

# 应用于特征提取方法的模糊差分嵌入投影

万鸣华<sup>1,2</sup>

(1. 南京审计大学工学院, 南京, 211815; 2. 东南大学信息科学与工程学院, 南京, 210096)

**摘要:** 局部线性嵌入(Locally linear embedding, LLE)算法对于分类的结果没有直接的关系。同时, 该算法受不同表情、光照以及姿态等因素的影响, 识别的效果会大大降低。为了能够很好地解决上述问题, 提出基于模糊的差分嵌入投影(Fuzzy difference embedding projection, FDEP)特征提取算法。FDEP算法首先在模糊数学的思想指导下, 通过模糊隶属度(Fuzzy sets)的形式表示; 然后分别构造模糊局部近邻图与模糊全局方差图来表征局部与全局结构信息, 采用最大间距准则函数来构造目标函数避免“小样本”问题; 最后, 通过拉格朗日乘子解决约束条件下的优化问题。FDEP算法既可以最大化地模糊全局数据之间的非局部散度, 又可以保持模糊近邻数据之间的内在联系。在 ORL, Yale 和 AR 人脸图像库的实验结果表明, FDEP 算法具有较好的识别性能。

**关键词:** 特征提取; 局部线性嵌入; 最大间距准则; 模糊隶属度; 流形

**中图分类号:** TP391.41      **文献标志码:** A

## Fuzzy Difference Embedding Projection for Feature Extraction

Wan Minghua<sup>1,2</sup>

(1. School of Technology, Nanjing Audit University, Nanjing, 211815, China; 2. School of Information Science and Engineering, Southeast University, Nanjing, 210096, China)

**Abstract:** Locally linear embedding (LLE) algorithm has no direct relationship with the classification. Meanwhile, the recognition effect is decreased when the LLE algorithm is affected by different facial expressions, illumination and pose, etc., and the distribution of the original sample is usually nonlinear and complex. Therefore, an efficient dimensional reduction and classification algorithm is presented, that is fuzzy difference embedding projection (FDEP) algorithm. The FDEP algorithm constructs different radiograms to characterize the local and the global structure information using fuzzy membership degree (fuzzy sets) under fuzzy thinking, and then uses the maximum margin criterion (MMC) to construct the objective function for avoiding the "small-size sample" problem. Finally, the algorithm solves the constrained optimization by Lagrange operators. The FDEP algorithm maintains the original neighbor relations for neighboring data points of the same class and is also crucial to keep away neighboring data points of different classes. The results of face recognition experiments on ORL, Yale and AR face databases demonstrate the effectiveness of the FDEP algorithm.

**Key words:** feature extraction; locally linear embedding; maximum margin criterion; fuzzy sets; manifold learning

## 引 言

目前,众多的特征提取算法在模式识别、机器学习和数据挖掘的研究领域中受到了广泛的关注。在这些特征提取算法中,最为经典的特征提取算法有线性鉴别分析方法(Linear discriminant analysis, LDA)<sup>[1]</sup>和主成分分析(Principle component analysis, PCA)<sup>[2]</sup>两种。这两种特征提取算法在各类图像识别中取得了较好的识别效果。PCA和LDA算法所得到的图像特征是线性的特征,因此,它们本质上都是线性的特征提取方法。

为了发现非线性数据中的流形结构,众多学者已经提出了许多流形学习相关算法。目前,流形学习的研究方法有主流形<sup>[3]</sup>、谱分析<sup>[4]</sup>、变分法<sup>[5]</sup>和互信息<sup>[6]</sup>。当前基于流形学习的特征提取算法有等距离映射(Isometric mapping, ISOMAP)<sup>[7]</sup>、局部线性嵌入(Locally linear embedding, LLE)<sup>[8]</sup>和拉普拉斯算子特征映射(Laplacian eigemap, LE)<sup>[9]</sup>等。

由于样本缺少泛化能力,上述的流形学习以及拓展方法不适合应用在图像识别中。为了解决这个问题,He<sup>[10]</sup>等人提出了局部保持投影(Locality preserving projections, LPP)方法并将其成功应用于人脸识别。由于LPP算法是无监督的,众多研究者通过增加样本类别信息相继提出了边缘Fisher分析(Marginal fisher analysis, MFA)<sup>[11]</sup>和局部鉴别嵌入(Local discriminant embedding, LDE)<sup>[12]</sup>等相关的学习算法<sup>[13]</sup>。

