

其中, $I_p \in \mathbb{R}^{p \times p}$ 为单位矩阵.同时,根据文献[29],可得:

$$m_{G_i} = \frac{t_{G_i}}{\|t_{G_i}\|_2} \max\left(\|t_{G_i}\|_2 - \frac{\lambda_1 w_i^x}{\mu}, 0\right) \quad (8)$$

其中, $t = a - \frac{\tau_1}{\mu}$.最后,乘子的更新准则为

$$\tau_1 = \tau_1 + \mu(m - a), \tau_2 = \tau_2 + \mu(\mathbf{1}^T a - 1) \quad (9)$$

算法 1 给出了针对变量 a 的子问题(3)的求解过程.

算法 1. 使用 ADMM 求解子问题(3).

输入:图像集 X, Y ,表示系数 b ,参数 λ_1 .

输出: a, m .

- (1) 初始化 $\mu=0.9, \rho=1.1, \varepsilon=10^{-3}, \mu_{\max}=10^4, b_i=1/q$.
- (2) **while** 不收敛 **do**
- (3) 使用公式(7)更新 a .
- (4) 使用公式(8)更新 m .
- (5) 使用公式(9)更新乘子 τ_1 和 τ_2 .
- (6) 更新 $\mu=\max(\rho, \mu_{\max})$.
- (7) 检测收敛条件 $\|a^k - m^k\|_{\infty} < \varepsilon$ 并且 $\max(\|a^k - a^{k-1}\|_{\infty}, \|m^k - m^{k-1}\|_{\infty}) < \varepsilon$.
- (8) **end while**

其中,第(6)步由于固定的 μ 会导致算法收敛速度较慢,因此通过引入变量 ρ 构建 μ 的递增序列以加速收敛^[30].根据对称性,易得变量 b 的更新算法(这里省略了 b 的求解过程).算法 2 给出了采用 ADMM 交替更新 a, b 以求解 DSRID-E 模型的过程.算法 2 为两变量的 ADMM,其收敛性得到保证^[28].

算法 2. 使用 ADMM 求解问题(2)的 DSRID-E 模型.

输入:图像集 X, Y ,参数 λ_1, λ_2 .

输出: a, b .

- (1) **while** 不收敛 **do**
- (2) 使用算法 1 求解系数 a .
- (3) 使用与算法 1 类似的对称算法求解系数 b .
- (4) **end while**

4 流形空间图像集距离学习

图像集结构的复杂性,导致不同研究者对其结构采用了不同的假设.流形假设是一类重要的假设.在该假设下,一个集合被认为是流形上的一个点.不同于传统的欧式度量,黎曼度量通过计算流形上任意两点间最短路径,即测地距离,表示两点间关系.拓展于黎曼流形,两类对图像集分类最常用的流形是对称正定矩阵流形和格林斯曼流形.本节讨论 DSRID 框架在这两类流形上的实现.

4.1 SPD流形上的图像集距离学习

SPD 流形由对称正定矩阵张成,为使其建模每个图像集,同时考虑到每个图像集内部的多个子结构,这里采用高斯混合模型(GMM)分析每个集合,进而得到的所有协方差矩阵构成了一个 SPD 流形.同时,为保证正定性,与文献[8]类似,可以在每个矩阵的对角线元素上加一个小的扰动.假设每个集合含有 g 个高斯成分,则有图像集 $X = \{X_i\}_{i=1}^g$,且 $X_i \in \mathcal{S}_+^d$ 为对应的协方差矩阵.SPd 上有两个应用广泛的度量,分别为仿射不变度量(affine-invariant metric,简称 AIM)^[19]和对数欧式度量(log-Euclidean metric,简称 LED)^[20].基于运算量与实验结果,这里选用 LED 代替公式(1)中的距离度量函数,其定义为

$$d(\mathbf{X}_i, \mathbf{X}_j) = \|\log(\mathbf{X}_i) - \log(\mathbf{X}_j)\|_F \quad (10)$$

其中, $\log(\cdot)$ 为矩阵对数.

