求解耗时的大规模特征分解问题。文献[79]中的 LLC 算法避免了上述问题，并提出了一种适用于一般情形的全局对齐算法。这一算法的不足在于，其优化目标采用了与 LLE 算法一致的代价函数，即最小化局部邻域样本的线性重构误差。同时，LLC 算法还需要满足与 LLE 相同的约束条件，即单位协方差矩阵约束。如文献[56][59]所指出，这一约束条件在一定程度上会导致流形数据全局尺度信息的丢失，并带来其整体结构的变形。因此，本文进而提出一种有效的局部模型对齐算法，称之为基于地标点的全局对齐 (Landmarks-based Global Alignment, LGA)。算法动机同样来自于直观的几何观察，其基本思想如下：首先，从每个 MLP 中随机稀疏采样一定数量的地标点(landmarks)，利用基于测地距离的 MDS 算法将这些 landmarks 投影到一个保距的(保持无穷小的距离和角度)全局坐标空间中；接下来，利用这些“局部-全局”对齐的 landmarks 样本作为控制点，计算每个局部模型（即线性子空间）到上述全局空间的显式可逆的线性变换关系（包括平移、尺度和旋转变换），进而完成局部模型的全局对齐。将这些线性变换函数综合起来，LGA 算法最终可以得到从样本空间到全局嵌入空间的一个分段线性的、可逆的映射函数，从而可以自然地应用于处理训练数据与未知测试数据的降维运算。

从上述介绍可以看出，MLE 算法的核心包括两个方面：局部线性分块 MLP 和全局对齐算法 LGA，接下来的一节对其进行详细介绍。

### 2.3 最大线性嵌入算法

本节首先介绍 MLP 的定义方式，并给出建立 MLP 的层次聚类算法；之后详细介绍 MLE 算法的学习流程包括局部模型的构建、全局对齐算法 LGA 以及流形内在本质结构

的描述；最后，分析对比 MLE 与其它相关算法的区别并给出算法的复杂度分析。

### 2.3.1 最大线性分块

流形学习可以看作是从一组观察数据中推导其产生式模型的过程。给定一组观察数据集合  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i \in \mathbb{R}^D$ , 其中  $N$  为样本数量,  $D$  为观察数据的维数。假设这些观察数据采样自一个本质维数  $d$  ( $d < D$ ) 的平滑流形, 那么流形学习的目标可以看作是寻找一个从观察空间到低维嵌入空间的映射:  $F(\mathbf{X}) \rightarrow \mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ ,  $\mathbf{y}_i \in \mathbb{R}^d$ , 以及其一对一的逆映射:  $F^{-1}(\mathbf{Y}) \rightarrow \mathbf{X}$ , 这一映射函数应该同时保持流形数据的全局结构与局部几何关系。如前所述, 本文 MLE 算法通过拼接不同的分块局部线性模型来近似待求的非线性映射函数  $F$ , 其中每个局部线性模型对应于从一个 MLP 中学习得到。

#### 2.3.1.1 分块建立算法

直观看来, 最大线性分块 MLP 的字面意思包含了如下两个准则: (1) 线性准则: 针对分块内的每对样本点, 它们的欧氏距离与测地距离的差别应该尽可能小, 从而保证该分块可以展开为一个近似线性的子空间; (2) 最大准则: 分块的样本规模应该尽可能大, 从而使得任意新样本的加入都将破坏上述线性准则。

针对 MLP 分块的构建, 本文从实际需求出发, 采用灵活高效的层次聚类算法思想。这种聚类方式可以得到不同层次级别的聚类层次树(dendrogram), 用户可以根据需要选用合适的聚类层次[18][40]。传统的层次聚类算法有两种: 自顶向下的层次分裂聚类(hierarchical divisive clustering, HDC) 和自底向上的层次融合聚类(hierarchical agglomerative clustering, HAC)。由于在本文实际应用中, 合适的聚类类别个数(即 MLP 个数)远小于原始样本的个数, 考虑到算法计算效率, 采用分裂式的 HDC 聚类框架。

图 2.3 形象地给出了 HDC 聚类算法的流程。在聚类第一层, 所有样本初始化为一个类别(即一个 MLP); 之后, 在新增加的每一个聚类层次中, 上一层当中非线性程度最大的一个 MLP 分块分解为两个规模较小且非线性程度也减小的 MLP 分块。最终, 算法构造得到一个多层次的聚类树, 不同层级上的 MLP 分别对应不同程度的非线性偏差。

形式上, 算法的目标是将原始数据集  $\mathbf{X}$  划分为一组互不相交的 MLP 分块  $\mathbf{X}^{(i)}$  组合:

$$\mathbf{X} = \bigcup_{i=1}^P \mathbf{X}^{(i)}, \quad \mathbf{X}^{(i)} \cap \mathbf{X}^{(j)} = \emptyset \quad (i \neq j, \quad i, j = 1, 2, \dots, P), \quad (2-1)$$

$$\mathbf{X}^{(i)} = \{\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{N_i}^{(i)}\} \quad \left( \sum_{i=1}^P N_i = N \right).$$

这里,  $P$  是 MLP 分块的个数,  $N_i$  是第  $i$  个分块  $\mathbf{X}^{(i)}$  中的样本个数。

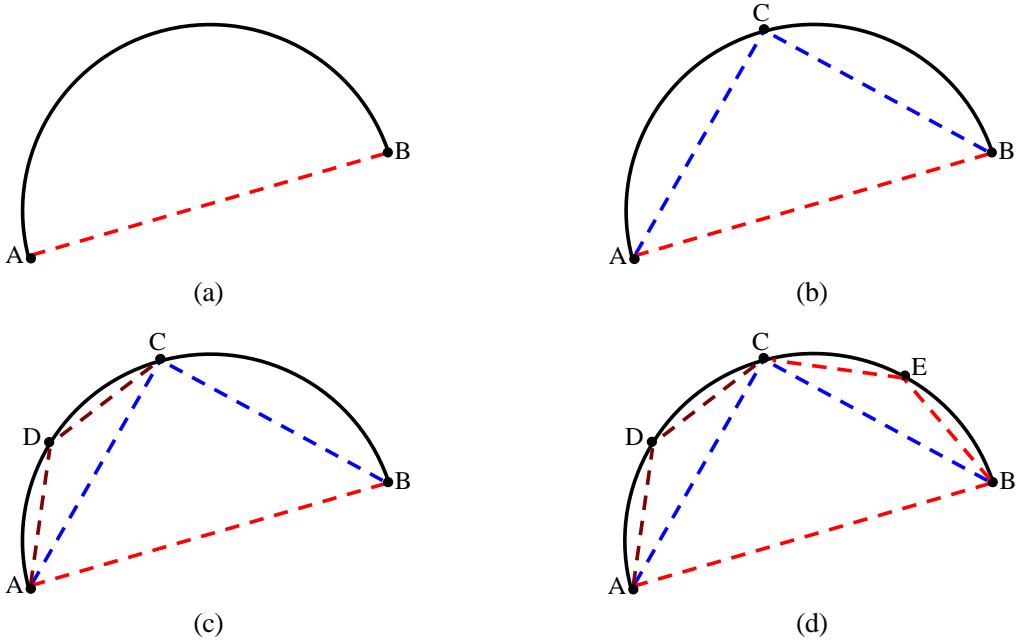


图 2.3 层次分裂聚类算法(HDC)示意图。图中实线圆弧  $\widehat{AB}$  表示一维流形, (a)~(d)分别给出了聚类算法前四层分裂得到的 MLP 结果。在聚类的第一层(a), 整段圆弧  $\widehat{AB}$  初始化为一个 MLP 分块; 在第二层(b),  $\widehat{AB}$  分裂为两段较小的圆弧片段  $\widehat{AC}$  与  $\widehat{BC}$ , 相应地这两个小片段的非线性程度也得到降低; 在第三层(c)和第四层(d),  $\widehat{AC}$  与  $\widehat{BC}$  又分别继续分裂为更小的 MLP 片段。图中虚线段的长度对应于两点之间的欧氏距离, 实线圆弧则对应于其测地距离。

首先, 与 ISOMAP 算法[80]类似, 基于  $k$ -NN 邻域图, 采用 Dijkstra 算法计算样本任意点对之间的测地距离和欧氏距离, 从而得到测地距离矩阵  $\mathbf{D}_G$  与欧氏距离矩阵  $\mathbf{D}_E$ ; 同时, 计算任意点对之间两个距离的比值得到如下矩阵:  $\mathbf{R}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{D}_G(\mathbf{x}_i, \mathbf{x}_j) / \mathbf{D}_E(\mathbf{x}_i, \mathbf{x}_j)$ 。这三个矩阵都是  $N \times N$  规模的方阵。由于两个样本之间的测地距离始终大于等于欧氏距离, 因此矩阵  $\mathbf{R}$  的所有元素都满足  $\mathbf{R}(\mathbf{x}_i, \mathbf{x}_j) \geq 1$ 。此外, 算法需要保存一个  $k \times N$  规模的邻域索引矩阵  $\mathbf{H}$ , 该矩阵的每一列  $\mathbf{H}(:, j)$  ( $j = 1, 2, \dots, N$ ) 对应存放样本  $\mathbf{x}_j$  的  $k$  近邻标号。需要说明的是, 作为计算矩阵  $\mathbf{D}_G$  和  $\mathbf{D}_E$  的副产品, 邻域矩阵  $\mathbf{H}$  不需要额外的计算开销。基于上述计算得到的若干矩阵, 下面可以定义一个“非线性分值函数”, 用来度量 MLP 分块  $\mathbf{X}^{(i)}$  的非线性程度:

$$S^{(i)} = \frac{1}{N_i \cdot N_i} \sum_{m=1}^{N_i} \sum_{n=1}^{N_i} \mathbf{R}(\mathbf{x}_m^{(i)}, \mathbf{x}_n^{(i)}), \quad (2-2)$$

表 2.1 所示的算法 1 给出了本文提出的 HDC 算法具体实现流程, 该算法的运行结果即为  $P$  个互不相交的 MLP 分块。注意步骤 3 中的阈值  $\delta$  指定了算法的终止条件, 从而决定了最终聚类 MLP 的数目与其对应的非线性程度。不难得出, 由于所有的  $S^{(i)}$  均大于等于 1, 当  $\delta$  被指定为任意小于 1 的数值时, 算法可以得到完整的聚类层次树。

表 2.1 算法 1：层次分裂聚类算法(HDC)流程

---

**输入：**流形观察数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$

**输出：**最大线性分块集合  $\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(P)}\}$

---

**1** 初始化：  $\mathbf{X}^{(1)} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $P = 1$ ; 根据公式(2-2)计算非线性分值  $S^{(1)}$ 。

**2** 选择当前非线性分值最大的分块  $\mathbf{X}^{(i)}$  ( $i \in \{1, 2, \dots, P\}$ ) 作为待分裂的父分块。

**3** 如果满足 ( $S^{(i)} > \delta$ ), 则

**3.1** 根据测地距离矩阵  $\mathbf{D}_G$ , 从分块  $\mathbf{X}^{(i)}$  中选择距离最远的两个样本  $\mathbf{x}_l$  和  $\mathbf{x}_r$  作为种子点。

初始化左右两个子分块：  $\mathbf{X}_l^{(i)} = \{\mathbf{x}_l\}$ ,  $\mathbf{X}_r^{(i)} = \{\mathbf{x}_r\}$ ; 更新：  $\mathbf{X}^{(i)} \leftarrow \mathbf{X}^{(i)} \setminus \{\mathbf{x}_l, \mathbf{x}_r\}$ 。

**3.2** 当满足条件  $\mathbf{X}^{(i)} \neq \emptyset$  时, 重复进行如下步骤 3.2.1~3.2.3。

**3.2.1** 针对当前左子分块  $\mathbf{X}_l^{(i)}$ , 构建其邻域点集合  $\mathbf{C}_l$ 。具体地, 依据邻域标号矩阵  $\mathbf{H}$  的记录, 集合  $\mathbf{C}_l$  包含了  $\mathbf{X}_l^{(i)}$  中所有样本的  $k$  近邻标号。

**3.2.2** 针对当前右子分块  $\mathbf{X}_r^{(i)}$ , 构建其邻域点集合  $\mathbf{C}_r$ 。具体构造方式与 3.2.1 类似。

**3.2.3** 依次更新左右子分块与父分块：

$$\mathbf{X}_l^{(i)} \leftarrow \mathbf{X}_l^{(i)} \cup (\mathbf{C}_l \cap \mathbf{X}^{(i)}), \quad \mathbf{X}^{(i)} \leftarrow \mathbf{X}^{(i)} \setminus (\mathbf{C}_l \cap \mathbf{X}^{(i)});$$

$$\mathbf{X}_r^{(i)} \leftarrow \mathbf{X}_r^{(i)} \cup (\mathbf{C}_r \cap \mathbf{X}^{(i)}), \quad \mathbf{X}^{(i)} \leftarrow \mathbf{X}^{(i)} \setminus (\mathbf{C}_r \cap \mathbf{X}^{(i)}).$$

**3.3** 父分块  $\mathbf{X}^{(i)}$  分裂为两个子分块  $\mathbf{X}_l^{(i)}$  和  $\mathbf{X}_r^{(i)}$ 。

更新：  $P \leftarrow P + 1$ ; 计算非线性分值  $S_l^{(i)}$  和  $S_r^{(i)}$ 。

**4** 否则返回当前的聚类结果, 算法终止。

**5** 跳转至步骤 2。

---

### 2.3.1.2 算法细节讨论

针对上述 MLP 构造算法, 有几点值得深入分析讨论。首先是关于 MLP 线性准则的约束函数定义。公式(2-2)中采用了分块内部所有样本点对的测地距离与欧氏距离的比值平均值作为非线性程度度量。除此之外, 别的一些可能的度量准则也可以考虑, 比如分块内所有点对的测地距离之和与其欧氏距离之和的比值, 或者利用两种距离的差值替代其比值等等。本文认为这些准则在一定程度上都是等价的。

算法的另一个特点是其层次聚类方式。如何来确定最终采用的聚类类别即 MLP 分块的个数  $P$  呢? 以图 2.1 所示的“V 型”流形数据为例。利用算法 1 获取该流形的 MLP 层次聚类结果, 并计算聚类每一级所有 MLP 分块的非线性分值的平均值。图 2.4(a)给出了该平均分值随聚类数目增加的变化曲线。可以看出, 随着聚类层级和 MLP 分块数量

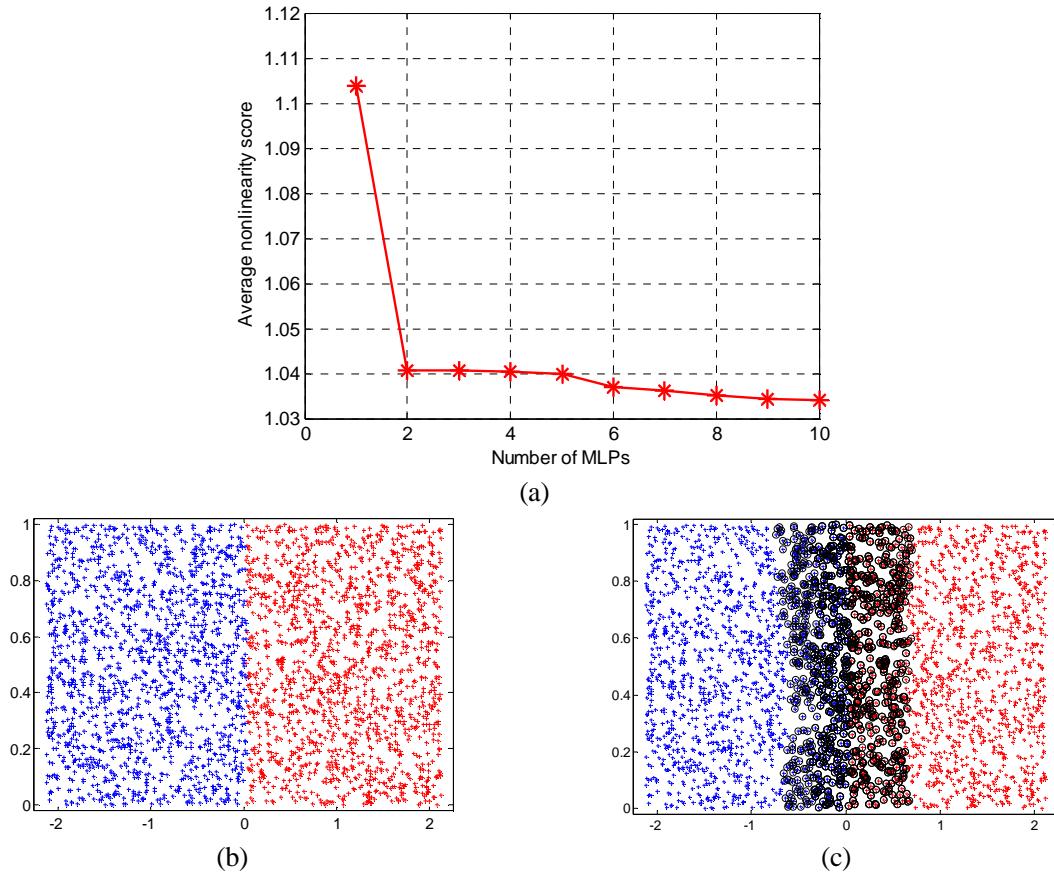


图 2.4 V 型流形的 HDC 算法聚类结果。(a) 聚类每一层 MLP 分块的平均非线性分值曲线; (b) (c) 分别给出 XY 视角下的进行软划分前后得到的 MLP 分块结果, 不同的颜色对应不同的 MLP。与算法的期望结果一致, 流形上的两个线性平面对应得到两个 MLP (图示 b), 经过软划分之后, 在不同 MLP 分块的邻接处又得到了一些额外的分块 (图示 c 的黑色圆圈所示)。

的增多, 整体的非线性程度逐渐下降, 这与算法的预期输出是一致的。幸运的是, 这一下降曲线正好为合适的聚类个数的选取提供了一个重要的参考。一种简单且有效的选择是当曲线下降到某个分值 (其典型取值为 1.1) 后, 曲线上的拐点(elbow), 即纵轴取值变化趋势开始趋于平缓的位置。在拐点位置之后, 即使聚类数目发生大幅度增加, 其平均非线性分值的变化也相对非常轻微。从图 2.4(a)可以看出, “V 型” 流形的理想分块个数应为 2, 图 2.4(b)给出了对应的两个 MLP 分块聚类结果, 这一结果与预期也是一致的。

进一步地, 从图 2.4(b)所示的两个独立不相交的 MLP 引出一个新的问题: 两分块连接处的样本与它们的近邻可能被分别划分到了不同的分块中去, 这部分样本称之为边界样本。本质上, 算法 1 得到的 MLP 聚类结果是对流形样本集合的一个硬划分, 即不同聚类类别之间没有重叠。为了减轻这种分块不重叠对最终结果稳定性的潜在影响, 可以对上述得到的硬划分结果进行软扩展, 在任意相邻分块边界处构造一些新的分块与之重叠, 从而将相邻的分块连接起来。具体做法是, 分别采用上述每个边界点为种子点初始化一个新的 MLP 分块, 通过邻域扩张的方式不断增加该分块内的样本规模, 直至分块的非线性程度达到上述层次分裂算法硬划分得到的既有 MLP 分块的非线性程度, 详细

的算法流程请参考本文前期工作[89]中提出的一站式聚类算法。图 2.4(c)展示了最终经过软划分得到的 MLP 分块结果。

很显然，扩展后的软划分结果是对原始数据集  $\mathbf{X}$  的一个相对平滑的分解，从而可以带来更稳定的低维嵌入空间，并且在一定程度上保证了学习所得映射函数的连续性。在后续章节中，采用  $M$  标记软划分得到的 MLP 个数。

### 2.3.2 局部线性模型

如前所述，接下来需要为每个 MLP 分块构建其局部模型表示。由于其简单有效的特点，本文中采用 PCA 子空间方法来完成这个建模过程。形式上，针对 MLP 分块  $\mathbf{X}^{(i)}$  中的每个样本  $\mathbf{x}_m^{(i)}$ ，其 PCA 投影通过如下公式得到：

$$\mathbf{z}_m^{(i)} = \mathbf{W}_i^T \cdot (\mathbf{x}_m^{(i)} - \bar{\mathbf{x}}^{(i)}), (m=1, 2, \dots, N_i, i=1, 2, \dots, M) \quad (2-3)$$

其中，样本均值

$$\bar{\mathbf{x}}^{(i)} = \frac{1}{N_i} \sum_{m=1}^{N_i} \mathbf{x}_m^{(i)} \quad (2-4)$$

和  $D \times d$  的主成分矩阵

$$\mathbf{W}_i = [\mathbf{p}_1^{(i)}, \mathbf{p}_2^{(i)}, \dots, \mathbf{p}_d^{(i)}] \quad (2-5)$$

共同刻画了分块  $\mathbf{X}^{(i)}$  对应学习到的局部模型  $\mathcal{M}_i$ 。

事实上，每个局部模型  $\mathcal{M}_i$  对应于输入样本空间的一个局部的  $d$  维笛卡尔坐标系统，该坐标系以  $\bar{\mathbf{x}}^{(i)}$  为原点，每一维坐标轴平行于主成分矩阵  $\mathbf{W}_i$  的一个列向量。这里，主成分的维数  $d$  可以通过 PCA 计算中保持最大方差的比例来确定，而且由于所有 MLP 分块来自于同一个低维流形，它们应该对应一个共同的维数  $d$ 。后续的 2.3.3.4 节中将详细介绍  $d$  的估计方法。

分块  $\mathbf{X}^{(i)}$  中的样本对应的低维局部模型表示可以记为：

$$\mathbf{Z}^{(i)} = \{\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}, \dots, \mathbf{z}_{N_i}^{(i)}\}, (i=1, 2, \dots, M). \quad (2-6)$$

### 2.3.3 全局对齐算法

采用上述 PCA 建立的若干局部模型可以很好的保持流形上样本间的局部几何关系，如何寻找一个统一的坐标空间以保持这些局部模型之间的全局拓扑关系，即流形的全局结构，成为接下来要解决的问题。

#### 2.3.3.1 地标点样本集构造

事实上，流形的全局结构可以通过不同局部模型对应的样本均值和主成分坐标轴来刻画。因此，全局低维嵌入空间的一个自然的选择便是利用这些样本均值和沿着各局部

模型主成分坐标轴采样的一些样本点进行 MDS 计算得到的保距坐标空间。这些均值点和沿主成分坐标轴采样的样本统称为地标点(landmarks)。显然，这里的 MDS 计算应该基于样本间的测地距离，因为不同局部模型之间的关系反映的正是流形的全局非线性结构。

理论上，每个局部模型只需要其样本均值点和沿着其  $d$  维主成分坐标轴分别采样的一个样本，即总计  $d+1$  个地标点，就可以完全表示。而实际上，每个局部模型内的样本均值往往不是真实存在的训练样本，这种情况下可以采用离均值最近的训练样本来代替均值点，该训练样本称之为重心点(centroid)。与之类似，样本均值之外的其余地标点也不必刻意沿着主成分坐标轴的方向选取，只需要在各局部模型内部随机采样若干样本并保证其数量略大于  $d$  即可，以保证算法结果的可靠性。

具体地，针对每个 MLP 分块  $\mathbf{X}^{(i)}$ ，随机选取一定数量比如  $ni$  ( $ni \geq d+1$ ) 个样本作为地标点，并将这些地标点集合记为：

$$\mathbf{X}_L^{(i)} = \{\mathbf{x}_{L(1)}^{(i)}, \dots, \mathbf{x}_{L(ni)}^{(i)}\}. \quad (2-7)$$

这里， $L(k)$  ( $k=1, 2, \dots, ni$ ) 是各个地标点在原始分块  $\mathbf{X}^{(i)}$  中的样本标号，具体请参考公式(2-1)。为了后续表示方便，各分块的第一个地标点  $\mathbf{x}_{L(1)}^{(i)}$  选用其重心点。

将所有分块的地标点合并起来，构成如下集合：

$$\mathbf{X}_L = \bigcup_{i=1}^M \mathbf{X}_L^{(i)} = \{\mathbf{x}_{L(1)}^{(1)}, \dots, \mathbf{x}_{L(n1)}^{(1)}; \dots; \mathbf{x}_{L(1)}^{(M)}, \dots, \mathbf{x}_{L(nM)}^{(M)}\}. \quad (2-8)$$

相应地，这些地标点在各自的局部模型内部的坐标表示构成如下集合：

$$\mathbf{Z}_L = \bigcup_{i=1}^M \mathbf{Z}_L^{(i)} = \{\mathbf{z}_{L(1)}^{(1)}, \dots, \mathbf{z}_{L(n1)}^{(1)}; \dots; \mathbf{z}_{L(1)}^{(M)}, \dots, \mathbf{z}_{L(nM)}^{(M)}\}. \quad (2-9)$$

对于第  $i$  个 MLP 分块，如前所述，当其样本均值  $\bar{\mathbf{x}}^{(i)}$  不是训练集中的真实样本点时，需要采用其最近邻训练样本（该样本记作  $\mathbf{x}_n^{(i)}$ ）代替。为了表述的一致性，PCA 局部模型的坐标原点也需要相应地平移到  $\mathbf{z}_n^{(i)}$ ，即  $\mathbf{Z}^{(i)}$  内所有样本的坐标需要同时减去  $\mathbf{z}_n^{(i)}$ 。经过这些处理，可以得出，公式(2-9)中的  $\mathbf{z}_{L(1)}^{(i)}$  变为一个  $d$  维零向量，即

$$\mathbf{z}_{L(1)}^{(i)} = [0, 0, \dots, 0]^T, (i=1, 2, \dots, M). \quad (2-10)$$

为表示方便，仍采用  $\bar{\mathbf{x}}^{(i)}$  表示分块的重心样本点。

### 2.3.3.2 基于地标点的全局对齐

基于上述地标点集合，就可以利用 MDS 求解全局坐标空间了。然后，通过计算各局部模型到该全局空间的线性变换关系完成全局对齐。图 2.5 以“V 型”流形为例形象地描述了这一对齐过程。

给定地标点集合  $\mathbf{X}_L$  和集合内所有样本点对的测地距离（这些测地距离可以通过 2.3.1.1 节所述的  $\mathbf{D}_G$  矩阵得到），经典 MDS 算法就可以将这些地标点唯一地嵌入到一个  $d$  维欧氏空间  $\mathcal{E}$  中。得益于 MDS 算法所具有的良好的距离保持属性，这里得到的空间  $\mathcal{E}$  便

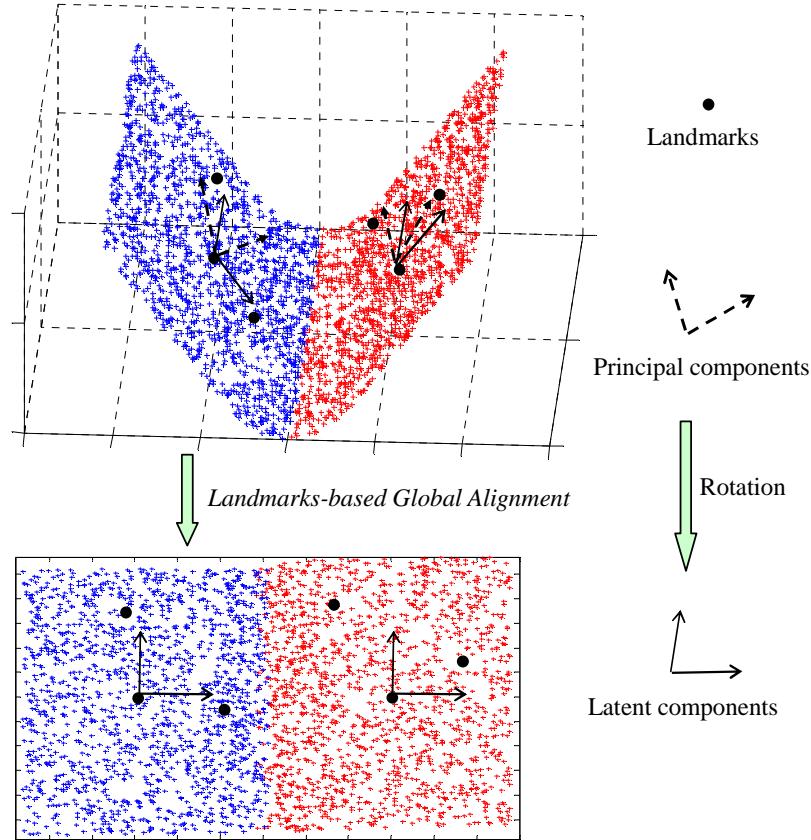


图 2.5 基于地标点的全局对齐算法(LGA)示意图。首先, 针对地标点进行 MDS 运算得到全局坐标空间; 然后, 利用地标点样本在各局部模型内部的坐标和它们在全局空间的坐标来计算局部模型与全局空间之间的线性变换关系; 最后, 针对各局部模型应用相应的线性变换将训练样本投影到全局坐标空间中从而完成降维。从各个局部线性模型到全局坐标空间的线性变换的核心是从 PCA 的主成分基向量到语义成分(Latent components)基向量的旋转变换, 这些语义成分基向量反映了流形本质变化模式。

可以作为一个理想的目标空间用来保距地嵌入整个训练样本集合  $\mathbf{X}$ 。地标点集合在这个全局空间  $\mathcal{E}$  中的坐标表示如下:

$$\tilde{\mathbf{Y}}_L = \bigcup_{i=1}^M \tilde{\mathbf{Y}}_L^{(i)} = \{\tilde{\mathbf{y}}_{L(1)}^{(1)}, \dots, \tilde{\mathbf{y}}_{L(n1)}^{(1)}; \dots; \tilde{\mathbf{y}}_{L(1)}^{(M)}, \dots, \tilde{\mathbf{y}}_{L(nM)}^{(M)}\}, \text{ 其中} \quad (2-11)$$

$$\tilde{\mathbf{Y}}_L^{(i)} = \{\tilde{\mathbf{y}}_{L(1)}^{(i)}, \dots, \tilde{\mathbf{y}}_{L(ni)}^{(i)}\}. \quad (2-12)$$

只需要计算出各局部模型与上述统一的全局空间  $\mathcal{E}$  之间的线性变换关系, 就可以完成整个训练集数据的低维嵌入投影。首先分析一下这些局部模型与该全局空间的关系。一方面, 对采样自各局部模型的所有地标点进行 MDS 分析, 等价于分别对各局部模型内部的地标点进行单独的 MDS 分析 (二者之间相差一个线性变换)。另一方面, 针对各局部模型 (即 MLP) 进行的 MDS 分析近似等价于对其进行 PCA 分析 (二者之间同样相差一个线性变换) [71], 因为根据公式(2-2)中的非线性程度的约束, 每个 MLP 内部的测地距离是近似等于欧氏距离的。综上可以得出, 针对每个 MLP 分块, 其样本的基于 PCA

局部模型的低维表示 (即  $\mathcal{M}_i$ ) 近似等价于它们在全局空间  $\mathcal{E}$  中的低维表示, 这两种表示之间同样相差一个线性变换。接下来, 就需要针对各 MLP 分块求解这一线性变换的三个参数, 即旋转、平移和尺度参数, 从而将它们的 PCA 局部模型表示  $\mathcal{M}_i$  变换到其在全局空间中  $\mathcal{E}$  相应的局部嵌入表示。这里, 将第  $i$  个 MLP 分块在空间  $\mathcal{E}$  中的局部嵌入表示记为  $\mathbf{E}_i$ 。与  $\mathcal{M}_i$  不同的是, 这些局部嵌入  $\mathbf{E}_i$  已经在全局空间中得到了一致的对齐。

不难理解, 对于每个 MLP 分块,  $\tilde{\mathbf{y}}_{L(1)}^{(i)}$  对应于其样本在全局空间  $\mathcal{E}$  中局部嵌入表示的中心。因此, 从  $\mathbf{E}_i$  中所有样本点的坐标表示中减去  $\tilde{\mathbf{y}}_{L(1)}^{(i)}$  就可以去除平移参数的影响。对于尺度参数, 只需要将地标点样本的局部嵌入表示  $\mathbf{E}_i$  的坐标缩放某个尺度, 使得它们在全局空间  $\mathcal{E}$  中的欧氏距离等于它们的高维表示  $\mathbf{X}_L^{(i)}$  在原始空间中的测地距离即可。

