



# 基于视图内字典原子不一致的多视图字典学习算法

## 摘要

传统的多视图字典学习算法旨在利用多视图数据间的相关性,未能考虑多视图数据的差异性,这可能会降低字典的学习性能.受此启发,提出一种基于视图内字典原子不一致的多视图字典学习算法.该算法为每个视图学习类属字典和共享字典,同时,引入编码系数方差的最小化约束,以降低视图间字典的差异性;此外,通过每个视图编码系数与所有视图编码系数均值之间距离的加权的最小化来约束相应特征的贡献度;然后,施加视图内字典原子的不一致性约束以降低视图内字典的冗余.最后,在两个数据集(AR和Extended Yale B数据集)上的实验验证了所提算法的有效性.

## 关键词

多视图字典学习;相关性;差异性;人脸识别

中图分类号 TP181

文献标志码 A

收稿日期 2019-05-16

资助项目 国家自然科学基金重点项目(61432008);国家自然科学基金项目(61876087)

## 作者简介

田泽,男,硕士生,研究方向为字典学习. zetian\_edu@126.com

杨明(通信作者),男,博士,教授,主要研究方向为机器学习及计算机视觉. m.yang@njnu.edu.cn

1 南京师范大学 计算机科学与技术学院,南京,210023

2 河海大学 计算机与信息学院,南京,210098

## 0 引言

多视图字典学习算法按照字典原子组成形式上的不同可分为2大类:第1大类由训练样本组成的字典<sup>[1-4]</sup>,第2大类为紧凑且表示能力较强的字典<sup>[5-13]</sup>.Yuan等<sup>[1]</sup>假设视图间共享相同的稀疏模式,但是其假设条件过于强烈.针对这一问题,Zhang等<sup>[2]</sup>提出联合动态的稀疏表示并鼓励不同视图间的稀疏模式共享在组级别上;Yang等<sup>[3]</sup>鼓励不同视图间的表示系数在总体上尽可能相似;Shekhar等<sup>[4]</sup>利用不同特征的编码系数进行联合分类.然而,由原始训练样本组成的字典<sup>[1-4]</sup>会降低字典的鉴别能力,主要是因为未能充分挖掘字典中的有效信息,比如如何降低视图间字典的冗余、如何增强字典原子的表示能力以及如何利用视图间字典的相似性等.

紧凑且表示能力较强的多视图字典学习算法<sup>[5-13]</sup>大体上分为3大类.第1大类侧重于利用不同视图间编码系数的相关性.Gao等<sup>[5]</sup>通过图模型融合不同视图的特征以去除特征的重叠部分并使用结构组稀疏编码系数以学习多视图的相关性并用于行为识别;Jia等<sup>[6]</sup>施加 $l_{1,\infty}$ 正则化到编码系数以获得结构化的稀疏编码系数.类似于文献<sup>[6]</sup>,Zheng等<sup>[7]</sup>同样使用 $l_{1,\infty}$ 正则化约束编码系数以获得组稀疏编码系数.第2大类侧重于为每个视图学习鉴别和表示能力较强的字典.基于多模态稀疏表示的分类算法<sup>[8]</sup>(MSRC)旨在为每个模态选择最具有判别能力的样本,并且鼓励子字典尽可能不同,以增强字典的判别能力;Tosic等<sup>[9]</sup>为每个视图(模态)学习过完备的字典,从而增强字典的表示能力.为降低视图间字典的冗余,Jing等<sup>[10]</sup>嵌入不相关约束到字典学习中,以增强字典的表示能力;Yao等<sup>[11]</sup>提出为每个视图学习类属字典和共享类字典以增强字典的判别能力,同时,对稀疏编码系数进行局部约束,从而保留了局部特征之间的内在关系.第3大类侧重于学习视图间共享字典的多视图字典学习算法.Zheng等<sup>[12]</sup>提出为每个视图学习结构化类属字典,视图间学习一个共享字典;Patel等<sup>[13]</sup>采用域适应算法解决训练和测试数据分布不一致的情形并且数据可以由一个共享字典所表示,但其共享字典存在大量的冗余.针对文献<sup>[13]</sup>的不足,Wu等<sup>[14]</sup>提出为所有视图学习一个不相关的共享字典,从而降低共享字典间的冗余.但是,以上算法<sup>[5-14]</sup>未能充分利用特征间的相关性和差异性.

针对文献<sup>[5-14]</sup>的缺陷,本文提出一种基于视图内字典原子不一

致的多视图字典学习算法,并将其用于多视图下的人脸识别问题.本文的主要贡献如下:

第一,字典由类属字典和共享字典所组成,以挖掘视图内的类属信息和类间共享信息;

第二,引入编码系数方差的最小化约束,从而有效利用视图间数据的相关性;

第三,通过每个视图编码系数与所有视图编码系数均值之间距离的加权之和的最小化来约束相应特征的贡献度;

第四,嵌入视图内字典原子的不一致性约束来降低视图内字典的冗余.