4.1.1 目标函数

采用 LED 度量的 DSRID 在 SPD 流形的实现简称为 DSRID-S. 其初衷是找到集合中具有代表性的 SPD 矩阵, 并以 LED 作为距离度量函数. 为了便于应用 DSRID 框架, 首先把 SPD 矩阵映射到欧式空间, 然后把映射矩阵拉伸成向量形式有 $\tilde{\mathbf{X}} = [\text{vec}(\log(\mathbf{X}_1)), \text{vec}(\log(\mathbf{X}_2)), \dots, \text{vec}(\log(\mathbf{X}_g))]$. 同时, 对于 $\tilde{\mathbf{Y}}$ 有类似定义. 下面给出 DSRID-S 的目标函数:

$$\begin{cases} \min_{a,b} \frac{1}{2} \|\tilde{\mathbf{X}}\mathbf{a} - \tilde{\mathbf{Y}}\mathbf{b}\|_2^2 + \lambda_1 \|\mathbf{a}\|_1 + \lambda_2 \|\mathbf{b}\|_1 \\ \text{s.t. } \sum_{i=1}^g a_i = 1, \sum_{j=1}^g b_j = 1 \end{cases} \quad (11)$$

其中, $\|\mathbf{a}\|_1 = \sum_{i=1}^g |a_i|$, $\|\mathbf{b}\|_1 = \sum_{j=1}^g |b_j|$.

该目标函数与公式(2)类似, 由于 $\tilde{\mathbf{X}}$ 和 $\tilde{\mathbf{Y}}$ 中每一列已经代表了一个子结构, 因此用稀疏约束代替了公式(2)中的组稀疏约束.

4.1.2 优化

问题(11)的优化方法与第3节中 DSRID-E 模型的优化方法类似, 具体可参考第3.2节的算法1与算法2, 两者最大的不同是对于辅助变量 m 的求解. 在 DSRID-S 中, 由于对 m 施加了稀疏约束, 对比公式(8), 应用稀疏求解的收缩算子^[29]. 因此, 对 m 的更新操作为

$$\mathbf{m} = \text{sign}(t) \max\left(|t| - \frac{\lambda_1}{\mu}, 0\right) \quad (12)$$

其中, $t = \mathbf{a} - \frac{\tau_1}{\mu}$. 可以看出, 公式(8)为公式(12)在组稀疏上的扩展.

4.2 Grassmann流形上的图像集距离学习

另一类常见的图像集建模方法是假设每个集合内含有 1 个或多个线性子空间, 一个线性子空间可以认为是 Grassmann 流形上的一个点. 为了识别集合内的多个子空间结构, 可以采用子空间聚类算法^[31]. 假设每个集合内含有 g 个 s 维子空间, 则有 $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^g$, 令 $\hat{\mathbf{X}}_i \in \mathbb{R}^{d \times s}$ 为第 i 个图像子集 \mathbf{X}_i 对应的子空间的 s 个正交基, 且 $\hat{\mathbf{X}}_i^T \times \hat{\mathbf{X}}_i = \mathbf{I}_s$. 与 SPD 流形上的 DSRID 实现类似, 这里引入投影度量(projection metric)^[9]来计算两个子空间的测地距离, 其定义为

$$d(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j) = \|\hat{\mathbf{X}}_i \hat{\mathbf{X}}_i^T - \hat{\mathbf{X}}_j \hat{\mathbf{X}}_j^T\|_F \quad (13)$$

由于投影度量最终采用欧式距离测量子空间距离, 因此为了便于应用 DSRID 框架, 这里把投影矩阵拉伸成向量形式, 有 $\hat{\mathbf{X}} = [\text{vec}(\hat{\mathbf{X}}_1 \times \hat{\mathbf{X}}_1^T), \text{vec}(\hat{\mathbf{X}}_2 \times \hat{\mathbf{X}}_2^T), \dots, \text{vec}(\hat{\mathbf{X}}_g \times \hat{\mathbf{X}}_g^T)]$. 同时, 对于 $\hat{\mathbf{Y}}$ 有类似定义.

