

# 低维格拉斯曼流形鉴别分析

曾青松, 钟闰禄

(广州番禺职业技术学院信息工程学院, 广东 广州 511483)

**摘 要:** 图像集匹配需要解决如何对集合模型并度量模型之间的相似性的问题, 为此提出一种维数约减的格拉斯曼流形鉴别分析方法并用于对象和人脸识别。首先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, 为克服高维矩阵在小样本条件下不能有效描述样本分布的缺陷, 通过投影度量学习对子空间的正交基矩阵降维得到一个低维、紧致的格拉斯曼流形以获得图像集更好地表达。最后投影到再生核希尔伯特空间中进行分类。在公开的视频数据库中的实验结果证明, 该方法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; 度量学习

中图分类号: TP 391.41

DOI: 10.11996/JG.j.2095-302X.2017010069

文献标识码: A

文章编号: 2095-302X(2017)01-0069-07

## Low Dimension Discriminant Analysis on Grassmann Manifold

ZENG Qingsong, ZHONG Runlu

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

**Abstract:** The key issues of video based face recognition are the way to model facial images precision and high efficiency and measure the similarity between two sets. To end this, a Grassmann manifold dimension reduction method is proposed to improve the performance of image set matching. Firstly, a subspace constructed by an image set is presented as a point in a Grassmann manifold with a projection matrix. Then, a projection metric learning approach is applied to reduce the dimension of the orthogonal basis matrix to obtain a lower dimension and tighten Grassmann manifold. Finally, a kernel function mapped the orthogonal basis matrix from a Grassmann manifold to Euclidean space for classification. 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.

**Keywords:** subspace; set matching; Grassmann manifold; projection metric; metric learning

基于图像集合的人脸识别方法通过视频中的多个图像帧构建人脸的图像集合表示模型。多帧图像能够弥补由于分辨率、姿态变化等信息的损失, 因而能有效地提高模型的准确性, 近年来得到了广泛的研究<sup>[1-3]</sup>。基于图像集合的人脸识别, 可以分为参

数方法和非参数方法两大类。参数方法一般使用参数分布对图像集合建立模型, 应用 Kullback-Leibler 散度量两个模型的相似性<sup>[4-6]</sup>, 其方法一般很难得到分布模型的准确估计<sup>[7-9]</sup>。非参数方法主要通过子空间<sup>[8,10]</sup>、流形<sup>[11-15]</sup>、仿射包<sup>[9,16]</sup>、凸包<sup>[16]</sup>、协方差

收稿日期: 2016-06-28; 定稿日期: 2016-09-13

基金项目: 广东省自然科学基金项目(2015A030313807); 广州市属高校科研项目(1201610059); 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); 广东省高等职业教育教学改革项目(201401181)

第一作者: 曾青松(1976-), 男, 湖南邵东人, 教授, 博士。主要研究方向为模式识别、数据挖掘。E-mail: qingsongzeng@163.com

矩阵<sup>[17-18]</sup>和支持向量域描述<sup>[19-20]</sup>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,将图像集投影到低维线性子空间,用典型相关度量两个子空间的相似性<sup>[10]</sup>,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<sup>[21]</sup>。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再生核希尔伯特空间,然后在新的空间中学习一个分类器<sup>[22]</sup>。协方差鉴别学习(covariance discriminative learning, CDL)<sup>[17]</sup>考虑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-欧式距离(Log-Euclidean distance, LED)<sup>[23]</sup>的核函数将这些矩阵映射到欧式空间上,并应用核鉴别分析进行分类。流形判别分析方法(manifold discriminant analysis, MDA)<sup>[15]</sup>依据最大化流形间隔完成流形判别分析的目的。虽然基于非线性流形的方法可以同时建模集合里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,这些方法达不到理想的效果。