## 1 相关工作

### 1.1 多任务联合稀疏表示(MTJSRC)

Yuan 等<sup>[1]</sup>于 2010 年提出一种多任务联合稀疏表示(Multi-Task Joint Sparse Representation, MTJSRC)用于视觉分类.该算法将 SRC 用于多任务学习,提出利用联合稀疏表示系数  $l_{2,1}$  学习视图间的相关性,其目标函数如下:

$$\min_{\mathbf{X}} \frac{1}{2} \sum_{k=1}^V (\|\mathbf{y}_k - \sum_{j=1}^J \mathbf{D}_k^j \mathbf{X}_k^j\|_2^2 + \lambda \sum_{j=1}^J \|\mathbf{X}^j\|_2), \quad (1)$$

其中  $V$  表示视图的数量,  $J$  表示类别的数量,  $\lambda$  是正的参数;  $\mathbf{y}_k$ ,  $\mathbf{D}_k$ ,  $\mathbf{X}_k$  分别表示每个视图的测试样本、字典和编码系数;  $\mathbf{X}^j = [\mathbf{X}_1^j, \mathbf{X}_2^j, \dots, \mathbf{X}_k^j]$  表示第  $j$  类的所有视图;  $\mathbf{D}_k^j, \mathbf{X}_k^j$  表示第  $k$  个视图的第  $j$  类. MTJSRC 算法有以下缺陷:第一,它仅仅考虑视图间的相关性而未考虑每个视图的差异性,这会降低字典的表示能力;第二,它将带有少量噪声信息的训练样本视为字典,这会降低字典的鉴别能力.

### 1.2 宽松的协同表示(RCR)

传统的多视图学习算法只考虑视图间的相似性,没有考虑视图间的差异性.针对这一问题, Yang 等<sup>[3]</sup>于 2012 年提出一种宽松的协同表示(Relaxed Collaborative Representation, RCR)算法,其算法所构造的目标函数如下:

$$\min_{\omega_k, \mathbf{X}_k} \sum_{k=1}^V (\|\mathbf{y}_k - \mathbf{D}_k \mathbf{X}_k\|_2^2 + \lambda \|\mathbf{X}_k\|_2^2 + \beta \omega_k \|\mathbf{X}_k - \bar{\mathbf{X}}\|_2^2), \quad (2)$$

s.t. prior $\{\omega_k\}$

其中,  $V$  表示视图的数量,  $\lambda$ ,  $\beta$  是正的参数,  $\omega_k$  表示先验权重,权重的大小意味着每个视图的贡献不同以反映视图间的相似性和差异性;  $\mathbf{y}_k, \mathbf{D}_k, \mathbf{X}_k$  分别表示每个视图的测试样本、字典和编码系数;  $\bar{\mathbf{X}} = \sum_{k=1}^V \omega_k \mathbf{X}_k / \sum_{k=1}^V \omega_k$  表示所有视图的平均编码系数.第 1

项表示样本可以由字典协同表示,第 2 项为表示系数,第 3 项通过每个视图编码系数与所有视图编码系数均值之间距离的加权之和的最小化来约束相应特征的贡献度. RCR 算法存在以下不足:第一,将带有噪声信息的训练样本视为字典,会影响字典的鉴别能力;第二,字典原子可能存在冗余,从而降低字典的表示能力.

## 2 基于视图内字典原子不一致的多视图字典学习算法(MDLIAI)

### 2.1 MDLIAI 模型构建

针对传统的多视图字典学习算法未能有效利用视图间的相似性和差异性的问题,本节提出一种基于视图内字典原子不一致的多视图字典学习算法(Multi-view Dictionary Learning based on Intra-view Atom Incoherence, MDLIAI).该算法为每个视图学习类属字典和共享类字典,同时,引入编码系数方差的最小化约束,以降低视图间字典的差异性;此外,通过每个视图编码系数与所有视图编码系数均值之间距离的加权之和的最小化来约束相应特征的贡献度. MDLIAI 模型的定义形式如下:

$$\min_{\substack{D_1, \dots, D_V \\ D_0^1, \dots, D_0^V \\ X_1, \dots, X_V \\ X_0^1, \dots, X_0^V}} \sum_{k=1}^V \left( \sum_{c=1}^C \sum_{i \in Q_c} \left( \|\mathbf{y}_k^i - \bar{\mathbf{D}}_k \widetilde{\mathbf{X}}_k^i\|_2^2 + \|\mathbf{y}_k^i - \mathbf{D}_k \widetilde{\mathbf{X}}_k^i\|_2^2 + \lambda \|\widetilde{\mathbf{X}}_k^i\|_2^2 + \tau \omega_k \|\widetilde{\mathbf{X}}_k^i - \bar{\mathbf{X}}\|_2^2 \right) + \eta \left\| \sum_{h=0}^C \sum_{j=0, j \neq h}^C (\mathbf{D}_k^h)^T \mathbf{D}_k^j \right\|_F^2 \right), \quad (3)$$