形式上, 将去除了尺度和平移参数之后的地标点在空间  $\mathcal{E}$  中的坐标表示记为:

$$\hat{\mathbf{Y}}_L^{(i)} = \{\hat{\mathbf{y}}_{L(1)}^{(i)}, \dots, \hat{\mathbf{y}}_{L(ni)}^{(i)}\}, (i=1, 2, \dots, M), \text{ 其中} \quad (2-13)$$

$$\hat{\mathbf{y}}_{L(k)}^{(i)} = s_i \cdot (\tilde{\mathbf{y}}_{L(k)}^{(i)} - \tilde{\mathbf{y}}_{L(1)}^{(i)}), (k=1, 2, \dots, ni). \quad (2-14)$$

$s_i$  表示尺度因子。由于所有地标点通过 MDS 运算统一嵌入在同一个全局空间  $\mathcal{E}$  内, 因此对于不同的 MLP 分块, 它们的尺度因子应该是相同的。为了表示方便, 后续讨论中假设  $s_i = 1$ 。经过尺度和平移参数变换之后, 公式(2-13)中的  $\hat{\mathbf{y}}_{L(1)}^{(i)}$  变为一个  $d$  维零向量。此时, 地标点样本在  $\mathbf{Z}_L^{(i)}$  与  $\hat{\mathbf{Y}}_L^{(i)}$  中的坐标表示仅相差一个旋转变换。根据线性代数理论, 这个旋转变换实质上是通过一个  $d \times d$  的过渡矩阵  $\mathbf{T}_i$  刻画的, 该过渡矩阵理论上应该是一个正交矩阵且满足如下的坐标变换公式:

$$[\mathbf{z}_{L(1)}^{(i)} \cdots \mathbf{z}_{L(ni)}^{(i)}]_{d \times ni} = \mathbf{T}_i \cdot [\hat{\mathbf{y}}_{L(1)}^{(i)} \cdots \hat{\mathbf{y}}_{L(ni)}^{(i)}]_{d \times ni}. \quad (2-15)$$

令  $\mathbf{A}_i = [\mathbf{z}_{L(1)}^{(i)} \cdots \mathbf{z}_{L(ni)}^{(i)}]_{d \times ni}$ ,  $\mathbf{B}_i = [\hat{\mathbf{y}}_{L(1)}^{(i)} \cdots \hat{\mathbf{y}}_{L(ni)}^{(i)}]_{d \times ni}$ ,  $\mathbf{T}_i$  可以通过下式求解:

$$\mathbf{T}_i \doteq \mathbf{A}_i \mathbf{B}_i^\dagger = \mathbf{A}_i \mathbf{B}_i^T (\mathbf{B}_i \mathbf{B}_i^T)^{-1}. \quad (2-16)$$

这里,  $(\cdot)^\dagger$  表示 Moore-Penrose 伪逆。对每个 MLP 分块  $\mathbf{X}^{(i)}$ , 矩阵  $\mathbf{T}_i$  刻画了其局部 PCA 模型表示  $\mathcal{M}_i$  到其在全局统一空间  $\mathcal{E}$  中的局部嵌入表示  $\mathbf{E}_i$  之间的变换关系。

需要注意, 公式(2-16)的计算中需要求解  $d \times d$  的矩阵  $\mathbf{B}_i \mathbf{B}_i^T$  的逆。幸运的是, 如 2.3.3.1 节所讨论, 算法随机选取的  $ni$  ( $ni \geq d+1$ ) 个地标点通常可以保证  $\text{rank}(\mathbf{B}_i) \geq d$ , 从而满足  $\mathbf{B}_i \mathbf{B}_i^T$  的非奇异性。

利用上面计算的各局部线性模型到全局坐标空间的线性变换关系, 只需要对各模型内部的样本进行相应的投影, 就可以完成对全体训练样本的统一降维。这一变换过程仅包含如下简单的代数运算:

$$\begin{aligned} \mathbf{y}_m^{(i)} &= \mathbf{T}_i^{-1} \cdot \mathbf{z}_m^{(i)} + \tilde{\mathbf{y}}_{L(1)}^{(i)} = \mathbf{T}_i^{-1} \cdot (\mathbf{W}_i^T \cdot (\mathbf{x}_m^{(i)} - \bar{\mathbf{x}}^{(i)})) + \tilde{\mathbf{y}}_{L(1)}^{(i)}, \\ (m &= 1, 2, \dots, Ni, i = 1, 2, \dots, M). \end{aligned} \quad (2-17)$$

将各局部模型内样本的降维结果综合起来, 根据这些样本在原始训练集中的标号,

就可以得到训练集降到  $d$  维空间后的坐标表示:  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ ,  $\mathbf{y}_i \in \mathbb{R}^d$ 。有一点需要注意的是, 在 2.3.1.2 节中, 针对那些分块边界点进行了软划分扩展, 从而使得同一个边界点隶属于多个局部线性分块。针对这些边界样本, 目前算法采用的是一种简单的处理方式, 即通过计算它们在对应的多个局部模型降维结果的平均值得到其最终降维坐标。

经过本小节介绍的算法流程处理, 目前已经可以得到显式的参数式降维映射函数:

$F = \{F_1, F_2, \dots, F_M\}$ , 其中函数  $F_i (i=1, 2, \dots, M)$  通过公式(2-17)中的参数组  $\{\bar{\mathbf{x}}^{(i)}, \mathbf{W}_i, \mathbf{T}_i, \tilde{\mathbf{y}}_{L(i)}^{(i)}\}$  来刻画。

表 2.2 算法 2: 可视化算法 (前向映射) 流程

**输入:** 观察空间的测试样本  $\mathbf{x}_t \in \mathbb{R}^D$

**输出:** 嵌入空间的  $d$  维投影  $\mathbf{y}_t \in \mathbb{R}^d$

**1** 根据如下最小重构误差准则寻找  $\mathbf{x}_t$  对应的局部线性模型的索引  $j$ :

$$j = \arg \min_i (\|\mathbf{x}_t - (\mathbf{W}_i \cdot \mathbf{W}_i^T \cdot (\mathbf{x}_t - \bar{\mathbf{x}}^{(i)}) + \bar{\mathbf{x}}^{(i)})\|), (i=1, 2, \dots, M) \quad (2-19)$$

**2** 计算嵌入空间的坐标:

$$\mathbf{y}_t = \mathbf{T}_j^{-1} \cdot \mathbf{W}_j^T \cdot (\mathbf{x}_t - \bar{\mathbf{x}}^{(j)}) + \tilde{\mathbf{y}}_{L(i)}^{(j)} \quad (2-20)$$

表 2.3 算法 3: 重构算法 (后向映射) 流程

**输入:** 嵌入空间的测试样本  $\mathbf{y}_t \in \mathbb{R}^d$

**输出:** 观察空间的  $D$  维重构  $\mathbf{x}_t \in \mathbb{R}^D$

**1** 根据如下最近中心准则寻找  $\mathbf{y}_t$  对应的局部线性模型的索引  $j$ :

$$j = \arg \min_i (\|\mathbf{y}_t - \tilde{\mathbf{y}}_{L(i)}^{(i)}\|), (i=1, 2, \dots, M) \quad (2-21)$$

**2** 计算观察空间的重构样本:

$$\mathbf{x}_t = \bar{\mathbf{x}}^{(j)} + \mathbf{W}_j \cdot \mathbf{T}_j \cdot (\mathbf{y}_t - \tilde{\mathbf{y}}_{L(i)}^{(j)}) \quad (2-22)$$

### 2.3.3.3 解析映射函数

公式(2-17)的映射函数显式地给出了从高维观察空间到低维嵌入空间的前向映射, 其逆映射可以通过完全反向的方式推导得出:

$$\mathbf{x}_m^{(i)} = \bar{\mathbf{x}}^{(i)} + \mathbf{W}_i \cdot \mathbf{T}_i \cdot (\mathbf{y}_m^{(i)} - \tilde{\mathbf{y}}_{L(i)}^{(i)}), (m=1, 2, \dots, N_i, i=1, 2, \dots, M). \quad (2-18)$$

从公式(2-17)和(2-18)可以看出 MLE 算法的一个显著的优点: 当映射函数  $F$  学习完成之后, 后续应用就不再需要训练数据集合, 这一点可以极大节省算法的计算与存储开销。

具体地, 由于高维观察空间和低维嵌入空间之间的映射关系是通过一组混合的分块线性变换反映的, 针对新来的测试数据, MLE 算法只需要通过一定的准则寻找该测试样本所对应的局部模型进而利用相应的线性变换完成投影即可。表 2.2 和 2.3 形式化地给出了针对测试数据分别来自高维观察空间和低维嵌入空间时, 对应的前向/后向投影算法, 分别称为可视化算法和重构算法。

#### 2.3.3.4 流形内在结构描述

针对流形内在本质结构的描述, 本节探讨这方面的两个核心问题: 流形本质维数的估计和相应的内在变化模式的刻画。

流形的本质维数  $d$  反映了观察数据的本质自由度。理论上, 基于流形分布的假设, MLE 算法中得到的局部 PCA 子空间与全局统一嵌入空间的维数应该都是  $d$ 。如 2.3.2 节所述, PCA 的维数  $d$  可以通过保持最大方差的比例来粗略设定。另一方面, 利用 MDS 求解全局嵌入空间时, 类似于 ISOMAP 文章的做法[80], 维数  $d$  则可以通过观察距离保持残差曲线的拐点来估计。在本文中, MLE 算法将 PCA 与 MDS 的优点结合起来, 采用一种 validation-feedback 的方式来综合两种方法估计的结果。

首先, 通过 PCA 和 MDS 的估计结果, 可以得到对维数  $d$  取值区间的一个初略估计, 如  $[d_{\min}, d_{\max}]$ 。然后, 利用公式(2-15)中的变换误差作为代价函数来评估上述取值区间中的每个候选值。换言之, 本文提出的维数估计算法的目标是最小化地标点样本的局部 PCA 模型表示与其全局 MDS 空间表示之间的线性变换偏差。具体地, 针对每个 MLP 分块, 由于过渡矩阵  $\mathbf{T}_i$  通过公式  $\mathbf{A}_i = \mathbf{T}_i \cdot \mathbf{B}_i$  求解得到, 那么理论上  $\mathbf{B}_i^* = \mathbf{T}_i^{-1} \cdot \mathbf{A}_i$  应该尽可能接近  $\mathbf{B}_i$ 。因此, 维数  $d$  的最优取值估计可以通过如下的目标函数得到:

$$d^* = \arg \min_d \sum_{i=1}^M \sum_{k=1}^{n_i} \left\| \mathbf{T}_i^{-1} \cdot \mathbf{z}_{L(k)}^{(i)} - \hat{\mathbf{y}}_{L(k)}^{(i)} \right\|, (d_{\min} \leq d \leq d_{\max}). \quad (2-23)$$

这里  $\mathbf{T}_i \in \mathbb{R}^{d \times d}$ ,  $\mathbf{z}_{L(k)}^{(i)}, \hat{\mathbf{y}}_{L(k)}^{(i)} \in \mathbb{R}^{d \times 1}$ 。通过公式(2-23)的优化函数, 就可以将 PCA 与 MDS 的维数估计结果有效的结合起来进行最终的决策。由于在 PCA 和 MDS 两个算法的降维过程中, 随着结果维数的增加, 之前得到的低维坐标保持不变, 因此公式(2-23)中针对不同候选  $d$  值进行评价时, 只需要 PCA 和 MDS 算法运行一次即可, 整个过程非常高效。

完成了本质维数的估计之后, 进入另一个更为困难的问题, 即如何推断隐藏在流形观察数据中的内在变化模式, 这里的每一种变化模式分别对应本质维数的一个维度。针对这一问题, 前人工作如[66][80][93]通常采用的做法是将训练数据投影到低维嵌入空间中, 通过可视化分析观察数据的分布来间接地推断数据变化模式, 而无法进行显式建模。

本文尝试显式地建模流形的内在变化模式。首先回顾一下公式(2-15), 该式本质上反映了从各局部 PCA 子空间到统一的全局嵌入空间的坐标变换。相应地, 通过各自的基变换公式, 可以将 PCA 模型对应的一组基向量  $\mathbf{W}_i$  (图 2.5 所示的虚线坐标轴) 变换到另

一组基向量  $\mathbf{W}_i^E$  (图 2.5 所示的实线坐标轴), 具体可以写为如下的公式:

$$\mathbf{W}_i^E = \mathbf{W}_i \cdot \mathbf{T}_i = [\mathbf{q}_1^{(i)}, \mathbf{q}_2^{(i)}, \dots, \mathbf{q}_d^{(i)}], \quad (i=1,2,\dots,M). \quad (2-24)$$

如图 2.5 所示, 经过这一变换之后, 不同局部模型对应的基  $\mathbf{W}_i^E$  在嵌入空间中得到了一致的对齐。可以看出,  $d \times d$  的过渡矩阵  $\mathbf{T}_i$  刻画的正是图 2.5 中的旋转操作。将公式(2-24)代入(2-18)并经过简单的变换, 可以得到下式:

$$\mathbf{x}_m^{(i)} - \bar{\mathbf{x}}^{(i)} = \mathbf{W}_i^E \cdot (\mathbf{y}_m^{(i)} - \tilde{\mathbf{y}}_{L(1)}^{(i)}). \quad (2-25)$$

可以看到, 该式中  $\mathbf{W}_i^E$  所起的作用与因子分析[19][25]中因子载荷矩阵的功能非常类似。由于基向量  $\mathbf{W}_i^E$  直接反映了流形高维观察数据的语义变化模式, 类似于 PCA 方法中将各维基向量称为主成分(principal component), 本文类似地将基向量  $\mathbf{q}_1^{(i)}, \mathbf{q}_2^{(i)}, \dots, \mathbf{q}_d^{(i)}$  称为语义成分(Latent Components, LCs), 其中每一个成分  $\mathbf{q}_j^{(i)}$  ( $j=1,2,\dots,d$ ) 对应于流形嵌入空间的一维坐标轴。这一新的理论发现在当前流形学习研究中尚未提及, 由于其显式地描述了流形数据的变化模式, 可以期待并预见的是, 语义成分可以广泛应用于计算机视觉和模式识别的很多实际问题。

## 2.3.4 讨论

### 2.3.4.1 与前人工作对比

可以看出, 本文提出的 MLE 算法采用了与全局协调算法(global coordination)[67]以及其后续工作[10][79][85][102][103]相类似的算法框架。总的来说, 这些方法都采用了文献[67]中最初提出的思想, 即: 将全局非线性流形分解为一组局部线性模型组合, 然后通过某些准则将这些局部模型一致地对齐到统一的全局嵌入空间。

文献[67]和[85]均采用了期望最大化算法(expectation-maximization, EM)来同时进行局部线性模型的拟合与对齐。尽管与文献[67]相比, 文献[85]对局部模型进行了更强的约束从而在一定程度上改进了算法的训练流程, 然而 EM 迭代算法的采用仍然在很大程度上带来了训练流程的低效。此外, 作者在文献[67]中也指出, 基于 EM 迭代的优化算法经常受到局部极值的干扰, 因此[67]和[85]中的训练流程均需要利用其它算法(如 LLE 或者 ISOMAP)的结果作为初始化, 从而监督其 EM 迭代优化的过程。

与[67]和[85]中的算法不同, charting[10], LLC[79]以及本文提出的 MLE 算法都可以看作是采用了先拟合后对齐的学习方式, 即局部模型的对齐是在完成了模型的拟合之后进行的[79]。通过将局部模型的拟合与对齐两个阶段分解开来, 这三个方法均通过非迭代的方式获得了算法的闭形式解(closed-form)。采用不同的凸代价函数, 这三个方法有效避免了对齐阶段的局部极值问题。在计算效率方面, charting 方法的局部模型是针对每个训练样本逐个建立的, 因此其求解的特征分解问题规模是与样本个数相同的, 这一点与 LLE 和 ISOMAP 算法相同, 因此其时间开销比较大[56]。与之不同, LLC 与 MLE 二者求解的特征分解问题都是与局部模型个数相关, 其运算规模远远小于训练样本的个数。

接下来分析 MLE 与 LLC 的关系。在文献[79]中, LLC 算法主要是作为一种通用的局部模型对齐算法进行研究的。该算法的第一阶段, 即局部模型拟合阶段, 采用的局部模型实例为混合因子分析(mixture of factor analyzers, MFA)[25]。由于 MFA 模型的拟合需要利用 EM 算法迭代优化, 如文献[59]所述, 这一过程一方面可能会陷入局部极值, 另一方面可能受到噪声数据的影响。此外, LLC 算法不仅需要优化局部模型 MFA 的参数, 还需要优化局部模型的个数。与之相比, 本文提出的基于 MLP 的线性模型, 尽管无法保证得到最优的局部线性模型, 但是其具有的显式非线性程度度量可以大大方便选择合适数量的局部模型。针对第二阶段, 即模型对齐阶段, 由于 LLC 算法中施加的单位协方差矩阵约束, 可能会带来流形全局尺度的缩放和变形。与之相比, MLE 采用的 LGA 对齐算法可以看作是采用了 ISOMAP 算法中基于保距思想的凸代价函数, 因此其嵌入空间可以尽可能完整地保持流形全局尺度信息。

此外, 局部切空间对齐算法(local tangent space alignment, LTSA) [103]和局部多维尺度变换算法(locally multi-dimensional scaling, LMDS) [102]也采用了 LLC 算法中类似的对齐算法框架。其核心思想是, 非线性流形的全局结构可以通过对互相重叠交叉的若干局部结构的分析来获取[103]。然而, LTSA 和 LMDS 算法中采用的局部模型仍然都是基于  $k$ -NN 的邻域结构, 正如作者在[102][103]中指出的, 邻域参数的设置对算法结果的合理性有很大的影响。与 charting 算法[10]类似, LTSA 算法中的局部模型采用了极度稠密采样的方式, 即针对每个训练样本建立一个局部模型。为了去除这种过度稠密所带来的冗余计算代价, LMDS 算法尝试寻找一组近似最小的重叠式局部模型组合。更进一步, LTSA 和 LMDS 这两个算法都没有学习得到参数式的映射函数用于处理 out-of-sample 问题。尽管 LMDS 算法中讨论了一种非参数式的解决思路, 然而该方法面临与文献[7]的非参数映射函数相同的计算复杂度问题。此外, 文献[102]的作者也没有给出具体的实验以验证其算法的非参数映射扩展功能。

### 2.3.4.2 算法复杂度分析

总体来说, MLE 算法的计算复杂度集中在如下四个阶段:

(1) 计算  $N \times N$  规模的三个矩阵  $\mathbf{D}_E$ ,  $\mathbf{D}_G$  和  $\mathbf{R}$ 。任意点对的欧氏距离矩阵  $\mathbf{D}_E$  的计算复杂度为  $O(N^2)$ 。测地距离矩阵  $\mathbf{D}_G$  的计算中, 采用基于 Fibonacci heaps 的 Dijkstra 算法的复杂度为  $O(N^2 \log N + kN^2 / 2)$  [56], 其中参数  $k$  为  $k$ -NN 邻域图中的邻域大小。矩阵  $\mathbf{R}$  的计算复杂度为  $O(N^2)$ 。

(2) 基于算法 1 构建 MLP 分块。从表 2.1 所示的算法流程可以看出, 该算法的大部分步骤都是对已有矩阵元素的访问操作。最主要的运算集中于步骤 3.3, 即根据公式(2-2)计算每个 MLP 分块的非线性分值  $S^{(i)}$ 。由于精确的复杂度分析很难得到, 为简便考虑, 假设步骤 3.3 中分裂得到的左右两个子分块  $\mathbf{X}_l^{(i)}$  和  $\mathbf{X}_r^{(i)}$  总是具有相同的样本规模。因此, 算法 1 的总复杂度为  $O(\sum_{p=1}^{\lfloor \log N \rfloor} (2^p (N/2^p)^2)) \approx O(N^2)$ 。

(3) 建立局部 PCA 模型。针对每个 MLP 分块  $X^{(i)}$ ，其对应的样本数据矩阵规模为  $D \times Ni$ ，PCA 建模的主要运算是对  $D \times D$  的协方差矩阵进行特征分解。由于在实际问题中，样本的观察维数往往大于样本个数，即  $D \gg Ni$ ，这里的特征分解可以在  $Ni \times Ni$  规模的矩阵上进行，然后再通过一些简单的矩阵运算来实现。因此，针对  $M$  个 MLP 分块，每个分块建立 PCA 模型的复杂度为  $O(\min(D, Ni)^3)$ 。

(4) 基于 MDS 进行局部模型的对齐。如前所述，全局对齐算法中的 MDS 运算是针对地标点集合进行的，每个局部模型中选取的地标点数量最小值为  $d+1$ ，因此这里 MDS 的复杂度为  $O(M^3 d^3)$ ，与局部模型的个数  $M$  成正比。与 ISOMAP 和 LLE 算法中  $O(N^3)$  的复杂度相比，这里采用的基于地标点的 MDS 处理具有很高的计算效率。MDS 运算完成之后，对齐算法剩余的操作主要是公式(2-16)和(2-17)中的一些简单的矩阵乘法，其中公式(2-16)的矩阵逆运算只需要求解  $d \times d$  规模的特征分解，这一步也非常高效。

综合上述分析，MLE 算法的总计算复杂度可以表示为上述四个阶段的总和，近似为

$$O(N^2 \log N + kN^2 + \sum_{i=1}^M \min(D, Ni)^3 + M^3 d^3) .$$

一般而言，参数  $Ni$ ， $M$  和  $d$  要远小于  $N$ ，因此算法总复杂度进一步的粗略近似为  $O(N^2 \log N + kN^2)$ ，可以看出，整个算法的计算量集中在第一阶段的欧氏距离和测地距离矩阵计算。

关于算法测试阶段的复杂度，从公式(2-19)至(2-22)可以看出，其主要的计算均为简单的线性代数运算，时间开销非常小。

## 2.4 实验结果

本节通过一系列实验，分别在人工数据集和真实数据集上，针对 MLE 算法在数据降维与重构方面的有效性，进行了详细的评估对比。

### 2.4.1 三维人工数据库实验

本小节中，首先选用两个基准的三维流形：瑞士卷(swiss-roll)和 S 曲线(s-curve)，来验证算法的有效性。其中，两个流形各自从原始的三维流形曲面均匀随机采样 3000 个数据点用作训练样本。MLE 算法中的主要参数包括：(i)近邻参数  $k$ ，(ii)硬划分 MLP 分块个数  $P$ ，以及(iii)每个 MLP 分块中的地标点样本个数  $ni$ 。本节实验中针对两个数据库采用了完全相同的参数调节方式。需要注意的是，如前 2.3.1.2 节所述，经过了软扩展之后最终得到的 MLP 分块个数为  $M$ 。

按照 2.3.1 节描述的算法流程，在计算欧氏距离和测试距离矩阵时，设置近邻参数  $k=12$ ；之后运行算法 1 产生硬划分的层级 MLP 分块。图 2.6 和 2.7 分别给出了两个数据集所得的层次分裂聚类结果。根据图示中的非线性分值下降曲线，在后续实验中针对两个数据集分别设定参数  $P$  值为 20 和 16，并选取大约 10% 的训练样本作为地标点。

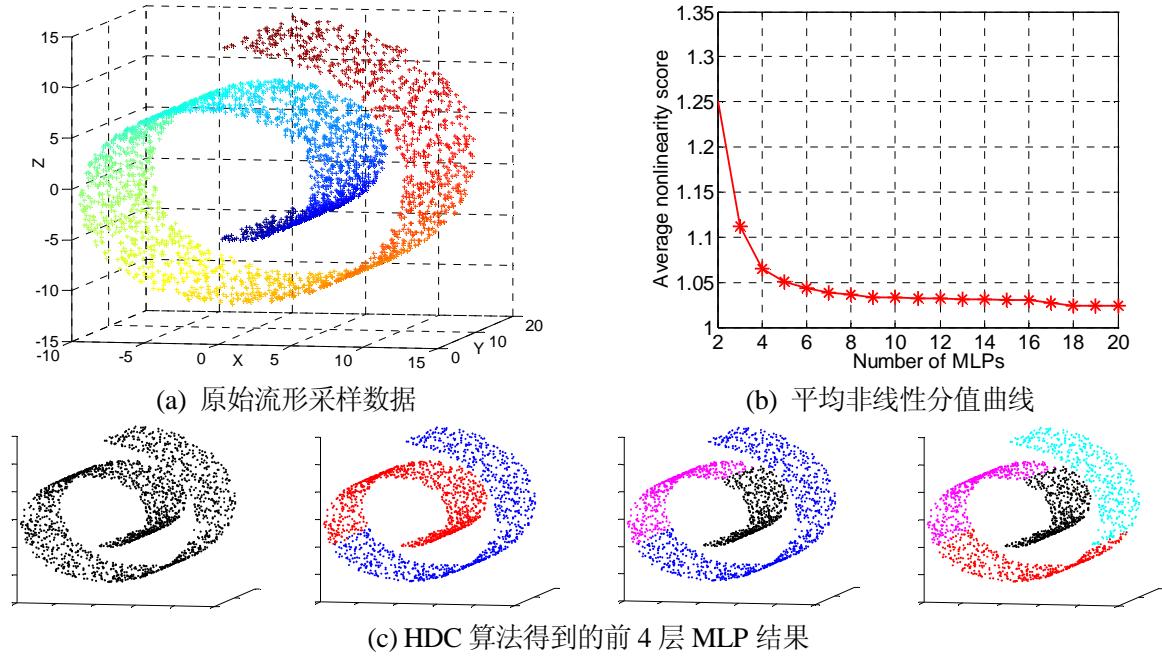


图 2.6 瑞士卷(swiss-roll)流形的 HDC 算法聚类结果

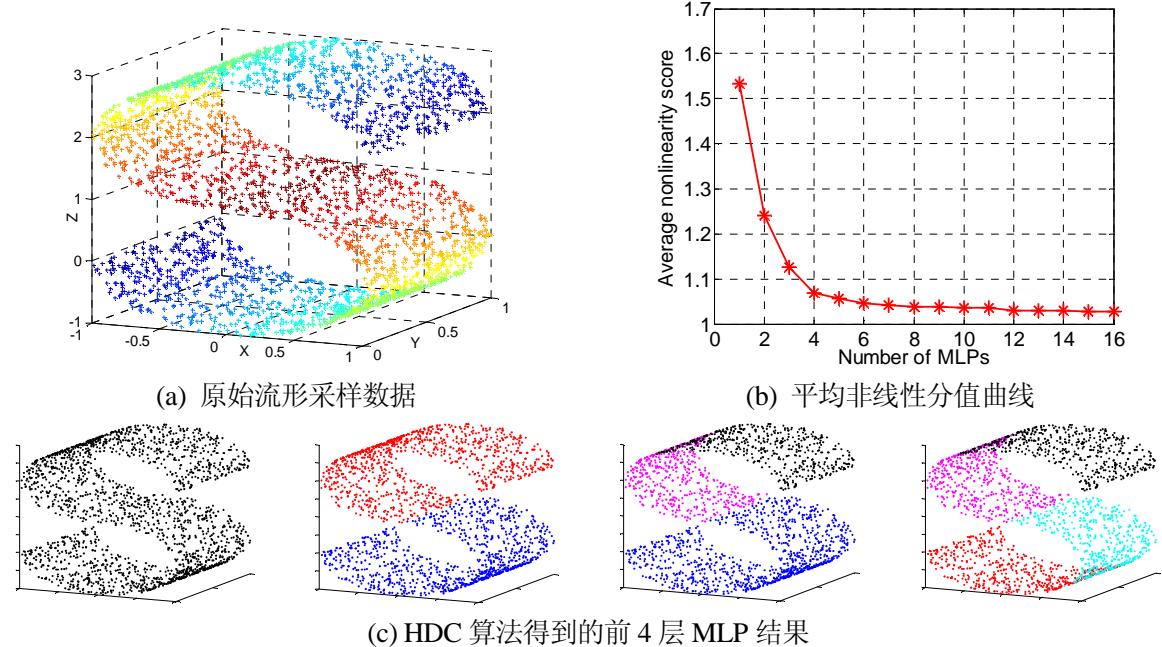
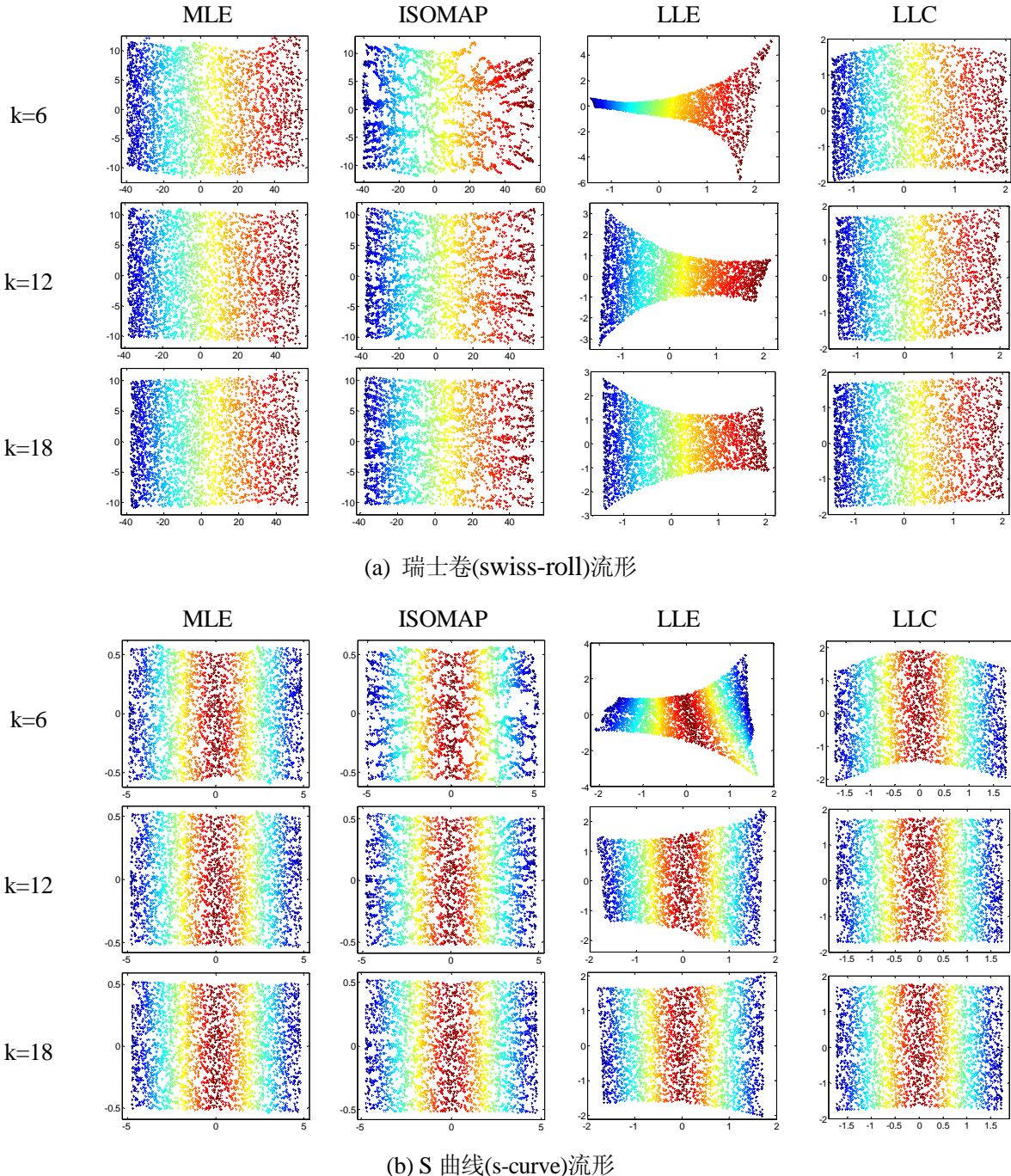