上述讨论的流形学习以及拓展方法只考虑流形上的邻近点,没有考虑全局流形结构几何特性。即当流形上距离比较远的数据样本点之间的联系较弱或者邻域之间的数据样本点交叠不够的时候,会出现高维空间数据样本点映射错误的情况。同时不同人脸之间在高维空间也存在结构相似的问题,导致识别率降低。解决该问题比较有效的方法有Fuzzy Fisherface<sup>[14]</sup>方法,但是该方法仅在计算样本类内中心样本时考虑了样本分布信息,并没有充分利用样本分布信息。

针对上述问题,在模糊数学和流形学习的思想指导之下,通过模糊隶属度(Fuzzy sets)<sup>[15]</sup>和局部线性嵌入<sup>[16]</sup>表示的形式,将测试数据样本重新划入所在的类别之中,提出一种新的局部与全局保持嵌入(Fuzzy difference embedding projection, FDEP)算法。FDEP算法用模糊隶属度来表示原始样本的分布信息;其次,利用局部线性嵌入分别构建模糊局部邻图与模糊全局方差图来表示局部邻与全局结构信息;最后,通过拉格朗日乘法解决了FDEP算法约束条件下的优化问题。在ORL, Yale和AR三个标准人脸库上的实验结果验证了FDEP算法的有效性。

## 1 相关算法

将数据样本集 $\mathbf{X}=\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{x}_i \in \mathbf{R}^n$ 从高维特征空间中寻求一个最优的投影矩阵 $\mathbf{V}=[\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$ 映射到一个相对低维的特征空间 $\mathbf{R}^d$ ,  $d \leq n$ 。这样, $\mathbf{Y}=\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ ,  $\mathbf{y}_i = \mathbf{V}^T \mathbf{x}_i$ 可以表述为在新特征空间中的数据样本,其中投影矩阵 $\mathbf{V}$ 由 $d$ 个最佳鉴别矢量 $\mathbf{v}_i$ 构成。

### 1.1 局部线性嵌入(LLE)

LLE算法分为如下3个步骤:

**Step 1** 计算数据样本集 $\mathbf{X}$ 中 $\mathbf{x}_i$ 的 $k$ 个邻域点 $\{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_k}\}$ , 其中 $k$ 为大于零的常数。

**Step 2** 通过定义误差函数式(1)来计算数据样本点的局部重建权重矩阵 $\mathbf{W}_{ij}$

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

式中:  $\sum_{j=1}^k W_{ij} = 1$ , 当  $\mathbf{x}_j \notin \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik}\}$  时,  $W_{ij} = 0$ 。

**Step 3** 通过最小化式(2)将所有的样本点从高维空间映射到低维空间中。映射条件为

$$\epsilon(\mathbf{Y}) = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^k W_{ij} \mathbf{y}_j \right\|^2 \quad (2)$$

根据式(2)求得  $\mathbf{x}_i$  的最优  $d$  维投影向量  $\mathbf{y}_i$ 。必须同时满足  $\sum_{i=1}^n \mathbf{y}_i = 0$  以及  $\frac{1}{n} \mathbf{Y} \mathbf{Y}^T = \mathbf{I}$  来保证式(2)有唯一解。根据 Rayleitz-Ritz 定理<sup>[13]</sup>求解  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ 。

### 1.2 模糊 $K$ 近邻方法

设  $\mathbf{U} = [\mathbf{U}_{ij}]$  为模糊隶属度矩阵, 其中  $i=1, 2, \dots, c, j=1, 2, \dots, N$ , 矩阵  $\mathbf{U}$  中的元素  $U_{ij}$  表示第  $j$  个训练样本对于第  $i$  类的依赖程度, 并且  $U_{ij}$  必须满足式(3,4)的两个条件。

$$\sum_{i=1}^c U_{ij} = 1 \quad (3)$$

$$0 < \sum_{j=1}^n U_{ij} < N \quad (4)$$

可以通过下面 3 个步骤得到矩阵  $U_{ij}$ 。

**Step 1** 构成一个  $N \times N$  的距离矩阵。通过计算训练数据样本集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}, \mathbf{x}_i \in \mathbf{R}^n$  为任意两个数  $\mathbf{x}_i$  和  $\mathbf{x}_j$  距样本点之间的欧氏距离。

**Step 2** 将步骤 1 中得到的  $N \times N$  距离矩阵的对角线上的元素置为无穷大, 并将距离矩阵的每一列或者每一行按照距离值从小到大排列。

**Step 3** 隶属度矩阵计算。根据式(5)计算第  $j$  个样本跟第  $i$  类的隶属度。

$$U_{ij} = \begin{cases} 0.51 + 0.49 \times (n_{ij}/k) & i \text{ 等于 } j \text{ 所属的类别} \\ 0.49 \times (n_{ij}/k) & \text{其他} \end{cases} \quad (5)$$