s.t. prior $\{\omega_k\}$ ,

其中  $V$  表示视图的数量,  $Q_c$  表示第  $c$  类样本的集合,  $i \in Q_c$  表示第  $c$  类样本中的一个样本,  $\widetilde{\mathbf{D}}_k = [\mathbf{D}_k, \mathbf{D}_k^0]$ ,  $\bar{\mathbf{D}}_k = [\mathbf{D}_k^c, \mathbf{D}_k^0]$ ,  $\widetilde{\mathbf{X}}_k^i = [\mathbf{X}_k^i, \mathbf{X}_k^{0,i}]$ ,  $\bar{\mathbf{X}}$  表示所有视图的平均编码系数,  $\omega_k$  表示每个视图编码系数的权重,  $\lambda, \tau, \eta$  表示平衡参数,  $\mathbf{D}_k$  表示第  $k$  个视图的类属字典,字典的大小为  $K = K_1 + \dots + K_c + \dots + K_C$ ,  $\mathbf{D}_k^0$  表示第  $k$  个视图的共享字典,字典的大小为  $K_0$ . 训练样本  $\mathbf{y}_k^i$  可以由类属字典  $\mathbf{D}_k$  和共享字典  $\mathbf{D}_k^0$  共同表示,所对应的表示系数分别为  $\mathbf{X}_k^i$  和  $\mathbf{X}_k^{0,i}$ . 目标函数的第 1 项希望每个视图第  $c$  类样本尽可能由类属字典  $\mathbf{D}_k$  和共享类字典  $\mathbf{D}_k^0$  共同表示;第 2 项希望每个视图第  $c$  类样本尽可能由第  $c$  类类属字典  $\mathbf{D}_k^c$  和共享类字典  $\mathbf{D}_k^0$  共同表示;第 3 项为协同表示系数;第 4

项希望减少表示系数的方差,促使视图间尽可能相似,同时调节权重  $\tau$  以刻画视图间的差异性,如果  $\mathbf{y}_k^i$  与其他视图的样本尽可能相似,那么  $\omega_k$  的值应该足够大以促使  $\widehat{\mathbf{X}}_k^i$  尽可能接近于  $\bar{\mathbf{X}}$ ,如果  $\mathbf{y}_k^i$  与其他视图的样本不同,那么  $\omega_k$  的值应该足够小以促使  $\widehat{\mathbf{X}}_k^i$  与  $\bar{\mathbf{X}}$  不同;第5项为字典不一致性约束项,以减少同一视图间字典间的冗余.

## 2.2 MDLAI 模型求解

目标函数(3)是一个非凸问题,当  $\omega_k$  为强先验时,其求解过程采用以下策略:通过固定其他视图的变量,更新第  $k$  个视图的变量;固定字典  $\mathbf{D}_k$  和  $\mathbf{D}_k^0$ ,通过逐类更新编码系数  $\mathbf{X}_k^i$  和  $\mathbf{X}_k^{0,i}$  以更新编码系数  $\mathbf{X}_k$  和共享系数  $\mathbf{X}_k^0$ ;固定变量  $\mathbf{X}_k, \mathbf{X}_k^0$  和  $\mathbf{D}_k^0$ ,通过逐类计算  $\mathbf{D}_k^c$  来更新类属字典  $\mathbf{D}_k$ ;固定变量  $\mathbf{X}_k, \mathbf{X}_k^0$  和  $\mathbf{D}_k$ ,更新共享字典  $\mathbf{D}_k^0$ .

### 1) 更新编码系数 $\mathbf{X}_k$ 和共享类系数 $\mathbf{X}_k^0$

固定字典  $\mathbf{D}_k$  和  $\mathbf{D}_k^0$ ,通过逐个更新编码系数  $\mathbf{X}_k^i$  和  $\mathbf{X}_k^{0,i}$  来更新编码系数  $\mathbf{X}_k$  和共享类系数  $\mathbf{X}_k^0$ ,目标函数(3)可以转化为如下形式:

$$\min_{\widehat{\mathbf{X}}_k^i} \|\mathbf{y}_k^i - \widehat{\mathbf{D}}_k \widehat{\mathbf{X}}_k^i\|_2^2 + \|\mathbf{y}_k^i - \bar{\mathbf{D}}_k \widehat{\mathbf{X}}_k^i\|_2^2 + \lambda \|\widehat{\mathbf{X}}_k^i\|_2^2 + \tau \omega_k \|\widehat{\mathbf{X}}_k^i - \bar{\mathbf{X}}\|_2^2. \quad (4)$$

类似于文献[11],式(4)可以等价于如下形式:

$$\min_{\widehat{\mathbf{X}}_k^i} \|\widehat{\mathbf{y}}_k^i - \widehat{\mathbf{D}}_k \widehat{\mathbf{X}}_k^i\|_2^2 + \lambda \|\widehat{\mathbf{X}}_k^i\|_2^2 + \tau \omega_k \|\widehat{\mathbf{X}}_k^i - \bar{\mathbf{X}}\|_2^2, \quad (5)$$

