

格拉斯曼流形降维及应用研究

曾青松¹ 黄晓宇² 钟闰禄¹

(广州番禺职业技术学院信息工程学院 广州 511483)¹ (华南理工大学经济与贸易学院 广州 510006)²

摘要 视频人脸识别的核心问题是如何准确、高效地构建人脸模型并度量模型的相似性,为此提出一种维数约减的格拉斯曼流形鉴别分析方法以提高集合匹配的性能。首先通过子空间建模图像集合,引入投影映射将格拉斯曼流形上的基本元素表示成对应的投影矩阵。然后,为解决高维矩阵计算开销大以及在小样本条件下不能有效描述样本分布的缺陷,引入二维主成分分析方法对子空间的正交基矩阵降维。通过QR分解正则化降维后的矩阵,得到一个低维、紧致的格拉斯曼流形以获得图像集更好的表达。最后将其投影到高维核空间中进行分类。在公开的视频数据库中的实验结果证明,提出的方法在降低计算开销的同时能够获得较高的正确率,是一种有效的基于集合的对象匹配和人脸识别方法。

关键词 子空间,集合匹配,格拉斯曼流形,投影度量,二维主成分分析

中图分类号 TP391.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.07.058

Study on Grassmann Manifold Dimension Reduction and Its Application

ZENG Qing-song¹ HUANG Xiao-yu² ZHONG Run-lu¹

(School of Information and Technology, Guangzhou Panyu Polytechnic, Guangzhou 511483, China)¹

(School of Economics and Commerce, South China University of Technology, Guangzhou 510006, China)²

Abstract The key issues of video based face recognition is how to model facial images and measure the similarity between two models. To this end, a dimension reduction method in the Grassmann manifold was proposed to improve the performance of set matching. Firstly, an image set is modeled with a subspace, and the basic element of the Grassmann manifold is presented as the projection matrix by projection mapping. Then, to solve the problem of computational overhead with high dimension matrix, while the model cannot strictly describe the distribution with fewer samples, a two dimensional principal component analysis is implemented to reduce the dimension of the orthogonal basis matrix. By applying QR decomposition on the matrix, a lower dimension and tighten Grassmann manifold is obtained, which can be better to model the image set. Finally, a kernel function that mapped the orthogonal basis matrix from a Grassmann manifold to Euclidean space is used to classify image sets. Extensive experimental results on shared video based dataset show that the proposed method is an effective object matching and face recognition method based on set-to-set matching, and it outperforms other state of the art set-based matching methods with lower computational cost.

Keywords Subspace, Set matching, Grassmann manifold, Projection metric, 2DPCA

1 引言

在模式识别和机器学习领域,视频、图像集合等输入数据通常以矩阵形式组织。例如,基于视频的人脸识别直接通过视频或者视频中的多个图像帧构建数学模型,能有效提高模型的准确性,特别地,基于图像集合的方法由于不需要考虑视频的时序,近年来得到了广泛的研究^[1-2]。基于图像集合的人脸识别,可以通过参数方法和非参数方法两大类来对矩阵数据进行建模。参数方法使用一个参数分布对图像集合建立模型,并且应用 Kullback-Leibler 散度量两个模型的相似

性^[3-4]。参数估计方法的主要缺点是如果获取到的图像数量过少会导致估计的模型不准确。非参数方法主要通过子空间^[5]、流形^[6-8]、仿射包^[9-10]、凸包^[9]、协方差矩阵^[11-12]和支持向量域描述^[13]等对图像集合建模,然后度量两个模型之间的相似性。

一般地,可以对图像集合张成的子空间建模,将图像集投影到低维线性子空间,使用典型相关度量两个子空间的相似性^[14];或者计算图像集合张成的仿射子空间,使用两个仿射包中最近邻点之间的几何距离来度量集合之间的相似性^[15],在这个基础上, Hu 等人引入稀疏表达,通过仿射包对图像集

到稿日期:2016-06-11 返修日期:2016-09-26 本文受广东省自然科学基金(2015A030313807),广东省科技计划项目(2016ZC0039),广州市属高校科研项目(1201610059),广东省公益研究与能力建设项(2015A030402003),广州市教育系统创新团队建设计划(1201610034),广州番禺职业技术学院“十三五”科研项目(2016X002),华南理工大学中央高校基本科研业务项目(2015QNXM20),第二批广州市教育系统创新学术团队(13C18)资助。

曾青松(1976-),男,博士,副教授,主要研究方向为模式识别、数据挖掘, E-mail: qingsongzeng@163.com; 黄晓宇(1977-),男,博士,副教授,主要研究方向为机器学习; 钟闰禄(1979-),男,硕士,讲师,主要研究方向为模式识别。