图 2.7 S 曲线(s-curve)流形的 HDC 算法聚类结果

**实验 1：测试参数  $k$  的影响。**针对不同的邻域规模大小  $k$ ，对比如下几个流形学习算法：MLE, ISOMAP, LLE 和 LLC。实验中分别将各算法的  $k$  值依次设定为区间[6, 18]内的各个数值，图 2.8 给出了各自在两个数据库上的降维结果。由于 LLC 算法需要随机初始化其 MFA 局部模型，因此该算法在同样的数据和参数设置条件下，每次运行结果都不相同。为了进行公平的比较，本实验中针对相同的参数设置，多次运行该算法并选

图 2.8 不同算法对邻域参数  $k$  的敏感性对比结果

取最佳的结果与其余算法对比。从图中可以看出, 本文的 MLE 算法与经典算法 ISOMAP 均能很好的保持流形的全局尺度信息并得到更加可靠的低维嵌入表示, 主要原因应该归结于这两个全局性算法学习过程中所具有的距离尺度保持属性。与之对比, 另外两个局部性算法 LLE 和 LLC, 由于其学习过程中施加的单位协方差约束, 则完全丢失了流形的尺度比例信息。观察发现, LLE 的算法性能对参数  $k$  的取值最为敏感, 由于其完全基于局部属性的学习算法, 流形的全局结构信息很难得到较好的保持。与之相比, LLC 算法

的结果则有明显改进，这主要归功于该算法中采用的局部模型表示方式以及对这些局部模型的统一全局对齐。不过，从图示仍然可以看到，LLC 算法的降维结果在  $k$  取值较小时表现出一些变形。接下来，对比全局算法 ISOMAP 和 MLE 的差别。从图 2.8 可以看出，在保持流形的局部结构方面，MLE 比 ISOMAP 的结果更加可靠。虽然实验中的数据集是在原始流形曲面上均匀采样而来，但是 ISOMAP 算法的嵌入空间却出现了一些比较明显的大的孔洞，从而破坏了原始数据的局部结构，这一现象在文献[102]的实验中也有类似报告。MLE 算法能够在同时保持流形的全局和局部信息方面展示出良好的效果，主要得益于其高效的 MLP 局部模型表示和基于 LGA 算法的全局对齐流程。

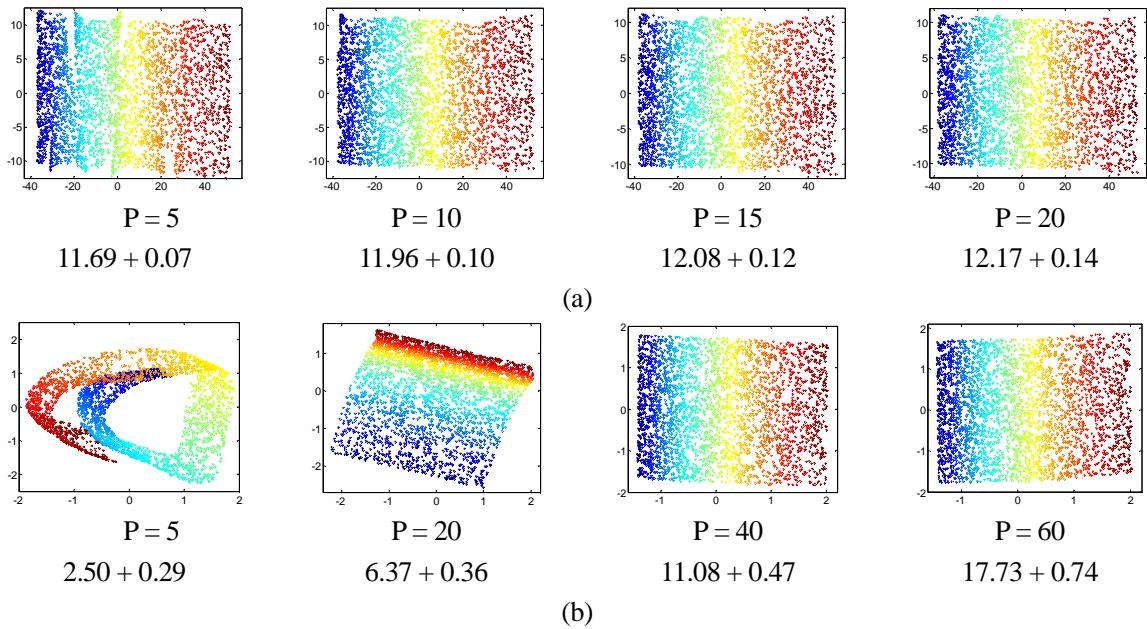


图 2.9 算法 MLE 与 LLC 采用不同数量局部模型的对比结果。(a)为 MLE 算法结果，(b)为 LLC 算法结果，图中的三行依次为算法得到的二维嵌入结果，局部模型的个数，算法的计算时间(单位：秒)。其中，计算时间包括算法运行的两个阶段，即“第一阶段(模型拟合)”+“第二阶段(模型对齐)”。

**实验 2：测试参数  $P$  的影响。**不难看出，局部线性模型的个数  $P$  控制着算法的计算时间与表示精度之间的折衷。换句话说，局部模型(即 MLP 分块)的个数  $P$  越少，则算法的计算时间也越少，但相应地每个 MLP 分块的非线性程度或者偏差也越大；反之亦然。以上述瑞士卷(swiss-roll)流形数据为例。根据图示 2.6(b)中的非线性分值曲线，实验中测试 MLE 算法当  $P$  取值分别为 5, 10, 15, 20 时的结果。为保证与 LLC 算法进行公平的对比，本实验中仅采用层次分裂聚类得到的硬划分 MLP 分块结果进行后续的 LGA 对齐算法运算。相应地，针对 LLC 算法，设定与 MLE 相同的邻域参数  $k=12$ ，测试其在不同的局部模型个数条件下的结果。图 2.9 给出了两个算法各自得到的低维嵌入结果与相应的计算时间。从图中可以看出，随着局部模型的不断增多，MLE 算法的结果越来越趋于稳定。即使采用很少的局部模型，比如当  $P=5$  时，MLE 仍能得到比较合理的流形低维嵌入。与之相比，LLC 算法则表现出对局部模型个数  $P$  较为敏感的特点。只有当其

采用的局部模型 MFA 的个数达到较大规模时, LLC 才能得到相对合理的结果。该实验对比再次验证了本文提出的 MLP 局部线性分块的经济性与高效性。从计算时间的方面分析, 两个算法的时间开销均集中在局部模型拟合这一阶段, 而对齐阶段所花费的时间则极少。从图 2.9 给出的时间统计看出, 当局部模型个数变化相同的情况下, 比如从 5 增长到 20, MLE 算法的时间开销增量远小于 LLC 算法的时间增量。其主要原因在 2.3.4.2 节有所讨论, MLE 最主要的计算开销在于测地距离的计算, 其层次聚类算法 HDC 所花的时间几乎可以忽略; LLC 所花费的时间则随局部模型的增多成比例上升, 而其每个局部模型(factor analyzer)的拟合都是通过相对耗时的 EM 算法迭代优化得到的。

**实验 3: 测试参数  $ni$  的影响。**针对实验 2 中的 swiss-roll 流形, 当参数  $P$  取值 20 时, 经过软划分扩展之后最终得到的 MLP 分块个数为  $M = 57$ 。根据 4.3.1 节所述, 算法从各局部模型中采样的地标点个数需满足  $ni \geq d + 1$ , 这里的 swiss-roll 流形中  $d = 2$ 。为了考察不同地标点数量对 MLE 算法降维结果的影响, 实验中依次增加各局部模型内采样的地标点数目, 用以计算流形的二维嵌入空间。参考 ISOMAP 文献[80]的做法, 本实验中采用嵌入空间相对于原始空间的距离保持残差作为评价指标。该残差的具体计算方式为: 将流形中所有样本点对在原始空间中的测地距离串行构成一个向量, 类似地将这些点对在嵌入空间的欧氏距离也串行构成另一个向量, 计算这两个向量的相关系数就可以得到最终的距离保持残差。实验中, 分别测试了分块内地标点数量从 3 依次增加到 6 的结果, 如图 2.10 所示。观察发现, 算法得到的低维嵌入随着地标点个数增加而趋于稳定, 这主要是由于地标点的数量越多, MDS 算法的输入信息和所附加的约束就越多, 从而得到的嵌入空间也就越稳定。值得指出的是, 即使算法采用理论上的地标点个数的最小值, 即  $d + 1 = 3$  个地标点, 其降维结果仍然可以接受, 低维嵌入中仅包含轻微的形变。当参数  $ni = 5$  时, MLE 得到的二维嵌入残差达到与 ISOMAP 同等的量级  $5 \times 10^{-4}$ , 而此时 MLE 采用的地标点总数为  $57 \times 5 = 285$ , 即只利用了 10% 左右的训练样本, 其求解的特征分解问题规模为  $285 \times 285$ ; 与之相比, ISOMAP 算法则需要求解  $3000 \times 3000$  矩阵的特征分解。

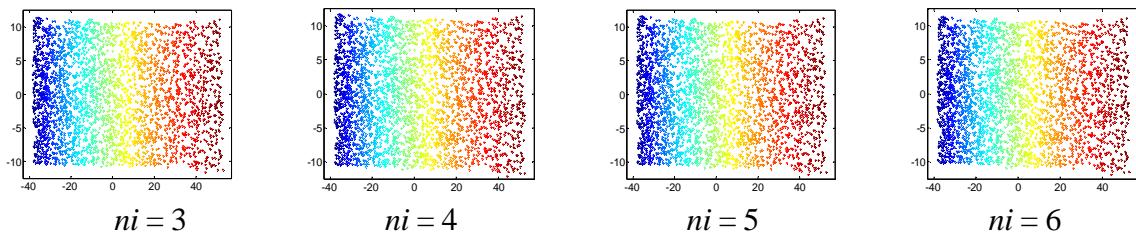


图 2.10 MLE 算法采用不同数量地标点样本的运行结果

接下来, 针对 MLE 算法的几个理论环节, 仍以 swiss-roll 流形为例, 设计相应的实验进行验证。

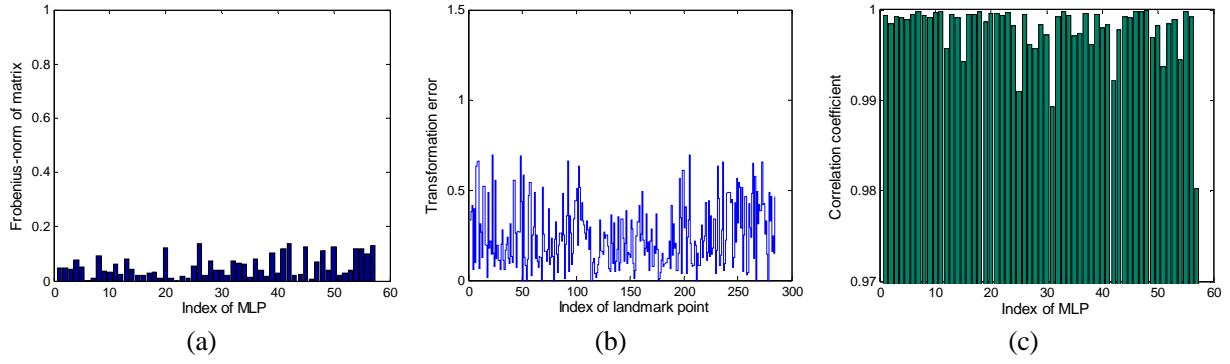


图 2.11 验证 MLE 算法的三个理论环节。(a) 矩阵  $(\mathbf{T}_i)^T \mathbf{T}_i - \mathbf{I}$  的 F-范数, (b) 地标点样本的变换误差, (c) 语义成分向量与流形变化模式方向向量之间的相关系数。请参考 2.4.1 节最后三段的详细描述。

(1) 过渡矩阵  $\mathbf{T}_i$  的正交性验证。针对上面得到的 57 个局部线性模型, 计算矩阵  $(\mathbf{T}_i)^T \mathbf{T}_i - \mathbf{I}$  的 F-范数 (Frobenius-norm), 其中矩阵  $\mathbf{T}_i$  为公式(2-15)中的过渡矩阵,  $\mathbf{I}$  为单位矩阵。矩阵的 F-范数定义为矩阵元素平方和的开方。理论上, 如果  $\mathbf{T}_i$  为正交矩阵, 则  $(\mathbf{T}_i)^T \mathbf{T}_i - \mathbf{I}$  的 F-范数为 0。实际计算结果如图 2.11(a)所示, 可以看出, 绝大部分范数值都与目标值 0 非常接近, 从而验证了本文关于过渡矩阵  $\mathbf{T}_i$  正交属性的分析。

(2) 内在维数  $d$  的有效估计。实验中采用公式(2-23)的目标函数估计 swiss-roll 流形的内在维数, 算法估计得到正确的结果  $d^* = 2$ 。此时, 计算每个地标点样本的变换误差, 即公式(2-23)求和公式的各累加项。图示 2.11(b)给出了相应的结果。考虑到图 2.8(a)中 MLE 低维嵌入空间的坐标尺度, 这里得到的变化误差尺度相对非常小。事实上, 通过计算发现, 所有 285 个地标点的变换误差平均值为 0.247, 与之相比, 嵌入空间中所有样本点与其最近邻之间的距离平均值则为 0.389, 二者接近, 从而表明维数估计值的正确性。

(3) 语义成分(Latent Component, LC)的有效性验证。从图示 2.6(a)可以看出, swiss-roll 流形有两个变化模式, 第一个变化模式是沿着 Xoz 坐标平面内的卷曲缠绕方向, 第二个变化模式是沿着 Y 坐标轴平行的深度方向。虽然不同的局部模型对应的卷曲缠绕方向各不相同, 但它们都对应相同的深度方向, 即沿着观察空间中向量  $[0, 1, 0]^T$  的方向。根据公式(2-24)的分析, MLE 算法得到的第二维语义成分  $\mathbf{q}_2^{(i)}$  ( $i = 1, 2, \dots, 57$ )应该正好对应于上述深度向量  $[0, 1, 0]^T$ 。实验中通过计算二者的线性相关系数来量化评估  $\mathbf{q}_2^{(i)}$  的有效性, 图示 2.11(c)给出了相应的结果。从图示结果看出, 各局部模型对应得到的相关系数都相当接近于理想值 1, 即语义成分与流形本质的变化模式具有高度相关性, 这一结果验证了本文算法可以显式刻画流形内在变化模式的有效性。

## 2.4.2 合成人脸数据库实验

本小节实验通过高维人脸图像集合形成的表观流形来验证算法的性能。选取的数据集是 ISOMAP 文献[80]中的人脸图像库, 实验中称之为 ISOFace 数据集。该集合包含了

人工合成的同一个人的 698 张具有不同姿态和光照变化的人脸，其图像大小为  $64 \times 64 = 4096$  像素。这些人脸图像本质上采样于一个三维的人脸表观流形，该流形通过三个参数来刻画，即左右姿态、上下姿态和光照方向[80]。该数据库在流形学习算法的相关研究文献中得到了广泛应用，如[86][102][103]。实验中，将数据集中的 698 个样本分为训练集和测试集两部分，其中训练集包括前 650 个样本，测试集则包括剩余的 48 个样本。需要注意的是，原始的 698 个样本是无序排列的，因此，实验划分的训练集和测试集都广泛覆盖到了数据库中的姿态和光照变化范围。

(1) 定性评价。实验采用算法 ISOMAP 的结果作为基准，用来评价本文 MLE 算法的结果，包括其针对高维流形数据的降维性能与其参数映射功能。依据文献[80]的参数和数据设置方案，实验中将邻域参数设为  $k=6$ ，利用全部 698 个样本进行 ISOMAP 算法学习，从而得到文献[80]报告的基准结果。与之对比，MLE 算法采用上述 650 个样本的训练集进行流形学习。MLE 算法得到  $M=27$  个 MLP 分块，设定参数  $ni=7$ ，从而采用  $27 \times 7=189$  个地标点样本（即大约 30% 的训练数据）进行嵌入空间的学习。两个算法最终都得到了正确的三维人脸流形嵌入空间，其前两维可视化结果如图 2.12 所示。可以看出，与基准算法 ISOMAP 相比，尽管采用了较少的训练数据，本文的 MLE 算法也能成功地展开高维观察数据的低维流形，并得到与基准结果相接近的保持了全局流形结构的低维嵌入空间。

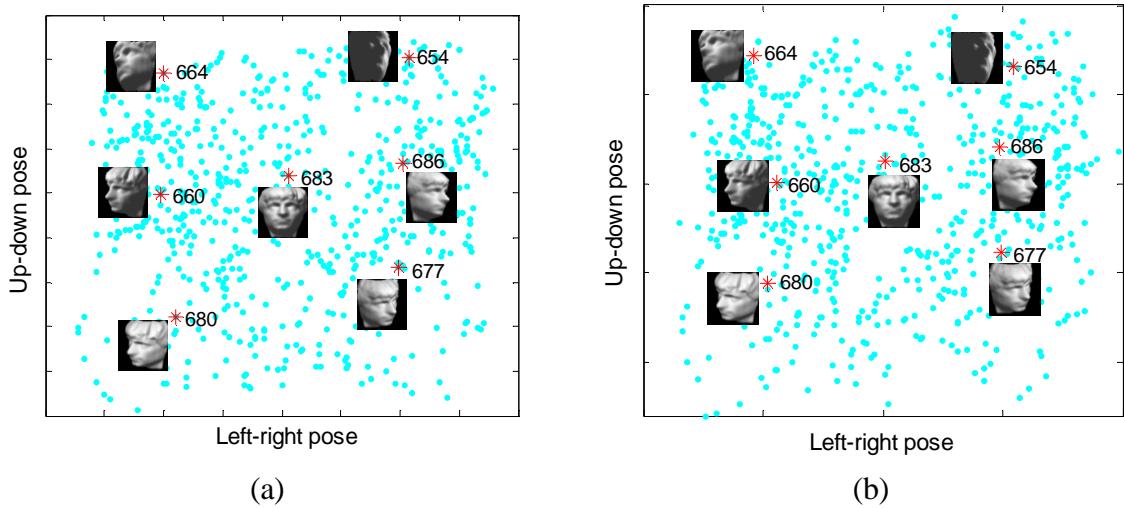


图 2.12 ISOFace 人脸数据库上的低维嵌入结果。(a) (b) 分别为 ISOMAP 和 MLE 算法的结果。图中红色“\*”为测试样本以及其对应的标号与图像。需要注意的是，(a)图所示 ISOMAP 算法结果，测试图像与训练图像通过 ISOMAP 的训练算法同时降维的。

MLE 算法利用训练数据完成对流形的低维嵌入空间学习之后，就可以同时学习得到显式的参数式映射函数，用来解决新来测试样本的 out-of-sample 问题。首先验证算法的前向映射功能，即表 2.2 的算法 2，从高维观察空间到低维嵌入空间的投影。实验数据采用上面划分得到的测试集，即序号为 651 到 698 的样本。利用前向映射将这些样本降维，



图 2.13 MLE 算法针对 ISOFace 数据库得到的重构人脸。图中每一行对应于算法沿着三维嵌入空间的每一维坐标轴重构生成的样本序列，从上到下分别为：左右姿态、上下姿态、光照方向的变化模式。



图 2.14 MLE 与 LLC 算法采用不同维数进行重构的结果对比

得到它们在嵌入空间的坐标表示。图示 2.12 给出了一些代表样本的结果。为便于对比，将这些样本对应的 ISOMAP 训练结果也在图中标注出来。可以看出，MLE 利用其参数投影成功地完成了这些测试样本的降维，找到了它们的内在属性，如左右姿态，上下姿态变化等；而在 ISOMAP 的结果中，这些测试样本是与训练样本同时利用训练算法降维的。接下来验证算法的后向映射功能，即表 2.3 的算法 3，从低维嵌入空间到高维观察空间的投影。实验中沿着三维嵌入空间的各坐标轴分别重构一系列虚拟样本，如图 2.13 所示。与预期一致，这些生成的虚拟样本相应的变化模式正好完全对应于各自坐标轴刻画的内在变化模式，即左右姿态、上下姿态和光照方向等。此外，需要指出的是，一些重构虚拟样本看起来没有原始样本的图像质量好，这主要有两方面原因：一方面是由于训练集的规模较小，部分区域比较稀疏；另一方面是由于这些虚拟样本仅仅利用了本质维数  $d=3$  的 3 维成分来重构。

(2) 定量对比。本实验进一步量化评估 MLE 算法在推广能力方面的表现，采用的对比算法为参数式算法 LLC。参考文献[86]中的实验设计，本实验采用重构误差来评价算法的重构性能。与上述 MLE 学习过程类似，LLC 算法也利用划分的 650 个训练样本

来学习参数映射函数，并且采用与 MLE 类似的参数设置，其中近邻参数设为  $k=6$ ，局部模型 MFA 的个数为  $P=30$ 。重构实验设计如下：分别利用两个算法的映射函数来对测试集的 48 个样本进行先降维后重构的过程。具体地，针对每个测试样本  $\mathbf{x}_n$ ，首先利用前向映射将其投影到嵌入空间得到其低维表示  $\mathbf{y}_n$ ，然后通过后向映射将  $\mathbf{y}_n$  投影回原始的高维观察空间得到重构样本  $\hat{\mathbf{x}}_n$ 。类似文献[86]的设置，利用下面的公式来度量样本重构误差

$$E_n = \frac{1}{\sqrt{D}} \|\mathbf{x}_n - \hat{\mathbf{x}}_n\| . \quad (2-26)$$

其中， $D$  为原始图像空间的维数，本实验中  $D = 4096$ 。直观分析，该重构误差可以理解为度量图像中单个像素的平均重构偏差值。本文实验中，所有图像的像素值均量化到[0, 255]的区间内。

由于 LLC 的训练算法要求先验指定流形的低维维数  $d$ ，实验中针对该算法测试了参数  $d$  取值为[3, 20]区间内不同数值的结果。最终利用 MLE 和 LLC 两个算法计算得到的重构误差如表 2.4 所示，表格中给出的各统计值为综合了 48 个测试样本重构误差的均值与标准差所得。其中，“MLE\_3”表示 MLE 算法利用参数  $d = 3$  训练，“LLC\_3/10/20”则表示 LLC 算法分别利用参数  $d = 3, 10, 20$  训练。图 2.14 给出了一些示例的原始样本与相应的重构图像。从这些结果可以看出，LLC 的重构误差随着维数  $d$  的增加而不断减小，当维数上升到  $d = 20$  时，才能取得与 MLE 可比的性能，而此时 MLE 算法采用的维数为流形的内在维数  $d = 3$ ；相比而言，LLC 在参数取值为  $d = 3$  时，则难以得到令人满意的结果，其重构误差非常高，这一点从图示 2.14 重构样本的质量也可以看出。相比于 LLC，MLE 算法的成功主要归结于其基于 MLP 局部线性分块的 PCA 建模算法和基于公式(2-23)的代价可控的 LGA 对齐算法。此外，需要说明的是，针对参数  $d$  的各个不同取值，LLC 算法需要求解不同规模的特征分解问题，因此需要反复多次的运行。从表 2.4 的计算时间对比可以发现，当处理的观察数据维数较高时（如本实验中的人脸图像），LLC 算法的局部模型拟合阶段非常耗时，这主要是由于基于 EM 算法的 MFA 模型估计的参数太多造成的。

表 2.4 ISOFace 数据库的重构误差与计算时间（单位：秒）对比

方法	MLE_3	LLC_3	LLC_10	LLC_20
误差	$18.9 \pm 7.3$	$47.7 \pm 20.1$	$35.1 \pm 18.0$	$22.2 \pm 6.8$
时间	11.5	27.6	35.0	49.5

### 2.4.3 真实人脸数据库实验

本小节实验采用另一个著名的人脸数据库验证 MLE 算法的功能。该数据库来自于 LLE 算法文献[66]，实验中称之为 LLEFace 数据库。与上述 ISOFace 库不同，LLEFace 数据库的样本来自于一段 1965 帧的真实人脸视频，视频格式为  $20 \times 28$  的灰度图像。该视频中的图像涵盖了一个人在摄像机镜头前的各种姿态与表情变化，并伴随一定程度的镜头抖动。该数据库同样被广泛应用于流形学习算法的研究和评估中，如[67][79][86]。

(1) 定性评价。实验首先采用全部 1965 个样本进行 MLE 算法学习。为了与 LLC 算法进行对比，MLE 的算法参数参考了 LLC 文献[79]中的相应设置。具体地，选用邻域参数  $k=36$ ，层级聚类算法得到  $P=10$  个硬划分 MLP 分块，经过软扩展之后得到的分块个数为  $M=26$ 。由于该数据库中的人脸图像来自于真实采集的视频，流形的本质维数无法精确获得，类似于文献[79]，在 MLE 算法中最终选用维数  $d=8$  进行学习。对齐过程中设定参数  $ni=15$ ，从而最终采用  $26 \times 15=390$  个地标点（约占训练样本的 20%）将人脸样本从原始的 560 维图像空间投影到 8 维的嵌入空间。MLE 降维结果的前两维可视化结果如图 2.15 所示，图中同时给出了利用算法的映射函数沿低维坐标轴重构生成的一系列虚拟样本。与前人工作[67][79][86]等在该数据库的结果类似，MLE 算法正确地展开了该人脸流形并发现其前两维最主要的变化模式，分别为姿态变化和表情变化。此外，少数几个重构生成的虚拟样本出现了明显的质量下降，主要原因仍然是由于训练数据的不均匀和稀疏性所造成。

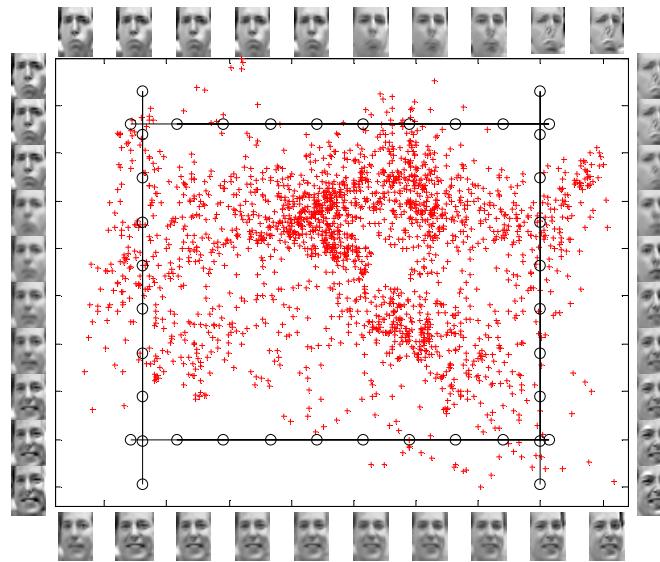


图 2.15 MLE 算法学习得到的 LLEFace 人脸流形低维结构。图中红色“+”对应于训练样本的低维坐标，矩形框四条边环绕的图像为 MLE 算法针对低维嵌入图中黑色“o”样本进行重构得到的虚拟样本。

(2) 关于语义成分(Latent Component)的进一步讨论。如前 2.3.3.4 节所述，图示 2.13 和 2.15 中的虚拟人脸事实上是通过公式(2-25)沿着语义成分的方向重构的。类似于人脸

识别领域中将 PCA 的主成分称为“特征脸”(Eigenface)，这里将 MLE 算法的语义成分称为“语义脸”(Latentface)。图示 2.16 中给出了 MLE 算法在 LLEFace 数据库上计算得到的其中一个局部线性模型对应的特征脸与语义脸实例对比。本质上，特征脸反映了高维数据空间中数据变化的最大方差对应的方向，语义脸则刻画高维数据所对应流形的内在语义(latent)变化模式的方向。因此，沿着 Latentfaces 的方向进行样本重构时，所得到的一系列虚拟样本正好展示出原始流形在这些语义成分方向上对应的变化模式。由于 Latentfaces 所具有的这些特性，可以期待在未来的工作中将其应用于解决人脸姿态估计、表情分析以及人脸识别等多种问题中。



图 2.16 特征脸(Eigenfaces)与语义脸(Latentfaces)对比示意图。对比结果来自于同一个局部模型，实验采用的降维维数为 8。

(3) 定量对比。与上述 ISOFace 数据库上进行的量化实验类似，本实验进一步对比 MLE 与 LLC 算法，重点考察算法的重构性能受训练数据规模和局部模型个数的影响。实验中将原始样本集的 1965 个样本随机划分为两组，分别作为训练数据和测试数据，其中训练数据的比例以 10% 的步长依次从 60% 增加到 90%。重构流程仍然为：利用训练数据学习映射函数，将测试样本依次正向投影到低维嵌入空间，再反向投影回观察空间。本实验的重构误差统计基于所有测试样本  $\mathbf{x}_n$  计算其误差均值，具体如下：

$$E_{rec} = \frac{1}{N\sqrt{D}} \sum_{n=1}^N \|\mathbf{x}_n - \hat{\mathbf{x}}_n\| . \quad (2-27)$$

与公式(2-26)相同，这里的  $D = 560$  为图像空间的维数， $N$  为测试样本的数量，针对不同的训练数据比例，其取值分别为 785, 589, 393, 196。

具体的算法设置为：MLE 与 LLC 均设置邻域参数  $k=36$ ，降维维数  $d=8$ ，局部模型的个数  $P$  则在 10 到 25 的范围内依次取值。同样为了公平对比，MLE 算法仅使用硬划分得到的 MLP 分块进行后续的对齐流程。表 2.5 所示列出了两个算法针对不同比例的训练数据和不同数量的局部模型得到的重构误差统计，表中各数值对应的均值和标准差来自于对各比例训练数据和测试数据的 5 组随机划分所得到的结果统计。从实验结果可以看出，MLE 和 LLC 算法的重构误差均随着训练数据的增加而逐渐下降，这一现象完全符合预期，逐渐增加的训练数据必然为学习算法带来更多的信息，从而使得学习所得的模型更加精确。同时，MLE 的误差也一致地小于 LLC 的误差，类似于 ISOFace 库的对比

结果, 这同样归因于 MLE 算法所采用的局部模型和全局对齐算法。进一步观察局部模型个数  $P$  对各算法性能的影响, 可以看出, MLE 的误差随着模型个数增加而不断下降, 这主要是由于不断增加的 MLP 分块所带来的更加精确的表示能力。与之相比, 当模型个数较多时, LLC 算法表现出一定的过拟合现象, 其原因类似于文献[86]中的分析, 较多的局部模型带来了更多的 MFA 参数估计, 从而容易造成模型的过拟合。

表 2.5 LLEFace 数据库的重构误差对比

$P$	10	15	20	25
60%	12.09 $\pm$ 0.29	11.53 $\pm$ 0.31	11.17 $\pm$ 0.17	10.83 $\pm$ 0.14
	13.15 $\pm$ 0.28	12.93 $\pm$ 0.06	13.20 $\pm$ 0.36	13.32 $\pm$ 0.44
70%	11.91 $\pm$ 0.28	11.48 $\pm$ 0.28	11.06 $\pm$ 0.29	10.77 $\pm$ 0.28
	12.99 $\pm$ 0.05	12.82 $\pm$ 0.40	12.83 $\pm$ 0.14	13.22 $\pm$ 0.12
80%	11.82 $\pm$ 0.17	11.31 $\pm$ 0.26	10.94 $\pm$ 0.32	10.78 $\pm$ 0.27
	12.87 $\pm$ 0.27	12.79 $\pm$ 0.46	12.74 $\pm$ 0.42	13.18 $\pm$ 0.05
90%	11.76 $\pm$ 0.11	11.12 $\pm$ 0.03	10.80 $\pm$ 0.18	10.54 $\pm$ 0.17
	12.76 $\pm$ 0.54	12.78 $\pm$ 0.28	13.13 $\pm$ 0.37	13.89 $\pm$ 0.77