其中  $\widehat{\mathbf{y}}_k^i = [\mathbf{y}_k^i; \mathbf{y}_k^i]$ ,  $\widehat{\mathbf{D}}_k = [\mathbf{D}_k \ \mathbf{D}_k^0; \mathbf{D}_k^c \ \mathbf{D}_k^0]$ ,  $\widehat{\mathbf{X}}_k^i = \widehat{\mathbf{X}}_k^i$ ,  $\bar{\mathbf{X}} = \sum_{k=1}^K \omega_k \mathbf{X}_k / \sum_{k=1}^K \omega_k$ . 类似于文献[3]的求解策略,式(5)转化为如下形式:

$$\widehat{\mathbf{X}}_k^i = \widehat{\mathbf{X}}_{0,k}^i + \tau \frac{\omega_k}{\sum_{k=1}^K \omega_k} \mathbf{P}_k \mathbf{Q} \sum_{k=1}^K \omega_k \widehat{\mathbf{X}}_{0,k}^i, \quad (6)$$

其中

$$\mathbf{P}_k = ((\widehat{\mathbf{D}}_k)^T \widehat{\mathbf{D}}_k + \mathbf{I}(\lambda + \tau \omega_k))^{-1}, \widehat{\mathbf{X}}_{0,k}^i = \mathbf{P}_k \widehat{\mathbf{D}}_k \widehat{\mathbf{y}}_k^i, \mathbf{Q} = (\mathbf{I} - \sum_{\eta=1}^K \tilde{\omega}_\eta \mathbf{P}_\eta)^{-1}, \tilde{\omega}_\eta = \tau \omega_\eta^2 / \sum_{k=1}^K \omega_k,$$

$\mathbf{I}$  为单位阵.具体的推导细节请参考文献[3].

### 2) 更新类属字典 $\mathbf{D}_k$

固定编码系数  $\mathbf{X}_k, \mathbf{X}_k^0$  和共享字典  $\mathbf{D}_k^0$ ,逐类更新类属字典  $\mathbf{D}_k^c$ ,更新类属字典  $\mathbf{D}_k$ ,目标函数(6)可以转化为如下形式:

$$\min_{\mathbf{D}_k^c} \sum_{i \in Q_c} \left( \left\| \mathbf{y}_k^i - \sum_{j=0, j \neq c}^C \mathbf{D}_k^j \mathbf{X}_k^{j,i} - \mathbf{D}_k^c \mathbf{X}_k^{c,i} \right\|_2^2 + \left\| \mathbf{y}_k^i - \mathbf{D}_k^0 \mathbf{X}_k^{0,i} - \mathbf{D}_k^c \mathbf{X}_k^{c,i} \right\|_2^2 \right) + 2\eta \sum_{j=0, j \neq c}^C \|(\mathbf{D}_k^c)^T \mathbf{D}_k^j\|_F^2, \quad (7)$$

其中

$$\mathbf{D}_k^c = [(\mathbf{d}_k^c)_1, \dots, (\mathbf{d}_k^c)_q, \dots, (\mathbf{d}_k^c)_{K_c}], \mathbf{X}_k^{c,i} = [(\mathbf{x}_k^{c,i})_1, \dots, (\mathbf{x}_k^{c,i})_q, \dots, (\mathbf{x}_k^{c,i})_{K_c}]^T.$$

参考文献[15],式(7)转化为更新字典中的每个原子.固定字典中其他每一列原子,更新其中的一列  $(\mathbf{d}_k^c)_q$ ,式(7)简化如下:

$$\min_{(\mathbf{d}_k^c)_q} \sum_{i \in Q_c} \left( \left\| \mathbf{y}_k^i - \sum_{j=0, j \neq c}^C \mathbf{D}_k^j \mathbf{X}_k^{j,i} - (\mathbf{d}_k^c)_q (\mathbf{x}_k^{c,i})_q \right\|_2^2 + \left\| \mathbf{y}_k^i - \mathbf{D}_k^0 \mathbf{X}_k^{0,i} - \sum_{h=1, h \neq q}^{K_c} (\mathbf{d}_k^c)_h (\mathbf{x}_k^{c,i})_h - (\mathbf{d}_k^c)_q (\mathbf{x}_k^{c,i})_q \right\|_2^2 \right) + 2\eta \sum_{j=0, j \neq c}^C \|((\mathbf{d}_k^c)_q)^T \mathbf{D}_k^j\|_F^2, \quad (8)$$

对式(8)求一阶导数即可得到解析解.

### 3) 更新共享字典 $\mathbf{D}_k^0$

固定编码系数  $\mathbf{X}_k, \mathbf{X}_k^0$  和类属字典  $\mathbf{D}_k$ ,更新共享字典  $\mathbf{D}_k^0$ ,目标函数(3)转化为如下形式:

$$\min_{\mathbf{D}_k^0} \sum_{i \in Q_c} \left( \left\| \mathbf{y}_k^i - \mathbf{D}_k \mathbf{X}_k^i - \mathbf{D}_k^0 \mathbf{X}_k^{0,i} \right\|_2^2 + \left\| \mathbf{y}_k^i - \mathbf{D}_k^c \mathbf{X}_k^{c,i} - \mathbf{D}_k^0 \mathbf{X}_k^{0,i} \right\|_2^2 \right) + 2\eta \|(\mathbf{D}_k^0)^T \mathbf{D}_k\|_F^2, \quad (9)$$