4.2.1 目标函数

采用投影度量的 DSRID 在 Grassmann 流形上的实现简称为 DSRID-G, 其对应的目标函数为

$$\begin{cases} \min_{a,b} \frac{1}{2} \|\hat{\mathbf{X}}\mathbf{a} - \hat{\mathbf{Y}}\mathbf{b}\|_2^2 + \lambda_1 \|\mathbf{a}\|_1 + \lambda_2 \|\mathbf{b}\|_1 \\ \text{s.t. } \sum_{i=1}^g a_i = 1, \sum_{j=1}^g b_j = 1 \end{cases} \quad (14)$$

该目标函数与公式(11)类似, 不同之处在于 $\hat{\mathbf{X}}$ 和 $\hat{\mathbf{Y}}$ 中每一列代表了一个子空间. 问题(14)的求解可参照第4.1.2节 DSRID-S 模型的优化方法.

5 分 类

根据以上 DSRID 在不同空间内的实现,可以获得任意两集间的距离,进而利用这些距离度量构建集合的特征表示.给定含有 n 个图像集的训练集 \mathcal{X} 和 c 个图像集的测试集 \mathcal{Y} ,针对其中任意第 i 个集合的特征表示 f_i 定义为

$$f_i = \left[\frac{d_{i,1}}{\|f_i\|_2}, \frac{d_{i,2}}{\|f_i\|_2}, \dots, \frac{d_{i,n}}{\|f_i\|_2} \right],$$

其中, d_{ij} 为第 i 个训练或测试图像集与第 j 个训练图像集的距离.这里采用最小二乘回归作为分类器,给定训练集的特征表示矩阵 $F \in \mathbb{R}^{n \times n}$,其中每一行为一个集合的特征表示,同时定义类标矩阵 $T \in \mathbb{R}^{n \times c}$,其中, $t_{ij}=1$ 表示第 i 个集合属于第 j 类(共有 c 个类),否则等于 0,则基于最小二乘回归的分类模型为

$$\min_W \|FW - T\|_F^2 + \lambda \|W\|_F^2 \quad (15)$$

其中, $W \in \mathbb{R}^{n \times c}$ 为映射矩阵,其闭解形式为 $W = (F^T F + \lambda I)^{-1} F^T T$.在以下实验中,参数 λ 均固定为 10^{-4} .给定任意测试集合的特征表示 f_i^{test} ,则其对应的预测向量为 $t_i^{test} = W^T f_i^{test}$,则其类标为 $j = \arg \max_j t_{i,j}^{test}$.

为利用不同度量下信息的互补性,这里我们融合了本文涉及的 3 种度量(DSRID-E, DSRID-S 和 DSRID-G)的分类结果(对应的预测向量为 $t_i^{test(E)}$, $t_i^{test(S)}$ 和 $t_i^{test(G)}$),采用平均预测向量作为最终的分类结果,即

$$j = \arg \max_j \frac{1}{3} (t_{i,j}^{test(E)} + t_{i,j}^{test(S)} + t_{i,j}^{test(G)}).$$

6 实 验

为了验证算法的有效性,我们比较了多种流行的图像集分类算法.这些算法按照图像集度量方式的不同分为 4 组:基于欧式空间的 AHISD^[6], CHISD^[6], SANP^[7] 和 RNP^[5];基于子空间假设的 DCC^[15], MMD^[10] 和 MDA^[9];基于 SPD 流形假设的 RSR^[11] 和 CDL^[8];基于组合度量的 HERML^[3] 和 DARG^[17].

各算法参数参照各文献中的实验设置.

本文提出的图像集距离框架 DSRID 共有 4 个实现:欧式空间上的 DSRID-E、SPD 流形上的 DSRID-S、Grassmann 流形上的 DSRID-G 以及它们的融合方法 DSRID-(E+G+S).模型中,两个正则化参数 λ_1 和 λ_2 的调试区间为 $\{10^{-3}, 10^{-2}, 10^{-1}\}$, DSRID 中的子结构数量的调整区间为 $\{1, 4, 8, 10\}$, DSRID-S 中的子空间维度的调试区间均为 $\{3, 5, 10, 12\}$,具体参数设置由交叉验证确定.