式中  $n_{ij}$  表示第  $j$  个样本的  $k$  个最近邻数据样本点中属于第  $i$  类的数据样本个数。

## 2 模糊的嵌入差分投影

由于 PCA 算法是一种基于全局结构信息的方法, 不能发现数据样本点的局部结构信息。而 LLE 算法是通过邻域内其他数据样本点的线性表示, 不能发现高维数据样本点的全局结构信息。因此, 本文利用 PCA 和 LLE 的优势, 结合模糊隶属度, 分别构造模糊全局方差图与模糊局部近邻图来表示全局与局部近邻结构信息。约束条件下 FDEP 目标函数的优化问题通过拉格朗日乘法来解决。FDEP 目标函数主要通过以下 3 个方面来实现。

### 2.1 模糊局部最小嵌入

模糊局部最小嵌入算法可分为 4 步来描述, 具体如下:

**Step 1** 计算每个测试样本数据点  $\mathbf{x}_i$  的  $k$  个距离最近的点集合  $\{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik}\}$ ;

**Step 2** 通过最小重构误差式(6)计算重建每个数据样本点  $\mathbf{x}_i$  的权值  $\mathbf{W} = [\tau_{ij}]_{N \times N}$ 。

$$\min J_L(\mathbf{W}) = \sum_{i=1}^N \left\| \mathbf{x}_i - \sum_{j=1}^N \tau_{ij} \mathbf{x}_j \right\|^2 \quad (6)$$

式(6)的最小重构误差可以化简如下

$$\xi_i = \left\| \mathbf{x}_i - \sum_{j=1}^N \tau_{ij} \mathbf{x}_j \right\|^2 = \left\| \sum_{j=1}^N \tau_{ij} (\mathbf{x}_i - \mathbf{x}_j) \right\|^2 = \sum_{j=1}^N \tau_{ij} (\mathbf{x}_i - \mathbf{x}_j) \sum_{h=1}^N \tau_{ih} (\mathbf{x}_i - \mathbf{x}_h) = \sum_{j=1}^N \sum_{t=1}^N \tau_{ij} \tau_{it} \mathbf{G}_{jt}^i \quad (7)$$

式中  $\mathbf{G}_{ji}^i = (\mathbf{x}_i - \mathbf{x}_j)^\top (\mathbf{x}_i - \mathbf{x}_i)$  称为局部 Gram 矩阵。最优系数  $w_{ij}$  可以通过函数  $\sum_{j=1}^N w_{ij} = 1$  求解最小二次平方得到。

$$w_{ij} = \frac{\sum_{i=1}^N (\mathbf{G}^i)_{ji}^{-1}}{\sum_{p=1}^N \sum_{q=1}^N (\mathbf{G}^i)_{pq}^{-1}} \quad (8)$$

重复 Step 1 和 Step 2 后得到权重矩阵  $\mathbf{W} = [w_{ij}]_{N \times N}$ 。

**Step 3** 计算模糊权值  $w_{ij}^G$

$$w_{ij}^G = \begin{cases} u_{ij}^G \mathbf{W} & i \in \{\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik}\} \\ 0 & \text{其他} \end{cases}$$

$$\text{其中 } u_{ij}^G = \begin{cases} 0.51 + 0.49 \times (\frac{n_{ij}}{k}) \\ 0.49 \times (\frac{n_{ij}}{k}) \end{cases}.$$

**Step 4** 将所有数据样本  $N$  映射到低维特征空间中, 根据权值  $w_{ij}^G$ , 通过映射条件为最小化式(9), 求得  $\mathbf{y}_i$ 。

$$\min J_L(\mathbf{Y}) = \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^N w_{ij}^G \mathbf{y}_j \right\|^2 \quad (9)$$

最后, 式(9)最小重构函数通过一个线性变换  $\mathbf{Y} = \mathbf{V}^\top \mathbf{X}$  化简如下

$$\begin{aligned} J_L(\mathbf{U}) &= \sum_{i=1}^N \left\| \mathbf{y}_i - \sum_{j=1}^N w_{ij}^G \mathbf{y}_j \right\|^2 = \sum_{i=1}^N \text{tr} \left\{ (\mathbf{y}_i - \sum_{j=1}^N w_{ij}^G \mathbf{y}_j) (\mathbf{y}_i - \sum_{j=1}^N w_{ij}^G \mathbf{y}_j)^\top \right\} = \\ &= \text{tr} \left\{ \sum_{i=1}^N (\mathbf{y}_i - \sum_{j=1}^N w_{ij}^G \mathbf{y}_j) (\mathbf{y}_i - \sum_{j=1}^N w_{ij}^G \mathbf{y}_j)^\top \right\} = \text{tr} \{ \mathbf{Y} (\mathbf{I} - (\mathbf{W}^G)^\top) (\mathbf{I} - (\mathbf{W}^G)^\top)^\top \mathbf{Y}^\top \} = \\ &= \text{tr} \{ \mathbf{Y} (\mathbf{I} - \mathbf{W}^G)^\top (\mathbf{I} - \mathbf{W}^G) \mathbf{Y}^\top \} = \text{tr} \{ \mathbf{V}^\top \mathbf{X} \mathbf{M} \mathbf{X}^\top \mathbf{V} \} \end{aligned} \quad (10)$$