其中

$$\mathbf{D}_k^0 = [(\mathbf{d}_k^0)_1, \dots, (\mathbf{d}_k^0)_q, \dots, (\mathbf{d}_k^0)_{K_0}], \mathbf{X}_k^{0,i} = [(\mathbf{x}_k^{0,i})_1, \dots, (\mathbf{x}_k^{0,i})_q, \dots, (\mathbf{x}_k^{0,i})_{K_0}]^T.$$

参考文献[15],式(9)转化为更新共享字典的每一列  $(\mathbf{d}_k^0)_q$  来更新共享字典  $\mathbf{D}_k^0$ .目标函数式(9)转化为如下形式:

$$\min_{(\mathbf{d}_k^0)_q} \sum_{i \in Q_c} \left( \left\| \mathbf{y}_k^i - \mathbf{D}_k \mathbf{X}_k^i - \sum_{h=1, h \neq q}^{K_0} (\mathbf{d}_k^0)_h (\mathbf{x}_k^{0,i})_h - (\mathbf{d}_k^0)_q (\mathbf{x}_k^{0,i})_q \right\|_2^2 + \left\| \mathbf{y}_k^i - \mathbf{D}_k^c \mathbf{X}_k^{c,i} - \sum_{h=1, h \neq q}^{K_0} (\mathbf{d}_k^0)_h (\mathbf{x}_k^{0,i})_h - (\mathbf{d}_k^0)_q (\mathbf{x}_k^{0,i})_q \right\|_2^2 \right) + 2\eta \|((\mathbf{d}_k^0)_q)^T \mathbf{D}_k\|_F^2, \quad (10)$$

对式(10)求一阶导数即可得到解析解.

目标函数(3)是一个非凸问题,当  $\omega_k$  为弱先验

或者中等先验时使用文献[3]进行优化.当 $\omega_k$ 为强先验时,其求解过程采用以下策略:通过固定其他视图的变量,更新第 $k$ 个视图的变量;固定字典 $D_k$ 和 $D_k^0$ ,通过逐类更新编码系数 $X_k^i$ 和 $X_k^{0,i}$ 以更新编码系数 $X_k$ 和共享系数 $X_k^0$ ;固定变量 $X_k, X_k^0$ 和 $D_k^0$ ,通过逐类计算 $D_k^c$ 来更新类属字典 $D_k$ ;固定变量 $X_k, X_k^0$ 和 $D_k$ ,更新共享字典 $D_k^0$ .

通过对上述变量的更新,输出 $X_k, X_k^0, D_k$ 和 $D_k^0$ .

### 2.3 MDLAI 模型的分类

在字典训练结束后,得到每个视图的类属字典 $D_1, \dots, D_V$ ,共享类字典 $D_1^0, \dots, D_V^0$ .类似于文献[16]的分类算法,求解测试样本 $z_k$ 的编码系数 $\alpha_1, \dots, \alpha_V$ 和 $\alpha_1^0, \dots, \alpha_V^0$ 的式子如下:

$$\min_{\alpha_1, \dots, \alpha_V, \alpha_1^0, \dots, \alpha_V^0} \sum_{k=1}^V \left( \|z_k - D_k \alpha_k - D_k^0 \alpha_k^0\|_2^2 + \rho \left\| \begin{pmatrix} \alpha_k \\ \alpha_k^0 \end{pmatrix} \right\|_2^2 \right). \quad (11)$$

将式(11)转化为如下形式:

$$\min_{\phi_1, \dots, \phi_V} \sum_{k=1}^V \left( \|z_k - \widetilde{D}_k \phi_k\|_2^2 + \rho \|\phi_k\|_2^2 \right), \quad (12)$$

其中 $\phi_k = [\alpha_k, \alpha_k^0]$ ,  $\widetilde{D}_k = [D_k, D_k^0]$ .对(12)求一阶导数后得到如下的推导公式:

$$\phi_k = (\widetilde{D}_k^T \widetilde{D}_k + \rho I)^{-1} \widetilde{D}_k^T z_k. \quad (13)$$

一旦获得 $\{\phi_1, \dots, \phi_V\}$ ,通过如下方式计算每个类的残差:

$$e_c = \sum_{k=1}^V \|z_k - \widetilde{D}_k^c \phi_k^c\|_2^2, \quad (14)$$

其中 $\phi_k^c$ 为 $\widetilde{D}_k^c = [D_k^c, D_k^{0,c}]$ 所对应的编码系数.最后,通过如下的方式进行分类:

$$\text{identity}(z_k) = \arg \min_c e_c, \quad (15)$$

其中 $c = 1, \dots, C$ 为类别信息.

### 2.4 MDLAI 算法

输入:每个视图的数据 $y_k$ ,参数 $\lambda, \eta, \tau, \rho$ .