所有实验结果如无特殊说明,均是 10 次测试的平均准确率和标准差.

6.1 基于图像集的人脸识别

针对基于图像集的人脸识别任务,我们选用了 3 个公共数据集: Honda/USCD^[32], Mobo^[33] 以及 YouTube Celebrities(YTC)^[34].

- Honda/USCD 涵盖了 20 类,由 59 个视频组成,一段视频中的若干帧可以构成一个集合,其中,人脸图像由 Viola-Jones 人脸检测器^[35]获得,并且缩放为 20×20 的大小.按照文献[5]的设置,数据采用直方图均衡化进行预处理,以消除不同光照的影响.在训练分类器时,20 个集合作为训练集(每个人选择一个集合),剩下的 39 个集合作为测试集.
- Mobo 数据集涵盖了 24 个人的 96 段动作视频,人脸图像的提取采用与 Honda 类似的方法,与在文献[5]中的设置类似,人脸图像被缩放为 40×40 ,并且采用局部二值模式(local binary pattern,简称 LBP)^[36]作为输入特征.在训练分类器时,每个人选择一个集合构成训练集,其他的作为测试集.
- YTC 数据集包含了 47 个人的 1 910 段视频.对于每个人,这些视频中包含了大量的光照、姿态、表情等变化.人脸图像仍然用 Viola-Jone 人脸检测器^[35]获得.每幅图像缩放到 20×20 ,并采用 LBP 直方图作为输入特征.与文献[7]的实验设置类似,数据集被平分为尽量不重叠的 5 部分,对于每个部分,每个人约

有 9 个图像集,其中,随机选取 3 个作为训练集,剩下的组成测试集.

图 2 给出了 3 个数据集的样例,分类结果在表 1 中(最好的结果用加粗标出).可以看出,由于 Honda 与 Mobo 两个数据集相对简单,所有算法都取得了较好的结果,其中,本文提出的 DSRID-S 与 DSRID-(E+G+S)算法取得了最好的结果.另一方面,由于 YTC 数据集的复杂性,使其更具有挑战性,基于欧式空间方法在此数据集上表现较好,融合类方法方法仍然优于其他类方法,其中,DSRID-(E+G+S)取得了最好的结果.

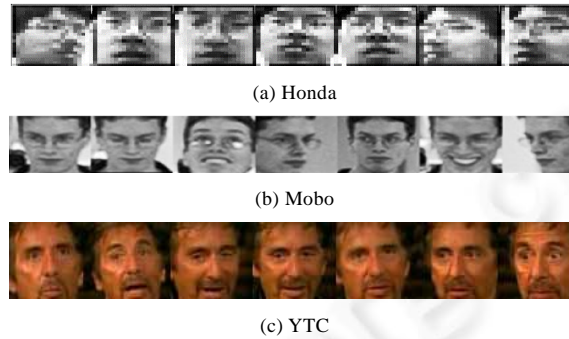


Fig.2 Examples of Honda, Mobo and YTC datasets

图 2 Honda, Mobo 和 YTC 的数据集样例

Table 1 Results for set-based face recognition

表 1 基于集合的人脸识别结果

类别	算法	Honda	Mobo	YTC
基于欧式空间	AHISD	93.81±2.97	97.70±1.43	73.42±2.78
	CHISD	94.62±2.54	96.76±1.30	73.73±3.90
	SANP	91.79±3.37	97.92±1.18	73.61±3.36
	RNP	96.92±2.64	98.06±1.87	74.08±3.87
基于子空间	DCC	92.01±3.21	91.25±1.61	69.12±3.81
	MMD	90.25±2.35	92.50±3.53	71.13±3.14
	MDA	92.82±4.80	96.66±1.63	75.82±3.95
基于 SPD 流形	CDL	99.49±1.08	88.86±3.10	66.04±3.89
	RSR	99.74±0.81	94.10±2.38	63.09±3.26
基于组合度量	HERML	98.97±1.32	97.31±2.33	76.50±3.34
	DARG	99.23±1.05	97.62±0.87	76.98±3.05
本文方法	DSRID-E	96.92±2.02	98.19±1.14	77.23±3.82
	DSRID-G	99.74±0.81	97.50±1.43	72.66±3.88
	DSRID-S	100.00±0.00	98.06±1.17	71.33±3.41
	DSRID-(E+G+S)	100.00±0.00	98.75±1.22	80.21±3.65