针对每种训练数据比例, 第一行结果为 MLE 的误差, 第二行为 LLC 的误差

## 2.5 本章小结

针对传统流形学习方法的不足, 本章提出一种新颖的基于最大线性嵌入(MLE)的学习算法。与经典算法 ISOMAP 和 LLE 相比, MLE 可以同时较好的保持流形的局部几何和全局结构, 并学习得到从观察空间到流形嵌入空间的显式参数映射。与 LLE 算法采用的基于  $k$ -NN 或者  $\epsilon$ -ball 的局部线性邻域不同, MLE 算法定义了一种更为高效合理的最大线性分块(MLP)作为线性嵌入框架的基础。由于 MLP 的定义与构造算法采用了测地距离的思想, 因此 MLE 算法同时融入了 ISOMAP 的核心理念。与其它参数式算法如 LLC 相比, MLE 在局部模型拟合和全局对齐两个阶段都有所改进, 具体如 2.3.4.1 节所述。第 2.4 小节的实验结果验证了 MLE 相对于 LLC 的优势, 包括采用较少的局部模型就可以得到更可靠的结果, 在类似的参数设置条件下取得更小的重构误差等。

关于算法可能的扩展思路, 针对局部模型拟合环节, 可以参考 LLC 算法[79]和 CFA[86]算法中的做法, 在 MLP 分块的 PCA 建模过程中引入概率模型, 从而将算法的映射过程以概率过程来刻画, 以期带来更加稳定与灵活的降维投影函数。针对局部模型

的全局对齐算法，目前采用的 LGA 算法采用了类似于 ISOMAP 的严格的保距投影约束，这一约束在保证降维结果尽可能保持原始数据结构的同时，也在一定程度上限制了算法的应用范围。接下来的工作可以参考文献[56]所采用的算法约束，在严格的保距约束与完全丢失全局尺度信息之间做一定的平衡。此外，后续工作中需要着重考虑流形学习算法普遍面临的两个问题，即噪声数据和采样密度对算法结果稳定性的影响。本文认为，这两个问题也是困扰当前流形学习算法走向更加实用化的核心障碍。

## 第三章 基于流形-流形距离的图像集合人脸识别

### 3.1 引言

在传统的人脸识别算法中，无论是分类器的训练还是测试均是基于每个人的单幅图像进行的。随着视频存储设备的发展，在很多新的应用场景如视频监控、视频检索等任务中，可以获取到人脸的多幅图像数据，从而为分类器的训练和测试阶段提供大量的样本。利用每个人的这些多幅图像，可以分别构建其 *gallery* 图像集合和 *probe* 图像集合，从而设计基于图像集合的人脸识别算法。

针对上述具体的应用问题，以上一章介绍的最大线性嵌入算法为理论基础，本章提出一种有效的解决方法，称为流形-流形距离(Manifold to Manifold Distance, MMD)。在 MMD 方法中，图像集合被建模为非线性的表观流形，从而将图像集合的匹配分类问题形式化为流形之间的距离计算问题。同样基于流形具有局部线性属性的特点，MMD 方法将全局非线性的流形表示为一组局部线性模型的组合，其中各局部模型采用子空间来刻画。这样，两个流形之间的距离就可以转化为它们各自对应的子空间对之间的距离的综合。与前人工作相比，本文提出的 MMD 方法主要有如下几点贡献：

(1) 本文首次将基于图像集合的物体识别问题形式化为流形-流形距离的计算问题，即计算训练所采用的 *gallery* 图像集对应的 *gallery* 流形和测试所采用的 *probe* 图像集对应的 *probe* 流形之间的距离。

(2) 注意到实际问题中的图像集合可能表示为三种模式层次，即点、子空间、流形，本文系统研究了这三种模式之间的各种距离，并将它们形式化在一个统一的多层次流形-流形距离框架下。

(3) 针对流形-流形距离的计算，本文从理论上和实验上探讨了其各组成模块的多种定义方案，包括局部线性模型的建立方法、局部模型之间的距离度量、局部模型距离的全局综合方法。

(4) 本文 MMD 方法在基于图像集合的人脸识别(face recognition with image set, FRIS)任务上进行了系统的实验验证。实验结果表明，流形-流形距离作为一种通用的图像集合相似性度量，一致地优于所对比的其它非判别式算法，并且达到了与当前最好的基于集合的判别式学习算法相当的性能。

本章接下来的安排如下：第 2 节系统调研点、子空间、流形三者之间的各种距离，并将它们形式化在一个统一的多层次流形-流形距离框架下；第 3 节详细介绍流形-流形距离的三个组成模块，深入探讨了各模块的多种可能定义方案；第 4 节重点讨论了算法的两个理论计算问题，并与相关的算法进行了对比讨论；第 5 节给出了算法在三个公开人脸数据库上与其它方法的对比结果；第 6 节对算法进行了总结讨论。

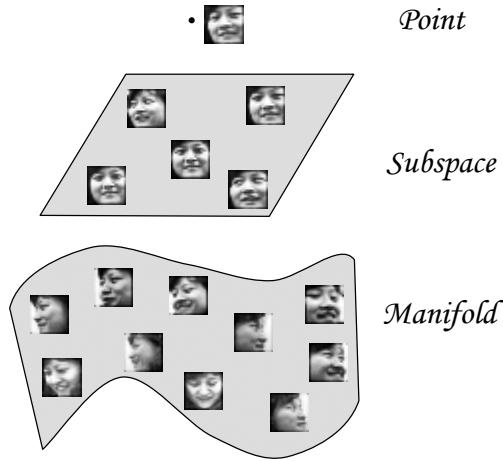


图 3.1 物体识别任务的三种模式层次（点、子空间、流形）所构成的层级结构

## 3.2 问题形式化框架

在基于图像集合的物体识别任务中，各图像集合一般包含属于同一物体类别的一定规模的图像样本，这些图像通常涵盖了物体广泛的非线性表观变化模式。然而，在一些特殊情况下，集合中可能仅包含具有轻微变化模式的少量样本，在极端情况下甚至只有孤立的一个样本。考虑到识别任务中的这些不同情况，可以从三个不同的模式层次对图像集合进行建模。这三个模式层次分别为：点（孤立的单样本）、子空间（少量样本张成的线性模型）、流形（大量样本在图像空间张成的低维非线性嵌入）。图 3.1 以人脸图像集合为例给出了三个模式层次的示意。从某种意义上来看，模式分类问题本质上可以看作上述三种模式之间的距离计算问题。目前，基于点和子空间的各种距离已经有了较多相对成熟的研究成果，而与流形距离相关的研究工作则比较少。

本节首先回顾基于点和子空间定义的各种距离形式，之后针对流形-流形距离，给出一个基础性的形式化定义。根据这一定义，上述三个模式层次之间定义的各种距离都可以在一个广义的多层次流形-流形距离框架下进行统一的形式化表述。

### 3.2.1 定义在点和子空间上的距离

如图 3.2 所示，与点和子空间相关的距离包括：点-点距离(point to point distance, PPD)，点-子空间距离(point to subspace distance, PSD)，子空间-子空间距离(subspace to subspace distance, SSD)。在后续章节中，将点记为  $x_i$ ， $y_i$ ，子空间记为  $S_i$ ，流形记为  $\mathcal{M}_i$ 。

**点-点距离：**两个点  $x_1$  到  $x_2$  的距离记为  $d(x_1, x_2)$ ，最常用的点-点距离即为传统的欧式距离：

$$d(x_1, x_2) = \|x_1 - x_2\| . \quad (3-1)$$

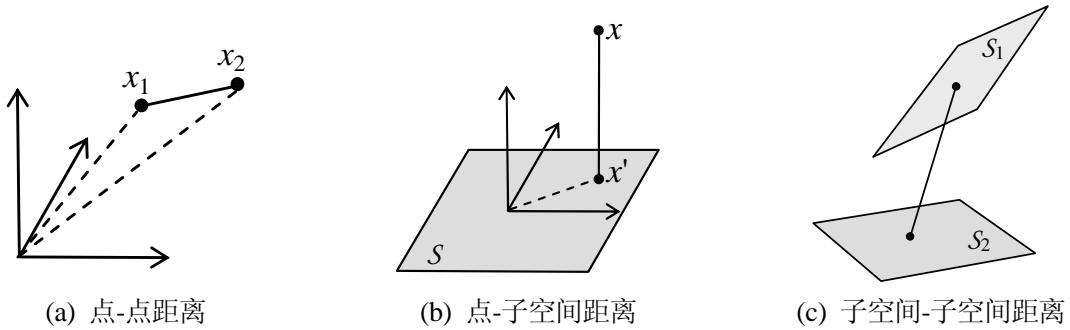


图 3.2 定义在点和子空间上的三种距离

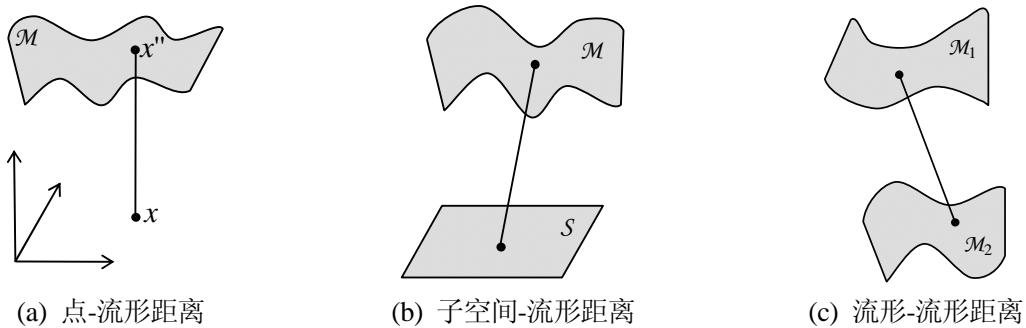


图 3.3 定义在流形上的三种距离

**点-子空间距离:** 点  $x$  到子空间  $S$  的距离记为  $d(x, S)$ ，通常采用所谓的 L2-Hausdorff 距离来计算：

$$d(x, S) = \min_{y \in S} \|x - y\| = \|x - x'\| . \quad (3-2)$$

事实上，点  $x'$  是点  $x$  在子空间  $S$  中的投影，即  $x'$  是子空间  $S$  中距离  $x$  最近的点。因此，点-子空间距离本质上是点  $x$  到其投影点  $x'$  的点-点距离。这一定义方式也称为“特征空间外的距离”(distance-from-feature-space, DFFS)[61][82]。

**子空间-子空间距离:** 两个子空间  $S_1$  与  $S_2$  之间的距离记为  $d(S_1, S_2)$ 。目前，关于子空间之间的距离尚未形成统一的定义方式。其中，主夹角方法(principal angles)以其在理论上具有的多种优点和实际应用中的良好性能，获得了最为广泛的应用[29][43][95][96][98]。此外，文献[91]从坐标空间基向量表示的角度提出了另外一种子空间距离定义，这种定义本质上可以看作是两个子空间各自的基向量组之间的 DFFS 距离之和。

根据线性代数的理论可知，孤立点  $x_i$  本质上可以看作是一个特殊的线性子空间，即所谓的平凡的零子空间(zero subspace)  $L\{\mathbf{0}\}$ ，该子空间的原点即为点  $x_i$ ，其维数为 0，基向量为空集[33]。从这个意义上来说，点-点距离和点-子空间距离都可以看作子空间-子空间距离的特例。举例来说，公式(3-1)中的点-点距离本质上是采用子空间的中心来定义了两个零子空间之间的距离。

### 3.2.2 定义在流形上的距离

如上一章所述，流形学习已经成为计算机视觉、模式识别和机器学习等领域的热点研究问题。一般来说，给定高维观察空间中采样得到的一组样本，如果它们的内在变化模式仅受少数几种因素影响，这组样本的内在结构就可以采用流形来建模表示[66][80]。目前，有关学习高维观察数据的低维流形嵌入表示的研究工作已经比较普及；然而，针对两个非线性流形之间的距离度量问题则尚未引起研究者关注。本节接下来就将针对点、子空间和流形之间的各种距离定义方式进行初步的探讨，具体包括如下三种距离：点-流形距离(point to manifold distance, PMD)、子空间-流形距离(subspace to manifold distance, SMD)、流形-流形距离(manifold to manifold distance, MMD)，如图 3.3 所示。

为了计算这几种距离，首要的问题是建立流形与子空间（点可以看作特殊的子空间）二者之间的表示关系。自然地，可以利用非线性流形所具有的局部线性属性的特点，即一个全局的非线性流形在其每一点的局部邻域内都与一个特定维数的欧氏线性子空间拓扑同胚[37]。根据这一基本思想，可以将流形表示为一组局部线性模型的组合，其中各局部模型采用子空间来刻画[66][67]。总的来说，流形可以看作是将子空间进行了扩展，以刻画反映更为广泛和复杂的数据变化模式。经过上述分析，与流形相关的距离计算就转化成了与子空间相关的距离计算。形式上，本文将流形  $\mathcal{M}$  定义为子空间的集合，集合的元素即为构成该流形的各成分子空间(component subspace)  $C_i$ ，具体如下式：

$$\mathcal{M} = \{C_i : i = 1, 2, \dots, m\} = \{C_1, C_2, \dots, C_m\} \quad (3-3)$$

其中， $m$  为流形中的局部线性子空间的个数。

**点-流形距离：**点  $x$  到流形  $\mathcal{M}$  的距离记为  $d(x, \mathcal{M})$ 。与“点-子空间距离”类似，可以将该距离定义为点  $x$  到流形  $\mathcal{M}$  上最近点的距离，即

$$d(x, \mathcal{M}) = \min_{C_i \in \mathcal{M}} d(x, C_i) = \min_{C_i \in \mathcal{M}} \min_{y \in C_i} \|x - y\| = \|x - x''\| \quad (3-4)$$

类似于“点-子空间距离”中的投影点  $x'$ ，这里将上式中的  $x''$  称为点  $x$  在流形  $\mathcal{M}$  上的投影。

**子空间-流形距离：**子空间  $S$  到流形  $\mathcal{M}$  的距离记为  $d(S, \mathcal{M})$ 。该距离可以定义为流形  $\mathcal{M}$  上距离子空间  $S$  最近的子空间与  $S$  之间的距离：

$$d(S, \mathcal{M}) = \min_{C_i \in \mathcal{M}} d(S, C_i) \quad (3-5)$$

可以看出，类似于将“点-子空间距离”化简为“点-点距离”，这里的“子空间-流形距离”最终化简成了“子空间-子空间距离”。

**流形-流形距离：**两个流形  $\mathcal{M}_1$  和  $\mathcal{M}_2$  之间的距离记为  $d(\mathcal{M}_1, \mathcal{M}_2)$ 。根据公式(3-3)中流形的表示方式，“流形-流形距离”可以转化为各流形对应的子空间对之间距离的综合。图 3.4 所示给出了形象的示意。

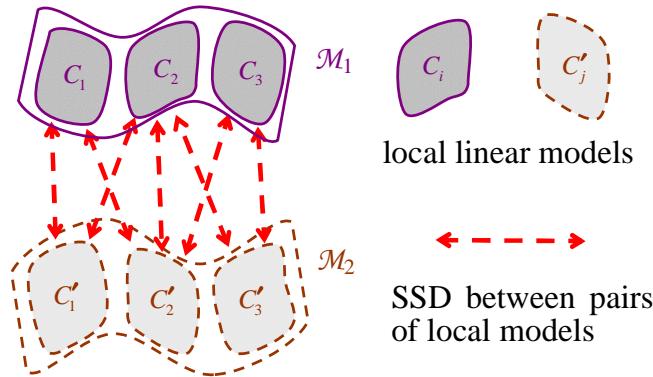


图 3.4 流形-流形距离(MMD)的计算示意图。两个流形  $\mathcal{M}_1$  与  $\mathcal{M}_2$  之间的距离通过综合其各自对应的局部模型对  $C_i$  与  $C'_j$  之间的距离得到。

给定两个流形的表示如下:  $\mathcal{M}_1 = \{C_i : i = 1, 2, \dots, m\}$ ,  $\mathcal{M}_2 = \{C'_j : j = 1, 2, \dots, n\}$ , 它们之间的距离可以形式化地表示为如下公式:

$$d(\mathcal{M}_1, \mathcal{M}_2) = \sum_{i=1}^m \sum_{j=1}^n f_{ij} d(C_i, C'_j), \text{ 其中} \\ \sum_{i=1}^m \sum_{j=1}^n f_{ij} = 1, \quad f_{ij} \geq 0. \quad (3-6)$$

在公式(3-6)的广义形式化框架中, “流形-流形距离”被定义为成对子空间之间距离即  $d(C_i, C'_j)$  的加权平均。

现在来回顾一下图 3.2 和 3.3 所示的定义在三种模式层次上的六种距离。前述讨论得到, 点可以看作子空间的特例。再进一步, 子空间同样可以看作流形的特例。根据公式(3-3)的定义, 子空间本质上是一种线性流形, 其自身就构成了该流形唯一的一个成分子空间。因此, 点、子空间和流形这三种模式形成了一种层级的结构化表示, 正如图 3.1 所示。在这一结构化关系下, 点和子空间都看作是流形的特例, 从而上述六种距离都可以形式化在广义的多层次流形-流形距离框架下。

### 3.3 流形-流形距离

根据图示 3.4 和公式(3-6)的定义, 可以看出流形-流形距离的计算具有三个核心模块:

(1) 局部线性模型的建立, 即计算成分子空间  $C_i, C'_j$ ; (2) 局部模型之间的距离度量, 即计算子空间-子空间距离  $d(C_i, C'_j)$ ; (3) 局部模型距离的全局综合, 即计算权重  $f_{ij}$ 。本节将详细讨论各组成模块的多种可能的定义方案。

#### 3.3.1 局部线性模型构建

针对非线性流形上局部线性模型的提取, 前人工作中一般采用经典的聚类算法, 如 K-均值(K-means)聚类[27][42][51]或者层次融合聚类(hierarchical agglomerative clustering,

HAC) [20][21][104]。这些方法主要有两个限制：第一，聚类的目标类别也就是局部模型的个数需要事先指定；第二，提取出的局部模型没有显式的线性属性约束。与这些经典聚类方法不同，本文上一章提出的层次分裂聚类算法 HDC 可以很好地克服上述不足，因此，在 MMD 方法中同样采用层次分裂聚类的方式自适应地建立具有显式线性程度约束的多层次局部模型，算法的基本思想请参考本文 2.3.1.1 小节。

形式上，给定流形  $\mathcal{M}$  与其采样得到的数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ，其中  $\mathbf{x}_i \in \mathbb{R}^D$  为  $D$ -维列向量， $N$  为样本个数。算法的目标是将原始数据集  $\mathbf{X}$  划分为一组互不相交的最大线性分块(MLP)  $\mathbf{X}^{(i)}$  的组合，即

$$\mathbf{X} = \bigcup_{i=1}^m \mathbf{X}^{(i)}, \text{ 其中 } \mathbf{X}^{(i)} \mid_{i=1}^m = \{\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_{N_i}^{(i)}\}, (\sum_{i=1}^m N_i = N). \quad (3-7)$$

之后，采用局部模型  $C_i$  来建模各 MLP 分块  $\mathbf{X}^{(i)}$ ，就可以得到公式(3-3)的流形表示。

可以看出，公式(3-7)与公式(2-1)本质是相同的，其中分块  $\mathbf{X}^{(i)}$  的构建即通过 2.3.1.1 小节介绍的 HDC 算法来完成。算法最终得到  $m$  个互不相交的 MLP 分块。在 3.5.4 节将通过具体实验对比本文采用的分裂式 HDC 算法与经典的融合式 HAC 算法和 K-均值聚类算法的性能和计算效率。

图示 3.5 给出了 HDC 算法在真实人脸数据集合上提取所得的局部模型。可以看出，这些局部模型中的样本仅呈现出轻微的表观变化模式（从而可以近似看作线性变化），并且都与模型的中心（即样本均值）具有相似的表观。本文将这些样本均值称为样例 (exemplar)，它们可以在一定程度上表征局部模型整体的表观模式。



图 3.5 人脸数据库对应的局部模型示意。图中每一行对应一个局部模型，其中第一列为该局部模型的样本均值即样例点(exemplar)，后六列为模型内部的一些代表性样本。前两行属于同一个人，后两行属于另外一个人。

### 3.3.2 局部模型距离度量

利用线性子空间来表示上述建立的局部模型，就可以采用“子空间-子空间距离”来度量这些局部模型之间的相似度。由于主成分分析(principal component analysis, PCA)方法具有简单高效等优点，与 MLE 方法相同，在 MMD 中也采用 PCA 来建模这些局部线性模型。针对每个局部模型  $C_i$ ，其样本均值记为  $\mathbf{e}_i$ ，主成分矩阵记为  $\mathbf{P}_i \in \mathbb{R}^{D \times d_i}$ ，矩阵的

各列向量构成子空间的一组标准正交基，其中  $d_i$  表示 PCA 子空间的维数。由于这里的子空间或者局部模型是由一组样本所张成的，因此， $\mathbf{e}_i$  与  $\mathbf{P}_i$  在刻画局部模型方面发挥着不同的作用。以图示 3.5 中的人脸图像为例。样本均值  $\mathbf{e}_i$  (图示每一列) 主要作为一种“范本” (prototype) 来反映局部模型内部样本公共的表观模式；而主成分矩阵  $\mathbf{P}_i$  则反映模型内部不同样本之间最主要模式差异 (图示后六列的样本表观差别)。换句话说，前者刻画数据样本本身，后者刻画数据的变化模式。

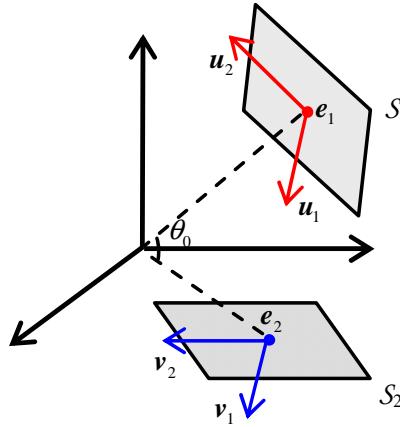


图 3.6 融合数据样例与变化模式度量的 SSD 距离示意。 $S_1$  与  $S_2$  为两个线性子空间，在图示中为二维平面。基于变化模式的距离通过子空间坐标轴  $\mathbf{P}_1 = [\mathbf{u}_1, \mathbf{u}_2]$  与  $\mathbf{P}_2 = [\mathbf{v}_1, \mathbf{v}_2]$  之间的主夹角度量，基于数据样例的距离通过样例点  $\mathbf{e}_1$  和  $\mathbf{e}_2$  之间的夹角  $\theta_0$  度量。

根据上述分析，为了合理而全面地度量两个子空间之间的 SSD 距离，应该同时考虑到子空间的主成分坐标轴  $\mathbf{P}_i$  和样本均值  $\mathbf{e}_i$  各自的差异。如图 3.6 所示， $\mathbf{e}_i$  给出了子空间在全局观察空间中的位置， $\mathbf{P}_i$  则进一步给出了子空间在全局空间中延展的方向。回顾 3.2.1 节中所讨论的 SSD 距离，可以看到前人的大部分工作[29][43][96][98]集中采用了主夹角方法来度量 SSD，这一度量仅利用了  $\mathbf{P}_i$  的信息，反映的是两个子空间公共的变化模式，本文称这类距离度量为“基于变化模式的度量” (variation based measure)。另一方面，其它一些方法[20][21][27]则仅利用样本均值  $\mathbf{e}_i$  来计算局部模型之间的距离而忽略了数据的变化模式的差异，本文称这类距离度量为“基于数据样例的度量” (exemplar based measure)。很显然，这两种度量各自分别刻画了数据变化模式的相似性和数据样本本身的相似性。基于主夹角方法的核心思想，本文将上述两类距离度量有效地融合在一起，提出一种完整的“子空间-子空间距离” (SSD)。

### 3.3.2.1 主夹角方法

给定两个子空间  $S_1$  和  $S_2$ ，它们各自的样例点和标准正交基分别为  $\mathbf{e}_1$ ， $\mathbf{e}_2$  和  $\mathbf{P}_1 \in \mathbb{R}^{D \times d_1}$ ， $\mathbf{P}_2 \in \mathbb{R}^{D \times d_2}$ ，其中  $d_1$ ， $d_2$  为子空间的维数。子空间  $S_1$  与  $S_2$  之间的主夹角  $0 \leq \theta_1 \leq \theta_2 \leq \dots \leq \theta_r \leq (\pi/2)$  有如下定义：

$$\begin{aligned} \cos \theta_k &= \max_{\mathbf{u}_k \in S_1} \max_{\mathbf{v}_k \in S_2} \mathbf{u}_k^T \mathbf{v}_k, \text{ 满足} \\ \mathbf{u}_k^T \mathbf{u}_k &= \mathbf{v}_k^T \mathbf{v}_k = 1, \quad \mathbf{u}_k^T \mathbf{u}_i = \mathbf{v}_k^T \mathbf{v}_i = 0, \quad i = 1, 2, \dots, k-1 \end{aligned} \quad (3-8)$$

其中,  $r = \min(d_1, d_2)$  [34]。 $\mathbf{u}_k$  与  $\mathbf{v}_k$  称为第  $k$  对典型向量(canonical vectors)。主夹角的余弦称为典型相关系数(canonical correlations)。

计算主夹角或者典型相关系数有很多方式, 其中最常用的是文献[9]提出的基于奇异值分解(Singular Value Decomposition, SVD)的算法, 该算法的一大优点是数值计算结果比较稳定。具体求解如下, 对矩阵  $\mathbf{P}_1^T \mathbf{P}_2$  进行 SVD 分解为

$$\mathbf{P}_1^T \mathbf{P}_2 = \mathbf{Q}_1 \mathbf{A} \mathbf{Q}_2^T, \quad (3-9)$$

其中,  $\mathbf{Q}_1$  和  $\mathbf{Q}_2$  为正交矩阵,  $\mathbf{A} = \text{diag}(\sigma_1, \dots, \sigma_r)$  为对角矩阵。SVD 分解得到的奇异值  $\sigma_1, \dots, \sigma_r$  即为典型相关系数:

$$\cos \theta_k = \sigma_k, \quad k = 1, 2, \dots, r. \quad (3-10)$$

对应的典型向量为  $\mathbf{U} = \mathbf{P}_1 \mathbf{Q}_1 = [\mathbf{u}_1, \dots, \mathbf{u}_{d_1}]$ ,  $\mathbf{V} = \mathbf{P}_2 \mathbf{Q}_2 = [\mathbf{v}_1, \dots, \mathbf{v}_{d_2}]$ 。如果两个子空间的最大主夹角比较小的话, 就说明这两个子空间彼此非常接近。直观地, 第一对典型向量对应于两个子空间最相似的变化模式, 第二对典型向量则对应于与第一对模式正交的剩余模式中两个子空间最相似的变化模式, 其它以此类推。

### 3.3.2.2 子空间-子空间距离的定义

关于子空间距离的相关研究工作中, 主夹角是应用最为广泛的基准度量。其中一项具有开创性的工作是互子空间方法(Mutual Subspace Method, MSM)[98], 该方法利用最小的主夹角  $\theta_1$  定义子空间距离, 称为最大相关距离(Max Correlation):

$$d_{\text{Max}}(S_1, S_2) = (1 - \cos^2 \theta_1)^{1/2} = \sin \theta_1. \quad (3-11)$$

与之对应的是, 利用最大的主夹角  $\theta_r$  定义的所谓最小相关距离(Min Correlation):

$$d_{\text{Min}}(S_1, S_2) = (1 - \cos^2 \theta_r)^{1/2} = \sin \theta_r. \quad (3-12)$$

如文献[29]所述, 最大和最小相关距离都严重依赖于主夹角的统计概率分布, 从而分别适用于不同的数据情形。更为稳定的距离方式应该是综合了上述最大和最小相关距离的特性, 利用所有主夹角信息定义的“投影距离”(projection metric), 具体形式如下:

$$d_p(S_1, S_2) = \left( \sum_{k=1}^r \sin^2 \theta_k \right)^{1/2} = \left( r - \sum_{k=1}^r \cos^2 \theta_k \right)^{1/2}. \quad (3-13)$$

投影距离的另一个特点是满足距离函数 metric 的三个要素条件, 即非负性、对称性与三角不等式[91]。关于其它一些子空间距离的详细介绍, 可以参考文献[29][91]。

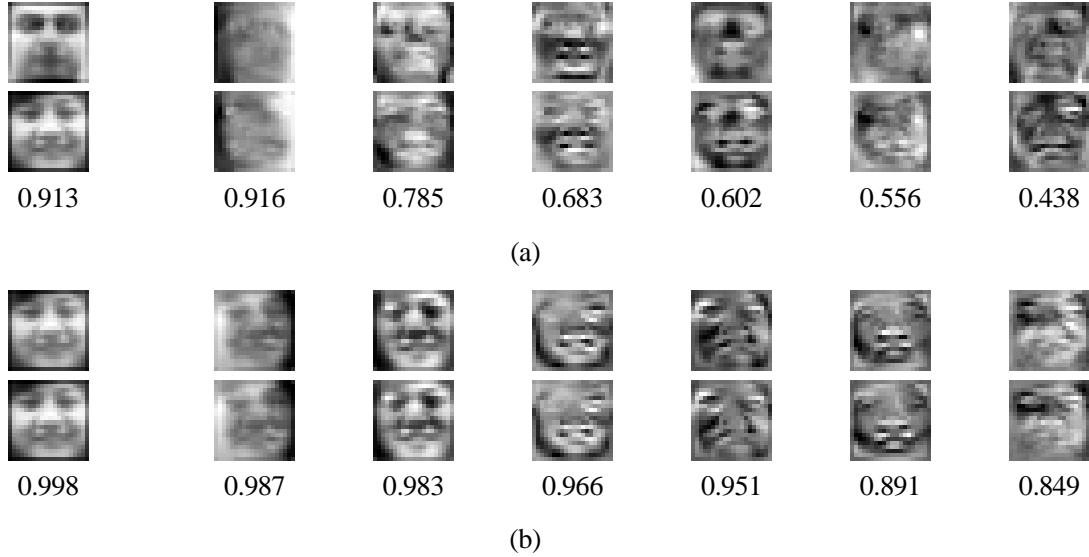


图 3.7 人脸数据对应局部模型之间的相似度。(a)为属于不同人的局部模型对, (b)为属于同一个人的局部模型对。图中第一列为局部模型的样例点, 剩余六列依次为主夹角方法计算得到的前六组典型向量对。每对图像下方的数值为该对图像之间的相关系数。可以看出, 典型向量对很好的捕捉到了不同局部模型对公共的变化模式。

公式(3-13)的投影距离可以看作是定义了两个子空间  $S_1$  和  $S_2$  之间的一种“变化模式距离度量”(variation distance measure), 这一距离仅利用了子空间的基向量  $\mathbf{P}_1$  和  $\mathbf{P}_2$  这部分信息。接下来需要考虑的是利用子空间的另一部分, 即样例点  $\mathbf{e}_1$  和  $\mathbf{e}_2$  的信息。度量两个样本点之间的距离, 最直接的方式应该是 3.2.1 小节公式(3-1)所给出的“点-点距离”, 即两点之间的经典欧氏距离。然而, 由于公式(3-1)与公式(3-13)分别是从向量范数和向量夹角的不同形式定义的, 这两种距离并不在同一个度量空间内, 因此很难将二者进行直接的融合。考虑两点之间基于向量夹角的另外一种距离定义, 即人脸识别中普遍采用的余弦相关距离[4][57]。如图 3.6 所示, 采用样例点之间的夹角  $\theta_0$  来定义它们的余弦距离:

$$d_E(S_1, S_2) = (1 - \cos^2 \theta_0)^{1/2} = \sin \theta_0, \text{ 其中 } \cos \theta_0 = \mathbf{e}_1^T \mathbf{e}_2 / \|\mathbf{e}_1\| \cdot \|\mathbf{e}_2\|. \quad (3-14)$$

本文将这一距离称为“数据样例距离度量”(exemplar distance measure)。

基于公式(3-13)与(3-14)定义的两种距离度量, 通过给各自施加同等的权重, 就可以将二者有机地融合起来, 从而得到完整的“子空间-子空间距离”的定义如下:

$$d(S_1, S_2) = (\sin^2 \theta_0 + \frac{1}{r} \sum_{k=1}^r \sin^2 \theta_k)^{1/2} = (2 - \cos^2 \theta_0 - \frac{1}{r} \sum_{k=1}^r \cos^2 \theta_k)^{1/2}. \quad (3-15)$$