1) 随机初始化每个视图的类属字典 $D_1, \dots, D_V$ ,共享类字典 $D_1^0, \dots, D_V^0$ ;

2) 固定每个视图的类属字典 $D_1, \dots, D_V$ ,共享字典 $D_1^0, \dots, D_V^0$ ,更新每个视图的编码系数 $X_1, \dots, X_V$ 和 $X_1^0, \dots, X_V^0$ ,使用式(6)求解;

3) 固定编码系数 $X_1, \dots, X_V$ 和 $X_1^0, \dots, X_V^0$ 以及 $D_1^0, \dots, D_V^0$ ,更新每个视图的类属字典 $D_k$ ,使用式(8)求解;

4) 固定编码系数 $X_1, \dots, X_V$ 和 $X_1^0, \dots, X_V^0$ 以及字典 $D_1, \dots, D_V$ ,更新 $D_1^0, \dots, D_V^0$ ,更新每个视图的共享字典 $D_k^0$ ,使用式(10)求解;

5) 重复步骤2到步骤4,直到达到收敛条件或者满足最大迭代次数.

输出:字典 $D_1, \dots, D_V$ 和 $D_1^0, \dots, D_V^0$ ,稀疏编码系数 $X_1, \dots, X_V$ 和 $X_1^0, \dots, X_V^0$ .

## 3 实验与结果

在MDLAI算法中, $\lambda$ 的取值为0.001、0.005、0.01、0.05、0.1, $\tau$ 的取值也为0.001、0.005、0.01、0.05、0.1, $\eta$ 的取值为0、0.0005、0.001、0.005、0.01、0.05, $\rho$ 的取值为0.0001、0.0005、0.001、0.005、0.01、0.05、0.1.主要的对比算法有:SVM<sup>[17]</sup>、SRC<sup>[18]</sup>、CRC<sup>[19]</sup>、LRC<sup>[20]</sup>、MTJSRC<sup>[1]</sup>、RCR<sup>[3]</sup>.对比算法的实验结果均来自文献[3].

### 3.1 The Extended Yale B 数据集

Extended Yale B数据集含有38个个体的2414张图像,其中每个个体受到光照、表情等外界因素的影响.每张图像都裁剪成 $54 \times 48$ 的大小(图1),再将裁剪后的图像分成4个区域块(图2),然后从每个个体随机选择10、15、20、25张图像用于训练,剩余图像用于测试.当训练样本为10、15、20、25时,MDLAI算法中总字典(类属字典和共享类字典)的大小分别设置为 $5 \times 38 + 190$ 、 $5 \times 38 + 380$ 、 $10 \times 38 + 380$ 、 $15 \times 38 + 380$ .实验重复5次以计算平均识别率和标准差.MDLAI算法的实验参数为 $\lambda = 0.05, \tau = 0.05, \eta = 0.001, \rho = 0.05$ .



图1 裁剪后的 Extended Yale B 图像  
Fig. 1 Cropped Extended Yale B image

图3—5展示Extended Yale B人脸数据集下训练样本为10且不同参数下的识别率,字典大小设置为 $5 \times 38 + 190$ 时,所提MDLAI算法在参数 $\lambda, \tau, \eta$ 值不同情况下的识别率.图3为当 $\tau = 0.005, \eta = 0.0005$ 时, $\lambda = 0.05$ 时取得最佳性能;图4为当 $\lambda = 0.05, \eta = 0.0005$ 时, $\tau = 0.05$ 时取得最佳性能;图5

为当  $\lambda = 0.05$ ,  $\tau = 0.05$  时,  $\eta = 0.001$  时取得最佳性能.

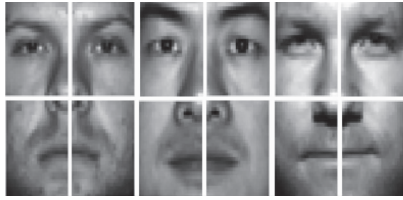


图2 分为4个区域的 Extended Yale B 图像  
Fig.2 Extended Yale B image divided into 4 areas

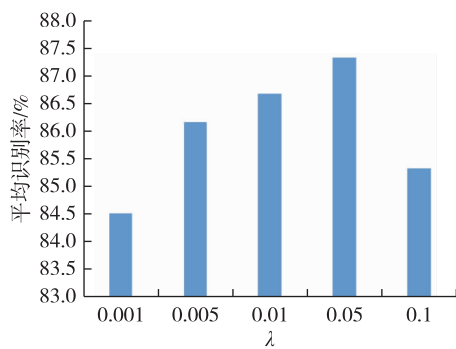


图3 不同  $\lambda$  值情况下的识别率  
Fig.3 Recognition rate under different  $\lambda$  values