6.2 基于图像集的动作识别

基于图像集的动作识别,我们选用了 UCF11 数据库^[37].该库涵盖了包含了投篮、骑车、跳水等 11 类动作,具体样例如图 3 所示.数据来源于真实视频,因此涵盖了角度、光影、遮挡等变化.每一类含有 25 组数据,每组含有 4 个以上的同类视频片段.在特征表示阶段,我们利用了该库提供的标注信息,截取出每个视频片段中每一帧的动作信息并缩放成 20×20 大小,进而提取其方向梯度直方图(histogram of oriented gradient,简称 HOG)特征,每个视频构成一个集合.测试方案与文献[37]相同,从每类中各取一组作为测试集,其余为训练集.25 次的平均结果见表 2(最好的结果用加粗标出).可以看出,基于 SPD 流形的方法要优于其他两类基于欧式空间和子空间的方法.DSRID 在 3 种空间下的实现均优于相应空间下的其他算法,DSRID-(E+G+S)算法仍然取得了最好的性能,这也验证了融合类方法的有效性.



Fig.3 Examples of UCF11 dataset

图 3 UCF11 数据集样例

Table 2 Results for set-based action recognition

表 2 基于集合的动作识别结果

类别	算法	UCF11
基于欧式空间	AHISD	61.37±6.18
	CHISD	64.51±6.64
	SANP	54.84±6.41
	RNP	68.32±5.82
基于子空间	DCC	70.94±5.78
	MMD	65.46±5.29
	MDA	72.74±6.47
基于 SPD 流形	CDL	83.14±5.04
	RSR	81.58±6.20
基于组合度量	HERML	85.01±5.69
	DARG	85.99±7.02
本文方法	DSRID-E	78.33±6.67
	DSRID-G	82.36±5.23
	DSRID-S	86.92±4.42
	DSRID-(E+G+S)	88.43±4.50

6.3 基于图像集的物体识别

针对基于图像集的物体识别任务,我们选用了两个公共数据集:ETH-80^[38]和 RGB-D^[39].

- ETH-80 涵盖 8 类物体,每个类别又分为 10 个子类,每个子类包含 41 幅多角度的 32×32 大小的图像.这里,使用 LBP 直方图作为输入特征.其中,每个子类形成一个集合,每个物体随机选取 5 个集合组成训练集,剩下 5 个集合为测试集.
- RGB-D 包含 51 类物体,每个而物体包含多角度的视频序列,一个序列为一个集合.这里,我们只使用了 RGB 信息而没有使用深度信息,每帧大小为 32×32,使用灰度图像作为输入特征,每个物体随机选取 3 个集合组成训练集,剩余为测试集.

图 4 给出了两个物体库的样例,实验结果见表 3(最好的结果用加粗标出).



(a) ETH-80



(b) RGB-D

Fig.4 Examples of ETH-80, RGB-D datasets

图 4 ETH-80,RGB-D 数据集样例

Table 3 Results for set-based object categorization**表 3** 基于集合的物体识别结果

类别	算法	ETH	RGB-D
基于欧式空间	AHISD	79.50±6.06	62.41±1.70
	CHISD	84.17±4.78	64.84±2.25
	SANP	83.50±3.48	44.90±2.15
	RNP	88.50±4.89	65.90±2.89
基于子空间	DCC	88.00±4.83	71.16±2.94
	MMD	84.85±4.93	52.52±1.00
	MDA	86.75±4.57	62.51±3.04
基于 SPD 流形	CDL	94.75±3.47	80.54±3.24
	RSR	92.25±4.05	78.91±2.03
基于组合度量	HERML	94.88±3.97	79.59±3.82
	DARG	95.75±3.53	81.08±2.72
本文方法	DSRID-E	87.28±4.28	67.35±3.87
	DSRID-G	91.63±4.08	79.93±2.40
	DSRID-S	96.37±2.36	83.61±2.39
	DSRID-(E+G+S)	96.88±2.79	83.88±2.87

可以看出,基于 SPD 流形的方法明显要优于基于欧式距离和子空间的方法,这也与文献[8]中的结论相一致.在这些方法中,本文提出的 DSRID-(E+G+S)取得了最高的准确率.