在智能监控环境中,由于光照、人的姿态、表情的变化,构建的流形或子空间通常是异构的。研究表明,通过学习一个公共的子流形或子空间可以找到异构流形的共性<sup>[17,24-25]</sup>,并通过度量学习来解决异构空间样本的匹配问题,得到更加符合实际应用的距离度量。本文首先介绍格拉斯曼流形的基础,通过投影矩阵将视频图像特征向量张成的线性子空间投影到格拉斯曼流形表示为一个点,然后通过投影度量学习得到一个低维的、更加紧致的格拉斯曼流形,在低维的格拉斯曼流形上应用核鉴别分析并进行分类。

## 1 格拉斯曼流形

### 1.1 格拉斯曼流形的定义

格拉斯曼流形  $\mathcal{G}(m, D)$ <sup>[26]</sup>是  $D$  维空间  $\mathcal{R}^D$  的所有  $m$ -维线性子空间  $\mathcal{R}^m$  ( $0 < m < D$ ) 构成的集合,是  $D \times m$  正交矩阵等价类集合,  $\mathcal{G}(m, D) = \{Y | Y = XV, \forall V \in \mathcal{O}_k\}$ , 式中,  $X, Y$  是  $D \times m$  正交矩阵,  $V$  是  $m \times m$  正交矩阵,  $\mathcal{O}_k = \{Q | Q^T Q = I, \det(Q) = 1\}$  为正交群。

一个线性子空间投影到  $\mathcal{G}(m, D)$  上对应应该空间的一个点,如果格拉斯曼流形上的两个点能够通过正交变换相互获得时,那么这两个点是等价的。即矩阵  $W_1 = W_2 Q$ , 式中  $Q$  是  $r \times r$  正交矩阵,并且列空间  $\text{span}(W_1) = \text{span}(W_2)$ , 从而  $W_1$  与  $W_2$  等价。

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  $\text{span}(U)$ 。其是一个一对一、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。

### 1.2 投影度量

假定  $X_1$  和  $X_2$  为两个大小为  $D \times m$  的正交基矩阵,那么两个线性子空间  $\text{span}(X_1)$  和  $\text{span}(X_2)$  之间的主夹角  $0 = \theta_1 \cdots \theta_p = \frac{\pi}{2}$  可以递归的定义为

$$\begin{cases} \cos(\theta_i) = \max_{u_i \in \text{span}(X_2)} \max_{v_i \in \text{span}(X_1)} u_i^T v_i \\ \text{s.t. } u_i^T u_i = 1, v_i^T v_i = 1 \\ u_i^T u_j = 0, v_i^T v_j = 0 \quad (j = 1, 2, \dots, i-1) \end{cases}$$

其中,  $\cos(\theta_i), 0 < i < m$  称为典型相关系数。主角可以通过  $U_i^T U_j$  的奇异值分解来求得

$$U_i^T U_j = P(\text{diag}(\cos \theta_1, \dots, \cos \theta_p)) Q^T$$

其中,  $P, Q$  为正交矩阵。令  $\Theta = [\theta_1, \dots, \theta_p]$  表示两个线性子空间  $\text{span}(X_1)$  与  $\text{span}(X_2)$  之间的主夹角序列,其之间的测地距离定义为

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

流形是一个嵌入在高维欧式空间中的一个光滑的曲面,黎曼流形本身具有的测度且允许测得流形上两点之间的距离。但是,格拉斯曼流形是一个非线性的空间,大部分的欧式空间的技术都不能直接用于处理格拉斯曼流形上的数据。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,推导出距离形式为

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

式中,  $\|\cdot\|_F$  表示矩阵的 Frobenius 范数,定义为:

$$\|X\|_F^2 = \sum_{i=1}^D \sum_{j=1}^D |x_{ij}|^2$$

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<sup>[22]</sup>。投影距离逼近于格拉斯曼流形上的测地距离。

## 2 格拉斯曼流形度量学习

$D$  维空间  $\mathcal{R}^D$  所有的  $m$  维线性子空间  $\text{span}(U_i)$

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, 并反映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, 得到样本分布的一个更加紧致的描述。