合建模,运用最近邻点来度量两个仿射包之间的相似性^[16]。

流形能够有效刻画样本分布的本质结构,并提供一种结构紧凑的表示。通常子空间可以解释为格拉斯曼流形上的一个点,通过核函数将流形上的特征映射到欧氏空间,然后在新的空间中学习一个分类器。格拉斯曼流形鉴别分析(Grassmannian Discriminant Analysis, GDA)^[17]使用核函数将欧氏空间的特征投影到再生核希尔伯特空间,在核 Fisher 鉴别分析中应用格拉斯曼核进行鉴别分析。Wang 等提出协方差鉴别学习(Covariance Discriminative Learning, CDL)^[11]的方法,CDL 考虑到所有图像集合的协方差矩阵位于黎曼流形之上,通过基于对数-欧氏距离(Log-Euclidean Distance, LED)的核函数将这些矩阵映射到欧氏空间上,并应用核鉴别分析进行分类。Wang 等提出一种流形判别分析方法(Manifold Discriminant Analysis, MDA)^[7]最大化流形间隔,从而达到在流形进行判别分析的目的。虽然基于非线性流形的方法可以同时建模集合中的数据的多种变化模式,并取得非常好的识别效果,但是当集合的数据非常有限时,这些方法有时会失效,而且由于特征处在相当高维的空间中,很难得到准确的密度估计,并且计算开销都比较大,很难在实用化嵌入式设备中得到应用。

2 格拉斯曼流形

2.1 图像集的格拉斯曼流形模型

定义 1 流形是一个嵌入在高维欧氏空间中的一个光滑的曲面。 D 维空间 \mathcal{R}^D 的所有 m -维线性子空间 $\mathcal{R}^m (0 \leq m \leq D)$ 构成格拉斯曼流形(Grassmann manifold),记作 $\mathcal{G}(m, D)$ ^[18],它是一个特殊的黎曼流形,一般用子空间的正交基来表示。

定义 2 格拉斯曼流形上两个点等价, $Y_1 \cong Y_2$, 当且仅当它们张成的线性子空间等价,即 $span(Y_1) \cong span(Y_2)$ 。

在投影映射 $\phi(U) = UU^T$ 的框架下,使用 UU^T 来表示格拉斯曼流形上的基本元素,即线性子空间 $span(U)$ 。它是一个一对一、连续、可微的映射,并且逆映射也是连续、可微的。因此,在格拉斯曼流形上的每个元素只对应唯一的一个投影矩阵。

给定 n 个视频人脸序列 $\{X_1, X_2, \dots, X_n\}$, 其中 $X_i (X_i \in \mathcal{R}^{D \times n_i})$ 表示包括 n_i 张图像帧的视频图像帧的集合,其中每张图像被表示成 D 维的列向量,那么图像集 X_i 可以由这些列向量构成的矩阵来描述。进一步,通过矩阵的正交基矩阵 $U_i (U_i \in \mathcal{R}^{D \times m})$ 张成的 m 维线性子空间 $span(U_i)$ 来表示,即 $X_i X_i^T \cong U_i \Lambda_i U_i^T$, 其中 Λ_i 和 U_i 分别对应最大的 m 个特征值以及相应特征向量构成的矩阵,其中 $U_i^T U_i = I_m$ 是大小为 $m \times m$ 的单位矩阵。

2.2 投影度量

定义 3 令 $\Theta = [\theta_1, \dots, \theta_p]$ 表示两个线性子空间 $span(X_1)$ 与 $span(X_2)$ 之间的主夹角序列,那么它们之间的测地距离定义为:

$$d_g(X_1, X_2) = \|\Theta\|_2$$

格拉斯曼流形是一个非线性的空间,大部分欧氏空间的技术都不能直接用于处理格拉斯曼流形上的数据。通过研究格拉斯曼流形特定的黎曼几何特征,引入投影映射框架将格

拉斯曼流形上的基本元素表示成对应的投影矩阵。

投影映射 $\phi(U)$ 是一个 $D \times D$ 的对称矩阵,对应的内积形式可以定义为: $\langle U_1, U_2 \rangle_\phi = \text{tr}(\phi(U_1)^T \phi(U_2))$, 这一内积对线性子空间具有不变性,可以推导出距离形式:

$$d_p(U_1, U_2) = \left(\sum_{i=1}^m \sin^2 \theta_i \right)^{1/2} = \frac{1}{\sqrt{2}} \|U_1 U_1^T - U_2 U_2^T\|_F$$

其中, $\|\cdot\|_F$ 表示矩阵的 Frobenius 范数,投影距离满足非负性、对称性和三角不等式,因此也称为投影度量^[17]。投影距离逼近于格拉斯曼流形上的测地距离。

3 格拉斯曼流形降维

D 维空间 \mathcal{R}^D 所有的 m 维线性子空间 $span(U_i)$ 均位于一个 D 维的格拉斯曼流形 $\mathcal{G}(m, D)$ 上,并一一对应于投影矩阵 $U_i U_i^T \in \mathcal{R}^{D \times D}$ 。通常情况下能够获得的图像数量小于样本的特征维数,这样得到的投影矩阵不能很好地建模原始样本的真实分布,为此可以将 $\mathcal{G}(m, D)$ 投影到一个更低维数的格拉斯曼流形上,在降低计算开销的同时能保持较高的匹配正确率。