6.4 子结构数量对准确率的影响

DSRID 中的一个重要参数是需要确定子结构的数量,图 5 给出了在 Mobo 数据库上,子结构数量变化对 DSRID 的 4 种实现算法(DSRID-E,DSRID-G,DSRID-S 和 DSRID-(E+G+S))的准确率的影响.可以看出,4 种算法的变化趋势大致相同,准确率随着子结构的数量增加先是增加,然后逐渐下降.这表明了以子结构为基础进行图像集距离学习的有效性.集合内多子结构的特点,使得较少数量的子结构设定无法反映集合真实的数据分布;而子结构的过划分,同样会破坏每个子结构的完整性.

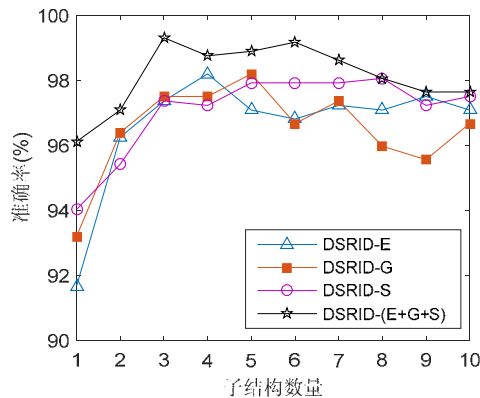


Fig.5 Effect of number of sub-structures to accuracy on Mobo dataset

图 5 Mobo 数据集上子结构数量对准确率的影响

6.5 运行时间对比

本节分析不同算法的运行时间,在表 4 中,我们列出了所有待比较的 15 种算法的训练时间和测试时间(s).所有算法均采用 Matlab 实现,并且运行于 4GHz 的 4 核机器上.可以看出,基于组合度量的方法(HERML,DARG,DSRID-(E+G+S))由于其需要融合多种度量方式,因此相对耗时较长.DSRID 在取得最好的分类准确率的基础上,在基于欧式空间的方法中,DSRID-E 的训练时间虽然高于 AHISD,CHISD,SANP 和 RNP,但测试时间用时最短.在基于子空间的方法中,DSRID-G 用时多于 DCC,但少于 MMD 和 MDA.在基于 SPD 流形的方法中,由于

DSRID-S 考虑了多个子结构,因此用时高于 CDL 与 RSR.

Table 4 Comparison of algorithms on running time for Mobo dataset
表 4 Mobo 数据集上各算法运行时间比较

类别	算法	训练时间(s)	测试时间(s)
基于欧式空间	AHISD	N/A	462.72
	CHISD	N/A	1 700.85
	SANP	N/A	1 844.45
	RNP	1.76	25.96
基于子空间	DCC	2.46	6.57
	MMD	51.76	234.02
	MDA	78.35	235.05
基于 SPD 流形	CDL	20.22	29.46
	RSR	19.40	27.31
基于组合度量	HERML	89.93	233.16
	DARG	35.66	116.34
本文方法	DSRID-E	13.54	23.63
	DSRID-G	12.12	24.12
	DSRID-S	38.64	99.47
	DSRID-(E+G+S)	64.42	159.43

7 结束语

本文提出了以集合内显著子结构为基础的图像集度量方法,并把该方法扩展到欧式空间和流形空间上,在以集合为基础的人脸识别、动作识别、物体分类任务中验证了其有效性.本文使用不同的度量方式,如何更加有效地确定子结构数量以及如何结合这些不同空间上的表示,成为下一阶段研究的重点.