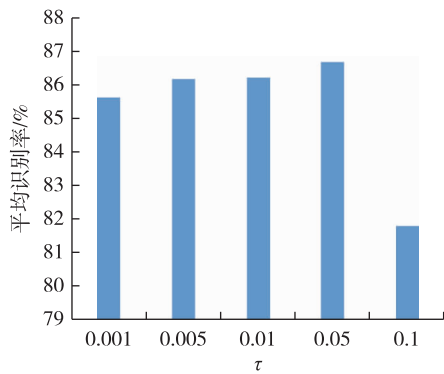


图4 不同  $\tau$  值情况下的识别率  
Fig.4 Recognition rate under different  $\tau$  values

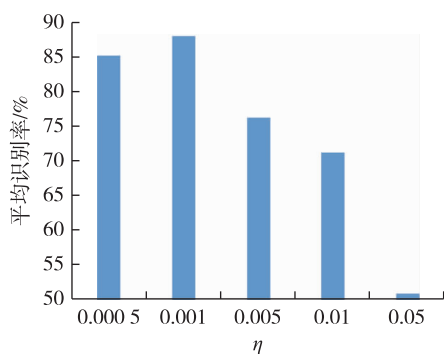


图5 不同  $\eta$  值情况下的识别率  
Fig.5 Recognition rate under different  $\eta$  values

表1展示了训练样本为10、15、20、25时, MDLAI算法比RCR算法的准确率至少高出1.2个百分点,这说明MDLAI算法能有效增强字典的表示能力并且适用于训练样本数较少情况下的人脸识别.表2展示训练样本数为20情况下的识别率和标准差.

表1 Extended Yale B数据集下的识别准确率

Table 1 Recognition rate on Extended Yale B dataset %

| 算法                    | 训练样本 |      |      |      |
|-----------------------|------|------|------|------|
|                       | 10   | 15   | 20   | 25   |
| SVM <sup>[17]</sup>   | 60.0 | 67.1 | 76.5 | 88.1 |
| SRC <sup>[18]</sup>   | 84.6 | 84.2 | 91.3 | 92.0 |
| CRC <sup>[19]</sup>   | 84.8 | 84.7 | 91.2 | 92.4 |
| LRC <sup>[20]</sup>   | 82.4 | 81.8 | 87.0 | 89.0 |
| MTJSRC <sup>[1]</sup> | 87.3 | 87.4 | 91.5 | 93.6 |
| RCR <sup>[3]</sup>    | 86.8 | 87.2 | 92.3 | 93.6 |
| MDLAI(本文)             | 88.0 | 89.2 | 94.3 | 94.9 |

表2 Extended Yale B数据集下的识别率和标准差

Table 2 Recognition rate and standard deviation on Extended Yale B dataset

| 算法    | (REC $\pm$ STD)/% |
|-------|-------------------|
| MDLAI | 94.3 $\pm$ 0.35   |

### 3.2 AR人脸数据集

AR数据集含有119个个体,每个个体包含26张图像.其中每个个体受到光照、表情等外界因素的影响.参考文献[3],将伪装和带有墨镜的图像剔除掉,使用AR数据集的子集,然后将每张图像都裁剪成60 $\times$ 44的大小(图6),再将裁剪后的图像分成4个区域块(图7),然后从每个个体选择前7张图像(第1阶段)用于训练,后7张图像(第2阶段)用于测试.MDLAI中总字典(类属字典和共享类字典)的大小都设置为6 $\times$ 100+100.实验重复5次以计算平均识别率.MDLAI算法的实验参数为 $\lambda = 0.05$ ,  $\tau = 0.1$ ,  $\eta = 0$ ,  $\rho = 0.01$ ,这说明参数 $\eta = 0$ 的情况下也能训练获得表示能力强和紧凑的字典.

表3为MDLAI算法在AR数据集下的识别率,

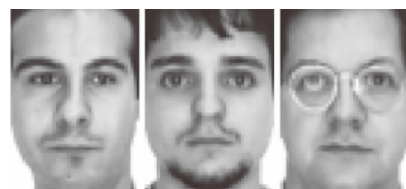


图6 裁剪后的AR图像  
Fig.6 Cropped AR image



图7 分为4个区域的AR图像  
Fig. 7 AR image divided into 4 areas

MDLAI 算法实现了最好的实验性能,比第二高的RCR 算法高出0.1个百分点.MTJSRC 算法的实验结果低于RCR 算法,这可能是由于MTJSRC 算法未能考虑视图间的差异性造成的.

表3 AR数据集下的识别率

| 算法                    | 识别率  |
|-----------------------|------|
| SVM <sup>[17]</sup>   | 87.1 |
| SRC <sup>[18]</sup>   | 93.7 |
| CRC <sup>[19]</sup>   | 93.3 |
| LRC <sup>[20]</sup>   | 76.4 |
| MTJSRC <sup>[1]</sup> | 95.8 |
| RCR <sup>[3]</sup>    | 95.9 |
| MDLAI(本文)             | 96.0 |

## 4 总结

本文提出一种基于视图内字典原子不一致的多视图字典学习算法.该算法学习每个视图下的类属字典和共享类字典,以增强类属字典的判别能力;引入编码系数方差的最小化约束,以降低视图间字典的差异性,从而利用视图间的相关性.此外,嵌入每个视图编码系数与所有视图编码系数均值之间距离的加权和的最小化来约束相应特征的贡献度,以刻画每个视图的相似性和差异性.最后,施加视图内字典原子的不一致性约束,从而降低视图内字典的冗余.实验结果表明该算法能有效解决多视图下的人脸识别问题.