3.1 二维主成分分析

Yang 等^[19]提出了二维主成分分析(Two-Dimensional Principal Component Analysis, 2DPCA), 2DPCA 本质上去除了图像矩阵各列之间的相关性,用于图像的降维和重构。

设矩阵 $V_i = U_i U_i^T \in \mathcal{R}^{D \times D}, i = 1, 2, \dots, n$, 定义投影矩阵的二维散度矩阵为:

$$G_i = \frac{1}{N} \sum_{i=1}^n (V_i - \bar{V})(V_i - \bar{V})^T$$

其中, $\bar{V} = \frac{1}{n} \sum_{i=1}^n V_i$ 表示 n 个投影矩阵的均值。

定义目标函数:

$$J(P) = P^T G_i P$$

最大化目标函数 $J(P)$ 获得最优的投影向量 P , 即满足正交约束 $P_i^T P_j = 0, \forall i \neq j, i, j = 1, 2, \dots, d$ 的条件下,求最大值问题:

$$\{P_1, \dots, P_d\} = \arg \max J(P)$$

最大的 q 个特征值对应的特征向量 P_i 组成 $D \times d$ 的投影矩阵 P 。定义 $g_j = V_j P$ 为样本 V_j 的 2DPCA 投影,每个样本均由 $D \times d$ 维特征矩阵表示。

3.2 格拉斯曼流形降维

广义映射 $f: \mathcal{G}(m, D) \rightarrow \mathcal{G}(m, d), d < D$ 定义为:

$$f(U_i U_i^T) = (P^T U_i)(P^T U_i)^T$$

其中, $P \in \mathcal{R}^{D \times d} (d < D)$ 是一个列满秩的投影矩阵, $P^T U_i \in \mathcal{R}^{d \times m}$, 如果 P 是正交矩阵,那么 $P^T U_i$ 是正交基矩阵,从而 $P^T U_i$ 构成子空间的正交基矩阵,即 f 将原始的格拉斯曼流形 $\mathcal{G}(m, D)$ 变换到一个更低维的格拉斯曼流形 $\mathcal{G}(m, d)$ 。

在投影映射 $\phi(U) = UU^T$ 的框架下, UU^T 是 $D \times D$ 的对称矩阵。假设 $\{X_i \in \mathcal{R}^{D \times n_i}\}_{i=1}^n$ 表示 n 个训练集合, $\{U_i \in \mathcal{R}^{D \times m}\}_{i=1}^n$ 表示集合 X_i 的 m -维线性子空间的正交基矩阵,那么投影矩阵 $U_i U_i^T \in \mathcal{R}^{D \times D}$ 是对称矩阵,将 $U_i U_i^T$ 看作一个 $D \times D$ 维的样本,通过求解二维主成分分析获得最优的投影向量 P ,但是此处得到的 P 不一定是正交矩阵,因此 $P^T U_i$ 也不能保证是正交矩阵。

定理 1 如果 $A = QR$ 表示 $A \in \mathcal{R}^{d \times m}$ 的 QR 分解, $A =$

$[a_1, \dots, a_m]$, $Q = [q_1, \dots, q_m]$ 是列分块矩阵, 那么, $span\{a_1, \dots, a_k\} = span\{q_1, \dots, q_k\}$, $k=1, \dots, m$ ^[20]。

对于所有的 i , 固定矩阵 P , 通过将矩阵 U_i 正则化为 U_i' , 可以使得 $P^T U_i'$ 是正交基矩阵。计算 $P^T U_i$ 的 QR 分解:

$$P^T U_i = Q_i R_i$$

其中, $Q_i \in \mathbb{R}^{d \times m}$ 是正交矩阵, 即 $Q_i^T Q_i = I_m$ 是大小为 $m \times m$ 的单位矩阵, $R_i \in \mathbb{R}^{m \times m}$ 是上三角可逆矩阵^[20]。由于 R_i 可逆, 并且 Q_i 正交, 因此下式使得 $P^T U_i'$ 是正交基矩阵。

$$Q_i = P^T (U_i R_i^{-1}) \rightarrow U_i' = U_i R_i^{-1}$$

根据定理 1 的结论, Q_i 与 $P^T U_i$ 具有相同的子空间结构, 因而具有等价的正交基矩阵。

通过上述变化, 得到一个低维的格拉斯曼流形 $\mathcal{G}(m, D)$ 。相对于高维空间的特征, 对于样本数量有限的情况, 低维逼近更容易得到准确的密度估计, 而且计算开销较小, 容易在嵌入式设备中得到应用。