### 2.1 投影矩阵学习

设  $U_i$  是格拉斯曼流形的正交基矩阵, 假设存在列满秩矩阵  $W, P=WW^T, P$  是一个秩为  $D \times D$  的对称半正定矩阵, 即  $P^T=P, P \geq 0$ 。定义  $V_i = W^T U_i$ , 那么  $V_i V_i^T = (W^T U_i)(W^T U_i)^T = W^T U_i U_i^T W$ ,  $V_i$  与  $V_j$  之间的投影距离定义为<sup>[27]</sup>

$$d_p(V_i V_i^T, V_j V_j^T) = \frac{1}{\sqrt{2}} \|V_i V_i^T - V_j V_j^T\|_F = \frac{1}{\sqrt{2}} \text{tr}(P A_{ij} A_{ij}^T P)$$

其中,  $A_{ij} = U_i U_i^T - U_j U_j^T$ , 与传统的马氏度量学习<sup>[28]</sup>类似, 这里的  $P$  可以看成是一个类马氏矩阵。

为了使得新的格拉斯曼流形上同类线性子空间的投影距离  $J_w(P)$  尽量地接近, 而异类的线性子空间之间的投影距离  $J_b(P)$  尽可能的被拉远, 设计目标函数  $J(P)$ , 求解最优的  $P$  矩阵<sup>[27]</sup>。

$$P^* = \arg \min_P (J_w(P) - \alpha J_b(P))$$

$$J_w(P) = \frac{1}{N_w} \sum_{i=1}^m \sum_{j: C_i=C_j} \frac{1}{\sqrt{2}} \text{tr}(P A_{ij} A_{ij}^T P)$$

$$J_b(P) = \frac{1}{N_b} \sum_{i=1}^m \sum_{j: C_i \neq C_j} \frac{1}{\sqrt{2}} \text{tr}(P A_{ij} A_{ij}^T P)$$

其中,  $N_w$  和  $N_b$  分别为来自同类样本对总数和来自异类样本对的总数;  $\alpha$  为权重参数;  $P$  是要学习的最优化的对称半正定矩阵, 问题的具体求解可阅读文献<sup>[27]</sup>。

### 2.2 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,

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

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

在投影映射  $\phi(U) = U U^T$  的框架下,  $U U^T$  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。通过投影度量学习得到投影矩阵  $P^* = W W^T$ , 其中  $W$  不一定是正交矩阵, 因此  $W^T U_i$  也不能保证是正交矩阵。

定理. 如果  $A = QR$  表示  $U_i$  的  $QR$  分解,  $A = [a_1, \dots, a_m]$ ,  $Q = [q_1, \dots, q_m]$  是列分块矩阵, 那么  $\text{span}\{a_1, \dots, a_k\} = \text{span}\{q_1, \dots, q_k\}$ ,  $k=1, \dots, m$ <sup>[29]</sup>。

对所有的  $i$ , 固定矩阵  $P$ , 通过将矩阵  $U_i$  正则化为  $U_i'$ , 可以使得  $W^T U_i'$  是正交基矩阵。计算  $W^T U_i$  的  $QR$  分解  $W^T U_i = Q_i R_i$ , 式中  $Q_i \in \mathcal{R}^{d \times m}$  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,  $R_i \in \mathcal{R}^{m \times m}$  是上三角可逆矩阵<sup>[29]</sup>。由于  $R_i$  可逆并且  $Q_i$  正交, 下式使得  $P^T U_i'$  是正交基矩阵

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

根据定理  $Q_i$  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。通过上述的变换得到了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, 当样本数量有限时, 低维逼近更容易得到准确的密度估计。

实际应用中, 可以直接使用查询集合或单独获取一组图像集合用于训练投影矩阵  $W$ , 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  $W$  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  $W^T U_i$  的  $QR$  分解  $W^T U_i = Q_i R_i$ , 得到降维后的线性子空间的正交基矩阵  $Q_i = W^T (U_i R_i^{-1}) \in \mathcal{R}^{d \times m}$ 。所有的  $m$  维线性子空间  $\text{span}(Q_i)$  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  $Q_i Q_i^T \in \mathcal{R}^{d \times d}$ 。