式中  $\mathbf{M} = (\mathbf{I} - \mathbf{W}^G)^\top (\mathbf{I} - \mathbf{W}^G)$ 。

## 2.2 模糊全局最大方差

数据样本最大全局散度的目标函数为

$$\max J_G(\mathbf{Y}) = \sum_{i=1}^N \left\| \mathbf{y}_i - \bar{\mathbf{y}} \right\|^2 \quad (11)$$

通过一个线性变换  $\mathbf{Y} = \mathbf{V}^\top \mathbf{X}$ , 式(10) 最大全局散度目标函数可以化简如下

$$J_G(\mathbf{U}) = \sum_{i=1}^N \left\| \mathbf{y}_i - \bar{\mathbf{y}} \right\|^2 = \sum_{i=1}^N \left\| \mathbf{V}^\top (\mathbf{x}_i - \bar{\mathbf{x}}) \right\|^2 = \sum_{i=1}^N \text{tr} \{ \mathbf{V}^\top (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^\top \mathbf{V} \} = \text{tr} \{ \mathbf{V}^\top \mathbf{S}_T \mathbf{V} \} \quad (12)$$

式中  $\bar{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i$ ,  $\bar{\mathbf{x}} = \frac{1}{C} \sum_{i=1}^C \bar{\mathbf{m}}_i$ ,  $\bar{\mathbf{m}}_i = \frac{\sum_{j=1}^m u_{ij} \mathbf{x}_j^i}{\sum_{j=1}^m u_{ij}}$ ,  $\mathbf{S}_T = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^\top$  为全局协方差矩阵。

## 2.3 FDEP 准则

通过同时满足 2.1 中的最小模糊局部散度  $J_L(\mathbf{W})$  和 2.2 中的最大模糊全局散度  $J_G(\mathbf{V})$  可以找到一个最优的投影方向  $\mathbf{V}$ , 因此可以采用  $J_L(\mathbf{W})$ ,  $J_G(\mathbf{V})$  和约束条件来解决最优问题

$$\begin{cases} \min \text{tr} \{ \mathbf{V}^\top \mathbf{X} \mathbf{M} \mathbf{X}^\top \mathbf{V} \} \\ \max \text{tr} \{ \mathbf{V}^\top \mathbf{S}_T \mathbf{V} \} \\ \text{s. t. } \mathbf{V}^\top \mathbf{X} \mathbf{D} \mathbf{X}^\top \mathbf{V} = \mathbf{I} \end{cases} \quad (13)$$

由式(13)可以看到,FDEP算法不仅能最大化非近邻样本点之间的距离,也能保持近邻数据样本点之间内在联系。为了避免类似LDA中出现的小样本问题,式(13)可以转化成以下形式

$$\begin{aligned} \min \operatorname{tr}\{\mathbf{V}^T((1-\alpha)\mathbf{XMX}^T - \alpha\mathbf{S}_T)\mathbf{V}\} \\ \text{s. t. } \mathbf{V}^T\mathbf{XDX}^T\mathbf{V} = \mathbf{I} \end{aligned} \quad (14)$$

式中 $\alpha(0 \leq \alpha < 1)$ 为平衡参数,可以利用Lagrange乘数法来解决式(14)最小化问题,即

$$\frac{\partial}{\partial \mathbf{V}} \operatorname{tr}\{\mathbf{V}^T((1-\alpha)\mathbf{XMX}^T - \alpha\mathbf{S}_T)\mathbf{V}_i - \lambda_i(\mathbf{v}_i^T\mathbf{XDX}^T\mathbf{v}_i - \mathbf{I})\} = 0 \quad (15)$$

令 $\frac{\partial L(\mathbf{V}, \lambda)}{\partial \mathbf{V}} = 0$ , 得到

$$[(1-\alpha)\mathbf{XMX}^T - \alpha\mathbf{S}_T]\mathbf{v}_i = \lambda_i\mathbf{XDX}^T\mathbf{v}_i \quad (16)$$

