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 effectivity 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