## 3 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, 投影核定义为<sup>[22]</sup>

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

核鉴别分析 (kernel discriminant analysis, KDA) 求解最优化目标函数

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

其中,  $K=[\kappa_{ij}]$ 是核矩阵,  $\kappa_{ij}=\kappa(Q_i, Q_j)$ , 相关系数  $H=[h_{ij}]$ 定义为: 当  $Q_i$  和  $Q_j$  同时属于第  $r$  类样本时,  $h_{ij}=1/n_r$ , 否则  $h_{ij}=0$ 。  $\alpha=[\alpha_1, \dots, \alpha_m]$ , 最优的  $\alpha$  可以通过求解特征值问题  $KHK\alpha=\lambda K\alpha$  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  $A$ 。

计算正交基  $U_i$  降维后得到的正交基矩阵  $Q_i$  在核空间上的投影  $B'_1, \dots, B'_n$ 。给定输入空间的一个样本  $Q_i \in \mathcal{R}^{d \times d}$ , 在鉴别子空间的  $q$  维投影可以通过公式  $B'_i = K_i A$  得到, 其中  $K_i = [\kappa_{i1}, \dots, \kappa_{im}]$ 。通过求解  $\arg \min_i \|B'_i - B'_i\|_F^2$  可以得到相应的类标  $i$ 。

算法 1 给出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  $X_t$ 。

##### 训练过程:

- (1) 计算所有训练图像集合 Train 的正交基矩阵;
- (2) 通过度量学习得到投影矩阵  $W$ ;
- (3) 对查询集合 Gallery, 计算每个集合的正交基矩阵  $U_i$  及  $W^T U_i$  的 QR 分解  $W^T U_i = Q_i R_i$ , 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  $KHK\alpha=\lambda K\alpha$ , 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  $B'_1, \dots, B'_n$ ;

(2) 对测试集  $X_t$ , 计算正交基矩阵  $U_t$ , 通过计算  $W^T U_t$  的 QR 分解  $W^T U_t = Q_t R_t$ , 得到低维格拉斯曼流形上的正交基矩阵  $Q_t$ ;

(3) 计算  $K_t$ , 通过公式  $B'_i = K_t A$  得到其在核空间上的投影;

(4) 求解  $\arg \min_i \|B'_i - B'_i\|_F^2$ , 得到类标  $i$ 。

##### 输出:

测试集  $X_t$  的类标。

## 4 实验分析

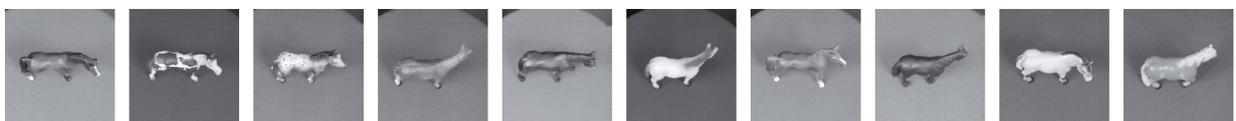
实验在 ETH80 对象数据库、ChokePoint 数据库、CMU Mobo 数据库和 YouTube Celebrities 视频数据库上进行基于集合的对象匹配和人脸识别实验。实验计算机配置为: Intel(R) Core (TM) i7-4710MQ CPU @ 2.5 GHz×2, 8 GB 内存, 安装 Windows 8.1, 程序使用 MATLAB 2013 编写。

### 4.1 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 个类别; 如图 1(b)所示, 其中每个类别中有 10 个对象, 总共 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取 41 个视图, 80 个对象一共 3 280 幅图片。图 1(b)显示每类物体的一个实例的不同视角下的 10 张图片。每一个实例的 41 张图片当成一个图像集合, 每类物体有 10 个图像集合。