References:

- [1] Harandi M, Salzmann M, Baktashmotlagh M. Beyond Gauss: Image-Set matching on the Riemannian manifold of PDFs. In: Proc. of the Int'l Conf. on Computer Vision. Piscataway: IEEE, 2015. 4112–4120. [doi: 10.1109/ICCV.2015.468]
- [2] Hayat M, Bennamoun M, An S. Deep reconstruction models for image set classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(4):713–727. [doi: 10.1109/TPAMI.2014.2353635]
- [3] Huang Z, Wang R, Shan S, Chen X. Hybrid Euclidean-and-Riemannian metric learning for image set classification. In: Proc. of the Asian Conf. on Computer Vision. Berlin: Springer-Verlag, 2014. 562–577. [doi: 10.1007/978-3-319-16811-1_37]
- [4] Harandi M, Hartley R, Shen C, Lovell B. Extrinsic methods for coding and dictionary learning on Grassmann manifolds. Int'l Journal of Computer Vision, 2015,114(2-3):113–136. [doi: 10.1007/s11263-015-0833-x]
- [5] Yang M, Zhu P, Van Gool L, Zhang L. Face recognition based on regularized nearest points between image sets. In: Proc. of the IEEE Int'l Conf. and Workshops on Automatic Face and Gesture Recognition. Piscataway: IEEE, 2013. 1–7. [doi: 10.1109/FG.2013.6553727]
- [6] Cevikalp H, Triggs B. Face recognition based on image sets. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2010. 2567–2573. [doi: 10.1109/CVPR.2010.5539965]
- [7] Hu Y, Mian AS, Owens R. Face recognition using sparse approximated nearest points between image sets. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(10):1992–2004. [doi: 10.1109/TPAMI.2011.283]
- [8] Wang R, Guo H, Davis LS, Dai Q. Covariance discriminative learning: A natural and efficient approach to image set classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012. 2496–2503. [doi: 10.1109/CVPR.2012.6247965]
- [9] Wang R, Chen X. Manifold discriminant analysis. In: Proc. of the IEEE Conf. Computer Vision and Pattern Recognition. Piscataway: IEEE. 2009. 429–436. [doi: 10.1109/CVPRW.2009.5206850]
- [10] Wang R, Shan S, Chen X, Gao W. Manifold-Manifold distance with application to face recognition based on image set. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587719]