## 参考文献

### References

[ 1 ] Yuan X T, Yan S C. Visual classification with multi-task joint sparse representation [ C ] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 13-18 June 2010, San Francisco, CA, USA, 2010: 3493-3500

[ 2 ] Zhang H C, Nasrabadi N M, Zhang Y N, et al. Multi-observation visual recognition via joint dynamic sparse rep-

resentation [ C ] // 2011 International Conference on Computer Vision, 6-13 Nov. 2011, Barcelona, Spain, 2011: 595-602

[ 3 ] Yang M, Zhang L, Zhang D, et al. Relaxed collaborative representation for pattern classification [ C ] // 2012 IEEE Conference on Computer Vision and Pattern Recognition, 16-21 June 2012, Providence, RI, USA, 2012: 2224-2231

[ 4 ] Shekhar S, Patel V M, Nasrabadi N M, et al. Joint sparse representation for robust multimodal biometrics recognition [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1): 113-126

[ 5 ] Gao Z, Zhang H, Xu G P, et al. Multi-view discriminative and structured dictionary learning with group sparsity for human action recognition [ J ]. Signal Processing, 2015, 112: 83-97

[ 6 ] Jia Y, Salzmann M, Darrell T. Factorized latent spaces with structured sparsity [ C ] // Proceedings of Neural Information Processing Systems (NIPS), 2010: 982-990

[ 7 ] Zheng S, Xie B, Huang K Q, et al. Multi-view pedestrian recognition using shared dictionary learning with group sparsity [ M ] // Zheng S, Xie B, Huang K Q, et al. eds. Neural Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011: 629-638. DOI: 10.1007/978-3-642-24965-5\_71

[ 8 ] Shi Y H, Gao Y, Yang Y B, et al. Multimodal sparse representation-based classification for lung needle biopsy images [ J ]. IEEE Transactions on Biomedical Engineering, 2013, 60(10): 2675-2685

[ 9 ] Tomic I, Frossard P. Dictionary learning for stereo image representation [ J ]. IEEE Transactions on Image Processing, 2011, 20(4): 921-934

[ 10 ] Jing X, Hu R, Wu F, et al. Uncorrelated multi-view discrimination dictionary learning for recognition [ C ] // Proceedings of Association for the Advancement of Artificial Intelligence (AAAI), 2014: 2787-2795

[ 11 ] Yao T T, Wang Z Y, Xie Z, et al. Learning universal multiview dictionary for human action recognition [ J ]. Pattern Recognition, 2017, 64: 236-244

[ 12 ] Zheng J J, Jiang Z L. Learning view-invariant sparse representations for cross-view action recognition [ C ] // 2013 IEEE International Conference on Computer Vision, 1-8 Dec. 2013, Sydney, NSW, Australia, 2013: 3176-3183

[ 13 ] Patel V M, Gopalan R, Li R N, et al. Visual domain adaptation: a survey of recent advances [ J ]. IEEE Signal Processing Magazine, 2015, 32(3): 53-69

[ 14 ] Wu F, Jing X Y, Yue D. Multi-view discriminant dictionary learning via learning view-specific and shared structured dictionaries for image classification [ J ]. Neural Processing Letters, 2017, 45(2): 649-666

[ 15 ] Kong S, Wang D H. A dictionary learning approach for classification: separating the particularity and the commonality [ M ] // Kong S, Wang D H. eds. Computer Vision-ECCV 2012. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 186-199. DOI: 10.1007/978-3-642-33718-5\_14

[ 16 ] Wu F, Jing X Y, You X G, et al. Multi-view low-rank dictionary learning for image classification [ J ]. Pattern Recognition, 2016, 50: 143-154

- [17] Heisele B, Ho P, Poggio T. Face recognition with support vector machines: global versus component-based approach [C] // Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001, 7-14 July 2001, Vancouver, BC, Canada, 2001: 688-694
- [18] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227
- [19] Zhang L, Yang M, Feng X C. Sparse representation or collaborative representation: which helps face recognition? [C] // 2011 International Conference on Computer Vision, 6-13 Nov. 2011, Barcelona, Spain, 2011: 471-478
- [20] Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(11): 2106-2112

## Multi-view dictionary learning based on intra-view atom incoherence algorithm

TIAN Ze<sup>1</sup> YANG Ming<sup>1</sup> CHEN Zhe<sup>2</sup> SHI Aiye<sup>2</sup>

1 College of Computer Science and Technology, Nanjing Normal University, Nanjing 210023

2 College of Computer and Information, Hohai University, Nanjing 210098

**Abstract** The traditional multi-view dictionary learning algorithm is designed to take advantage of the correlation between multi-view data and fails to consider the distinctiveness of the multi-view data, which may reduce the performance of dictionary. Inspired by this observation, we present a multi-view dictionary learning based on the intra-view atom inconsistency algorithm. The algorithm learns class-specific dictionaries and the shared class dictionary for each view and calculates the minimum of the coding coefficient variance to reduce the distinctiveness of inter-view dictionaries. In addition, the minimization of the weighted sum of the distance between the coding coefficients between each view and the mean of coding coefficients for all views restrict the contribution of the corresponding features. Then, we embed the inconsistency constraint into the intra-view dictionaries to reduce redundancy. Finally, two datasets (AR and Extended Yale B datasets) were used to validate the effectiveness of the proposed algorithm.

**Key words** multi-view dictionary learning; correlation; distinctiveness; face recognition