实际应用中, 可以直接使用查询集合或者单独获取一组图像集用于训练投影矩阵 P , 使得到的投影矩阵尽可能独立于查询数据库中的图像集合。使用投影矩阵 P 对测试集合 (Probe) 和查询集合 (Gallery) 的正交基矩阵 U_i 进行降维, 通过计算 $P^T U_i$ 的 QR 分解 $P^T U_i = Q_i R_i$, 得到降维后的线性子空间的正交基矩阵 $Q_i = P^T (U_i R_i^{-1}) \in \mathbb{R}^{d \times m}$ 。所有的 m 维线性子空间 $span(Q_i)$ 位于一个 d 维格拉斯曼流形 $\mathcal{G}(m, D)$ 上, 并一一对应于投影矩阵 $Q_i Q_i^T \in \mathbb{R}^{d \times d}$ 。

4 格拉斯曼流形核鉴别分析

设 Q_i, Q_j 是子空间的正交基矩阵, 基于投影度量的内积核函数 (投影核) 定义为^[17]:

$$\kappa(Q_i, Q_j) = \text{tr}(Q_i Q_i^T Q_j Q_j^T) = \|Q_i^T Q_j\|_F^2$$

核鉴别分析 (Kernel Discriminant Analysis, KDA) 求解最优化目标函数:

$$\alpha^* = \arg \max_{\alpha} \frac{\alpha^T K W K \alpha}{\alpha^T K K \alpha}$$

其中, $K = [\kappa_{ij}]$ 是核矩阵, $\kappa_{ij} = \kappa(Q_i, Q_j)$, 相关系数 $W = [w_{ij}]$ 定义为: 当 Q_i 和 Q_j 同时属于第 r 类样本时, $w_{ij} = 1/n_r$, 否则 $w_{ij} = 0$ 。 $\alpha = [\alpha_1, \dots, \alpha_n]$, 最优的 α 可以通过求解特征值问题 $K W K \alpha = \lambda K K \alpha$ 的最大特征值对应的特征向量得到。将最大的 q 个特征值对应的特征向量按照特征值大小排序组合在一起得到投影矩阵 A 。

给定输入空间的一个样本 $Q_i \in \mathbb{R}^{d \times d}$, 鉴别子空间的 q 维投影可以通过公式 $Q_i' = K_i A$ 得到, 其中 $K_i = [\kappa_{i1}, \dots, \kappa_{im}]$ 。算法 1 给出了维数约减的格拉斯曼流形鉴别分析的框架。

算法 1 维数约减的格拉斯曼流形鉴别分析

输入:

1. 训练图像数据集 Train;
2. 查询图像数据集 Gallery 以及对应的类别标签;
3. 测试集 X_i ;

训练过程:

1. 计算所有训练图像集合 Train 的正交基矩阵;
2. 求解最优化问题 $J(P) = P^T G_i P$, 计算 2DPCA 的投影矩阵 P ;
3. 对于查询集合 Gallery, 计算每个集合的正交基矩阵 U_i 及 $P^T U_i$ 的 QR 分解 $P^T U_i = Q_i R_i$, 得到低维格拉斯曼流形上的正交基矩阵 Q_i ;
4. 使用格拉斯曼核函数计算核矩阵 K ;

5. 求解广义特征值问题 $K W K \alpha = \lambda K K \alpha$, 取最大的 q 个特征值对应的特征向量, 得到核鉴别分析的投影矩阵 A ;

测试过程:

1. 对于查询集合 Gallery 中的每一图像集, 计算正交基 U_i 降维后得到的基矩阵 Q_i , 然后计算投影矩阵 $Q_i Q_i^T$ 在核空间上的投影 Q_i', \dots, Q_n' ;
2. 对测试集 X_i , 计算正交基矩阵 U_i , 通过计算 $P^T U_i$ 的 QR 分解 $P^T U_i = Q_i R_i$, 得到低维格拉斯曼流形上的正交基矩阵 Q_i ;
3. 计算 K_i , 通过公式 $Q_i' = K_i A$ 得到其在核空间上的投影;
4. 求解 $\arg \min_i \|Q_i' - Q_i'\|^2$, 得到类标 i ;

输出:

测试集 X_i 的类标

5 实验分析

本节在 ETH80 对象数据库、Honda&UCSD 视频数据库、YouTube 视频数据库、ChokePoint 数据库和 CMU Mobo 数据库上进行基于集合的对象匹配和人脸识别实验。实验计算机配置为: Intel(R) Xeon(R) E7-4807 双 1.87GHz CPU (2×6 核)、64GB 内存、Windows Server 2008。程序使用 MATLAB 2013 编写。

5.1 数据集

ETH80 数据库是由苏黎世理工大学提供的实物光场数据库。如图 1(a) 所示, 数据库中一共有 8 个类, 如图 1(b) 所示, 每个类中有 10 个对象, 共 80 个对象。8 个类分别为: 苹果、梨、西红柿、牛、狗、马、杯子和汽车。每个类的 10 个对象在采集时尽量选取了差异较大但又明显属于该类的对象。搭建光场时, 在对象实物的周围构造一个正八面体, 然后对其进行三级细分, 取上半部分, 在得到的顶点上设置照相机进行拍摄, 这样每个对象可以获得 41 个视图, 80 个对象一共 3280 幅图片。图 1 示出每类物体的一个实例及一个实例在不同视角下的 10 幅图片。将每一个实例的 41 幅图像作为一个图像集合, 每类物体有 10 个图像集合。