利用公式(3-15)的距离来比较两个图像集合时, 这两种距离以互补的方式分别从不同的方面来反映集合的差异: 数据样例距离主要描述两个集合中的图像在表观上的相似程度, 变化模式距离则描述两个集合中的图像在变化模式上的相似程度。以图 1.4 所示的三个

图像集合所对应的流形  $\mathcal{M}_1$ ,  $\mathcal{M}_2$  和  $\mathcal{M}^p$  举例说明。通过建立各流形的局部线性模型之后, 计算这些局部模型之间的距离。图示 3.7(a)给出了流形  $\mathcal{M}_1$  和  $\mathcal{M}^p$  (二者属于不同的人) 上最近的一对局部模型; 图示 3.7(b)给出了流形  $\mathcal{M}_2$  和  $\mathcal{M}^p$  (二者属于同一个人) 上最近的一对局部模型。可以看出, 上述两种距离度量均能在一定程度上区分出身份间的差异。

### 3.3.3 局部距离的全局综合

接下来进入 MMD 框架的第三个模块, 即公式(3-6)中权重  $f_{ij}$  的确定。从一般的意义上说, 这个问题可以看作是近年来引起研究者广泛关注的“多对多匹配”问题[41][64]。针对本文所研究的图像集合分类问题, 则需要考虑其具体的实际情形。以图 1.4 所示的人脸识别问题为例。通常来说, 属于同一个人的 gallery 与 probe 图像集合往往分别涵盖不同视角范围下的图像, 这些图像在视角上具有一定度的重叠, 但是从整体分布上看来可能存在较大差别。此时, 匹配属于同一类别的这样两个集合, 最有效的方式应该是寻找两个集合公共视角范围内的图像, 然后对这部分数据进行相似度计算[43]。从公式(3-6)所述的匹配流形上的局部模型对的角度来说, 应该更加关注两个流形上距离相近的那些局部模型对, 而并不期望将所有的局部模型对都能成功匹配上。

按照 3.2.2 节定义的变量标记方式, 给定两个流形  $\mathcal{M}_1 = \{C_i : i = 1, 2, \dots, m\}$ ,  $\mathcal{M}_2 = \{C'_j : j = 1, 2, \dots, n\}$ , 首先定义如下两个指示函数(indicator functions):

$$\begin{aligned} N(i) &= \arg \min_j d(C_i, C'_j), \quad j = 1, 2, \dots, n, \\ N'(j) &= \arg \min_i d(C_i, C'_j), \quad i = 1, 2, \dots, m. \end{aligned} \quad (3-16)$$

这里, 定义在流形  $\mathcal{M}_1$  上的指标函数  $N(i)$  表示局部模型  $C_i$  在另一个流形  $\mathcal{M}_2$  上的最近邻(nearest neighbor, NN)的标号。类似地, 定义在流形  $\mathcal{M}_2$  上的指标函数  $N'(j)$  表示局部模型  $C'_j$  在流形  $\mathcal{M}_1$  上的最近邻的标号。这样, 可以得到一个记录了所有最近邻局部模型对的集合  $A$  如下表示:

$$A = A_1 \cup A_2, \quad \text{其中 } A_1 = \{(C_i, C'_{N(i)})|_{i=1}^m\}, \quad A_2 = \{(C_{N'(j)}, C'_j)|_{j=1}^n\}. \quad (3-17)$$

需要注意的是, 尽管最近邻关系是一种非对称关系, 但  $A_1$  和  $A_2$  集合仍可能包含一些公共元素, 即  $A_1 \cap A_2 \neq \emptyset$ 。为了后续表示和讨论方便, 两个集合的这部分可能的公共元素仍然将区别对待, 因此, 集合  $A$  中的元素个数(即集合的势)始终为  $m+n$ 。基于上述定义, 针对局部距离的全局综合问题, 接下来讨论几种可能的方案并进行对比讨论。

#### 3.3.3.1 方案-1: 最小最近邻策略(Min NN)

为了匹配两个流形上的公共部分, 一种简单而直观的方法是计算它们最匹配的一对局部模型之间的相似度, 如图 3.8(a)所示。具体地, 就是计算集合  $A$  中子空间对之间的最小距离:

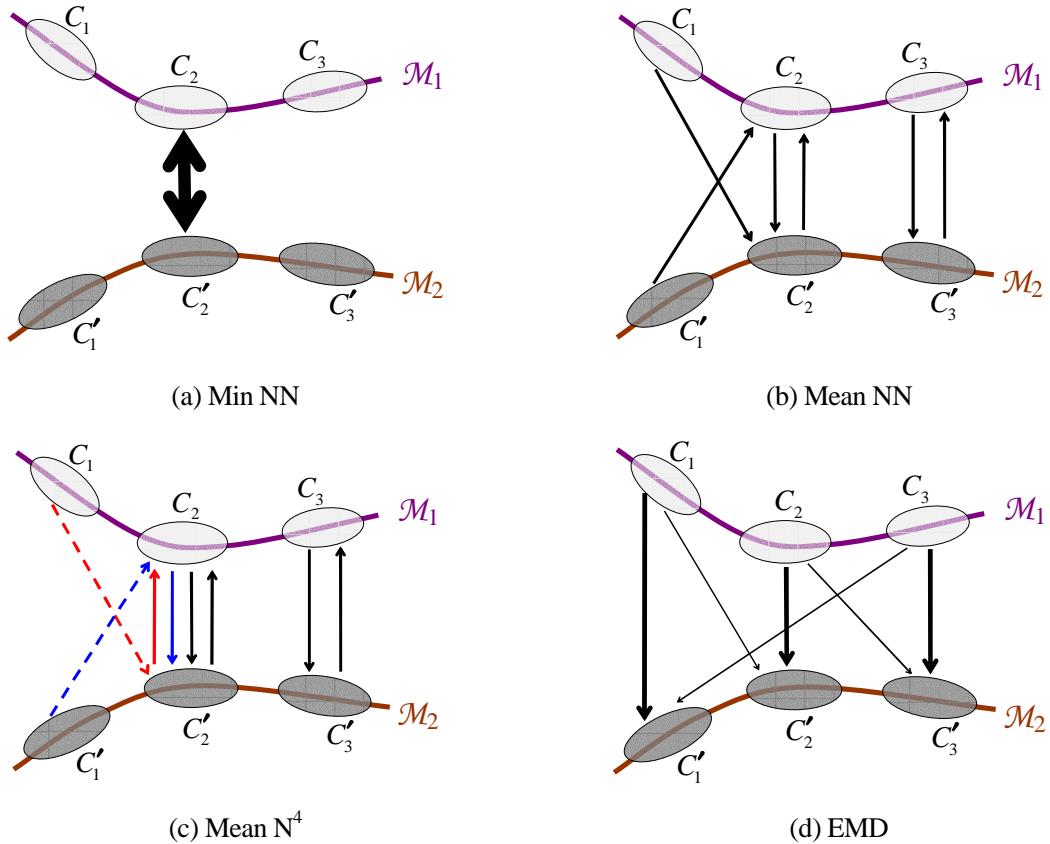


图 3.8 局部距离的全局综合方案示意。图中的两个流形各自包含三个局部模型，图(a)~(c)中，箭头的指向表示从一个局部模型到其最近邻的方向，比如  $C_1$  的最近邻为  $C'_2$ ，而  $C'_2$  的最近邻则为  $C_2$  而非  $C_1$ 。图(d)中，箭头指向表示 EMD 距离优化中从供应者到消费者的方向。所有图中实线箭头的粗细表示其相应的权重，(c)图的虚线箭头权重为 0。比如，(a)图箭头权重为 1，(b)(c)图中每个箭头的权重为 1/6，(d)图粗箭头和细箭头的权重分别为 1/4 和 1/12。

$$d_1(\mathcal{M}_1, \mathcal{M}_2) = \min_{C_i \in \mathcal{M}_1} \min_{C'_j \in \mathcal{M}_2} d(C_i, C'_j) = \min \{ d(C_i, C'_{N(i)}) |_{i=1}^m, d(C_{N'(j)}, C'_j) |_{j=1}^n \} . \quad (3-18)$$

这一方案是在本文关于 MMD 方法最初的研究工作[89]中所采用的匹配方式，初步实验中的良好性能表现验证了该方案的有效性。然而，由于这一方案中为最近邻的局部模型对赋值权重 1，而其它所有的模型对的权重都为 0，当流形中存在噪声与野点数据时，基于该方案的局部距离全局综合结果很容易受到影响，从而表现不够稳定。

### 3.3.3.2 方案-2：平均最近邻策略(Mean NN)

为了综合利用更多的最近邻局部模型对的信息，很容易想到的一种方案是计算集合  $A$  中所有子空间对距离的平均值：

$$d_2(\mathcal{M}_1, \mathcal{M}_2) = \frac{1}{m+n} \left( \sum_{i=1}^m d(C_i, C'_{N(i)}) + \sum_{j=1}^n d(C_{N'(j)}, C'_j) \right) . \quad (3-19)$$

这一方案中为所有的近邻模型对施加了同等的权重，而没有考虑到流形全局数据的实际分布情形。如图 3.8(b)所示，将所有的近邻模型对同等对待，而不区分各自的邻近程度，显然并不太合理。

### 3.3.3.3 方案-3：平均最近邻之最近邻策略(Mean NN's NN (N<sup>4</sup>))

事实上，上述两种方案分别对应于集合  $A$  中不同近邻模型对加权的两种极端情形。针对图像集合分类问题而言，基本的思想是距离越近的局部模型对相应的权重也应该越大。因此，针对方案-2 进行相应的改进，将距离较远的近邻对上的部分权重转移到距离较近的近邻对上来，如图 3.8(c)所示，这就是方案-3 的核心思想。

具体地，针对流形  $\mathcal{M}_1$  上的每个局部模型  $C_i$  ( $i=1, 2, \dots, m$ )，找到其最近邻(NN)  $C'_{N(i)}$ 。然后，从  $C'_{N(i)}$  出发再找到其最近邻  $C_{N'(N(i))}$ 。将  $C_{N'(N(i))}$  称为  $C_i$  的最近邻之最近邻 (NN's NN, N<sup>4</sup>)。将公式(3-19)中的  $(C_i, C'_{N(i)})$  替换为  $(C_{N'(N(i))}, C'_{N(i)})$ ，就将前者的权重转移到到了后者上。同样地，针对流形  $\mathcal{M}_2$  上的每个局部模型  $C'_j$  ( $j=1, 2, \dots, n$ )，分别找到其最近邻  $C_{N'(j)}$  和最近邻之最近邻  $C'_{N(N'(j))}$ 。然后，将公式(3-19)中的  $(C_{N'(j)}, C'_j)$  替换为  $(C_{N'(j)}, C'_{N(N'(j))})$ ，即可完成权重从较远近邻对到较近近邻对的转移。经过这些权重调整后，可以得到方案-3 的距离计算公式：

$$d_3(\mathcal{M}_1, \mathcal{M}_2) = \frac{1}{m+n} \left( \sum_{i=1}^m d(C_{N'(N(i))}, C'_{N(i)}) + \sum_{j=1}^n d(C_{N'(j)}, C'_{N(N'(j))}) \right) . \quad (3-20)$$

以图示 3.8(c)中的两个流形为例来分析方案-3 的计算过程。针对局部模型  $C_1$  (此时  $i=1$ )，其 NN 和 N<sup>4</sup> 分别为  $C'_{N(1)} = C'_2$  和  $C_{N'(N(1))} = C_2$ 。因此，距离较远的近邻对  $(C_1, C'_2)$  (图中红色虚线箭头所示) 上的权重就转移到了距离较近的近邻对  $(C_2, C'_2)$  (图中红色实线箭头所示) 上。类似地，针对局部模型  $C'_1$  (此时  $j=1$ )，近邻对  $(C_2, C'_1)$  (图中蓝色虚线箭头所示) 上的权重转移到到了  $(C_2, C'_2)$  (图中蓝色实线箭头所示) 上。

可以看出，方案-3 有效地综合了前两个方案各自的优点。与方案-1 相比，方案-3 综合利用了更多的数据信息，从而保证了对噪声数据的鲁棒性和结果的稳定。与方案-2 相比，方案-3 根据实际数据的分布自适应地调整不同局部模型对上的权重分布，从而更可靠地反映流形之间的近邻关系。

### 3.3.3.4 方案-4：EMD 匹配策略(Earth Mover's Distance, EMD)

如前所述，局部距离的全局综合问题可以看作传统意义上的多对多匹配问题，从而可以采用通用的多对多匹配方法来解决。近年来，在计算机视觉的很多问题中[41][64][97]获得广泛应用的一种方法称为 EMD 距离(Earth Mover's Distance, EMD)[68]。方案-4 即采用 EMD 来完成局部距离的全局综合。

具体地，利用 EMD 来自动计算公式(3-6)中的权重  $f_{ij}$ 。流形  $\mathcal{M}_1$  和  $\mathcal{M}_2$  分别看作 EMD 框架中的两个 signatures，各自的局部模型就对应于供应者(supplier)和消费者(consumer)

的角色, 具体表示为:  $\mathcal{M}_1 = \{(C_1, w_{C_1}), \dots, (C_m, w_{C_m})\}$ ,  $\mathcal{M}_2 = \{(C'_1, w_{C'_1}), \dots, (C'_n, w_{C'_n})\}$ , 其中,  $w_{C_i}$  和  $w_{C'_j}$  分别为局部模型  $C_i$  和  $C'_j$  的权重。EMD 距离的直观意义为: 给定多个供应者和多个消费者以及各自的供求能力, 计算从供应方到消费方的最佳物流路线。在 EMD 框架中, 权重  $w_{C_i} / w_{C'_j}$  分别对应于供应者的总供应能力与消费者的总消费能力, 其缺省设置分别为  $1/m$  和  $1/n$ , 称为归一化权重。由于其简单有效等特点, 这一参数设置在相关研究工作中[64][97]也得到了普遍使用。流形  $\mathcal{M}_1$  与  $\mathcal{M}_2$  之间的 EMD 距离计算如下:

$$d_4(\mathcal{M}_1, \mathcal{M}_2) = \sum_{i=1}^m \sum_{j=1}^n \hat{f}_{ij} d_{ij} / \sum_{i=1}^m \sum_{j=1}^n \hat{f}_{ij} . \quad (3-21)$$

其中, 局部模型 SSD 距离  $d_{ij} = d(C_i, C'_j)$  称为地面距离(ground distance), 目标权重  $\hat{f}_{ij}$  则反映了从供应方到消费方的最优流通路线。该优化问题通过如下线性规划问题[68][97]求解:

$$\begin{aligned} \hat{f}_{ij} &= \arg \min_{f_{ij}} \sum_{i=1}^m \sum_{j=1}^n f_{ij} d_{ij}, \quad \text{满足} \\ \sum_{i=1}^m \sum_{j=1}^n f_{ij} &= \min(\sum_{i=1}^m w_{C_i}, \sum_{j=1}^n w_{C'_j}); \quad f_{ij} \geq 0; \\ \sum_{j=1}^n f_{ij} &\leq w_{C_i}, 1 \leq i \leq m; \quad \sum_{i=1}^m f_{ij} \leq w_{C'_j}, 1 \leq j \leq n . \end{aligned} \quad (3-22)$$

$\hat{f}_{ij}$  可以理解为两个流形上局部模型之间的最佳匹配。由于公式(3-22)所施加的“供求平衡”约束, 基于归一化权重的 EMD 匹配将不可避免地将一些权重分配到非近邻的局部模型对上, 如图 3.8(d)所示。显然, 这种权重分配方式与匹配流形公共部分数据的初衷存在矛盾。采用其它较常用的 EMD 权重设置比如单位权重, 同样面临这个问题。由于权重  $w_{C_i} / w_{C'_j}$  的先验知识很难获取, 因此这一问题成为 EMD 能否得到符合期望匹配结果的最大障碍。

### 3.3.3.5 四种方案对比分析

方案-1 和方案-3 针对距离较小的近邻模型对均给予了更多关注, 二者表现出了类似性能, 且在绝大多数情况下都优于另外两种方案。在方案-4 中, 无论采用归一化还是单位权重设置, 均将流形中的各局部模型进行同等对待, 从而在所有方案中性能最差。这些讨论均通过 3.5.4 节的实验得到了验证。此外, 方案-1/3 还有一个优点值得关注。当其中一个流形仅由一个局部模型构成时, 即  $m=1$  或者  $n=1$ , 公式(3-6)中的流形-流形距离就变为了子空间与流形之间的距离, 而其退化后的形式与公式(3-5)中定义的子空间-流形距离的形式是完全一致的。方案-2/4 则不具有这个特点。

从公式(3-6)的加权策略的角度展开进一步分析。方案-1/2 可以看作是手工指定近邻模型对的权重, 二者分别采用 Dirac-delta 函数和均匀分布函数的形式。与之相比, 方案-3/4 则采用了自动确定权重的方式, 二者分别采用自适应的权重调整策略和通用的优化求解方式进行设置。

从计算代价方面来看, 方案-4 的 EMD 距离计算效率最低。假定两个流形中的局部模型个数相等, 即  $m = n$ , EMD 的复杂度为  $O(m^3 \log(m))$  [68] [97]。与之相比, 其余三个方案复杂度均为  $O(m^2)$ 。

最后, 需要说明的是, 针对流形-流形距离的计算, 除了上述讨论的四种方案, 还可以采用一些其它可能的策略。比如, 将方案-2“Mean NN”中的平均值替换为中值, 从而演变为“Median NN”, 由于均值与中值相似的性质, 这两种策略性能应该接近。另一种可能是采用近邻模型对距离的最大值, 即“Max NN”的方式, 则正好演变为经典的 Hausdorff 距离的形式[17][36]。这种方式本质是匹配两个集合中差异最大的近邻对, 与 3.3.3 节开始提到的图像集合匹配目标不一致, 因此并不适合于本文所研究的基于图像集合的物体识别问题。关于两个集合之间更多的距离定义方式, 请参考文献[17]的综述。

## 3.4 讨论

本节首先讨论流形-流形距离计算方面的两个理论问题, 然后从两个角度详细对比本文方法与相关工作的关系与差别。

### 3.4.1 关于度量准则的分析

所谓度量(metric)是指在空间  $X$  的元素间定义的距离函数, 且满足如下基本条件[29]:

**定义 1.** 实值函数  $d: X \times X \rightarrow \mathbb{R}$  称为一种度量, 如果满足

1. 非负性:  $d(x_1, x_2) \geq 0$ , 当且仅当  $x_1 = x_2$  时,  $d(x_1, x_2) = 0$  ;
2. 对称性:  $d(x_1, x_2) = d(x_2, x_1)$  ;
3. 三角不等式:  $d(x_1, x_2) + d(x_2, x_3) \geq d(x_1, x_3)$  ;

对于所有的元素  $x_1, x_2, x_3 \in X$  都成立。

容易验证, 无论采用 3.3.3 节中的哪一种综合方案, 本文提出的 MMD 距离均符合上述定义中的条件 1 和 2, 即非负性和对称性。针对条件 3 的三角不等式关系, 由于目前难以给出严格的理论证明, 这里尝试通过实例推导来讨论。给定三个流形:  $\mathcal{M}_1 = \{C_1\}$ ,  $\mathcal{M}_2 = \{C_1, C_2\}$ ,  $\mathcal{M}_3 = \{C_2\}$ , 其中  $\mathcal{M}_1$  和  $\mathcal{M}_3$  均为  $\mathcal{M}_2$  的子流形。分别采用四种方案来计算三个流形之间的距离, 在方案-4 的 EMD 距离中使用归一化权重设置。具体如表 3.1 所示:

表 3.1 MMD 采用不同全局综合方案对比

	$(\mathcal{M}_1, \mathcal{M}_2)$	$(\mathcal{M}_2, \mathcal{M}_3)$	$(\mathcal{M}_1, \mathcal{M}_3)$
$d_1$	0	0	$d(C_1, C_2)$
$d_2$	$1/3 \cdot d(C_1, C_2)$	$1/3 \cdot d(C_1, C_2)$	$d(C_1, C_2)$
$d_3$	0	0	$d(C_1, C_2)$
$d_4$	$1/2 \cdot d(C_1, C_2)$	$1/2 \cdot d(C_1, C_2)$	$d(C_1, C_2)$

在方案-1/2/3 中,  $d(\mathcal{M}_1, \mathcal{M}_2) + d(\mathcal{M}_2, \mathcal{M}_3) < d(\mathcal{M}_1, \mathcal{M}_3)$ , 因此均不满足三角不等式条件。考虑到流形匹配的实际需求, 这一结论并不出乎意料。在图像集合分类问题中, 如果其中一个集合仅与另一集合的子集部分的数据相似, 正如 3.3.3 节开始所讨论的情形, 那么对图像集合进行部分匹配的需求就非常必要。文献[83]也提出了类似的论点, 作者指出感知上的距离不一定需要满足度量的条件。

针对方案-4, 如果地面距离  $d(C_1, C_2)$  满足度量的条件, 比如采用公式(3-13)的投影距离, 则最终的 EMD 距离也满足度量条件[68]。至于公式(3-15)定义的完整 SSD 距离是否满足度量条件, 则仍需要进一步的研究和讨论。

值得注意的是, 当其中一个流形是另一个流形的子流形时, 方案-1/3 计算得到的距离均为 0, 这一结果与直觉上的期望也是一致的。与之相比, 方案-4 在相同情形下计算所得的距离是四种方案中最大的。这一对比也能在一定程度上解释下节实验中方案-4 效果最差的结果。

### 3.4.2 算法复杂度分析

流形-流形距离的算法复杂度主要由以下三部分构成:

(1) 基于 HDC 算法建立局部模型。根据本文 2.3.4.2 小节针对 MLE 方法的复杂度分析, 可知建立局部模型这一阶段的复杂度为  $O(\sum_{p=1}^{\lfloor \log N \rfloor} (2^p (N/2^p)^2)) \approx O(N^2)$ , 其中  $N$  为图像集合中的样本个数。

(2) 计算局部模型的 PCA 子空间表示。针对各局部模型  $C_i$  ( $i = 1, 2, \dots, m$ ), 其数据矩阵规模为  $D \times N_i$ 。假设所有局部模型的样本规模相同, 则  $N_i \approx N/m$ 。PCA 子空间计算的主要步骤为对  $D \times D$  的协方差矩阵进行特征分解。在实际问题中, 通常的情况是  $D > N > N_i$ , 此时特征分解问题可以在  $N_i \times N_i$  规模的矩阵上进行, 因此这一步的算法复杂度约为  $O((N/m)^3 \cdot m)$ 。

(3) 基于主夹角方法进行局部模型匹配。假设两个流形中的局部模型个数相同, 则总计需要进行  $m \times m$  次的模型匹配。总的时间复杂度为  $O(d^3 \cdot m^2)$ , 其中  $d$  为局部模型的子空间维数。

### 3.4.3 与前人工作对比

从集合匹配问题的角度分析, 本文的 MMD 方法与该领域的两个典型方法 MSM[98] 和“LLE + K-means”[27]在不同方面存在相似性。在集合建模方面, MSM 与“LLE + K-means”分别利用线性子空间与非线性流形来表示集合。然而, 后者在建模过程中没有考虑实际的数据分布情形, 只是简单地利用预先指定数量的局部模型来刻画实际分布各不相同的流形。与之不同, 本文的 MMD 方法则能够自适应地从流形中提取多尺度的局部模型表示。在所采用的子空间-子空间距离度量方面, MSM 与“LLE + K-means”分别采

用了基于变化模式和数据样例的距离度量,而 MMD 则将两种度量方式有机的融合起来,如 3.3.2.2 所述。总的来说, MMD 将这两种方法的优点有效地结合起来,作为一种通用的集合相似度度量可以应用在很多实际问题中。

针对另一个不同的问题—不变图像相似度的度量,即计算同一个物体在不同变化模式下得到的两幅图像之间的距离,有很多相关的研究工作,如联合流形距离(Joint Manifold Distance, JMD)[22],多分辨率流形距离(Multiresolution Manifold Distance, MRMD)[84],以及基于 DC 算法的流形距离(Manifold Distance using DC, MDDC)[47]。尽管这些方法与本文的流形-流形距离在名称上有相似之处,但它们在本质上有很多不同的地方,具体表现在如下几个方面:

1. 这些方法主要作为图像之间的相似度度量,用以解决图像匹配中针对参数式图像变换的不变性问题,其先驱性工作为经典的切距离算法(tangent distance, TD)[76]。本文 MMD 则主要是从非参数角度解决图像集合之间的相似度度量问题。
2. 这些方法中所谓的“流形距离”是指从参考点(reference point)到其变换流形的距离[47],具体是通过线性子空间中样本点的距离来定义的;而 MMD 则考虑到了通用流形上数据变化模式之间的距离。
3. 此外,这些方法中的距离计算都需要在一个较大的参数空间内通过迭代优化方式求解;而 MMD 的计算则通过闭形式求解,从而适用于在线图像集合匹配的应用问题。

## 3.5 实验结果

本节通过具体的基于图像集合的人脸识别应用问题来验证所提出的 MMD 方法的有效性。如前所述,每位注册人采用其一组人脸图像构成的集合来注册,这个图像集合建模为一个 *gallery* 流形;每位测试人则以同样的方式建模为一个 *probe* 流形。识别的过程即为利用 MMD 来进行 *probe* 流形与所有 *gallery* 流形匹配的过程,最终将测试人的类别标记为与其距离最近的注册人的类别。

### 3.5.1 实验数据库

为保证对比实验的充分性与广泛性,选用了三个具有不同特点的数据集合,包括两个基准数据库: Honda/UCSD [51], CMU MoBo [26],与一个富有挑战性的集合: YouTube Celebrities [45]。接下来简要介绍各数据库的构成。

**Honda/UCSD:** 该数据库广泛应用于基于图像集合或者视频的人脸识别问题[45][51][104]。本文选用了该数据库的第一个子集,其中包括 20 个人的 59 段视频(每人至少有 2 段)。每段视频大约含有 400 帧图像,涵盖了不同情形下的头部 2D/3D 姿态变化,同时包括面部表情的变化。实验中首先采用瀑布型人脸检测器[88]定位图像中的人脸,然后参照前人工作的设置,将裁减后的人脸统一缩放至  $20 \times 20$  的灰度图像,最后采用直方图均衡化消除不同图像中的光影响。部分实验用的人脸图像如图 3.9(a)所示。

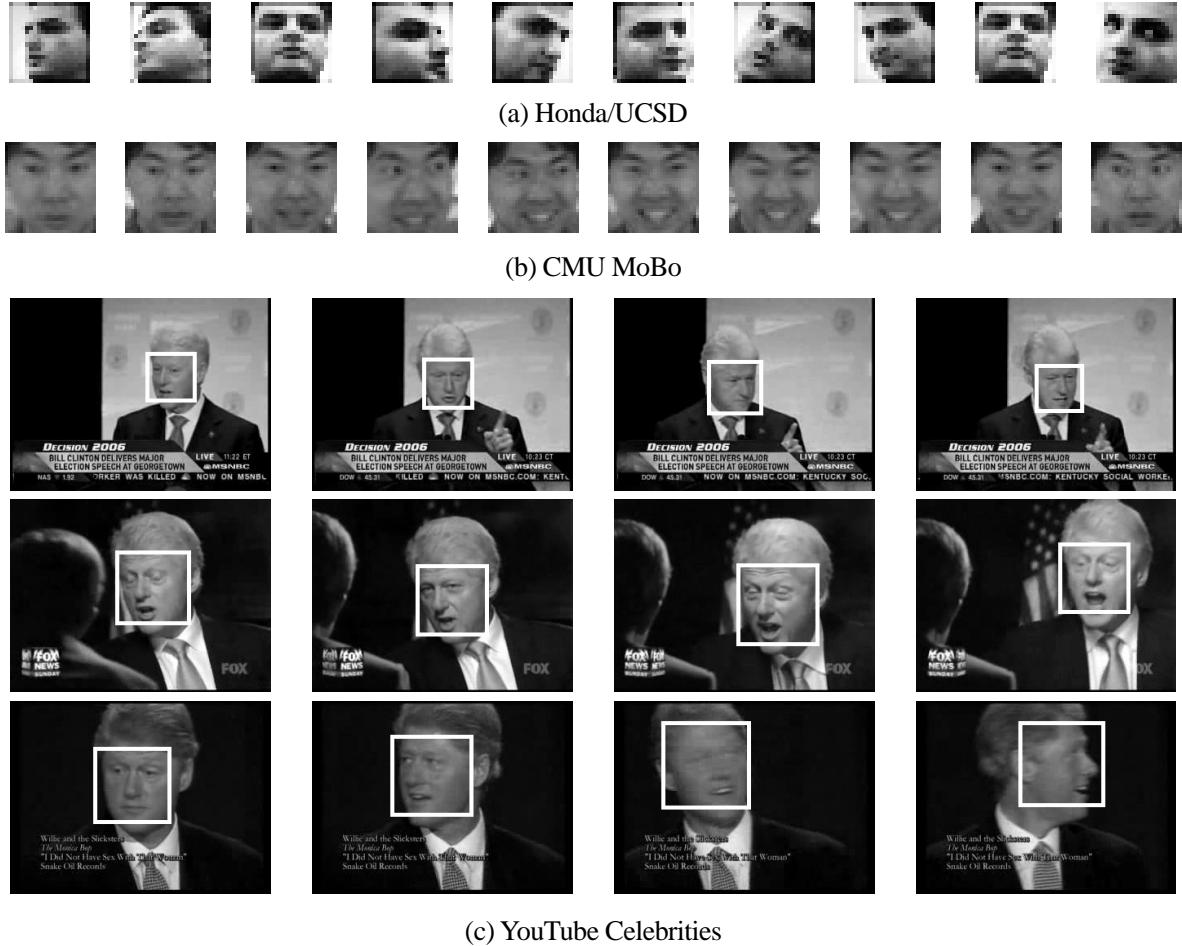


图 3.9 MMD 方法对比实验采用的人脸数据库示意。图中每一行对应于各数据库中一个图像集合的若干代表性样本。(c)图中给出了原始视频帧中的人脸检测结果。

**CMU MoBo:** 该数据库全名为 Motion of Body, 其最初的研究目的是远距离人物身份认证问题, 近些年来也广泛应用于基于视频的人脸识别研究工作中[27][58][106]。数据库中包含 25 个人的 100 段视频, 其中每人各有 4 段视频, 每段视频包含 300 帧图像。这 4 段视频分别对应于四种运动状态下采集的序列, 具体包括: 手中持球、快速行走、慢速行走、斜面上行走等。视频中人脸样本的提取方式与上一数据库相同, 并参考前人在该数据库的设置将图像归一化为  $30 \times 30$  大小。图示 3.9(b)给出了其中一段视频对应的部分人脸样本。

**YouTube Celebrities:** 该数据库由 M. Kim 等人[45]收集, 用于现实场景中的视频人脸跟踪与识别任务。该数据库包含 47 个人的 1 910 段视频, 均来自于 YouTube 网站上的演艺明星和政治人物的真实视频片段。其中, 每个人的视频片段可以分为 3 个小节(session), 每一小节对应于不同的采集时间与场景, 各小节中的视频片段数量不等。这些视频普遍具有较低的分辨率与较高的压缩率, 且受到不同程度的噪声影响。每段视频包含的图像数一般为数百幅。图示 3.9(c)给出了 Bill Clinton 的 3 段视频片段中的部分图

像，每段视频来自其一个小节。与 Honda 与 MoBo 库相比，该数据库更加具有挑战性，其视频数据变化模式更为复杂，包括更大的姿态变化、复杂的光照、多变的表情以及其它一些变化等。人脸样本的提取方式与上述两个库相同，并且归一化到  $30 \times 30$  大小。