式中 $\lambda_i$ 和 $\mathbf{v}_i$ 分别为 $\mathbf{X}(\mathbf{M} - \mu\mathbf{L}^b)\mathbf{X}^T$ 和 $\mathbf{XX}^T$ 的特征值及其对应的特征向量。所以, $\mathbf{V}_{\text{ULDP}} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_d]$ 由式(13)的特征方程分解的前 $d$ 个最小特征向量组成。

### 3 不同算法的相关性分析

通过对FDEP算法推导,可以看出FDEP算法形式上与LLE和NPE算法非常相似,但是又明显不同。下面分别讨论LLE,NPE和FDEP算法之间的相关性。

#### 3.1 LLE和NPE算法相关性分析

NPE算法是LLE算法的线性化。LLE算法对于新的测试数据样本集不能合适地嵌入,这是由于LLE算法仅仅定义在训练数据样本集上。相反,NPE算法则定义在训练数据样本集和测试数据样本集上。

在LLE算法中,矩阵 $\mathbf{XMX}^T$ 是半正定化和对称的,如下所示

$$\mathbf{YY}^T = \mathbf{I} \Rightarrow \mathbf{V}^T\mathbf{XX}^T\mathbf{V} = \mathbf{I} \quad (17)$$

最优投影矩阵 $\mathbf{V}$ 可以简化式(18)得到。

$$\min_{\mathbf{V}^T\mathbf{XX}^T\mathbf{V}=\mathbf{I}} \operatorname{tr}\{\mathbf{V}^T\mathbf{XMX}^T\mathbf{V}\} \quad (18)$$

即求解式(19)的广义特征方程前 $d$ 个最小的特征值所对应的特征向量可以得到投影矩阵 $\mathbf{V}$ 。

$$\mathbf{XMX}^T\mathbf{V} = \lambda_i\mathbf{XX}^T\mathbf{V} \quad (19)$$

#### 3.2 NPE和FDEP算法相关性分析

FEDP算法同时保持全局方差和局部近邻嵌入。因此,总体散度矩阵 $\mathbf{S}_T$ 也可以写成如下形式

$$\mathbf{S}_T = \frac{1}{2NN} \sum_{i,j} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T = \frac{1}{N}\mathbf{X}\left(\mathbf{I} - \frac{1}{N}\mathbf{ee}^T\right)\mathbf{X}^T \quad (20)$$

FEDP算法也可以写成如下带有约束最小化的问题

$$\min_{\mathbf{V}^T\mathbf{XDX}^T\mathbf{V}=\mathbf{I}} \operatorname{tr}\left\{\mathbf{V}^T\mathbf{X}\left((1-\alpha)\mathbf{M} + \alpha\left(\frac{1}{N^2}\mathbf{ee}^T - \frac{1}{N}\mathbf{I}\right)\right)\mathbf{X}^T\mathbf{V}\right\} \quad (21)$$

因此有

$$\hat{\mathbf{M}}\mathbf{X}^T\mathbf{v}_i = \lambda_i\mathbf{XDX}^T\mathbf{v}_i \quad (22)$$

式中: $\hat{\mathbf{M}} = (1-\alpha)\mathbf{M} + \alpha\tilde{\mathbf{M}}, \tilde{\mathbf{M}} = \frac{1}{N^2}\mathbf{ee}^T - \frac{1}{N}\mathbf{I}$ 。当 $\alpha=0, \mathbf{D}=\mathbf{I}$ 时,NPE算法是FDEP算法一个特例。

#### 3.3 LLE,NPE和FDEP算法相关性分析

NPE算法的本质是LLE算法的线性化,而FDEP算法是全局PCA算法和局部LLE算法的结合。式(22)中 $\hat{\mathbf{M}}$ 包含的 $\mathbf{M}$ 存在保持局部的性质,同时 $\tilde{\mathbf{M}}$ 存在全局的信息。FDEP算法整合局部和全局的信息用不同边缘权重排列的方法建立一个新的图像。FDEP算法比LLE算法和NPE算法具有更好的鲁

棒性和鉴别性。

## 4 实验

为了验证所提的 FDEP 算法与 PCA、LDA、LLE 和 MMC 等算法的分类识别性能,分别在 ORL、Yale 和 AR 标准人脸图像库上进行实验。所有算法均采用欧氏距离来度量和最近邻分类器进行分类,用 PCA 算法做预处理,并保持约 98% 的图像能量。实验环境:HP PC, CPU 为 Inter Athlon(tm) 64 Processor, 内存 2 GB, Matlab 7.01b。