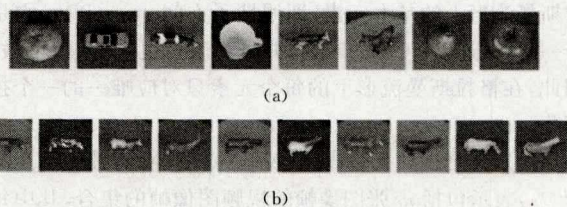


图 1 ETH80 对象分类数据库中的样本

Honda&UCSD 数据集是人脸识别领域的基准数据集, 它一共包含 19 个人的 59 段视频。该数据库的所有视频是在室内光照条件受到控制的情况下录制的, 视频相对比较清晰。每个视频片段内只有一个人, 视频包含该人的不同姿态和表情的变化。每段视频长度大约 300~500 帧, 被分割成多个视频小片段, 用于构建训练和验证集合。实验使用 P. Viola^[21] 算法逐帧检测出人脸区域。如图 2 所示, 检测到的人脸图像都接近正面人像。



图 2 Honda&UCSD 数据库人脸示意图

YouTube Celebrities(YTC)数据库^[22]收集来自于 YouTube 网站的 1910 个视频片段,一共包含 47 个人。数据库包含大范围的姿态、光照和表情变化。每个视频的第一帧的人脸位置已经人工标定。每个人的视频片段分为 3 个小节,每个小节对应不同的采集时间与场景。本文使用对象跟踪方法^[23]提取相应的人脸图像。图 3 示出了“Donald Trump”的第 10 和 150 帧的跟踪结果。

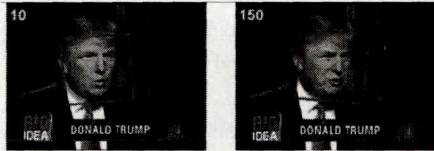


图 3 人脸跟踪检测示意图

ChokePoint 数据集主要为研究基于视频的人脸匹配、识别任务开发的数据库^[24],该数据集包含门口 1 录制的 25 个人的视频和门口 2 录制的 29 个人的视频,录制时进入门口和离开门口的动作分别录制。每个门口由有 3 台位于不同角度的摄像机同时录制,每个动作重复 4 次得到 4 个不同的视频序列。

CMU 的 Mobo(Motion of Body)数据库是卡内基梅隆大学(Carnegie Mellon University,CMU)机器人研究所于 2001 年 3 月开始收集的基于视频的人脸识别研究领域中使用最为广泛的数据库之一^[25]。Mobo 数据库包含 25 个人在跑步机上行走的视频,数据从 6 个不同视点采集,共记录慢走、快走、倾斜地面走和带球走 4 种运动。本文使用它的一个子集,包含 96 个视频序列,24 种不同的主题,每一个序列包含大约 300 帧。

5.2 比较的方法和设置

实验以原始论文公开的代码为基础,由于计算资源的限制,将人脸图像转换为灰度图,经直方图均衡化处理之后采样到像素。具体参数设置如下:互子空间方法(Constrained Mutual Subspace Method, CMSM)^[5]、流形鉴别分析(Manifold Discriminant Analysis,MDA)^[7]算法使用 PCA 降维,计算子空间时使用了最大的 10 个特征值对应的特征向量。格拉斯曼流形鉴别分析(Grassmann Discriminant Analysis,GDA)^[17]算法采用投影核。协方差鉴别学习(Covariance Discriminative Learning,CDL)^[11]参考原文实现,使用核鉴别分析(Kernel Discriminant Analysis,KDA)算法分类,KDA 使用文献^[26]提供的代码。

本文方法使用子空间的正交基矩阵对图像集合建模,通过 2DPCA 对正交基矩阵降维,然后应用 KDA 进行鉴别分析。2DPCA 降维之后在 ETH80, Honda&UCSD, ChokePoint, Mobo 和 YTC 5 个数据库上的维数分别为(查询集所包含的类别数量的 2 倍):16,38,50,50,28。实验给出 4 种组合的结果:基准,2DPCA、KDA 和 2DPCA+KDA,具体实验设置如下:1)基准,使用投影距离公式计算两个正交基矩阵之间的距离。2)2DPCA,使用 2DPCA 对正交基矩阵降维,并使用投影距离公式计算降维后的两个正交基矩阵之间的距离。3)KDA,计算投影核矩阵,并应用 KDA 进行鉴别分析。4)2DPCA+KDA,使用 2DPCA 对正交基矩阵降维,使用降维后的正交基矩阵计算投影核矩阵,并应用 KDA 进行鉴别分析。