- [11] Harandi MT, Sanderson C, Hartley R, Lovell BC. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In: Proc. of the European Conf. on Computer Vision. Berlin: Springer-Verlag, 2012. 216–229. [doi: 10.1007/978-3-642-33709-3_16]
- [12] Cherian A, Sra S. Riemannian sparse coding for positive definite matrices. In: Proc. of the European Conf. on Computer Vision. Berlin: Springer-Verlag, 2014. 299–314. [doi: 10.1007/978-3-319-10578-9_20]
- [13] Shakhnarovich G, Fisher JW, Darrell T. Face recognition from long-term observations. In: Proc. of the European Conf. on Computer Vision. Berlin: Springer-Verlag, 2002. 851–865. [doi: 10.1007/3-540-47977-5_56]
- [14] Arandjelovic O, Shakhnarovich G, Fisher J, Cipolla R, Darrell T. Face recognition with image sets using manifold density divergence. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2005. 581–588. [doi: 10.1109/CVPR.2005.151]
- [15] Kim TK, Kittler J, Cipolla R. Discriminative learning and recognition of image set classes using canonical correlations. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(6):1005–1018. [doi: 10.1109/TPAMI.2007.1037]
- [16] Chen S, Sanderson C, Harandi MT, Lovell BC. Improved image set classification via joint sparse approximated nearest subspaces. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012. 452–459. [doi: 10.1109/CVPR.2013.65]
- [17] Wang W, Wang R, Huang Z, Shan S, Chen X. Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 2048–2057. [doi: 10.1109/CVPR.2015.7298816]
- [18] Zhu P, Zuo W, Zhang L, Shiu SCK, Zhang D. Image set-based collaborative representation for face recognition. IEEE Trans. on Information Forensics and Security, 2014,9(7):1120–1132. [doi: 10.1109/TIFS.2014.2324277]
- [19] Arsigny V, Fillard P, Pennec X, Ayache N. Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications, 2007,29(1):328–347. [doi: 10.1137/050637996]
- [20] Pennec X, Fillard P, Ayache N. A Riemannian framework for tensor computing. Int'l Journal of Computer Vision, 2006,66(1): 41–66. [doi: 10.1007/s11263-005-3222-z]
- [21] Harandi MT, Sanderson C, Shirazi S, Lovell BC. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2011. 2705–2712. [doi: 10.1109/CVPR.2011.5995564]
- [22] Huang Z, Wang R, Shan S, Chen X. Learning Euclidean-to-Riemannian metric for point-to-set classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014. 1677–1684. [doi: 10.1109/CVPR.2014.217]
- [23] Zhu P, Zhang L, Zuo W, Zhang D. From point to set: Extend the learning of distance metrics. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Piscataway: IEEE, 2013. 2664–2671. [doi: 10.1109/ICCV.2013.331]
- [24] Huang Z, Wang R, Shan S, Chen X. Projection metric learning on Grassmann manifold with application to video based face recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 140–149. [doi: 10.1109/CVPR.2015.7298609]
- [25] Lu J, Wang G, Deng W, Moulin P. Simultaneous feature and dictionary learning for image set based face recognition. In: Proc. of the European Conf. on Computer Vision. Berlin: Springer-Verlag, 2014. 65–280. [doi: 10.1007/978-3-319-10590-1_18]
- [26] Lu J, Wang G, Deng W, Moulin P, Zhou J. Multi-Manifold deep metric learning for image set classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 1137–1145. [doi: 10.1109/CVPR.2015.7298717]
- [27] Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2002. 849–856.
- [28] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine Learning, 2010,3(1):1–122. [doi: 10.1561/22000000016]
- [29] Bach F, Jenatton R, Mairal J, Obozinski G. Structured sparsity through convex optimization. Statistical Science, 2012,27(4): 450–468. [doi: 10.1214/12-STS394]
- [30] Lin Z, Liu R, Su Z. Linearized alternating direction method with adaptive penalty for low-rank representation. In: Proc. of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2011. 612–620.

- [31] Elhamifar E, Vidal R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013,35(11):2765–2781. [doi: 10.1109/TPAMI.2013.57]
- [32] Lee KC, Ho J, Yang MH, Kriegman D. Video-Based face recognition using probabilistic appearance manifolds. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2003. 313–320. [doi: 10.1109/CVPR.2003.1211369]
- [33] Gross R, Shi J. The CMU motion of body database. Technical Report, CMU-RI-TR-01-18, Pittsburgh: Robotics Institute, Carnegie Mellon University, 2001. [doi: 10.1515/9781400843251-003]
- [34] Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2011. 529–534. [doi: 10.1109/CVPR.2011.5995566]
- [35] Viola P, Jones MJ. Robust real-time face detection. *Int'l Journal of Computer Vision*, 2004,57(2):137–154. [doi: 10.1023/B:VISI.000013087.49260.fb]
- [36] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002,24(7):971–987. [doi: 10.1109/TPAMI.2002.1017623]
- [37] Liu J, Luo J, Shah M. Recognizing realistic actions from videos in the wild. In: *Proc. of the Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2009. 1996–2003. [doi: 10.1109/CVPR.2009.5206744]
- [38] Leibe B, Schiele B. Analyzing appearance and contour based methods for object categorization. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2003. 409–415. [doi: 10.1109/CVPR.2003.1211497]
- [39] Lai K, Bo L, Ren X, Fox D. A large-scale hierarchical multi-view RGB-D object dataset. In: *Proc. of the Int'l Conf. on Robotics and Automation*. Piscataway: IEEE, 2011. 1817–1824. [doi: 10.1109/ICRA.2011.5980382]



刘博(1981 -),男,河北保定人,博士,主要研究领域为子空间聚类,半监督学习,人脸识别.



于剑(1969 -),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为聚类分析,图像处理.



景丽萍(1978 -),女,博士,教授,博士生导师,CCF 专业会员,主要研究领域为数据挖掘,文本挖掘,生物信息学,企业智能.