### 4.1 实验设计

(1) ORL 标准人脸库(<http://www.uk.research.att.com/facedatabase.html>) 总共有 400 张人脸, 分别由 40 人构成, 每人有 10 幅图像, 大小为  $56 \times 46$  灰度。图 1 是 ORL 标准人脸库中其中一个人的 10 幅图像。



图 1 ORL 标准人脸库中一个人的十张图

Fig. 1 Ten sample images of one person on the ORL face database

在 ORL 实验中随机选取  $l$  ( $l=2,3,4,5,6$ ) 幅人脸图像做为训练集, 剩余  $(10-l)$  幅人脸图像做为测试集。每次实验都运行 50 次, 最后的识别结果是 50 次的平均识别结果。

(2) Yale 标准人脸图像库(<http://www.cvc.yale.edu/projects/yalefaces/yalefaces.html>) 总共包含 165 幅灰度人脸图像, 由 15 个人构成, 每个人又包含 11 幅人脸图像。这些人脸照片含有不同的表情变化和光照变化。在 Yale 实验中, 图像被处理成  $50 \times 40$  灰度的形式。图 2 显示 Yale 标准人脸图像库中其中一个人的 11 幅图像。



图 2 Yale 标准人脸库一个人的 11 张图

Fig. 2 Eleven sample images of one person on the Yale face database

在 Yale 实验中随机选取  $l$  ( $l=2,3,4,5,6$ ) 幅人脸图像做为训练集, 剩余  $(11-l)$  幅人脸图像做为测试集。对于每次选取  $l$  图像运行 50 次实验, 最后的识别结果是 50 次的平均识别结果。

(3) AR 标准人脸图像库([http://cobweb.ecn.purdue.edu/~aleix/aleix\\_face\\_DB.html](http://cobweb.ecn.purdue.edu/~aleix/aleix_face_DB.html)) 总共包含 4 000 多张彩色人脸图像, 由 126 个人(70 位男性, 56 位女性)构成, 这些人脸图像在不同的光照、表情和遮挡情况下进行拍摄。在本次实验中采用 2 400 幅人脸图像, 图像被处理成  $50 \times 40$  灰度的形式。这些图像包含 120 个人(没戴围巾的)的人脸图像, 其中每人又包含 20 幅人脸图像。图 3 是 AR 标准人脸图像库中一个子类的 20 幅图像。

在 AR 实验中随机选取  $l$  ( $l=2,3,4,5,6$ ) 幅人脸图像做为训练集, 剩余  $(20-l)$  幅人脸图像做为测试集。对于每次选取  $l$  图像运行 10 次实验, 最后的识别结果是 10 次实验的平均识别结果。

### 4.2 实验结果

表 1~3 为 5 种算法在 ORL、Yale 和 AR 人脸库上的最大平均识别率及其对应的每个类所用训练

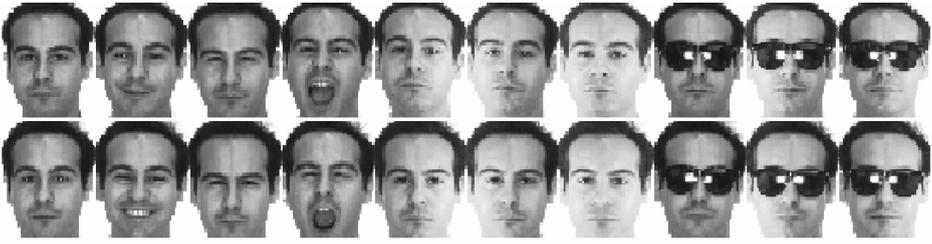


图 3 AR 人脸图像库中一个子类的 20 幅图像

Fig. 3 Images of one subject of the AR database

表 1 5 种算法在 ORL 标准人脸库上的最大平均识别结果及偏差

Tab. 1 The max average recognition accuracy of five algorithms on the ORL face database and the corresponding standard deviations and dimensions

Method	<i>l</i>					%
	2	3	4	5	6	
PCA	72.25±2.61 (22)	81.62±0.65 (22)	83.98±1.00 (22)	85.41±1.30 (22)	85.41±1.98 (22)	
LDA	76.35±1.07 (28)	83.40±1.69 (28)	84.13±2.03 (28)	86.07±1.16 (28)	88.60±0.78 (28)	
LLE	69.78±0.82 (32)	72.66±0.73 (30)	76.32±1.14 (36)	78.85±1.15 (20)	88.44±1.55 (28)	
MMC	75.01±1.77 (26)	83.96±0.48 (28)	84.37±2.48 (28)	87.47±0.85 (28)	90.49±1.17 (26)	
FDEP	79.35±0.78 (16)	84.85±0.82 (22)	88.76±1.12 (22)	89.30±0.82 (22)	93.63±0.62 (18)	