5.3 实验结果及分析

ETH80 数据集中每一类对象有 10 个图像集,实验随机选择 3,4,5 个图像集合作为训练集合,剩余的图像集合作为测试集合,实验重复 10 次,表 1 列出了随机实验的平均识别率和方差。

表 1 ETH80 数据集上的分类结果/%

方法	3 训练集合 分类准确率	4 训练集合 分类准确率	5 训练集合 分类准确率
CMSM ^[5]	84.64±2.94	84.17±4.52	89.25±2.37
MDA ^[7]	76.96±4.56	80.62±3.63	84.00±4.11
GDA ^[17]	87.14±3.12	88.33±2.23	91.00±2.41
CDL ^[11]	85.36±2.64	88.85±4.19	92.00±4.53
基准	82.32±4.79	82.91±3.07	87.50±6.00
2DPCA	82.14±5.25	82.08±4.41	88.75±6.79
KDA	86.07±2.76	90.00±3.77	92.75±4.47
2DPCA+KDA	84.82±3.49	88.33±5.39	91.75±4.25

Honda&UCSD 数据库中由于部分人只有 2 段视频,因此随机选择 1 段视频作为训练集合,其余的作为测试集合。YTC 数据库中部分人只有 1 段视频,根据本文实验的需要,选择 01 小节、满足每一个人至少有 5 段视频的一个子集,满足条件的一共有 14 个人的视频,随机选择 2 段(YTC 2 训练集合)和 3 段(YTC 3 训练集合)视频作为训练集,其余的作为测试集合。ChokePoint 数据库选择 2 段视频作为训练集合。Mobo 数据库提取图像的局部二值模式(Local Binary Pattern,LBP)特征,选择 2 段视频作为训练集合,剩余的作为测试视频。所有算法计算了 5 次随机实验的平均结果。计算 2DPCA 时使用全部的训练集合训练投影矩阵。

表 2 列出了在 Honda&UCSD,ChokePoint 和 Mobo 数据集上基于图像集合的人脸识别随机实验的平均识别率和方差。

表 2 基于图像集合的人脸识别结果/%

方法	Honda&UCSD 分类准确率	ChokePoint 分类准确率	Mobo 分类准确率
CMSM ^[5]	95.23±2.91	70.40±3.26	97.14±2.32
MDA ^[7]	99.52±1.06	98.56±1.63	98.09±1.99
GDA ^[17]	96.66±1.30	85.00±3.13	97.14±2.32
CDL ^[11]	97.14±2.60	88.08±4.10	96.32±3.02
基准	98.10±1.99	72.56±4.43	98.36±1.70
2DPCA	97.62±1.68	71.68±4.33	98.36±1.70
KDA	98.57±1.30	86.40±3.22	98.78±1.11
2DPCA+KDA	98.57±1.30	85.52±3.18	98.78±1.11

表 3 列出了 YTC 数据库上分别选择 2 段和 3 段测试视频的基于图像集合的人脸识别随机实验的平均识别率和方差。从实验结果可以看出,本文提出的方法的各种组合都能够达到相对较高的识别率。

表 3 基于图像集合的人脸识别结果/%

方法	YTC 2 训练集合 分类准确率	YTC 3 训练集合 分类准确率
CMSM ^[5]	86.79±0.00	87.18±0.00
MDA ^[7]	98.86±1.03	99.48±1.14
GDA ^[17]	79.62±0.84	87.69±1.14
CDL ^[11]	85.28±1.57	92.82±1.14
基准	96.22±0.00	89.74±0.00
2DPCA	95.85±0.84	92.30±0.00
KDA	96.98±1.68	94.35±1.14
2DPCA+KDA	96.60±1.57	93.84±2.29

表4列出了在Honda&UCSD和ChokePoint数据集上各个算法的平均运算时间,单位为秒。从实验结果分析可知,使用2DPCA对正交基矩阵降维之后计算投影距离比直接使用投影距离要节约时间,同样使用2DPCA对正交基矩阵降维之后计算投影核比直接使用正交基矩阵计算投影核要节约时间。根据表2和表3的实验结果分析,本文提出的“2DPCA+KDA”方法在识别率上虽然与MDA相当,但是表4的结果显示,本文算法的时间开销明显优于MDA等其他算法。

表4 Honda&UCSD和ChokePoint数据集上的平均运算时间/s

方法	Honda&UCSD	ChokePoint
CMSM ^[14]	46.83	466.2
MDA ^[7]	148.2	105.7
GDA ^[17]	13.63	78.5
CDL ^[11]	51.36	917.2
基准	14.79	70.27
2DPCA	10.44	19.46
KDA	12.82	27.48
2DPCA+KDA	11.28	21.11

结束语 图像集合匹配中的核心问题是如何对图像集合建模并比较两个模型的相似性,本文通过子空间对图像集合进行建模,并把它描述成格拉斯曼流形上的点。一般情况下,图像特征向量的维数大于能够提供的样本数量,得到的子空间的正交基不能很好地刻画图像集合的分布。应用二维PCA对子空间的正交基矩阵降维,得到一个更加低维、紧致的格拉斯曼流形,从而能更好地刻画集合的真实分布。但是本文依赖训练集合得到二维PCA的投影矩阵,没有充分考虑有监督样本的作用。后续可以利用有标签的样本通过度量学习^[27]得到投影矩阵,从而得到一个更加贴近实际应用的子流形。