实验中，将各数据库的每段视频提取得到的人脸样本作为一个图像集合。针对每个数据库，分别进行十组交叉验证，每组实验对应于一组随机选择的训练集与测试集标号组合。具体地，针对 Honda 和 MoBo 库，每个人选用一段视频作为训练集，其余视频作为测试集。针对 YouTube 库，在每组实验中从每个人的 3 个小节中随机选用共 9 段视频进行实验，其中每个小节选取 3 段视频。这 9 段视频中的 3 段作为训练集，6 段作为测试集。由于每个人不同小节的视频数据之间具有较大的变化，实验中考虑了两种训练/测试组合方案，分别称为 YouTube\_Easy 与 YouTube\_Hard。在 YouTube\_Easy 的方案中，3 段训练视频分别来自三个小节，同样地，6 段测试视频也分别来自三个小节，因此其训练与测试数据均涵盖了三个小节的不同场景。在 YouTube\_Hard 的方案中，3 段训练视频均来自同一个小节，而 6 段测试视频则来自另外两个小节，即训练数据与测试数据之间不存在数据采集场景的交叠。很显然，后一种方案比前一种方案的难度要大一些。

### 3.5.2 对比方法与参数设置

实验选取了以下几种代表性方法进行对比：

1. 基于单样本识别的代表方法：利用最近邻匹配的(Nearest Neighbor, NN) (i) Eigenface, (ii) Fisherface [4];
2. 基于数据样例距离识别的代表方法：LLE + K-means [27];
3. 基于变化模式距离识别的代表方法：Mutual Subspace Method (MSM) [98];
4. 基于集合的判别式学习的代表方法：Constrained MSM (CMSM) [23], Discriminant Canonical Correlations (DCC) [43];
5. 本文提出的 MMD 方法。

为了保证对比实验的公正性，实验中通过在训练集上的交叉验证来为各算法调节选择最优的参数组合。在 NN-Eigenface 方法中，特征子空间的维数选择为保留 95% 的数据能量。在 NN-Fisherface 方法中，首先采用 PCA 对原始数据降维以消除散度矩阵的奇异问题，之后 LDA 降维中将维数设置为类别数减一。在 LLE + K-means 方法中，选用与原始文献[27]相同的参数设置，每个图像集合对应的流形中提取  $k=5$  个数据样例作为局部模型表示。在识别阶段，按照与前人工作相同的测试流程，这三个方法均分别对 probe 图像集合中的每个样本进行单独分类，然后对所有样本的类别标签按照多数投票的策略确定整个 probe 集合的类别。

针对 MSM/CMSM/DCC 这三个方法，实验中重点参考文献[43]中的参数设置方式。具体地，首先采用 PCA 来学习各图像集合的线性子空间表示，通过保留 95% 的数据能量，各图像集合对应的子空间维数大约为 15。然后，针对 CMSM 和 DCC 两个方法学习

得到的判别子空间，其维数在各数据上设定为识别率最高点对应的数值。最后，在基于主夹角/典型相关系数计算集合相似度阶段，CMSM 与 DCC 均采用了全部相关系数的累加结果。与文献[43]中作者报告的实验现象类似，相关系数个数的选取对 MSM 方法的性能影响较大。为公平对比，本文中针对 MSM 测试了选取不同数量的典型相关系数分别对应的识别率，并在各数据库上报告其最佳结果。此外，需要注意的是，DCC 算法的训练阶段需要每一类至少包含两个图像集合以构造类内散度矩阵。然而，在 3.5.1 节的数据设置中，Honda 与 MoBo 这两个数据库针对每个人仅有一个集合作为训练集。为此，参照文献[43]作者的做法，实验中将每个人的单一训练图像集随机地划分为两个子集作为 DCC 每类的训练集合。

针对 MMD 方法，按照公式(3-6)的计算框架，首先需要确定局部模型的个数  $m, n$ 。根据 2.3.1.2 节所述，该参数可以根据 HDC 算法得到的非线性分值曲线上的拐点来选择，曲线在拐点处的一阶导数趋近于 0。图示 3.10 给出了上述三个人脸数据库中各自选取的一个集合得到的曲线。一般来说，包含有 300 到 400 个样本的集合得到的局部模型数为 5~8。接下来，针对局部模型  $C_i$  的表示，PCA 子空间的维数  $d_i$  选择为保留 95% 数据能量，取值大约在 5 维左右。局部模型之间的距离采用公式(3-15)定义的 SSD 来计算，局部距离的全局综合采用公式(3-20)的方案-3。

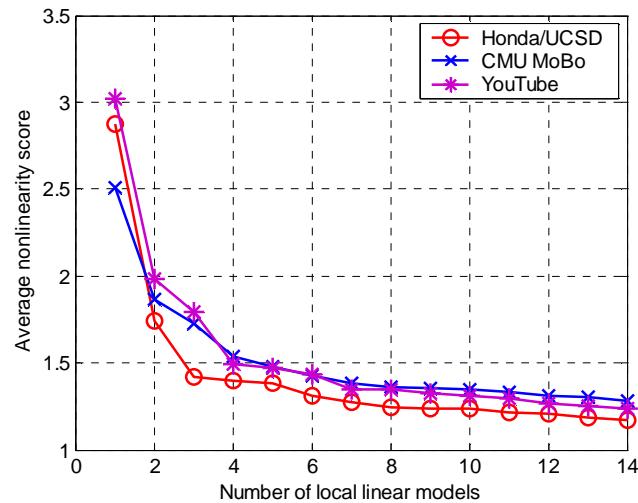


图 3.10 人脸图像集合的非线性分值曲线。图中每条曲线分别对应于三个数据库的一个图像集合。

### 3.5.3 实验对比结果与分析

表 3.2 给出了实验对比的不同方法在几个数据库上的结果，其中识别率的均值与标准差是十次实验统计所得。从该表的对比情况分析，可以发现如下一些现象和结论：

(1) 基于单样本的方法，如 Eigenface 和 Fisherface，通常比基于集合的方法在识别性能上要差一些，特别是在 Honda/UCSD 库上的对比更为明显，这一数据库包含的姿态变化相对于另两个库来说更加剧烈。事实上，由于这两类方法在识别机理上很不相同，

表 3.2 MMD 对比方法的识别结果

数据库	各对比方法 10 组随机实验的识别率均值与标准差						
	Eigenface	Fisherface	LLE+K-means	MSM	CMSM	DCC	MMD
Honda/UCSD	0.745 $\pm$ 0.02	0.789 $\pm$ 0.01	0.894 $\pm$ 0.03	0.923 $\pm$ 0.04	0.975 $\pm$ 0.02	0.980 $\pm$ 0.01	0.971 $\pm$ 0.02
CMU MoBo	0.813 $\pm$ 0.03	0.885 $\pm$ 0.01	0.879 $\pm$ 0.04	0.886 $\pm$ 0.03	0.918 $\pm$ 0.04	0.903 $\pm$ 0.05	0.935 $\pm$ 0.02
YouTube_Easy	0.535 $\pm$ 0.05	0.604 $\pm$ 0.03	0.592 $\pm$ 0.06	0.616 $\pm$ 0.04	0.650 $\pm$ 0.03	0.673 $\pm$ 0.03	0.652 $\pm$ 0.03
YouTube_Hard	0.479 $\pm$ 0.12	0.531 $\pm$ 0.08	0.534 $\pm$ 0.07	0.574 $\pm$ 0.06	0.638 $\pm$ 0.04	0.662 $\pm$ 0.02	0.635 $\pm$ 0.03

将它们进行对比并不公平。然而，从实验的结果看来仍然可以得出一些基本的结论。针对数据中较大的姿态和场景变化，基于单样本的方法显得更加敏感，其主要原因在于这类方法不能有效挖掘和利用图像集合提供的整体信息，因此不太适用于图像集合的分类任务，比如非受控条件下的真实监控场景。文献[27][43]也报告了类似的实验现象与结论。

(2) 基于集合的非判别式方法 LLE + K-means, MSM 和 MMD 由于其各自的特点，在实验中分别表现出不同的性能。在三个方法当中，LLE + K-means 的识别率最低。尽管该方法利用流形的思想来提取训练集中的数据样例，其测试流程却采用了相对有些简单的策略。针对待分类的图像集合，该方法采用类似于 NN-Eigenface 的方式分别单独对集合中的每个样本进行分类，最后通过简单的投票策略确定整个集合的类别。如文献[21]所述，这种方式无法全面地反映图像集合中数据的变化。通过选择最佳数量的主夹角，MSM 方法在识别率上比 LLE + K-means 有所提高。然而，在三个非判别式方法中性能表现最好的是本文提出的 MMD，这一结果也在意料之中，其主要原因是 MMD 将另外两种方法的优点有效地融合在了一起，具体请参考 3.4.3 节的讨论。

(3) 与两个判别式方法 CMSM 和 DCC 相比，MMD 取得了性能相当的识别结果，从而直接验证了 MMD 作为一种通用距离度量的有效性和其广泛的适用性。值得一提的是，在 CMU MoBo 库上，CMSM/DCC 的性能均不及 MMD，一方面是由于这两个方法采用线性子空间来建模图像集合，另一方面更重要的原因是它们在子空间距离度量中仅利用了基于变化模式的距离(variation-based SSD measure)。将变化模式与数据样例这两种距离度量融合在一起的优越性将在下一节实验中进一步验证。

(4) 综合不同方法在三个数据库上的性能表现，所有方法均在极具挑战性的 YouTube 库上给出了最低的识别率。针对该库的两种不同训练/测试集组合方案设置，LLE + K-means 与 MSM 都表现出了较大的性能偏差，它们的最好结果与 MMD 相比大约低 5 个百分点。与之相比，针对不同的实验场景设置，MMD 则表现出与判别式方法 CMSM/DCC 类似的稳定性。从 MMD 方法在不同数据库上表现出的相对较低的识别率标准差也可以印证这一结论。

### 3.5.4 算法各模块评估

本节通过进一步实验来评估 3.3 小节中讨论的 MMD 算法的三个构成模块。实验选取 MoBo 与 YouTube\_Easy 两个数据库方案来报告结果。

(1) 局部线性模型构建。实验对比了本文的 HDC 算法与另外两个常用的聚类算法 HAC 和 K-means。其中, HAC 算法实现采用文献[21]的框架, 聚类之间的距离度量采用测地距离。K-means 算法则采用经典的欧氏距离。实验中, 首先利用 HDC 对各图像集合进行聚类构建多层级不同非线性程度的局部模型。针对某个给定的非线性程度, 两个图像集之间的 MMD 距离就采用各图像集对应该分值的局部模型来计算。图示 3.11 给出了利用不同非线性程度的局部模型计算 MMD 所对应的识别性能曲线。需要注意的是, 如图 3.10 所示, 局部模型的个数直观反映了其非线性程度, 并且不同图像集合对应同一非线性程度的局部模型个数也不尽相同。为了进行对比, 与 HDC 算法类似, 同样利用 HAC 与 K-means 计算各图像集合的多层次局部模型表示, 并利用这些不同层级的局部模型计算 MMD。从图示 3.11 可以看出, 随着局部模型的个数增多(非线性分值随之下降), 三种聚类方法对应的性能均逐渐上升, 从而证明了本文采用局部线性模型来刻画图像集合流形的合理性和有效性。当算法中的局部模型达到一定数量时(通常小于 10), 局部模型的非线性程度变化开始趋于平缓, 相应地识别性能的提升也渐趋微弱。从图中对比曲线还可以进一步看出, HDC 和 HAC 算法性能要明显优于 K-means, 这主要是因为这两种方法采用流形上的测地距离来度量聚类之间的相似性, 从而再次验证了实验图像集合的非线性流形结构属性。

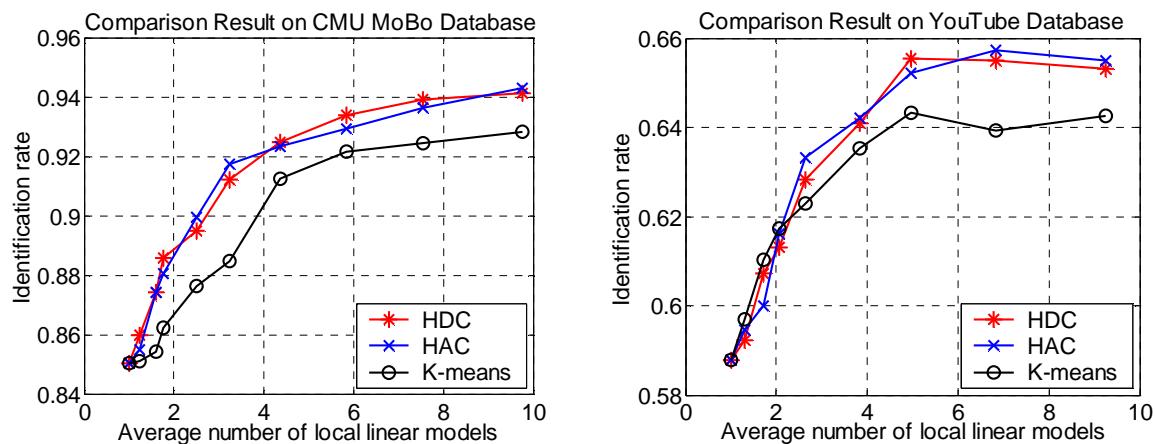


图 3.11 MMD 采用不同聚类方法建立局部模型的对比

表 3.3 详细对比了三种聚类算法的计算时间。实验中, 在 20 个各自包含 300 个样本的图像集合上, 分别将三个算法独立重复运行 20 次统计得到其时间平均值。需要注意的是, 在 HDC 算法的运行时间中包含了计算测地距离和欧氏距离的开销。从表中对比结果看出, HDC 算法最为高效, 特别是与 HAC 算法相比; 而在识别性能方面, 图示 3.11

中则显示出二者比较接近。这一对比结果表明了本文选择基于分裂的聚类算法进行局部模型构建的高效性。同时,从表 3.3 的结果还可以发现,在局部模型个数增量相同的情况下, HDC 与 HAC 算法的时间增量也接近,且都远小于 K-means 的时间增量。这主要是由于 HDC 和 HAC 都只需要将样本点对之间的距离计算一次,算法迭代过程的主要操作就是聚类的分裂或者融合操作。而 K-means 算法在其每次迭代过程中都需要计算所有样本与各聚类中心的距离。

表 3.3 不同聚类算法计算时间对比 (单位: 秒)

算法	局部模型数量				
	2	6	10	14	18
HDC	0.142	0.180	0.196	0.212	0.236
HAC	59.552	59.451	59.319	59.252	59.169
K-means	0.765	2.893	3.396	3.927	4.487

(2) 局部模型距离度量。本实验首先对比 3.3.2.2 节讨论的几种不同的子空间-子空间距离(SSD)度量方式,包括公式(3-14)的基于数据样例的 SSD,公式(3-13)的基于变化模式的 SSD,公式(3-15)的融合后的 SSD,分别记为“Exe.”,“Var.”和“Exe. + Var.”。此外,还将公式(15)的“Exe.”与公式(3-1)中的 PPD,记为“Exe.-ED”,进行了对比。图示 3.12 给出了不同度量方式的对比结果。从图中可以得出如下结论:第一,“Exe.”和“Var.”这两种单一的距离度量各有侧重,二者不存在哪一个绝对更优的问题。在 MoBo 库上,“Exe.”度量的识别率均值比“Var.”度量至少高 5 个百分点。这一对比现象同时也印证了上一节实验中在 MoBo 库上 MSM/CMSM/DCC 相对于 MMD 性能均偏低的结果,这三个对比算法均是仅采用了基于变化模式的 SSD 度量。在 YouTube 数据库上则不同,“Var.”度量的性能要略优于“Exe.”度量。第二,融合后的“Exe. + Var.”度量通常都比单一的度量有了性能的提升,特别是在 YouTube 数据库上的对比更加显著。这一结果表明,在实际应用中一种有效的方案就是将两种单一度量有机地融合起来,从而可以在不同数据集上取得更加稳定的结果。此外,通过对比两种基于数据样例的度量“Exe.”和“Exe.-ED”,可以发现它们的性能在各种情况下均非常接近,这也验证了本文采用公式(3-14)基于相关性的单样本距离度量取代公式(3-1)的 PPD 度量方式的可行性。

实验进一步评估了 MMD 算法采用不同数量主夹角时的性能变化趋势,并与 MSM 方法进行了对比。参考图示 3.12 两个子图中最右边的一列,实验中针对两个数据库分别将 MMD 中局部模型的个数固定为 7.54 和 6.84。MMD 的局部模型 PCA 子空间维数取值约为 5, MSM 的子空间维数约为 15,如前 3.5.2 节所述。图示 3.13 给出了相应的对比结果,其中“MMD-Exe.”/“MMD-Var.”/“MMD”分别对应 MMD 算法中局部模型之间的

SSD 距离度量采用“Exe.”/“Var.”/“Exe. + Var.”等不同方式。总的看来, MSM 算法的识别精度随着采用的主夹角个数变化而呈现出较大幅度的偏差, 这一点与文献[43]报告的结果一致。与之相比, MMD 的性能则受这一参数的影响较小, 特别是当把两种单一的 SSD 度量融合之后, MMD 的性能不仅得到了一致的提升而且也更加稳定。这主要归结于 MMD 采用一组局部模型来刻画图像集合流形, 可以在很大程度上抑制由于数据采集环境的类似而对图像集之间相似性度量所带来的影响。需要再次注意的是, 表 3.2 所报告的不同算法对比结果中, MSM 与 MMD 分别采用的是最佳数量的主夹角与全部主夹角。

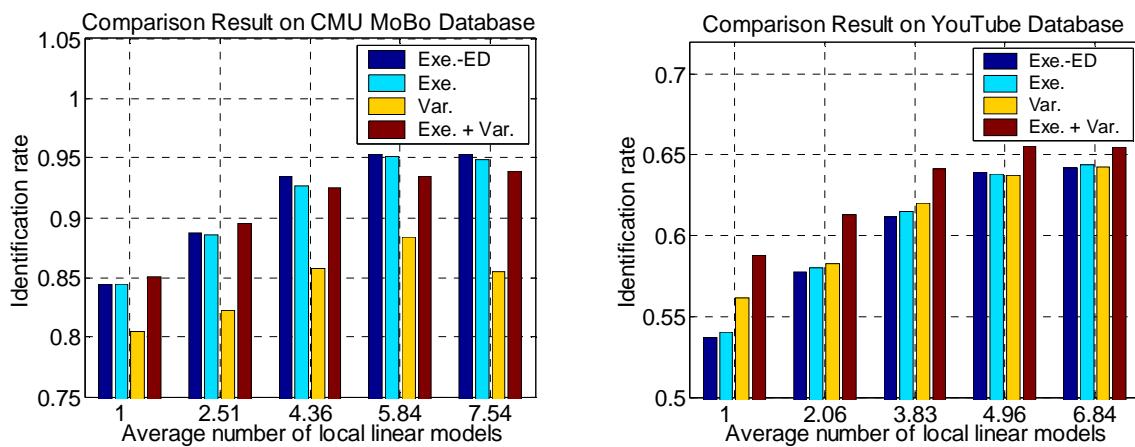


图 3.12 MMD 采用不同 SSD 距离度量的对比

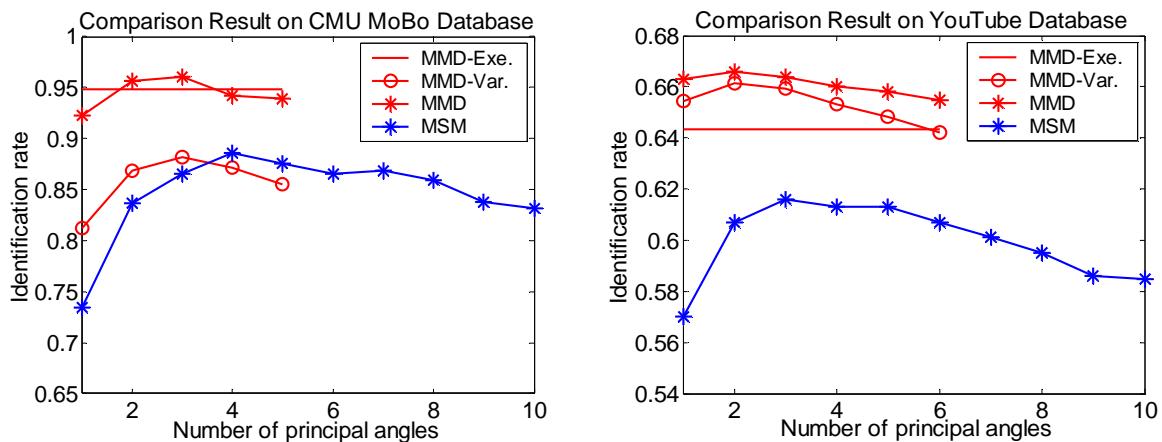


图 3.13 MMD 与 MSM 各自采用不同数量主夹角的性能对比

(3) 局部距离的全局综合。本实验中同样将 MoBo 库和 YouTube 库的局部模型个数分别固定为 7.54 和 6.84, 然后评估 MMD 算法的各种全局综合方案, 具体包括 3.3.3 节提出的方案-1~4 以及 3.3.3.5 节讨论的“Median NN”与“Max NN”方案, 后者即为 Hausdorff 距离。如前 3.3.3.4 节所述, EMD 距离方案中有两种权重选择, 即归一化权重与单位权重, 在实验中分别记为“EMD-Norm.”和“EMD-Unit”。图示 3.14 给出了各种方

案结合不同的 SSD 度量的对比结果。可以看出, 针对不同的局部距离 SSD 度量, 各种方案表现出基本一致的性能对比走势。具体来说, 通过方案-1~4 的对比发现, 方案-1(Min NN)与方案-3 (Mean N<sup>4</sup>)的性能通常优于方案-2(Mean NN)与方案-4(EMD), 其中以方案-3 性能最优, 而方案-4 性能最差。针对“Median NN”与“Max NN” (图例中记为 Hausdorff) 方案, 前者与方案-2 性能相近, 而后者则表现出了比其它所有方案都差得多的性能。这些实验对比结果与 3.3.3 节中的理论分析讨论完全一致。此外, 实验还发现基于归一化权重的“EMD-Norm.”性能通常优于基于单位权重的“EMD-Unit”, 这一点与文献[97]报告的结果类似。

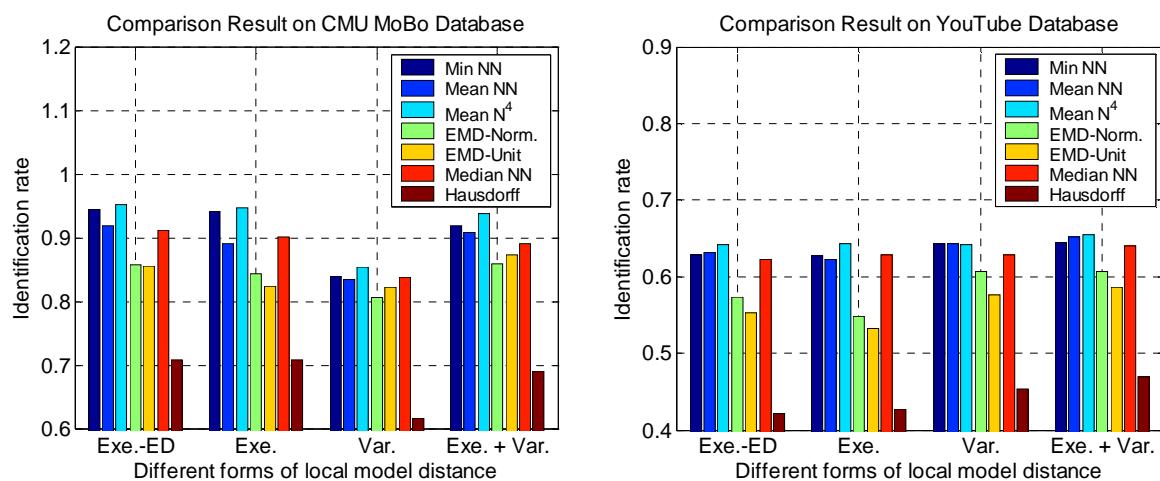


图 3.14 MMD 采用不同的局部距离全局综合方案的对比

### 3.6 本章小结

针对包含有复杂数据变化模式的图像集合进行比较和匹配是计算机视觉研究中的一个富有挑战性的问题。将每个图像集合建模表示为一个非线性流形, 图像集合的的匹配问题即可形式化为流形之间的距离度量问题。本章提出一种广义的流形-流形距离 (Manifold to Manifold Distance, MMD) 框架, 针对其实际计算问题, 探讨了多种解决方案。以基于图像集合的人脸识别问题为例, 本章通过系统深入的实验对提出的方法进行了验证。结果表明, 本文方法一致地优于其它对比的同类型方法, 特别是在性能上达到了与当前最优的基于集合的判别式学习方法同等可比的水平。

目前的 MMD 方法仅是作为一种通用的距离计算框架, 在后续研究工作中将考虑利用图像集合中的判别信息从而应用于监督式的分类学习问题。此外, 针对局部模型之间的距离度量, 如何挖掘利用不同主夹角的信息并且与基于数据样例的距离更加有效的融合也是值得深入研究的方向。

## 第四章 基于流形判别分析的图像集合人脸识别

### 4.1 引言

针对图像集合分类问题,上一章提出的流形-流形距离方法主要是从产生式的角度对问题进行形式化,将集合之间的关系转化为多流形之间的距离计算问题,并在所提出的距离框架内以无监督的方式来进行距离计算和流形匹配。一般来说,在分类问题中,基于训练数据的信息进行监督式学习通常更加有效。因此,本章主要从监督学习的角度出发,提出一种判别式学习方法,称为流形判别分析(Manifold Discriminant Analysis, MDA)。该方法以上一章的流形-流形距离为基本框架,通过学习判别式的嵌入空间以最大化不同类别流形之间的间隔同时增强流形内部的局部紧致性,从而进行更加稳定可靠的分类。

如前面章节所讨论的,当前流形学习方法主要仍集中于单流形的非线性降维问题。经典的方法如 ISOMAP [80], LLE [66], 和 Laplacian Eigenmaps [5] 主要解决流形数据的低维表示问题,而且主要是从无监督学习的角度建模单个流形数据集。由于这些方法均采用隐式的非线性映射函数,因此它们无法直接对新的测试数据进行处理。这一特点限制了这些方法在分类问题中的应用。针对测试数据的扩展问题,近年来提出了一些有效的方法可以在整个数据空间上建立映射函数,比如 LPP [32], UDP [101] 等。这些方法均采用了线性映射函数,计算也比较简单,在人脸识别、掌纹识别等问题中得到了应用。然而,这两个方法本质上仍是无监督式的,在很多实际问题中无法取得理想的效果[12]。

针对分类问题而言,线性判别分析(Linear Discriminant Analysis, LDA)是应用最为广泛的一种学习方法,在模式识别的很多实际问题中表现出了良好的性能。然而,作为一种参数式方法, LDA 的分类性能对数据的实际分布依赖性比较强,对于线性不可分的问题可能无法得到较好的效果。针对 LDA 的局限,很多方法对其进行了改进和扩展,比如非参数判别分析(Nonparametric Discriminant Analysis, NDA) [11], 子类判别分析(Subclass Discriminant Analysis, SDA) [107], 最大化间隔准则法(Maximum Margin Criterion, MMC) [54], 以及核判别分析(Kernel Fisher Discriminant, KFD) [60]。

近年来,涌现出一批新型的分类学习方法,它们的核心思想都是尝试从流形学习的角度来研究分类问题,相关工作包括:边界费舍尔分析(Marginal Fisher Analysis, MFA) [99], 局部判别嵌入(Local Discriminant Embedding, LDE) [12], 拉普拉斯支持向量机(Laplacian SVM, LapSVM) [6], 以及其它一些方法[77][100]。这些方法在传统的分类任务中表现出了良好的性能,然而类似于 LDA 的学习方式,它们的训练与测试框架仍是基于单个样本进行的。尽管数据的流形分布属性在这些方法中都有所反映和刻画,但它们很少从不同流形之间整体关系的角度来研究分类问题。换句话说,这些方法并不是特别针对集合分类问题而设计的。

本章主要以基于图像集合的物体识别问题为例，从判别分析的角度研究流形之间的匹配和分类，因此其问题出发点与上面所讨论的基于单样本的分类方法是不同的。接下来的内容安排如下：第2节简单介绍流形判别分析方法的基本思想；第3节详细介绍算法训练与测试阶段的各个环节，包括问题的形式化、局部模型的构建、判别函数的学习和测试算法流程；第4节讨论流形判别分析与其它相关方法的关系，第5节给出具体的实验对比结果，第6节对本章方法进行总结并对下一步研究方向进行讨论。

## 4.2 算法基本思想

在基于图像集合的物体识别问题中，每个图像集合通常包含属于同一类物体的多幅图像，这些图像涵盖了物体广泛的表观变化模式。识别的核心任务就是对未知类别的图像集合进行分类。由于可以利用图像集合的整体信息，因此基于集合输入的学习算法比基于单个样本输入的算法性能要更加稳定[43][89]。

在上一章介绍的流形-流形距离(MMD)方法中，采用流形来建模图像集合，并将图像集合分类问题形式化为流形之间的距离计算问题。然而，MMD方法本质上是一种无监督的方法，没有有效利用训练数据当中的判别信息，因此对于分类问题而言其识别性能还不是最优的。针对这一问题，本章提出一种有效的判别式学习方法，称为流形判别分析(Manifold Discriminant Analysis, MDA)。MDA方法的目标是最大化不同类别流形之间的间隔(maximize manifold margin)，同时增强每个流形内部局部区域的数据紧致性。

MDA方法的核心思想体现在如下两个方面：(1) 局部线性模型。类似于MMD方法中对流形的建模方式，MDA同样采用层次聚类方法将非线性表观流形表示为一组局部线性模型的集合，进而利用这些局部模型之间的关系来刻画流形之间的间隔。(2) 判别式学习。受线性判别分析方法LDA与最大化间隔准则方法MMC的启发，MDA通过学习一个线性判别投影函数将不同类别的多个流形投影到一个判别式的嵌入空间中。在该嵌入空间中，不同类别的流形所对应的局部模型之间能够得到更好的区分，同时每个局部模型内部的邻域关系得到保持从而保证局部数据的紧致性。在识别阶段，将待匹配的流形通过学习所得的投影函数变换到该嵌入空间中，就可以使得分类结果更加可靠。

## 4.3 流形判别分析

本节首先对图像集合的分类问题进行形式化，将其转化为面向分类任务的多流形学习的理论问题。之后，介绍流形判别分析方法的两个核心，即局部线性模型与判别式学习。最后，给出识别分类阶段的测试方法流程。

### 4.3.1 问题形式化

如上所述，MDA方法将图像集合建模为流形，进而将基于图像集合的物体识别问题形式化为面向分类任务的多流形学习问题。形式上，假定有如下 $M$ 个图像集合：

$\{X_1, X_2, \dots, X_M\}$ , 其中  $X_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,Ni}]$  ( $i = 1, 2, \dots, M$ ) 为第  $i$  个集合对应的图像数据矩阵,  $Ni$  为该集合中图像样本的个数。每个图像集合属于一种物体类别, 且不同集合可能属于同一类别, 集合  $X_i$  对应的类别记为  $\{L_i \mid L_i \in \{1, 2, \dots, P\}\}_{i=1}^M$ ,  $P$  为总的类别个数。将每个图像集合  $X_i$  建模为一个非线性流形  $\mathcal{M}_i$ , 并采用一组局部线性模型来刻画该流形:  $\mathcal{M}_i = \{C_{i,1}, C_{i,2}, \dots, C_{i,n_i}\}$ 。其中,  $n_i$  为第  $i$  个图像集合中局部模型的个数, 在大多数情况下,  $n_i$  远远小于  $Ni$ 。