(a) 数据库中的 8 个类别



(b) 每类 10 个对象样本

图 1 ETH80 对象分类数据库中的样本

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<sup>[30]</sup>, 该数据集包含门口 1 录制的 25 人的视频和门口 2 录制的 29 人的视频, 分别录制进入门口和离开门口的动作。每一个门口有 3 台摄像机在不同的角度同时录制, 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 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<sup>[31]</sup>。Mobo 数据库包含 25 人在跑步机上行走视频, 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。本文使用其中一个子集, 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。

YouTube Celebrities(YTC)数据库<sup>[32]</sup>收集来自于 YouTube 网站的 1 910 个视频片段, 一共包含 47 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, 每一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<sup>[33]</sup>提取相应的人脸图像。图 2 给出了“Donald Trump”的第 10 和 150 帧的跟踪结果。

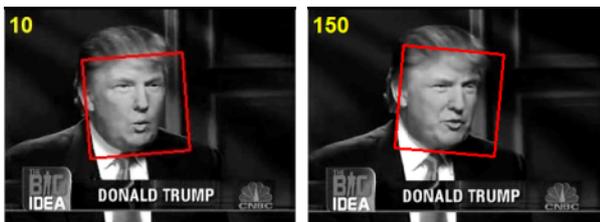


图 2 人脸跟踪检测示意图

#### 4.2 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, 人脸图像转换为灰度图, 经直方图均衡化处理后采样得到  $20 \times 20$  像素。具体参数设置如下: 互子空间方法(constrained mutual subspace method, CMSM)<sup>[34]</sup>、流形鉴别分析算法<sup>[15]</sup>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。协方差鉴别学习<sup>[17]</sup>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<sup>[35]</sup>提供的代码。格拉斯曼流形鉴别分析(Grassmann discriminant analysis, GDA)<sup>[22]</sup>算法采用投影核。投影度量学习(project metric learning, PML)<sup>[27]</sup>方法

在学习到的低维格拉斯曼流形上直接使用投影距离进行度量。

本文的方法使用子空间的正交基矩阵对图像集合建模, 将投影度量学习 PML 作为一个预处理过程对正交基矩阵降维得到一个低维的格拉斯曼流形, 在低维的格拉斯曼流形上应用 KDA 进行鉴别分析。

#### 实验结果及分析:

ETH80 数据集, 每一类对象有 10 个图像集, 实验随机选择 3、4、5 个图像集作为训练集合(即: 3 训练集合, 4 训练集合, 5 训练集合), 剩余的图像集作为测试集合, 实验重复 10 次, 表 1 报告了随机试验的平均识别率和方差, 图 3 显示了不同算法的识别率的对比曲线, 实验显示, 在 3 组随机实验中, 本文提出的方法都取得了最好的识别结果。

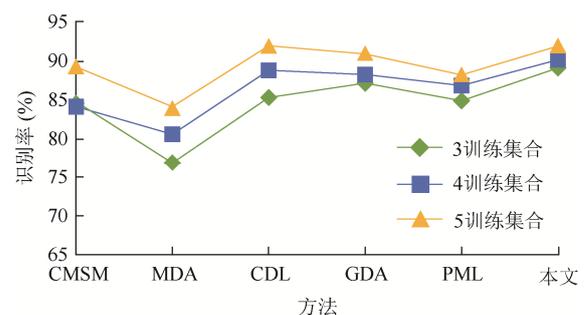


图 3 ETH80 数据集上的分类结果对比曲线

表 1 ETH80 数据集上的分类结果(%)

方法	分类准确率		
	3 训练集合	4 训练集合	5 训练集合
CMSM <sup>[10]</sup>	84.64±2.94	84.17±4.52	89.25±2.37
MDA <sup>[15]</sup>	76.96±4.56	80.62±3.63	84.00±4.11
CDL <sup>[17]</sup>	85.36±2.64	88.85±4.19	92.00±4.53
GDA <sup>[22]</sup>	87.14±3.12	88.33±2.23	91.00±2.41
PML <sup>[27]</sup>	85.00±4.47	86.88±3.81	88.25±4.37
本文方法	89.11±3.30	90.21±4.05	92.00±4.37