参考文献

- [1] HUANG Z, SHAN S, WANG R, et al. A benchmark and comparative study of video-based face recognition on COX face database[J]. IEEE Transactions on Image Processing, IEEE, 2015, 24(12): 5967-5981.
- [2] BARR J R, BOWYER K W, FLYNN P J, et al. Face recognition from video: A review[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2012, 26(5): 1266002.
- [3] WOLF L, HASSNER T, MAOZ I. Face recognition in unconstrained videos with matched background similarity[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA, IEEE, 2011: 529-534.
- [4] HARANDI M, SALZMANN M, BAKTASHMOTLAGH M. Beyond Gauss: Image-Set Matching on the Riemannian Manifold of PDFs[C]// Proc. ICCV. 2015: 1-19.
- [5] NISHIYAMA M, YAMAGUCHI O, FUKUI K. Face Recognition Using the Multiple Constrained Mutual Subspace Method [J]. IEEE Transactions on Information and Systems, 2005, 88(8): 1339-1348.
- [6] FAN W, YEUNG D. Locally Linear Models on Face Appearance Manifolds with Application to Dual-Subspace Based Classification[C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006). New York, NY, USA, IEEE Computer Society, 2006: 1384-1390.
- [7] WANG R, CHEN X. Manifold discriminant analysis[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009: 429-436.
- [8] WANG R, SHAN S, CHEN X, et al. Maximal Linear Embedding for Dimensionality Reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(9): 1776-1792.
- [9] CEVIKALP H, TRIGGS B. Face recognition based on image sets[J]. Computer Vision and Pattern Recognition, 2010, 26(2): 2567-2573.
- [10] HU Y, MIAN A S, OWENS R. Sparse approximated nearest points for image set classification[C]// 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011: 121-128.
- [11] WANG R, GUO H, DAVIS L S, et al. Covariance discriminative learning: A natural and efficient approach to image set classification[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2012: 2496-2503.
- [12] JAYASUMANA S, HARTLEY R, SALZMANN M, et al. Kernel Methods on the Riemannian Manifold of Symmetric Positive Definite Matrices[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2013: 73-80.
- [13] ZENG Q S, LAI J H, WANG C D. Multi-local model image set matching based on domain description[J]. Pattern Recognition, 2014, 47(2): 694-704.
- [14] YAMAGUCHI O, FUKUI K, MAEDA K. Face Recognition Using Temporal Image Sequence[C]// 3rd International Conference on Face & Gesture Recognition. Nara, Japan, IEEE Computer Society, 1998: 318-323.
- [15] YANG M, ZHU P, VAN G L J, et al. Face Recognition based on Regularized Nearest Points between Image Sets[C]// FG2013. Shanghai, China, IEEE, 2013: 1-7.
- [16] HU Y, MIAN A S, OWENS R. Face recognition using sparse approximated nearest points between image sets [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 1992-2004.
- [17] HAMM J, LEE D D. Grassmann discriminant analysis: a unifying view on subspace-based learning[C]// Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland, ACM, 2008: 376-383.
- [18] HARANDI M T, SANDERSON C, SHIRAZI S, et al. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2011: 2705-2712.
- [19] YANG J, ZHANG D, FRAGNI A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(1): 131-137.
- [20] YE J, LI Q. A two-stage linear discriminant analysis via QR-decomposition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(6): 929-941.
- [21] VIOLA P, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.

- [22] KIM M, KUMAR S, PAVLOVIC V, et al. Face tracking and recognition with visual constraints in real-world videos[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008). Anchorage, Alaska, USA; IEEE Computer Society, 2008.
- [23] ROSS D A, LIM J, YANG M H. Adaptive Probabilistic Visual Tracking with Incremental Subspace Update[C]// 8th European Conference on Computer Vision Computer Vision (ECCV 2004). Prague, Czech Republic; Springer, 2004: 470-482.
- [24] WONG Y, CHEN S, MAU S, et al. Patch-based probabilistic image quality assessment for face selection and improved video-

based face recognition[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop. Colorado Springs, CO, USA; IEEE, 2011: 74-81.

- [25] GROSS R, SHI J. The CMU Motion of Body (MoBo) Database [J]. Monumenta Nipponica, 2001, 45(4): 1-7.
- [26] CAI D, HE X, HAN J. Speed up kernel discriminant analysis [J]. Vldb Journal, 2011, 20(1): 21-33.
- [27] HUANG Z, WANG R, SHAN S, et al. Face recognition on large-scale video in the wild with hybrid Euclidean-and-Riemannian metric learning[J]. Pattern Recognition, Elsevier, 2015, 48(10): 3113-3124.