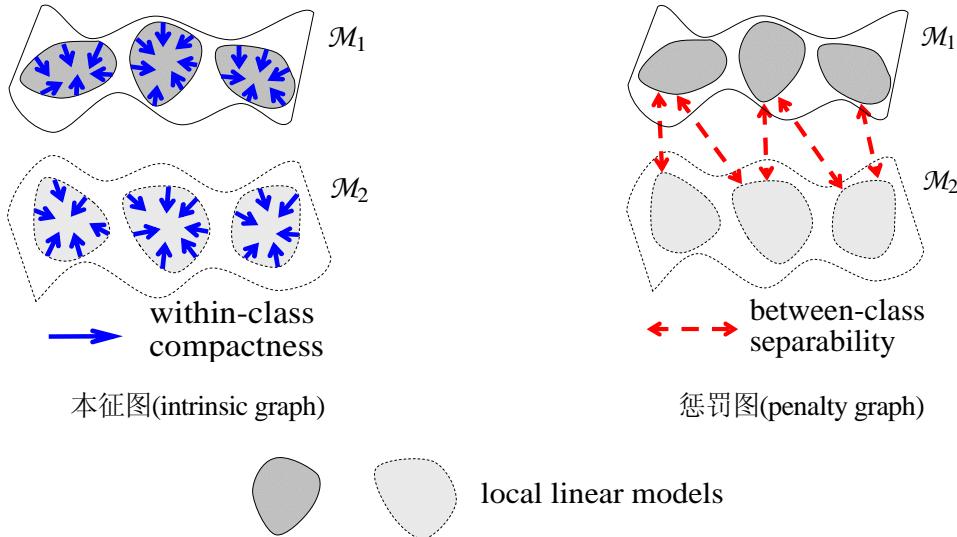


图 4.1 MDA 方法采用的图嵌入框架示意图

基于上述假设, 多流形学习问题可以转化为学习判别式嵌入空间以更好区分不同类别流形的问题。换言之, 学习的目标是最大化不同类别流形之间的间隔。由于在流形学习领域中, 关于不同流形之间的间隔尚未有相关的工作提及, 本文尝试从基于局部模型表示的流形建模方式出发, 将两个全局非线性流形之间的间隔转化为它们各自对应的局部线性模型之间的间隔。可以看出, 这种转化方式与上一章的 MMD 方法是一致的。

直观上, MDA 方法尝试学习一个线性判别投影函数, 以最大化不同类别流形之间的可分性, 同时增强每个流形类内的局部数据紧致性。根据图嵌入框架(graph embedding) [99]的思想, 流形之间的上述类内和类间优化目标可以通过设计两个图来分别刻画。具体而言, 通过本征图(intrinsic graph)来刻画流形的类内紧致性, 通过惩罚图(penalty graph)来刻画流形的类间可分性。如图 4.1 所示, 图中  $\mathcal{M}_1$  和  $\mathcal{M}_2$  为两个属于不同类别的流形。在本征图中, 每个局部模型内部的数据样本向模型的中心压缩, 从而增强类内的紧致性 (图中蓝色箭头示意); 在惩罚图中, 不同类别流形对应的近邻局部模型分别向相反的方向分散, 从而增加类间可分性 (图中红色箭头示意)。

### 4.3.2 局部线性模型构建

MDA 方法中采用的局部线性模型建立算法与 MMD 方法类似。形式上，采用上一小节的变量表示方式，给定流形  $\mathcal{M}_i$  与其数据集  $X_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,Ni}]$ ，局部模型构建的目标是将原始数据集  $X_i$  划分为一组互不相交的最大线性分块(MLP)  $C_{i,k}$  的组合，即

$$\mathcal{M}_i = \{C_{i,1}, C_{i,2}, \dots, C_{i,ni}\}, \text{ 其中 } C_{i,k} |_{k=1}^{ni} = \{\mathbf{x}_{i,1}^{(k)}, \mathbf{x}_{i,2}^{(k)}, \dots, \mathbf{x}_{i,a_k}^{(k)}\}, (\sum_{k=1}^{ni} a_k = Ni). \quad (4-1)$$

类似于公式(3-7)，公式(4-1)与公式(2-1)本质也是相同的，其中分块  $C_{i,k}$  的构建即通过 2.3.1.1 小节介绍的 HDC 算法来完成。算法最终得到  $ni$  个互不相交的 MLP 分块。

### 4.3.3 判别函数的学习

利用上述得到的局部线性模型，可以依据一定的准则来构建图示 4.1 中分别刻画流形类内与类间关系的两个图。针对刻画类内关系的本征图，每个局部模型内部的样本之间应该互相建立边连接关系。针对刻画类间关系的惩罚图，则需要在两个不同类别流形的近邻局部模型之间建立边连接关系。因此，如何度量局部模型之间的远近距离关系就成为建立图模型的一个前提条件。很显然，这里的局部模型距离度量就是上一章 MMD 方法中所讨论的子空间-子空间距离。出于计算简便的考虑，在 MDA 方法中仅利用基于数据样例的距离，即利用两个局部模型各自的样本中心来计算距离。

具体地，给定两个流形  $\mathcal{M}_i = \{C_{i,1}, C_{i,2}, \dots, C_{i,ni}\}$ ， $\mathcal{M}_j = \{C_{j,1}, C_{j,2}, \dots, C_{j,nj}\}$ ，其中  $C_{i,k} |_{k=1}^{ni} = \{\mathbf{x}_{i,1}^{(k)}, \mathbf{x}_{i,2}^{(k)}, \dots, \mathbf{x}_{i,a_k}^{(k)}\}$ ， $C_{j,l} |_{l=1}^{nj} = \{\mathbf{x}_{j,1}^{(l)}, \mathbf{x}_{j,2}^{(l)}, \dots, \mathbf{x}_{j,a_l}^{(l)}\}$ ，两个局部模型之间的距离计算公式如下：

$$d(C_{i,k}, C_{j,l}) = \|\mathbf{e}_{i,k} - \mathbf{e}_{j,l}\|, \text{ 其中 } \mathbf{e}_{i,k} = \frac{1}{a_k} \sum_{m=1}^{a_k} \mathbf{x}_{i,m}^{(k)}, \mathbf{e}_{j,l} = \frac{1}{a_l} \sum_{n=1}^{a_l} \mathbf{x}_{j,n}^{(l)}. \quad (4-2)$$

接下来，将不同流形对应的数据集综合起来，可以得到如下训练数据矩阵：  
 $X = [X_1, X_2, \dots, X_M] = [\mathbf{x}_{1,1}, \mathbf{x}_{1,2}, \dots, \mathbf{x}_{1,N1}; \dots; \mathbf{x}_{M,1}, \mathbf{x}_{M,2}, \dots, \mathbf{x}_{M,NM}]$ 。为表示方便，将所有样本的序号统一排列，从而得到  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ ，其中  $\mathbf{x}_i \in \mathbb{R}^d$  ( $i = 1, 2, \dots, N$ )，  
 $N = \sum_{i=1}^M Ni$  表示所有训练样本的个数。MDA 的训练算法可以形式化为如下三步：

(1) 建立图模型。采用  $G$  与  $G'$  分别标记本征图与惩罚图，其中图模型的顶点为所有训练样本。考虑每一对样本点  $\mathbf{x}_m$  与  $\mathbf{x}_n$ ，各自对应的局部模型为  $\mathbf{x}_m \in C_{i,k}$ ， $\mathbf{x}_n \in C_{j,l}$ 。如上所述，在本征图  $G$  中，属于同一个局部模型的样本点之间应该建立边连接，即如果同时满足  $i = j$  且  $k = l$ ，则在  $\mathbf{x}_m$  与  $\mathbf{x}_n$  之间建立边。针对惩罚图  $G'$ ，则需要在不同类别的近邻局部模型对所对应的样本间建立边连接，即如果同时满足  $L_i \neq L_j$  (二者分别为  $\mathbf{x}_m$  与  $\mathbf{x}_n$  所属的类别) 且  $C_{i,k}$  与  $C_{j,l}$  是对方的  $k'$ -近邻，则在  $\mathbf{x}_m$  与  $\mathbf{x}_n$  之间建立边。根据上述定义，当同一类别包含多个训练流形时，它们之间的近邻局部模型对并不会在惩罚图中建

立边连接关系。形式上，采用  $g_{m,n}$  与  $g'_{m,n}$  分别标记样本点  $\mathbf{x}_m$  和  $\mathbf{x}_n$  在图  $G$  与  $G'$  中的连接关系，采用  $N_{k'}(\cdot)$  表示局部模型之间的  $k'$ -近邻关系，则图中边邻接关系的定义如下式：

$$g_{m,n} = \begin{cases} 1 & \text{if } i = j, k = l \\ 0 & \text{else} \end{cases}, \quad (4-3)$$

$$g'_{m,n} = \begin{cases} 1 & \text{if } L_i \neq L_j, C_{i,k} \in N_{k'}(C_{j,l}) \text{ or } C_{j,l} \in N_{k'}(C_{i,k}) \\ 0 & \text{else} \end{cases}.$$

(2) 计算相似度矩阵。将本征图  $G$  对应的相似度矩阵记为  $W$ ，矩阵的元素  $w_{m,n}$  表示顶点  $\mathbf{x}_m$  与  $\mathbf{x}_n$  之间的边对应的权重。该权重可以定义为：如果  $\mathbf{x}_m$  与  $\mathbf{x}_n$  之间有边连接，则  $w_{m,n} = 1$ ，否则  $w_{m,n} = 0$ 。类似地，惩罚图  $G'$  对应的相似度矩阵  $W'$  也以同样的方式计算边上的权重。可以看出，矩阵  $W$  与  $W'$  均为对称阵。这里选用的权重定义是图模型当中经典的“simple-minded”方式[5]，另外一种常用的权重方式为“heat kernel”，定义为：如果  $\mathbf{x}_m$  与  $\mathbf{x}_n$  之间有边连接，则  $w_{m,n} = \exp[-\|\mathbf{x}_m - \mathbf{x}_n\|^2 / t]$ ，否则  $w_{m,n} = 0$ 。

(3) 计算判别式嵌入空间。针对判别嵌入空间的选择，本文目前采用的是学习简单的线性投影函数，即采用  $\mathbf{z} = \mathbf{V}^T \mathbf{x}$  的方式将原始样本通过线性变换映射到嵌入空间中。其中， $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l]$  是  $d \times l$  的线性变换矩阵， $d$  和  $l$  分别是原始空间和嵌入空间的维数，在多数情况下  $l \ll d$ 。根据上面计算的相似度矩阵，可以分别定义类内散度  $S_w$  和类间散度  $S_b$ ，二者分别刻画类内数据的紧致性和类间数据的可分性，具体形式如下：

$$S_w = \sum_{m,n} \|\mathbf{v}^T \mathbf{x}_m - \mathbf{v}^T \mathbf{x}_n\|^2 w_{m,n} = 2\mathbf{v}^T \mathbf{X} (\mathbf{D} - \mathbf{W}) \mathbf{X}^T \mathbf{v}, \quad (4-4)$$

$$S_b = \sum_{m,n} \|\mathbf{v}^T \mathbf{x}_m - \mathbf{v}^T \mathbf{x}_n\|^2 w'_{m,n} = 2\mathbf{v}^T \mathbf{X} (\mathbf{D}' - \mathbf{W}') \mathbf{X}^T \mathbf{v}. \quad (4-5)$$

上式中， $\mathbf{D}$  和  $\mathbf{D}'$  均为对角矩阵，其中  $d_{m,m} = \sum_n w_{m,n}$ ， $d'_{m,m} = \sum_n w'_{m,n}$ 。在谱图理论中[13]， $\mathbf{L}_w = \mathbf{D} - \mathbf{W}$  和  $\mathbf{L}_b = \mathbf{D}' - \mathbf{W}'$  分别为本征图  $G$  和惩罚图  $G'$  的拉普拉斯矩阵。回顾 4.2 节所述 MDA 的目标，即最大化类间可分性且同时增强类内紧致性，可以将其优化的目标函数表示为最大化上面类间散度与类内散度的商：

$$\text{Maximize} \quad J(\mathbf{v}) = \frac{|S_b|}{|S_w|} = \frac{\mathbf{v}^T \mathbf{X} \mathbf{L}_b \mathbf{X}^T \mathbf{v}}{\mathbf{v}^T \mathbf{X} \mathbf{L}_w \mathbf{X}^T \mathbf{v}}. \quad (4-6)$$

通过简单的代数运算，可以将上述目标函数转化为下式(4-7)中的广义特征值分解问题，其中，最优线性变换矩阵  $\mathbf{V}$  的前  $l$  列对应于最大的  $l$  个广义特征向量：

$$\mathbf{X} \mathbf{L}_b \mathbf{X}^T \mathbf{v} = \lambda \mathbf{X} \mathbf{L}_w \mathbf{X}^T \mathbf{v}. \quad (4-7)$$

### 4.3.4 基于流形判别分析的分类

采用训练阶段学习得到的最优变换矩阵  $V$ ，就可以在判别式嵌入空间中完成多流形的分类任务。在该判别式空间中，属于不同类别的流形之间的间隔变得更大，同时，每个流形局部线性模型也变得更为紧致，从而模型内部样本的中心可以更好地代表整个局部模型。

MDA 方法的测试流程可以进行如下形式化表述。给定两个待比较的图像集合，对应的流形表示为  $\mathcal{M}_i$  和  $\mathcal{M}_j$ 。类似于上一章中 MMD 的做法，图像集合的分类问题形式化为流形之间的距离计算问题。MDA 方法采用的距离计算公式如下：

$$d(\mathcal{M}_i, \mathcal{M}_j) = \min_{C_{i,k} \in \mathcal{M}_i} \min_{C_{j,l} \in \mathcal{M}_j} d_Z(C_{i,k}, C_{j,l}), \text{ 其中} \quad (4-8)$$

$$d_Z(C_{i,k}, C_{j,l}) = \|V^T \mathbf{e}_{i,k} - V^T \mathbf{e}_{j,l}\| = \|V^T \cdot (\mathbf{e}_{i,k} - \mathbf{e}_{j,l})\|.$$

具体地，首先采用表 2.1 中的 HDC 算法提取流形的局部模型  $C_{i,k}$  和  $C_{j,l}$ 。接下来需要计算局部模型之间在嵌入空间中对应的距离  $d_Z(C_{i,k}, C_{j,l})$ 。由于在原始空间中，局部模型之间的距离  $d(C_{i,k}, C_{j,l})$  采用了公式(4-2)所述的基于数据样例的距离度量方式。为了保持一致性，在嵌入空间中同样采用这种方式，即需要计算各局部模型的样本在嵌入空间中的均值，进而以这些均值之间的距离作为局部模型的距离。由于 MDA 采用的是线性判别投影函数，因此只需要将公式(4-2)中局部模型在原始空间中的均值通过变换矩阵投影即可得到嵌入空间中的局部模型样本均值。最后，两个流形之间的距离采用其最近邻的局部模型对之间的距离来度量，即 MMD 框架中“局部距离的全局综合”所采用的方案-1。经过流形在判别式空间的距离计算，就可以依据最小的流形-流形距离来确定测试图像集合的类别。综合上述分析可以看出，流形判别分析 MDA 的测试流程非常高效。

## 4.4 讨论

针对本文 MDA 方法与其它相关研究工作的关系，本节分别从基于单样本和基于集合的分类方法两个方面展开讨论。

### 4.4.1 与基于单样本的方法对比

对比方法：MFA [99], LDE [12], KFD [60]。这三个方法都是基于单样本的分类方法，在它们的分类学习框架中，集合的信息很少被有效利用，因此，它们并不是针对特定的集合分类任务设计的。如文献[101]所分析，MFA 与 LDE 本质上是相同的，二者都可以看作是针对 LDA 方法的一些局限，在散度矩阵局部化与数据分布假设非参数化两个方面所进行的改进和扩展。MDA 与这两个方法的共同点在于，它们均是从图嵌入框架[99]的思想出发来建模分类问题的，通过构建本征图与惩罚图的方式来分别刻画类内紧致性与类间可分性这两个基本属性。所不同的是，MDA 方法本质是基于从流形上提取得到

的局部模型来进行学习和分类的，无论是在训练阶段构建图模型还是在测试阶段度量集合的相似度。而 MFA 与 LDE 中的图模型都是基于单样本之间的邻接关系建立的，其测试框架也是基于单样本的匹配和相似度计算设计的。此外，MDA 与 KFD [60]相比，它们的共同点是均为基于数据分布非线性的假设来设计分类器。不同的是，KFD 作为 LDA 方法的非线性扩展，是从核方法的思想出发，将数据映射到隐式的高维特征空间进行线性分类。而 MDA 则是采用显式学习的方式得到低维嵌入空间，将原始数据投影到该空间中进行多个非线性流形的分类。

#### 4.4.2 与基于集合的方法对比

对比方法: MMD [89], DCC [43]。与 MDA 相同, MMD 和 DCC (Discriminant-analysis of Canonical Correlations)都是针对图像集合的分类问题所设计的分类方法。如 4.2 节所讨论, MMD 方法主要是针对流形之间的距离计算提出一种框架性的解决方案, 作为一种无监督方法, MMD 所涉及的距离计算和匹配都是在数据的原始空间进行的。DCC 方法则是基于监督学习的思想, 利用训练数据学习判别式的投影函数, 在其学习框架中数据之间的相似度采用典型相关分析(canonical correlation analysis, CCA) [9]来计算。从分类思想的角度来看, MDA 可以看作是在一个判别式的特征空间中采用 MMD 的距离计算框架来进行多流形的匹配, 而其判别式空间的学习则与 DCC 类似, 二者均是基于 LDA 的“最大化类间散度同时最小化类内散度”的思想。与 DCC 方法不同的是, MDA 采用相对更一般的流形来建模图像集合, 而 DCC 则采用线性子空间的表示。当图像集合包含尺度较大的表观变化时, 后者的建模方式就显得相对简单。此外, DCC 在学习判别空间时采用的是迭代优化的方式, 而 MDA 则采用特征分解的方法求解闭形式的最优解。

### 4.5 实验结果

为了对所提出的方法进行验证, 实验中同样以基于图像集合的物体识别问题为例, 在人脸识别和物体分类两个任务上, 将 MDA 与相关的经典方法进行对比。具体地, 这两个任务均被看作多流形的分类问题, 每个已知类别的图像集合建模为 *gallery* 流形, 每个待测试类别的图像集合则被建模为 *probe* 流形, 采用 4.3.4 节介绍的测试流程完成分类。

#### 4.5.1 实验数据库

针对人脸识别任务, 实验选取了与上一章 MMD 对比实验中相同的数据库, 即两个基准数据库: Honda/UCSD [51], CMU MoBo [26], 与一个富有挑战性的集合: YouTube Celebrities [45]。针对物体分类任务, 实验选取了基准数据库 ETH-80 [52]。由于三个人脸数据库在上一章已有详细介绍, 接下来仅对物体分类数据库 ETH-80 进行介绍。

**ETH-80:** 该数据库包含 8 个不同物体类别, 每个类别 10 个不同的个体, 如图 4.2 所示, 图中每一行为一个物体类别对应的 10 个个体。其中, 每个个体对应于一个图像集合, 集合中包含了该个体不同视角下的 41 张图像。物体分类的任务是将未知类别的一个

图像集合划分到所对应的物体类别。与人脸识别任务类比，可以将 8 个物体类别看作 8 个不同的人，而每类的 10 个个体则看作每个人对应的 10 个图像集合。实验中将原始图像下采样并作灰度化处理，得到  $30 \times 30$  的灰度图像，用于后续的算法训练和测试。



图 4.2 物体分类数据库 ETH-80 的样本示意图

实验中针对各数据库的训练集/测试集的划分方案与上一章 3.5.1 节的设置基本相同。针对每个数据库，分别进行十组交叉验证，每组实验对应于一组随机选择的训练集与测试集标号组合。具体地，针对 Honda 和 MoBo 库的划分与 3.5.1 节完全相同。针对 YouTube 库，同样在每组实验中从每个人的 3 个小节(session)中随机选用共 9 段视频进行实验，其中每个小节选取 3 段视频。这 9 段视频中的 3 段作为训练集，6 段作为测试集。所不同的是，本章的实验不再考虑训练集和测试集所属的小节是否存在交叠，即不再细分 YouTube\_Easy 与 YouTube\_Hard 这两种测试场景。针对 ETH-80 数据库，每组实验对应地从每类物体中随机选择 5 个个体作为训练集，另外 5 个个体作为测试集。

### 4.5.2 对比方法与参数设置

根据 4.4 节的讨论，实验中选取以下两类方法进行对比：

- 基于单样本的方法
  1. 基于最近邻匹配的 LDA 方法，即 Fisherface [4]，作为单样本方法的基准；
  2. Kernel Fisher Discriminant (KFD) [60]；
  3. Marginal Fisher Analysis (MFA) [99]；
- 基于集合的方法

4. Mutual Subspace Method (MSM) [98], 作为集合方法的基准;
5. Manifold-Manifold Distance (MMD);
6. Discriminant-analysis of Canonical Correlations (DCC) [43];
7. 本文提出的 MDA 方法。

类似于 3.5.2 小节的实验设置, 为了保证对比实验的公正性, 实验中通过在训练集上的交叉验证来为各对比方法调节选择最优的参数组合。针对基于单样本的方法(LDA, MFA, KFD), 为了避免在计算特征分解问题时遇到的矩阵奇异问题, 参照文献[4][99]中的做法, 采用 PCA 对原始图像数据进行预处理, 在降维过程中保留 95% 的数据能量。KFD 方法的核函数选择通用的高斯核。LDA 与 KFD 的维数均设置为类别数减一, MFA 的维数一般选择范围在[50, 70]区间内。按照与前人工作中相同的测试流程, 三个方法均采用多数投票的策略确定整个 *probe* 集合的类别。针对基于集合的方法(MSM, MMD, DCC), 实验方案与 3.5.2 小节的设置完全相同。

针对 MDA 方法, 其主要的参数包括: (i) 公式(4-1)中的局部模型个数  $ni$ , 该参数的设置与 MMD 方法完全相同; (ii) 算法训练阶段构建惩罚图时采用的类间  $k$ -NN 局部模型的个数, 即公式(4-3)中的  $k'$ ; (iii) 算法测试阶段所选用的嵌入空间的维数  $l$ 。表 4.1 给出了四个数据库上各参数的典型取值。

表 4.1 MDA 方法的参数设置

数据库	算法参数		
	$ni$	$k'$	$l$
Honda/UCSD	6	3	70
CMU MoBo	8	3	50
YouTube	9	10	70
ETH-80	2	5	10

表 4.2 MDA 对比方法的识别结果

数据库	各对比方法 10 组随机实验的识别率均值与标准差						
	LDA	KFD	MFA	MSM	MMD	DCC	MDA
Honda/UCSD	$0.789 \pm 0.01$	$0.815 \pm 0.01$	$0.838 \pm 0.01$	$0.923 \pm 0.04$	$0.971 \pm 0.02$	$0.980 \pm 0.01$	$1.000 \pm 0.00$
CMU MoBo	$0.885 \pm 0.01$	$0.898 \pm 0.03$	$0.885 \pm 0.02$	$0.886 \pm 0.03$	$0.935 \pm 0.02$	$0.903 \pm 0.05$	$0.965 \pm 0.02$
YouTube	$0.573 \pm 0.03$	$0.607 \pm 0.01$	$0.594 \pm 0.02$	$0.595 \pm 0.04$	$0.640 \pm 0.04$	$0.667 \pm 0.04$	$0.672 \pm 0.04$
ETH-80	$0.673 \pm 0.02$	$0.811 \pm 0.02$	$0.801 \pm 0.02$	$0.833 \pm 0.04$	$0.850 \pm 0.07$	$0.908 \pm 0.05$	$0.890 \pm 0.02$

### 4.5.3 实验对比结果与分析

表 4.2 给出了实验对比的不同方法在四个数据库上的结果，其中识别率的均值与标准差是十次实验统计所得。总的看来，MDA 方法在绝大多数对比中显示出最优的性能。从表中报告的结果可以得出如下的一些分析和结论：

(1) 类似于 3.5.3 节的分析，基于单样本的方法通常比基于集合的方法在识别性能上要差一些。主要的原因还是由于这类方法无法有效挖掘和利用图像集合提供的整体信息，因此并不特别适用于图像集合的分类任务。

(2) 基于集合的四个方法由于其各自的特点，在实验中分别表现出不同的性能。MSM 和 MMD 都是在原始空间中进行图像集合的匹配，由于它们没有利用数据中的判别信息，因此其性能明显不如判别式的 DCC 和 MDA。此外，MSM 的识别率与 MMD 相比平均相差 5 个百分点左右，这主要是由于前者采用线性子空间的方式来建模图像集合。这一原因同样可以在一定程度上解释 MDA 性能优于 DCC 的实验结果。

(3) 综合对比不同方法在四个数据库上的性能，可以看出所有方法在 YouTube 库的识别率都是最低的。这是由该数据库的特点所决定的，大范围的姿态变化、低质量的图像、广泛的数据来源等等，都大大增加了识别问题的难度。在该数据库上，本文提出的 MDA 给出了所有方法中最高的识别率 67.2%，而在原文献[45]中，作者所提出的方法报告的结果是 71.24%。事实上，文献[45]的方法是基于视频进行识别的，其中利用了相关的跟踪技术和视频帧间的运动动态信息。此外，[45]的作者所采用的实验设置方案与本文的实验方案也不尽相同。

(4) 针对 ETH-80 库的对比结果，值得指出的是，本文实验中 MSM 和 DCC 的识别性能与文献[43]报告的结果是相当的。DCC 在这个数据库上给出了最高的识别率，而本文 MDA 方法识别率则略低一些，这主要是由该库特定的数据构成所决定的。如前 4.5.1 节所介绍的，该数据库每类物体的每个个体（即每个图像集合）仅包含 41 幅图像，数据的稀疏性在一定程度上使得 MDA 方法中流形分布的假设不能得到很好满足。尽管如此，与剩余其它对比方法相比，MDA 仍然具有识别性能的优势。

## 4.6 本章小结

针对图像集合的分类任务，本章将其形式化为面向分类的多流形学习问题，并提出一种有效的判别式学习方法，称为流形判别分析(MDA)。基于最大化流形间隔的思想，该方法在理论上具有良好的属性：通过构建图模型的方式，可以有效地刻画多流形之间的类间可分性和类内紧致性关系；基于其优化目标函数，可以通过广义特征分解解析地求得闭形式的最优解。算法的有效性通过系统的实验对比进行了验证，在人脸识别与物体分类两个任务上，算法表现出良好的性能。与同类型的基于集合的分类方法相比，本文方法取得了与当前最优方法相当甚至更优的性能，从而表明其可以有效地解决图像集合的分类问题。

针对算法的进一步研究，其中一个可能的改进方案是将算法进行核化扩展，通过嵌入核函数的方式，使得算法可以利用数据中更高阶的判别信息。针对算法中采用的子空间距离度量，根据上一章“流形-流形距离”的分析讨论，综合利用子空间样本中心与变化模式的距离应该是更合理而全面的距离度量，因此在 MDA 框架中融合 DCC 的思想也是值得尝试的一个改进方向。进一步地，将当前方法的训练方式进行在线式的扩展，从而应用于增量式学习问题，同样具有很重要的理论意义和实用价值。

回到最初的流形分类问题，目前的流形判别分析方法分别从最大化异类流形间的可分性且同时保持同类流形内的紧致性两方面来设计目标函数。在实际问题中，数据分布的原始流形结构可能并不适合分类，此时如果出于保持流形本质结构的目的去设计目标函数则可能对于最终的分类目的并无实际意义。反过来，将原始流形基于训练类别信息进行一定的扭曲变形则可能更有利于区分不同类别的流形。这一思路同样值得进行进一步的理论探讨和实验分析。



## 第五章 结束语

### 5.1 本文工作总结

流形学习作为一种有效的数据统计分析方法，在机器学习、模式识别和计算机视觉等多个领域中具有重要的理论意义和应用价值。特别是针对人脸识别这一具体的视觉任务而言，人脸图像集合所普遍具有的高数据维数、高度非线性分布等特点，为流形学习方法的应用研究提供了广阔的平台。近年来研究者们在推进流形学习方法理论研究的同时，针对其在人脸识别等相关计算机视觉问题中的实用化也进行了很多积极的探索。

本文正是在这样的背景下，针对流形学习方法和基于图像集合的人脸识别应用，从理论研究和应用研究两个层面开展工作。在理论研究层面，本文主要从流形学习的角度研究新型高效的非线性降维方法，特别是在流形的局部线性模型表示、解析映射函数学习、内在变化模式刻画等方面开展相关研究。在应用研究层面，本文针对基于图像集合的人脸识别问题，主要从图像集合的流形建模、流形之间的距离定义和计算、多流形判别式分类学习等几个角度开展研究。本文旨在推动流形学习算法理论与应用研究，特别是在扩展流形学习在人脸识别问题中的应用前景方面做了一些有益的尝试。本文取得的主要贡献如下：

(1) 提出了一种最大线性嵌入流形学习方法，用以解决非线性降维问题。该方法通过学习一个参数式可逆的解析映射函数，可以将流形数据从原始高维观察空间保距地投影到一个全局低维嵌入空间中。算法从几何直觉的角度，引入一种合理有效的局部线性分块的定义，称为最大线性分块，分块的非线性程度通过块内样本间测地距离与欧氏距离的偏差来度量。首先，原始的流形采样数据集合被分解为一组最大线性分块的组合，每个分块采用一个局部线性模型来建模表示。接下来，算法从每个局部模型内部随机选取一定数量的地地标点，利用其测地距离进行多维尺度变换运算，进而得到最终的保距低维嵌入空间并完成局部模型的全局对齐。这一对齐过程称为基于地地标点的全局对齐算法。该对齐算法避免了传统算法中存在的局部极值与大规模迭代优化这两个不足，并且只需要求解小规模的特征分解问题就能得到有效的闭形式解。针对流形内在结构的描述，本文方法不仅提供了一种快速有效的本质维数估计方法，而且可以显式建模流形观察数据的内在变化模式，从而可以广泛应用于多种实际问题。

(2) 提出了流形-流形距离的理论框架，将其应用于基于图像集合的人脸识别问题。在该理论框架中，属于每一类别的图像集合采用非线性表观流形来建模，图像集合的分类问题从而形式化为两个流形之间的距离计算问题，即流形-流形距离。注意到实际问题中的图像集合可能表示为三种模式层次，即点、子空间、流形，本文系统研究了这三种模式之间的各种距离，并将它们形式化在一个统一的多层次流形-流形距离框架下。具体

地, 将流形表示为一组局部线性模型的组合, 其中各局部模型采用子空间来刻画。这样, 两个流形之间的距离就可以转化为它们各自对应的子空间对之间的距离综合。针对流形-流形距离的计算, 本文从理论上和实验上探讨了算法各个模块的多种不同的定义方案, 包括局部线性模型构建、局部模型距离度量、局部距离的全局综合等。通过图像集合人脸识别的实验验证表明, 本文提出的流形-流形距离作为一种通用的图像集合相似性度量, 一致地优于其它对比的非判别式方法, 并且达到了与当前领先的基于集合的判别式学习方法可比的性能。

(3) 提出了一种流形判别分析方法, 将其应用于基于图像集合的人脸识别问题。该方法从监督学习的角度, 将图像集合分类问题形式化为面向分类的多流形学习问题, 通过在流形-流形距离框架内引入“最大化流形间隔”的思想来学习判别式的特征空间, 最终在该空间内对不同类别的流形进行匹配来完成图像集合的分类。具体地, 流形判别分析方法采用图嵌入的框架来构建图模型, 采用本征图和惩罚图分别刻画流形的类内紧致性与类间可分性。通过设计的最优化目标函数来学习判别式的嵌入空间, 使得具有不同类别标号的流形之间能够更好地区分, 同时每个流形局部区域的数据紧致性得到增强。在分类识别阶段, 通过将待测试的流形数据投影到该判别式空间中, 就可以进行更加可靠的分类。通过人脸识别和物体分类两个任务的实验对比表明, 本文提出的流形判别分析方法取得了与当前领先的方法相当的性能, 同时具有很高的训练和测试效率。

## 5.2 下一步研究方向

本文针对流形学习方法及其在人脸识别中的应用进行了深入而细致的研究工作, 并取得了一定的研究成果。然而, 流形学习算法和图像集合人脸识别应用作为机器学习、模式识别和计算机视觉领域内新兴的研究课题, 其理论研究和实用化推广还有很多问题需要解决。本文所做的工作仅仅是针对这两个问题分别从流形的局部模型构建与全局对齐、图像集合的流形建模和匹配的角度进行了相应的探索。下一步的研究工作可以围绕以下几个方面来开展:

1. 针对最大线性嵌入流形学习方法, 主要考虑在其局部模型拟合环节中引入概率模型, 从而将算法的映射过程以概率过程来刻画, 以期带来更加稳定与灵活的降维投影函数。同时, 算法采用的保距投影约束可以进行适当放松, 可以参考相关文献在严格的保距约束与完全丢失全局尺度信息之间做一定的平衡。此外, 后续工作中需要着重考虑流形学习算法普遍面临的两个问题, 即噪声数据和采样密度对算法结果稳定性的影响。这两个问题也是困扰当前流形学习算法走向更加实用化的核心障碍。

2. 针对流形-流形距离框架, 主要考虑改进其局部模型之间的距离度量, 挖掘利用不同主夹角的信息并且与基于数据样例的距离进行更加有效的融合是值得深入研究的方向。同时, 针对不同的实际问题需求可以考虑采用不同于本文中提出的几种全局综合方案。针对两个流形之间的距离度量, 本文主要是从流形局部线性建模的角度来分解问题,

后续工作可以考虑采用流形学习中更为灵活一般的流形建模方式，比如统计概率建模。此外，当前的算法框架依赖于图像集合的大规模数据量，当样本比较稀少从而流形假设很难满足时，就需要采用更为稳定的集合建模方式。这一问题在流形判别分析方法中同样存在。

3. 针对流形判别分析方法，其中一个可能的改进方案是将算法进行核化扩展，通过嵌入核函数的方式，使得算法可以利用数据中更高阶的判别信息。针对算法中采用的子空间距离度量，根据流形-流形距离理论框架，综合利用子空间样本中心与变化模式的距离应该是更合理而全面的距离度量，因此可以将当前的算法判别学习框架与基于主夹角判别分析方法的思想进行融合，期望获得更加稳定和准确的识别性能。此外，将当前方法的训练方式进行在线式的扩展，从而应用于增量式学习问题，同样具有很重要的理论价值和实用意义。



## 参考文献

- [1] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, “Face Recognition with Image Sets Using Manifold Density Divergence,” *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 581-588, 2005.
- [2] M. Balasubramanian and E. L. Schwartz, “The IsoMap Algorithm and Topological Stability,” *Science*, 295(5552):7, January 2002.
- [3] P. Baldi and K. Hornik, “Neural Networks and Principal Component Analysis: Learning from Examples without Local Minima,” *Neural Networks*, vol. 2, pp. 53-58, 1989.
- [4] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997.
- [5] M. Belkin and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,” *Advances in Neural Information Processing Systems*, vol. 14, pp. 585-591, 2001.
- [6] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples,” *Journal of Machine Learning Research*, vol. 7, pp. 2399-2434, 2006.
- [7] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Roux, and M. Ouimet, “Out-of-Sample Extensions for LLE, ISOMAP, MDS, Eigenmaps, and Spectral Clustering,” *Advances in Neural Information Processing Systems*, vol. 16, 2004.
- [8] C.M. Bishop, M. Svensen, and C.K. I. Williams, “GTM: The Generative Topographic Mapping,” *Neural Computation*, vol. 10, pp. 215-234, 1998.
- [9] Å. Björck and G. H. Golub, “Numerical Methods for Computing Angles between Linear Subspaces,” *Math. Computation*, vol. 27, no. 123, pp. 579-594, 1973.
- [10] M. Brand, “Charting a Manifold,” *Advances in Neural Information Processing Systems*, vol. 15, pp. 961-968, 2003.
- [11] M. Bressan and J. Vitrià, “Nonparametric Discriminant Analysis and Nearest Neighbor Classification,” *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2743-2749, 2003.
- [12] H.-T. Chen, H.-W. Chang, and T.-L. Liu, “Local Discriminant Embedding and Its Variants,” *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 846-853, 2005.