YTC 数据库部分人只有 1 段视频, 根据本文实验的需要, 选择 01 小节、需满足每人至少有 5 段视频子集, 满足条件的视频共有 14 人, 随机选择 2 段视频作为训练集, 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。所有算法报告了 5 次随机试验的平均结果。表 2 为在 ChokePoint、

Mobo 和 YTC 数据集上基于图像集合的人脸识别随机试验的平均识别率和方差。从实验结果可以看出,本文提出的方法在 ChokePoint 数据上仅次于 MDA;在 YTC 数据上仅次于 MDA 和 YTC,而在 Mobo 数据库上取得了最好的识别结果。

表 2 基于图像集合的人脸识别结果(%)

方法	分类准确率		
	ChokePoint	Mobo	YTC
CMSM <sup>[10]</sup>	70.40±3.26	97.14±2.32	86.79±0.00
MDA <sup>[15]</sup>	98.56±1.63	98.09±1.99	98.86±1.03
CDL <sup>[17]</sup>	88.08±4.10	96.32±3.02	85.28±1.57
GDA <sup>[22]</sup>	85.00±3.13	97.14±2.32	79.62±0.84
PML <sup>[27]</sup>	82.16±4.18	98.37±1.70	93.21±1.03
本文方法	91.36±2.84	99.18±1.11	92.45±0.00

表 3 为 ChokePoint、CMU Mobo 和 YTC 数据集上各个算法的平均运算时间。由于本文方法使用了 PML 作为预处理过程,然后在低维的格拉斯曼流形上应用 GDA 算法。本质上讲,本文方法结合了 PML 与 GDA 算法。从实验结果分析,改进的算法识别率比单独使用 GDA 和 PML 均有较大提高,但是由于需要学习度量矩阵,因此较其他算法耗时。但是大部分的时间花在距离学习上,而学习过程可以离线进行。因此,实际应用中,测试阶段消耗的时间会大幅地缩减。

表 3 对比算法的平均运算时间(s)

方法	平均运算时间		
	ChokePoint	Mobo	YTC
CMSM <sup>[10]</sup>	154.5	167.1	117.5
MDA <sup>[15]</sup>	45.1	158.3	113.0
CDL <sup>[17]</sup>	411.9	2750.0	213.9
GDA <sup>[22]</sup>	18.8	110.3	11.9
PML <sup>[27]</sup>	189.2	793.1	91.9
本文方法	198.0	845.5	95.5

## 5 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,并将其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。通过投影度量学习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,从而更好地刻画集合的真实分布。本文的方法整合了度量学习和格拉斯曼流形鉴别分析方法,并且取得比单独应用 GDA 和 PML 更好地识别准确率。根据本文的思路,可以对现有的一些经典的基于图像集合的匹配方法进行改造,通过度量学习得到一个领域相关的度量矩阵,在降低空间的维数得到数据更好的描述的同时改进现有算法的精度。