(上接第317页)

监督图论方法。在提出的方法中,图像由属性关系图建模;然后,档案中的图像与查询图像通过新奇非精确图匹配策略进行匹配;最后,将与查询图像最相似的图像排列,按照图的相似性值进行检索。为匹配相应的图像以及按照图的相似点实现图像检索,采用了结合子图同构算法和光谱图嵌入技术的新型非精确图像匹配策略。实验结果显示,与其他两种无监督遥感图像检索方法相比,所提方法的检索性能得到了明显改善。

参考文献

- [1] NIU L, NI L, MIAO Y, et al. The compressed domain based on ROI multi spectral remote sensing image retrieval Chinese [J]. Journal of Image and Graphics, 2005, 10(10): 1212-1217. (in Chinese)
牛蕾,倪林, MIAO Y, 等. 基于 ROI 的压缩域多波段遥感图像的检索[J]. 中国图象图形学报, 2005, 10(10): 1212-1217.
- [2] ZHU J L, LI S J, WAN D S, et al. Remote sensing image retrieval, feature selection and semi supervised learning based on China [J]. Journal of Image and Graphics, 2011, 16(8): 1474-1482. (in Chinese)
朱佳丽, 李士进, 万定生, 等. 基于特征选择和半监督学习的遥感图像检索[J]. 中国图象图形学报, 2011, 16(8): 1474-1482.
- [3] YANG J, LIU J B, DAI Q. An Improved Remote Sensing Image Retrieval Method Based on Bag of Word Framework[J]. Geomatics and Information Science of Wuhan University, 2014, 38(9): 1109-1113. (in Chinese)
杨进, 刘建波, 戴芹. 一种改进包模型的遥感图像检索方法[J]. 武汉大学学报(信息科学版), 2014, 39(9): 1109-1113.
- [4] TANG X H, QIN K, MENG L K. Qualitative description model of directional relations based on topological constraints [J]. Journal of Surveying and Mapping, 2014, 43(4): 396-403. (in Chinese)
唐雪华, 秦昆, 孟令奎. 基于拓扑约束的方向关系定性描述模型[J]. 测绘学报, 2014, 43(4): 396-403.
- [5] XU Y X, CHEN F. Recent research progress of local image descriptor [J]. China Journal of Image and Graphics, 2015, 20(9): 1133-1150. (in Chinese)
许允喜, 陈方. 局部图像描述符最新研究进展[J]. 中国图象图形学报, 2015, 20(9): 1133-1150.
- [6] DEMIR B, BRUZZONE L. A Novel Active Learning Method in

Relevance Feedback for Content-Based Remote Sensing Image Retrieval[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(5): 2323-2334.

- [7] DU Z, LI X, LU X. Local structure learning in high resolution remote sensing image retrieval[J]. Neurocomputing, 2016, 207: 813-822.
- [8] SEVILLA J, JIMENEZ L I, PLAZA A. Sparse Unmixing-Based Content Retrieval of Hyperspectral Images on Graphics Processing Units[J]. IEEE Geoscience & Remote Sensing Letters, 2015, 12(12): 2443-2447.
- [9] CHAUDHURI B, DEMIR B, BRUZZONE L, et al. Region-Based Retrieval of Remote Sensing Images Using an Unsupervised Graph-Theoretic Approach [J]. IEEE Geoscience & Remote Sensing Letters, 2016, 13(7): 987-991.
- [10] JIAO L, TANG X, HOU B, et al. SAR Images Retrieval Based on Semantic Classification and Region-Based Similarity Measure for Earth Observation[J]. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2015, 8(8): 1-16.
- [11] JIANG J, MA X, CAI Z, et al. Sparse Support Regression for Image Super-Resolution [J]. IEEE Photonics Journal, 2015, 7(5): 1.
- [12] CHEN X, HAN G, ZHANG Z. Real-time video super-resolution reconstruction based on GPU acceleration[J]. Journal of Computer Applications, 2013, 33(12): 3540-3543.
- [13] SEDAGHAT A, EBADI H. Remote Sensing Image Matching Based on Adaptive Binning SIFT Descriptor[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 53(10): 5283-5293.
- [14] DEMIR B, BRUZZONE L. Hashing-Based Scalable Remote Sensing Image Search and Retrieval in Large Archives[J]. IEEE Transactions on Geoscience & Remote Sensing, 2015, 54: 1-13.
- [15] YANG Y, NEWSAM S. Geographic Image Retrieval Using Local Invariant Features[J]. IEEE Transactions on Geoscience & Remote Sensing, 2013, 51(2): 818-832.
- [16] ZHOU Y G, WANG P, GAO Y H. Remote Sensing Image Classification with Bag-of-Visual-Words Model [J]. Journal of Chongqing University Technology (Natural Science), 2015, 29(5): 71-77. (in Chinese)
周宇谷, 王平, 高颖慧. 基于视觉词袋模型的遥感图像分类方法[J]. 重庆理工大学学报(自然科学), 2015, 29(5): 71-77.