- [13] F.R.K. Chung, “Spectral Graph Theory,” *Proc. Regional Conf. Series in Math.*, no. 92, 1997.
- [14] T.F. Cox and M.A.A. Cox, *Multidimensional Scaling*. Chapman and Hall, 2001.
- [15] D. DeMers and G. Cottrell, “Nonlinear Dimensionality Reduction,” *Advances in Neural Information Processing Systems*, vol. 5, pp. 580- 587, 1993.
- [16] D.L. Donoho and C. Grimes, “Hessian Eigenmaps: New Locally Linear Embedding Techniques for High-Dimensional Data,” *Proc. Nat'l Academy of Sciences*, vol. 100, no. 10, pp. 5591-5596, May 2003.
- [17] M. P. Dubuisson and A. K. Jain, “A Modified Hausdorff Distance for Object Matching,” *Proc. 12th Int'l Conf. Pattern Recognition*, vol. 1, pp. 566-568, 1994.
- [18] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley, 2000.
- [19] B.S. Everitt, *An Introduction to Latent Variable Models*. Chapman and Hall, London, 1984.
- [20] W. Fan and D.-Y. Yeung, “Face Recognition with Image Sets Using Hierarchically Extracted Exemplars from Appearance Manifolds,” *Proc. 7th IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 2006.
- [21] W. Fan and D.-Y. Yeung, “Locally Linear Models on Face Appearance Manifolds with Application to Dual-Subspace Based Classification,” *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 1384-1390, 2006.
- [22] A. W. Fitzgibbon and A. Zisserman, “Joint Manifold Distance: A New Approach to Appearance based Clustering,” *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 26-33, 2003.
- [23] K. Fukui and O. Yamaguchi, “Face Recognition Using Multi-Viewpoint Patterns for Robot Vision,” *Proc. Int'l Symp. Robotics Research*, pp. 192-201, 2003.
- [24] K. Fukunaga and D.R. Olsen, “An Algorithm for Finding Intrinsic Dimensionality of Data,” *IEEE Trans. Computers*, 20(2): 176-193, 1971.
- [25] Z. Ghahramani and G.E. Hinton, “The EM Algorithm for Mixtures of Factor Analyzers,” Technical Report, Univ. of Toronto, 1996.
- [26] R. Gross and J. Shi, “The CMU Motion of Body (MoBo) Database,” Technical Report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, June 2001.
- [27] A. Hadid and M. Pietikäinen, “From Still Image to Video-Based Face Recognition: An Experimental Analysis,” *Proc. Sixth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 813-818, 2004.

- [28] J. Hamm, D. Lee, S. Mika, and B. Schölkopf, “A Kernel View of the Dimensionality Reduction of Manifolds,” *Proc. Int'l Conf. Machine Learning*, pp. 47-54, 2004.
- [29] J. Hamm and D. D. Lee, “Grassmann Discriminant Analysis: a Unifying View on Subspace-Based Learning,” *Proc. Int'l Conf. Machine Learning*, pp. 376-383, 2008.
- [30] T. Hastie and W. Stuetzle, “Principal Curves,” *J. Am. Statistical Assoc.*, vol. 84, pp. 502-516, 1989.
- [31] X. He, D. Cai, S. Yan, and H.-J. Zhang, “Neighborhood Preserving Embedding,” *Proc. Tenth IEEE Int'l Conf. Computer Vision*, vol. 2, pp. 1208-1213, 2005.
- [32] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, “Face Recognition Using Laplacianfaces,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328-340, Mar. 2005.
- [33] K. Hoffman, R. Kunze, *Linear Algebra*. Second Ed., Prentice-Hall, Inc, Englewood Cliffs, New Jersey, pp. 35-45, 1971.
- [34] H. Hotelling, “Relations between Two Sets of Variates,” *Biometrika*, vol. 28, no. 34, pp. 321-372, 1936.
- [35] D.R. Hundley and M. J. Kirby, “Estimation of Topological Dimension,” *Proc. SIAM Int'l Conf. Data Mining*, 2003.
- [36] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, “Comparing Images Using the Hausdorff Distance,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, pp. 850- 863, 1993.
- [37] <http://en.wikipedia.org/wiki/Manifold>
- [38] I.T. Jolliffe, *Principal Component Analysis*. Springer-Verlag, 1986.
- [39] N. Kambhatla, T. K. Leen, “Dimension Reduction by Local Principal Component Analysis,” *Neural Computation*, vol. 9, pp. 1493-1516, 1997.
- [40] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York. 1990.
- [41] Y. Keselman, A. Shokoufandeh, M. F. Demirci, and S. Dickinson, “Many-to-Many Graph Matching via Metric Embedding,” *Proc. Computer Vision and Pattern Recognition Conf.*, vol. 1, pp. 850-857, 2003.
- [42] T.K. Kim, O. Arandjelović, and R. Cipolla, “Boosted Manifold Principal Angles for Image Set-based Recognition,” *Pattern Recognition*, vol. 40, no. 9, pp. 2475-2484, 2007.
- [43] T.K. Kim, J. Kittler, and R. Cipolla, “Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations,” *IEEE Trans. Pattern Analysis and Machine*

- Intelligence*, vol. 29, no. 6, pp. 1005-1018, June 2007.
- [44] T.K. Kim, J. Kittler and R. Cipolla, “Incremental Learning of Locally Orthogonal Subspaces for Set-based Object Recognition,” *Proc. British Machine Vision Conference*, pp. 559-568, 2006.
- [45] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, “Face Tracking and Recognition with Visual Constraints in Real-World Videos,” *Proc. Computer Vision and Pattern Recognition Conf.*, 2008.
- [46] T. Kohonen, *Self-Organizing Maps*, Third ed. Springer-Verlag, 2001.
- [47] E. Kokiopoulou and P. Frossard, “Minimum Distance between Pattern Transformation Manifolds: Algorithm and Applications,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1225-1238, Jul. 2009.
- [48] E. Kokiopoulou and Y. Saad, “Orthogonal Neighborhood Preserving Projections: A Projection-Based Dimensionality Reduction Technique,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2143- 2156, Dec. 2007.
- [49] S. Lafon and A.B. Lee, ‘Diffusion Maps and Coarse-Graining: A Unified Framework for Dimensionality Reduction, Graph Partitioning, and Data Set Parameterization,’ *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 9, pp. 1393-1403, Sep. 2006.
- [50] M.C. Law and A.K. Jain, “Incremental Nonlinear Dimensionality Reduction by Manifold Learning,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 377-391, Mar. 2006.
- [51] K. Lee, M. Yang, and D. Kriegman, “Video-Based Face Recognition Using Probabilistic Appearance Manifolds,” *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 313-320, 2003.
- [52] B. Leibe and B. Schiele, “Analyzing Appearance and Contour Based Methods for Object Categorization,” *Proc. Computer Vision and Pattern Recognition Conf.*, 2003.
- [53] E. Levina and P.J. Bickel, “Maximum Likelihood Estimation of Intrinsic Dimension,” *Advances in Neural Information Processing Systems*, vol. 17, pp. 777-784, 2005.
- [54] H. Li, T. Jiang, and K. Zhang, “Efficient and Robust Feature Extraction by Maximum Margin Criterion,” *IEEE Trans. Neural Networks*, vol. 17, no. 1, pp. 157-165, 2006,.
- [55] R.S. Lin, C.B. Liu, M.H. Yang, N. Ahuja, and S. Levinson, “Learning Nonlinear Manifolds from Time Series,” *Proc. Ninth European Conf. Computer Vision*, vol. 2, pp. 245-256, May 2006.

- [56] T. Lin and H. Zha, “Riemannian Manifold Learning,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 796-809, May. 2008.
- [57] C. Liu, “Capitalize on Dimensionality Increasing Techniques for Improving Face Recognition Grand Challenge Performance,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725-737, May 2006.
- [58] X. Liu and T. Chen, “Video-Based Face Recognition Using Adaptive Hidden Markov Models,” *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 340-345, 2003.
- [59] L. Maaten, E. Postma, J. Herik, “Dimensionality Reduction: A Comparative Review,” Tilburg University Technical Report, TiCC-TR 2009-005, 2009.
- [60] S. Mika, G. Rätsch, and K.-R. Müller, “A Mathematical Programming Approach to the Kernel Fisher Algorithm,” *Advances in Neural Information Processing Systems*, vol. 13, pp. 591-597, 2000.
- [61] B. Moghaddam and A. Pentland, “Probabilistic Visual Learning for Object Representation,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, July 1997.
- [62] K. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, “An Introduction to Kernel-Based Learning Algorithms,” *IEEE Trans. Neural Networks*, vol. 12, pp. 181-201, 2001.
- [63] J. Park, Z. Zhang, H. Zha, and R. Kasturi, “Local Smoothing for Manifold Learning,” *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 452-459, 2004.
- [64] Y. Peng and C.-W. Ngo, “EMD-Based Video Clip Retrieval by Many-to-Many Matching,” *Proc. Int'l Conf. Image and Video Retrieval*, pp. 71-81, 2005.
- [65] K. Pettis, T. Bailey, A. K. Jain, R. Dubes, “An Intrinsic Dimensionality Estimator from Near-Neighbor Information,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 1, no. 1, pp. 25-36, Jan. 1979.
- [66] S.T. Roweis and L.K. Saul, “Nonlinear Dimensionality Reduction by Locally Linear Embedding,” *Science*, vol. 290, no. 22, pp. 2323- 2326, Dec. 2000.
- [67] S.T. Roweis, L.K. Saul, and G.E. Hinton, “Global Coordination of Local Linear Models,” *Advances in Neural Information Processing Systems* 14, pp. 889-896, 2002.
- [68] Y. Rubner, C. Tomasi, and L. Guibas, “The Earth Mover's Distance as a Metric for Image Retrieval,” *Int'l J. Computer Vision*, vol. 40, no. 2, pp. 99-121, Nov. 2000.
- [69] J.W. Sammon, “A Nonlinear Mapping for Data Structure Analysis,” *IEEE Trans. Computers*, vol. 18, no. 5, pp. 401-409, May 1969.
- [70] S. Satoh, “Comparative Evaluation of Face Sequence Matching for Content-Based Video

- Access,” *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 163-168, 2000.
- [71] L. K. Saul and S. T. Roweis, “Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds,” *J. Machine Learning Research*, vol. 4, pp. 119-155, 2003.
- [72] B. Schölkopf, A.J. Smola, and K.-R. Müller, “Nonlinear Component Analysis as a Kernel Eigenvalue Problem,” *Neural Computation*, vol. 10, pp. 1299-1319, 1998.
- [73] H.S. Seung and D.D. Lee, “The Manifold Ways of Perception,” *Science*, vol. 290, pp. 2268-2269, Dec. 2000.
- [74] F. Sha and L.K. Saul, “Analysis and Extension of Spectral Methods for Nonlinear Dimensionality Reduction,” *Proc. 22nd Int'l Conf. Machine Learning*, pp. 785-792, 2005.
- [75] G. Shakhnarovich, J.W. Fisher, and T. Darrell, “Face Recognition from Long-term Observations,” *Proc. European Conf. Computer Vision*, pp. 851-868, 2002.
- [76] P. Simard, Y.L. Cun, J. Denker, and B. Victorri, “Transformation In-variance in Pattern Recognition-Tangent Distance and Tangent Propagation,” *Neural Networks: Tricks of the Trade*, pp. 239-274, 1998.
- [77] Y. Song, F. Nie, and C. Zhang, “Semi-supervised Sub-manifold Discriminant Analysis,” *Pattern Recognition Letters*, vol. 29, Issue 13, October 2008, pp. 1806-1813.
- [78] J. Stallkamp, H.K. Ekenel, R. Stiefelhagen, “Video-based Face Recognition on Real-World Data,” *Proc. 11th IEEE Int'l Conf. Computer Vision*, 2007.
- [79] Y.W. Teh and S.T. Roweis, “Automatic Alignment of Hidden Representations,” *Advances in Neural Information Processing Systems*, vol. 15, pp. 841-848, 2003.
- [80] J. Tenenbaum, V. Silva, and J. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction,” *Science*, vol. 290, no. 22, pp. 2319-2323, Dec. 2000.
- [81] R. Tibshirani, “Principal Curves Revisited,” *Statistics and Computing*, vol. 2, pp. 183-190, 1992.
- [82] M. Turk and A. Pentland, “Face Recognition Using Eigenfaces,” *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 586-591, 1991.
- [83] A. Tversky, “Features of similarity,” *Psychological Review*, 84(4): 327-352, 1977.
- [84] N. Vasconcelos and A. Lippman, “A Multiresolution Manifold Distance for Invariant Image Similarity,” *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 127-142, Feb. 2005.
- [85] J. Verbeek, N. Vlassis, and B. Kröse, “Coordinating Principal Component Analyzers,” *Proc. Int'l Conf. Artificial Neural Networks*, vol. 12, pp. 914-919, 2002.
- [86] J. Verbeek, “Learning Nonlinear Image Manifolds by Global Alignment of Local Linear

- Models,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1236-1250, Aug. 2006.
- [87] V. Colin de Verdiére and J.L. Crowley, “Visual Recognition Using Local Appearance,” *Proc. European Conf. Computer Vision*, vol. 1, pp. 640-654, 1998.
- [88] P. Viola and M. Jones, “Robust Real-Time Face Detection,” *Int'l J. Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [89] R. Wang, S. Shan, X. Chen, and W. Gao, “Manifold-Manifold Distance with Application to Face Recognition based on Image Set,” *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, 2008.
- [90] R. Wang and X. Chen, “Manifold Discriminant Analysis,” *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, 2009.
- [91] L. Wang, X. Wang, and J. Feng, “Subspace Distance Analysis with Application to Adaptive Bayesian Algorithm for Face Recognition,” *Pattern Recognition*, vol. 39, pp. 456-464, 2006.
- [92] J. Wang, Z. Zhang, and H. Zha, “Adaptive Manifold Learning,” *Advances in Neural Information Processing Systems*, vol. 17, 2005.
- [93] K.Q. Weinberger and L.K. Saul, “Unsupervised Learning of Image Manifolds by Semidefinite Programming,” *Proc. Int'l Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 988-995, 2004.
- [94] G. Wen, L. Jiang, and N.R. Shadbolt, “Using Graph Algebra to Optimize Neighborhood for Isometric Mapping,” *Proc. 20th Int'l Joint Conf. Artificial Intelligence*, pp. 2398-2403, 2007.
- [95] M. Werman and D. Weinshall, “Similarity and Affine Invariant Distances between 2D Point Sets,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 810-814, Aug. 1995.
- [96] L. Wolf and A. Shashua, “Learning over Sets Using Kernel Principal Angles,” *J. Machine Learning Research*, vol. 4, no. 10, pp. 913-931, 2003.
- [97] D. Xu and S.-F. Chang, “Video Event Recognition Using Kernel Methods with Multilevel Temporal Alignment,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 30, no. 11, pp. 1985-1997, 2008.
- [98] O. Yamaguchi, K. Fukui, and K. Maeda, “Face Recognition Using Temporal Image Sequence,” *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pp. 318-323, 1998.

- [99] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang and S. Lin, “Graph Em-bedding and Extension: A General Framework for Dimensionality Reduction,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40-51, Jan. 2007.
- [100] S. Yan, H. Zhang, Y. Hu, B. Zhang, Q. Cheng, “Discriminant Analysis on Embedded Manifold,” *Proc. Eighth European Conf. Computer Vision*, vol. 1, pp. 121-132, 2004.
- [101] J. Yang, D. Zhang, J. Y. Yang, and B. Niu, “Globally Maximizing, Locally Minimizing: Unsupervised Discriminant Projection with Ap-plications to Face and Palm Biometrics,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 650-664, Apr. 2007.
- [102] L. Yang, “Alignment of Overlapping Locally Scaled Patches for Multidimensional Scaling and Dimensionality Reduction,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(3), pp. 438-450, Mar. 2008.
- [103] Z. Zhang and H. Zha, “Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment,” *SIAM J. Scientific Computing*, vol. 26, no. 1, pp. 313-338, 2004.
- [104] Y. Zhao, S. Xu, and Y. Jia, “Discriminant Clustering Embedding for Face Recognition with Image Sets,” *Proc. Asian Conf. Computer Vision*, vol. 2, pp. 641-650, 2007.
- [105] D. Zhao and L. Yang, “Incremental Isometric Embedding of High-Dimensional Data Using Connected Neighborhood Graphs,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(1), pp. 86-98, 2009.
- [106] S. Zhou, V. Krüeger, and R. Chellappa, “Probabilistic Recognition of Human Faces from Video,” *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 214-245, 2003.
- [107] M. Zhu and A.M. Martinez, “Subclass Discriminant Analysis,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1274-1286, Aug. 2006.
- [108] K.Q. Weinberger, F. Sha, and L.K. Saul, “Learning a Kernel Matrix for Nonlinear Dimensionality Reduction,” *Proc. The 21st Int'l Conf. Machine Learning*, 2004.
- [109] J. Venna and S. Kaski, “Visualizing Gene Interaction Graphs with Local Multidimensional Scaling,” *Proc. The 14th European Symposium on Artificial Neural Networks*, pp. 557-562, 2006.

## 攻读博士学位期间发表的论文

### 【期刊论文】

1. 王瑞平, 陈杰, 山世光, 陈熙霖, 高文. 基于支持向量机的人脸检测训练集增强. *软件学报*, 2008, 19(11): 2921-2931.
2. Jie Chen, Ruiping Wang, Shengye Yan, Shiguang Shan, Xilin Chen, Wen Gao, "Enhancing Human Face Detection by Resampling Examples Through Manifolds," *IEEE Transactions on Systems, Man, and Cybernetics-Part A*, vol. 37, no. 6, pp.1017-1028. 2007.11.
3. Jie Chen, Xilin Chen, Jie Yang, Shiguang Shan, Ruiping Wang, Wen Gao, "Optimization of a Training Set For More Robust Face Detection," *Pattern Recognition*, vol. 42, pp. 2828-2840, 2009.

### 【会议论文】

4. Ruiping Wang, Xilin Chen, "Manifold Discriminant Analysis," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 429-436, Miami Beach, Florida, June 20-25, 2009.
5. Ruiping Wang, Shiguang Shan, Xilin Chen, Wen Gao, "Manifold-Manifold Distance with Application to Face Recognition based on Image Set," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, pp. 2940-2947, Anchorage, Alaska, June 24-26, 2008.
6. Ruiping Wang, Jie Chen, Shiguang Shan, Xilin Chen, Wen Gao, "Enhancing Training Set for Face Detection," *Proceeding of International Conference on Pattern Recognition (ICPR 2006)*. vol.3, pp. 477-480, Hong Kong, 2006.8.
7. Ruiping Wang, Jie Chen, Shengye Yan, Wen Gao, "Face Detection based on the Manifold," *Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, LNCS 3546, pp. 208-218, Springer-Verlag.
8. Jie Chen, Ruiping Wang, Shiguang Shan, Xilin Chen, Wen Gao, "ISOMAP Based on the Image Euclidean Distance," *Proceeding of International Conference on Pattern Recognition (ICPR 2006)*. vol.2, pp. 1110-1113, Hong Kong, 2006.8.
9. Jie Chen, Ruiping Wang, Shengye Yan, Shiguang Shan, Xilin Chen, Wen Gao, "How to Train a Classifier Based on the Huge Face Database?," *Proceeding of IEEE International Workshop on Analysis and Modeling of Faces and Gestures, (AMFG 2005)*, LNCS 3723, pp. 84-95, 2005.



## 致 谢

博士论文即将完成之际，回首过去的一长段日子，每一件事情都依旧那么清晰，仿佛自己刚刚拿到先进人机通信联合实验室免试录取的通知，却发现硕博连读六年多的时光已经过去了。很庆幸自己一直保持了积极乐观向上的心态，以长久的激情终于达成了自己最初的目标。在这个长长的过程中，有老师们的谆谆教诲和信任鼓励，有师兄弟师姐妹的相互支持和无私帮助，有挚友们的雪中送炭和锦上添花，有亲人们的默默关心和温馨港湾。正是因为这些，求学的路上才有了无尽的动力源泉，谢谢所有关心、帮助、支持、鼓励过我的人！

首先，我要衷心感谢我的导师高文教授。从课题的选择到论文的完成每一步都离不开高老师的悉心指导和亲切关怀。在学业上，高老师鼓励我积极大胆创新，探索研究领域中的“金问题”，寻找前人没有涉足的领域。每当我研究工作中遇到疑难问题时，总是能够得到导师及时认真的指导和富有建设性的意见。作为实验室的开创者、领导者和学术带头人，高老师广阔的视野、渊博的学识、严谨民主的学风、高超的领导才能和在事业上不断追求的精神，都给我留下了深刻的印象，并将使我终身受益。

感谢陈熙霖教授对我的悉心培养，陈老师教给了我要有自信的心态。在科研道路上，我所取得的每一个成绩都饱含了陈老师的耐心指导和全力支持。衷心感谢陈老师一直以来对我的信任和鼓励，每次与陈老师交谈，他总是平易近人地给我提出富有建设性的意见，而针对我的不足和欠缺，陈老师也总是循循善诱，耐心指导我如何改进自己。正是在陈老师的指引下，我从一开始怯生生的新手逐渐成长和成熟起来。陈老师广博的学识、敏锐的思维、对待科学谦逊的精神和乐观的生活态度，是我一直钦佩和学习的榜样。

感谢我的课题组长山世光博士，山师兄教给了我如何做正确的事。正是在山师兄的带领下，我才开始进入本文的研究课题，在研究工作中取得的每一点进步无不凝聚着山师兄的汗水与辛劳。每一个创新思路的提出与讨论、每一组实验结果的报告与分析、每一篇学术论文的撰写与修改、每一次审稿意见的斟酌与反馈，这所有贯穿科研工作的每一步都见证了山师兄的倾心付出。特别感谢山师兄一直以来对我的信任与宽容，正是在宽松的环境下我才能够不急不躁的稳步开展工作。山师兄对待科学的研究的真挚热爱、敏锐犀利的科研素养、孜孜以求的探索精神、亲历亲为的实践方式，更是为我们树立了优秀青年科研工作者的典范。

感谢师兄陈杰博士，陈杰师兄教给了我如何正确的做事。在从书本学习转移到实验室研究实践这个过程中，是陈杰师兄将我带上了正确的道路，引领着我度过了研究的童年时期，他告诉了我如何以激情的态度面对科研的过程，以精细的方式操作科研的细节，以平和的心态看待科研的结果。与陈杰师兄一起工作的日子是快乐而难忘的，是他手把手教会我如何调试程序，如何分析实验结果，如何撰写和修改论文，如何合理规划自己

的研究工作和时间安排。特别感激师兄在毕业进入新的工作岗位之后，仍然时时关注关心我的工作进展，不断地给予我及时而真诚的鞭策和激励。在生活中，陈杰师兄乐观开阔的胸怀同样深深影响了我，感谢他与我分享他的生活阅历与心得。

感谢师兄苗军博士，他是一位真诚、热情而朴实的师兄。感谢他在我进入实验室最初一年多时间里给予我的具体指导和一直以来的无私帮助。感谢师兄张宝昌博士，他敏捷的思维方式、勤奋的工作态度、对科研的热爱与追求，都给我留下了深刻的印象。感谢师兄张文超博士，他踏实努力的工作态度、坚持不懈的钻研精神、诚恳谦虚的为人方式，都值得我认真学习。感谢他一直以来对我的鼓励和帮助。

感谢师姐卿来云博士、柴秀娟博士，师兄王建宇博士、张洪明博士、曹波博士、马丙鹏博士、闫胜业博士、刘亚洲博士等的帮助。他们从先行者的角度给了我无数的建议和关心，所有这些都使我有幸避免了很多的弯路，为我能顺利毕业提供了莫大的帮助。

感谢我的同学苏煜、牛志恒、李博、杨非、黄倩、张莉、唐杰，和他们在一起度过了快乐的时光，与他们的交流讨论使我受益匪浅。特别感谢苏煜在我确定求职方向和准备求职材料过程中对我自始至终的鼓励支持和真诚无私的帮助。正因为有了这些帮助，我才能够以更加积极的心态去面对所有已知和未知的困难。

感谢已毕业的师弟房佑寒，与他的每一次讨论都充满思维的乐趣和想象力的挑战。他对科研的执着热爱、广阔而深刻的理解洞察力、高效率的工作方式，都深深影响了我。衷心感谢他为我的研究工作提出的很多富有价值的建议。

感谢师弟洪晓鹏，从他身上我学习到很多做事情的方式，更加理解了坚持、乐观、拼搏的精神。衷心感谢他给予我的很多热心诚恳的建议，它们毫无疑问及时而有效的帮助我消除了诸多困惑。同样真诚感谢他为我的研究工作提出的很多有价值的启发和建议。

感谢人脸组这个团结向上、开放包容的集体。感谢同组的索津莉、谢术富、李安南、方芳、韩琥、郑伟、马志国、王丹、翟德明、翟艺涛、阚美娜、任昊宇、路亮、蒋长征、杨涛、都大龙、缑丹、陈静、吉娜烨、池晨、赵小伟、崔振。在和他们朝夕相处的日子里，一起共同进步。大家共同营造了人脸组团结友好、轻松愉快的良好氛围。

感谢办公室蔡光辉老师、王晓彪老师，他们的辛勤劳动为实验室提供了良好的研究环境。感谢研究生部周世佳老师、李丹老师，每次在我遇到困难时都给予热情的帮助。

感谢我的师兄范东睿博士，是他当年以无私而执着的胸怀来帮助我，才使得我有机会进入计算所继续追寻自己的梦想。感谢他多年来对我一直的信任、鼓励和帮助！

感谢陪我一路走来的各位好兄弟好朋友们。感谢张凯当年帮助我培养自信力，从保送面试到玉泉路学习，从编程技术到英文听说，从论文修改到审稿反馈，这所有的环节一路走来都饱含了他无尽的无私付出。感谢徐寰然在精神上一直默默支持鼓励我，无论在我得意还是失意的时候都给予我最诚挚的理解。感谢隗立涛、霍嘉、王晓、李鹏这许多兄弟们一直以来对我的无私帮助，他们与我一起分担困难，分享快乐！十年是一个节点，真心祝愿我们的友谊在下一个十年、二十年……

最后，我要特别感谢我的家人，他们的支持、鼓励和期许是我人生前进道路上最强大的源动力。感谢我的母亲和姐姐这许多年来辛勤操持我们的家庭，让我们的家欣欣向荣！感谢我的弟弟对我的理解和支持，帮助我分担家庭的责任！感谢我的女友李楠，艰辛的求学之路因为有了她携手相伴，而充满了七彩的阳光！最后，我谨将本文献给已经远去七年，为了我的成长付出毕生心血的最亲爱的父亲，并作为对我在一九九八年初写下的一段文字的回答！



## 作者简介

姓名：王瑞平 性别：男 出生日期：1981.11.8 籍贯：山西省盂县

### 【教育背景】

2003.9--至今 中国科学院计算技术研究所 计算机应用技术专业硕博连读  
1999.9--2003.7 北京交通大学信息与计算科学系 获理学学士学位

### 【参与的科研项目】

- [1] 2009 - 2013 面向视频编码的视觉计算模型与方法研究, 国家重点基础研究发展计划 (973) 项目 (编号 2009CB320902)
- [2] 2009 - 2012 面向视频监控的生物特征识别关键理论与技术研究, 国家自然科学基金 (NSFC) 项目 (编号 U0835005)
- [3] 2007 - 2009 基于视感知机理的新型人脸识别方法研究, 国家高技术研究发展计划 (863) 项目 (编号 2007AA01Z163)
- [4] 2007 - 2009 面像识别研究, 银晨网讯科技有限公司资助项目

### 【攻读博士学位期间的获奖情况】

- [1] 2006 年度 ACM KDD Cup 竞赛单项任务第二名 (第一完成人)
- [2] 2006 年度 先进人机通信联合实验室(JDL) “优秀个人”
- [3] 2007 年度 中国科学院研究生院 “三好学生”
- [4] 2008 年度 IEEE CVPR Best Student Poster Award Runner-up(最佳学生张贴论文亚军)
- [5] 2008 年度 中国科学院研究生院 “三好学生标兵”
- [6] 2008 年度 中国科学院计算技术研究所 “所长奖学金特别奖”
- [7] 2008 年度 先进人机通信联合实验室(JDL) “优秀个人”
- [8] 2009 年度 中国科学院研究生院 “三好学生”