## 参考文献

- [1] HUANG Z W, SHAN S G, WANG R P, et al. A benchmark and comparative study of video-based face recognition on COX face database [J]. IEEE Transactions on Image Processing, 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, World Scientific, 2012, 26(5): 1-56.
- [3] WANG H, WANG Y, CAO Y. Video-based face recognition: a survey [J]. World Academy of Science, Engineering and Technology, 2009, 2: 293-302.
- [4] 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. New York: IEEE Press, 2011: 529-534.
- [5] ARANDJELOVIC O, CIPOLLA R. Face recognition from face motion manifolds using robust kernel resistor-average distance [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshop. New York: IEEE Press, 2004: 88.
- [6] HANDANDI M, SALZMANN M, BAKTASHMOTLAGH M. Beyond Gauss: image-set matching on the Riemannian manifold of PDFs [C]//IEEE Conference on Computer Vision (ICCV 2015). New York: IEEE Press, 2015: 1-19.
- [7] WANG R, SHAN S, CHEN X, et al. Manifold-manifold distance and its application to face recognition with image sets [J]. IEEE Transactions on Image Processing, 2012, 21(10): 4466-4479.
- [8] KIM T K, KITTLER J, CIPOLLA R. Discriminative learning and recognition of image set classes using canonical correlations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1005-1018.
- [9] HU Y, MIAN A S, OVENS R. Sparse approximated nearest points for image set classification [C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. New York: IEEE Press, 2011: 121-128.
- [10] YAMAGUCHI O, FUKUI K, MAEDA K. Face recognition using temporal image sequence [C]//3rd International Conference on Face & Gesture Recognition. New York: IEEE Press, 1998: 318-323.
- [11] 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). Los Alamitos: IEEE Computer Society Press, 2006, 2: 1384-1390.
- [12] 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.
- [13] 曾青松. 保局投影在图像集匹配中的应用[J]. *计算机应用与软件*, 2015, 6(6): 304-307.
- [14] 曾青松. 黎曼流形上的保局投影在图像集匹配中的应用[J]. *中国图象图形学报*, 2014, 19(3): 414-420.
- [15] WANG R, CHEN X. Manifold discriminant analysis [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2009: 429-436.
- [16] CEVIKALP H, TRIGGS B, TRIGGS W. Face recognition based on image sets [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2010, 365(2): 2567-2573.
- [17] WANG R, GUO H, DAVIS L S, et al. Covariance discriminative learning: a natural and efficient approach to image set classification [C]//*Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on. New York: IEEE Press, 2012: 2496-2503.
- [18] 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. New York: IEEE Press, 2013: 73-80.
- [19] 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.
- [20] 曾青松. 基于支持向量域描述的图像集匹配[J]. *模式识别与人工智能*, 2014, 8(8): 735-740.
- [21] YANG M, ZHU P, VANGOOL L, et al. Face recognition based on regularized nearest points between image sets [C]//2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). New York: IEEE Press, 2013: 1-7.
- [22] 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*. New York: ACM Press, 2008: 376-383.
- [23] ARSIGNY V, COMMOWICK O, PENNEC X, et al. A log-euclidean framework for statistics on diffeomorphisms [C]//*Medical Image Computing and Computer-Assisted Intervention-MICCAI 2006*. Berlin: Springer, 2006: 924-931.
- [24] HUANG Z, WANG R, SHAN S, et al. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification [J]. *JMLR Workshop and Conference Proceedings*, 2015, 37: 720-729.
- [25] 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.
- [26] HARDNI M T, SANDERSON C, SHIRAZI S, et al. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching [C]//*IEEE Conference on Computer Vision and Pattern Recognition Workshop*. New York: IEEE Press, 2011: 2705-2712.
- [27] HUANG Z, WANG R, SHAN S. et al. Projection metric learning on Grassmann manifold with application to video based face recognition [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2015: 140-149.
- [28] 沈媛媛, 严严, 王菡子. 有监督的距离度量学习算法研究进展[J]. *自动化学报*, 2014, 40(12): 2673-2686.
- [29] 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.
- [30] 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*. New York: IEEE Press, 2011: 74-81.
- [31] GROSS R, SHI J. The CMU motion of body (MoBo) database [R]. Pittsburgh, PA: -, 2001(CMU-RI-TR-01-18).
- [32] 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)*. Los Alamitos: IEEE Computer Society Press, 2008: 1-8.
- [33] ROSS D, LIM J, YANG M H. Adaptive probabilistic visual tracking with incremental subspace update [C]//*Computer Vision-ECCV 2004, 8th European Conference on Computer Vision*. Prague, Czech Republic: Springer, 2004: 470-482.
- [34] 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.
- [35] CAI D, HE X, HAN J. Speed up kernel discriminant analysis [J]. *The International Journal Very Large Data Bases*, 2011, 20(1): 21-33.