表 2 5 种算法在 Yale 人脸库上的最大平均识别结果及偏差

Tab. 2 The max average recognition accuracy of five algorithms on the Yale face database and the corresponding standard deviations and dimensions

Method	<i>l</i>					%
	2	3	4	5	6	
PCA	77.29±1.20 (18)	80.20±1.27 (20)	83.99±1.38 (18)	84.72±1.24 (20)	86.21±0.80 (22)	
LDA	80.45±1.48 (8)	84.55±1.06 (8)	87.85±0.45 (8)	87.20±1.64 (8)	88.54±0.82 (8)	
LLE	83.55±1.38 (14)	83.58±0.59 (6)	85.90±0.75 (6)	88.37±1.63 (8)	88.97±1.56 (8)	
MMC	80.58±0.71 (12)	82.84±0.88 (6)	85.87±1.12 (6)	86.83±0.37 (6)	87.79±0.50 (6)	
FDEP	85.85±1.89 (18)	88.15±0.94 (18)	91.50±1.81 (18)	93.35±1.25 (14)	94.52±1.46 (16)	

表 3 5 种算法在 AR 人脸库上的最大平均识别结果及偏差

Tab. 3 The max average recognition accuracy of five algorithms on the AR face database and the corresponding standard deviations and dimensions

Method	<i>l</i>					%
	2	3	4	5	6	
PCA	66.68±1.11 (85)	70.21±1.62 (85)	77.60±1.27 (85)	79.03±0.81 (80)	81.54±1.04 (85)	
LDA	71.50±0.71 (70)	75.58±0.76 (70)	82.53±0.81 (70)	87.12±0.33 (70)	87.58±0.75 (70)	
LLE	70.40±0.89 (85)	74.31±1.16 (75)	83.35±1.49 (70)	84.78±0.76 (80)	86.98±0.40 (85)	
MMC	69.39±1.15 (80)	75.71±0.59 (80)	82.98±0.62 (85)	85.27±0.94 (80)	87.86±0.82 (80)	
FDEP	73.13±1.25 (95)	78.86±0.98 (85)	86.72±1.63 (85)	89.50±1.16 (80)	93.85±0.92 (80)	

样本数目(括号内的数字代表最大平均识别率所对应的特征向量维数)。

图 4 为在 ORL、Yale 和 AR 标准人脸图像库上 5 种算法的平均识别率随特征向量维数的变化关

系。图 4(a)显示在 ORL 标准人脸图像库中实验时,随机选择每个人的前  $l=4$  幅人脸图像作为训练样本集,剩余的人脸图像作为测试样本集。图 4(b)显示在 Yale 人脸图像库中试验时,随机选择每个人前  $l=6$  幅人脸图像作为训练样本集,剩余人脸图像作为测试样本集。图 4(c)显示在 AR 人脸图像库中实验时,随机选择每个人前  $l=5$  幅人脸图像作为训练样本集,剩余的人脸图像作为测试样本集。其中,在 ORL 和 Yale 标准人脸图像库进行 50 次重复实验取其平均值,AR 标准人脸图像库进行 10 次重复实验取其平均值。

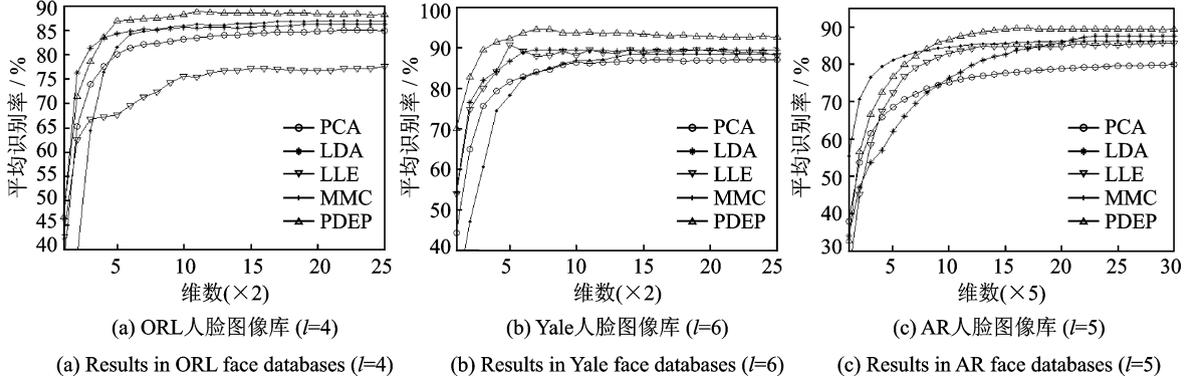


图 4 FDEP 算法平均识别率随特征向量维数的变化情况

Fig. 4 Average recognition rates of FDEP versus the corresponding varied dimensions

### 4.3 实验分析

从 4.2 节实验结果可以得到以下的结论:

(1) 从表 1~3 可以得到,随着训练样本的增加,5 种方法的最大平均识别率也增加。在 3 个人脸数据库上的实验看出本文 FDEP 算法的结果优于其他算法。

(2) 从图 4 可以得到,FDEP 算法在特征向量维数增大的情况下,平均识别率也一直增大;且在特征向量维数比较高时,FDEP 算法的平均识别率优于 PCA、LDA、LLE 和 MMC 等算法的平均识别率。

(3) 由于 FDEP 算法既考虑局部近邻结构信息分布,又考虑全局结构信息的分布,使得该算法平均识别率远远优于 LLE 算法平均识别率。因此,FDEP 算法可以通过保持局部嵌入和全局方差得到最优映射矩阵。

## 5 结束语

在机器学习和模式识别中,特征提取技术被广泛地应用在人脸识别和分类上。本文主要针对流形学习 LLE 算法未考虑数据样本全局信息的问题,提出一种新的特征提取的方法——模糊差分嵌入投影(FDEP)算法,并将其应用在人脸库上。所提出的 FDEP 算法可以看作是 PCA 算法和 LLE 算法的结合,同时融合了模糊全局保持嵌入与模糊局部方法的优点,保持了局部信息和全局信息的最优。因此,相对 PCA 算法和 LLE 算法,FDEP 算法可以获得更好的分类性能。在 ORL, Yale 和 AR 三个标准人脸库上的实验结果验证了 FDEP 算法的有效性。下一步工作将研究 FDEP 算法在监督和半监督学习时的性能。

### 参考文献:

- [1] Swets D L, Weng J. Using discriminant eigenfeatures for image retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, 18(8): 831-836.
- [2] Kirby M, Sirovich L. Application of the KL procedure for the characterization of human faces [J]. *IEEE Transactions on Pat-*

tern Analysis and Machine Intelligence,1990,12(1): 103-108.

- [3] Chang K Y,Ghosh J. A unified model for probabilistic principal surfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2001,23(1): 22-41.
- [4] Hinton G,Dayan P,Revow M. Modeling the manifolds of images of handwritten digits[J]. IEEE Transactions on Neural Network,1997,8(1): 65-74.
- [5] Gomes J,Mojsilovic A. A variational approach to recovering a manifold from sample points[C]// 7th European Conference on Computer Vision. Copenhagen:[s. n.],2002: 3-17.
- [6] Saul L K,Roweis S T. Think globally,fit locally: Unsupervised learning of low dimensional manifolds [J]. Journal of Machine Learning Research,2003,4(1): 119-155.
- [7] Tenenbaum J B,de Silva V,Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000(290): 2319-2323.
- [8] Roweis S T,Saul L K. Nonlinear dimensional reduction by locally linear embedding [J]. Science,2000(290): 2323-2326.
- [9] Belkin M,Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation [J]. Neural Computation, 2003,15(6): 1373-1396.
- [10] He X,Yan S,Hu Y,et al. Face recognition using Laplacianfaces [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2005,27(3): 328-340.
- [11] Yan S,Xu D,Zhang B,et al. Graph embedding and extensions: A general framework for dimensionality reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2007,29(1): 40-51.
- [12] Chen H T,Chang H W,Liu T L. Local discriminant embedding and its variants[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. [S. l.]:IEEE Computer Society, 2005: 846-853.
- [13] 李忠国,侯杰,王凯,等. 模糊支持向量机在路面识别中的应用[J]. 数据采集与处理,2014,29(1):146-151.  
Li Zhongguo,Hou Jie,Wang Kai,et al. Application of fuzzy support vector machine on road type recognition [J]. Journal of Data Acquisition & Processing,2014,29(1):146-151.
- [14] Kwak K C,Pedrycz W. Face recognition using a fuzzy fisher classifier [J]. Pattern Recognition,2005,38(10): 1717-1732.
- [15] Bezdek J C,Keller J,Krishnapuram R. Fuzzy models and algorithms for pattern recognition and image processing[M]. Dordrecht: Kluwer Academic Publishers,1999.
- [16] Wan M,Zheng W. Fuzzy two-dimensional local graph embedding discriminant analysis (F2DLGEDA) with its application to face and palm biometrics [J]. Neural Computing and Applications,2013,23(1): 201-207.

#### 作者简介:



万鸣华(1978-),男,副教授,研究方向:模式识别和图像处理,E-mail: wmh36@sina.com。

(编辑:张黄群)

