软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2015,26(11):2897−2911 [doi: 10.13328/j.cnki.jos.004894]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

视频人脸识别中判别性联合多流形分析*

于 谦[1,2], 高 阳[2], 霍 静[2], 庄韫恺[2]

[1](山东女子学院 信息技术学院,山东 济南 250300)

[2](计算机软件新技术国家重点实验室(南京大学),江苏 南京 210023)

通讯作者: 于谦, E-mail: yuqian_sd@sdwu.edu.cn

摘 要: 将基于视频的人脸识别转换为图像集识别问题,并提出两种流形来表示每个图像集:一种是类间流形,表示每个图像集的平均脸信息;另一种是类内流形,表示每个图像集的所有原始图像的信息.类间流形针对图像集之间的区别提取整体判别信息,作用是选出几个与待识别图像集较为相似的候选图像集.类内流形则考虑图像集内各原始图像之间的关系,负责从候选图像集中找出最为相似的一个.不同于现有的非线性流形方法中每幅图像对应流形中的一个点,采用分片技术学习两种流形的投影矩阵,每个分片对应流形中的一个点,所学到的特征更具有判别性,进而使流形边界更加清晰,同时解决了传统非线性流形方法中的角度偏差和不充分采样问题.还提出了与分片技术相匹配的流形之间的距离度量方法.最后在几个广为研究的数据集上进行了实验,结果表明:新方法的识别准确率高,尤其适用于不受控环境下的视频识别,而且不受视频段长短的影响.

关键词: 基于视频的人脸识别;图像集;分片;多流形;相似性度量

中图法分类号: TP391

中文引用格式: 于谦,高阳,霍静,庄韫恺.视频人脸识别中判别性联合多流形分析.软件学报,2015,26(11):2897−2911. http://www.jos.org.cn/1000-9825/4894.htm

英文引用格式: Yu Q, Gao Y, Huo J, Zhuang YK. Discriminative joint multi-manifold analysis for video-based face recognition. Ruan Jian Xue Bao/Journal of Software, 2015,26(11):2897−2911 (in Chinese). http://www.jos.org.cn/1000-9825/4894.htm

Discriminative Joint Multi-Manifold Analysis for Video-Based Face Recognition

YU Qian[1,2], GAO Yang[2], HUO Jing[2], ZHUANG Yun-Kai[2]

[1](College of Information Technology, Shandong Women's University, Ji'nan 250300, China)

[2](State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

Abstract: In this paper, video-based face recognition (VFR) is converted into image set recognition. Two types of manifolds are proposed to represent each gallery set: one is inter-class manifold which represents mean face information of this set, and the other is intra- class manifold corresponding to original images information of this set. The inter-class manifold abstracts discriminative information of the whole image set so as to select a few candidate gallery sets relevant to query set. The intra-class manifold chooses the most similar one from candidate sets by considering the relationships among all original images of each gallery set. Existing nonlinear manifolds methods project each image into low dimensional space as a point, thus suffer from cluster alignment and un-sufficient sampling. In order to avoid the above flaws and make the margin clearer between manifolds, projecting matrices in new method are gotten by means of partitioning image into un-overlapping patches so that features extracted this way can be more discriminative. In addition, a method of similarity measure between manifolds is proposed in accordance with the patching scheme. Finally, extensive experiments are conducted on several widely studied databases. The results demonstrate that new method achieves better performance than those state-of-the-art VFR methods, and it especially works well in un-controlled videos without being affected by the length of video.

* 基金项目: 国家自然科学基金(61035003, 61175042, 61305068 , 61432008); 山东省高等学校青年骨干教师国内访问学者项目; 山东省高等学校科技计划(J15LN58); 山东女子学院数据挖掘科研创新团队基金

收稿时间: 2015-05-19; 修改时间: 2015-07-14; 定稿时间: 2015-08-26

Key words: video-based face recognition; image set; patch; multi-manifold; similarity measure

1 引 言

摄像机的广泛使用和网络社交的蓬勃兴起,使人们可以轻易地获取大量的视频段,这为基于视频的人脸识别技术的研究提供了便利.而在当今生活中,基于视频的人脸识别技术也起着越来越重要的作用,我们可以利用公共场所的视频段查询和确认犯罪份子的身份、帮助寻找走失和被拐儿童,还可以利用视频信息完成网上银行开户、在线支付等功能,保障和促进信息安全、互联网金融等领域的快速发展.因此,基于视频的人脸识别在近年来成为研究热点.相对于在受控环境下进行的传统人脸识别任务,来自不受控环境的视频人脸识别更具有挑战性,需要解决分辨率低、光照和姿态变化剧烈、遮挡等问题,如何充分利用视频中的人脸信息以解决这些问题,一直是本领域研究的核心内容[1].

1.1 研究现状

在基于视频的人脸识别中,度量待识别的视频段与视频库中的各视频段之间的距离是核心问题,目前,解决该问题的方法主要分成 3 种,即,基于视频序列的(sequence-based)、基于图像集的(image set-based)、基于稀疏表示和字典学习的(sparse and dictionary-based).表 1 中对各种方法的代表文献进行了总结.

- 基于视频序列的方法[2−4]

充分利用视频的空间时间序信息,通过训练模型将同一视频序列中不同表情或姿态的人脸组合在一起,能够很好地刻画人脸的动态变化[5],但识别效果易受视频序列长短或光照的影响.

- 基于图像集的方法

将每段视频都看成一个无序的图像集,待识别视频段的图像集称为疑问集(query set),视频库中的称为画廊集(gallery set),每个已知视频段都对应一个画廊集.先对每个图像集进行建模,再度量疑问集与每个画廊集之间的距离.从建模的角度讲,本方法可分为有参数[6,7]和无参数[8−18]两种.有参数模型中,每个图像集都用一个带参数的分布函数来描述,一般用高斯模型,两个分布之间的 KL 距离(kullback-leibler divergence)代表了两个图像集的相似度,当相比较的图像集都服从同一分布而且是指定的模型时效果是好的;否则,识别的效果将变差.在无参数模型中,每个图像集被表示成一个线性子空间(linear subspace)或非线性流形(nonlinear manifold),用子空间之间的距离和流形到流形的距离(manifold–to-manifold distance,简称 MMD)来度量两个图像集的相似度.子空间的方法易被噪声和遮挡干扰,而且当光照或表情姿态等变化时,即使属于同一段视频中的图像也会表现出明显的非线性[19,20].基于 MMD 的方法在很多被广泛研究的人脸库中都取得了很好的效果,但在每个非线性流形中的子线性片聚类时,存在角度偏差现象[16].例如,在一个图像集的 Manifold 中,侧脸角度 15º~30º 的聚为一个线性片;但在另一个图像集中,有可能是 25º~40º 的聚在一起,因此影响了距离计算.为解决这个问题,Cui 等人[16]提出了一个参照图像集(reference set),尽量涵盖所有的角度.每个画廊集和疑问集都按参照图像集的方式聚类产生子线性片,解决了角度偏差问题,但是一个“好”的参照图像集并不好设定.而且在基于 MMD 的方法中,当图像集中所包含的图像非常少,即,不充足采样时,不能很好地进行聚类,会影响识别的效果.

- 基于稀疏表示和字典学习的方法

近年来,越来越多的研究者通过稀疏表示和字典学习来解决基于图像集的人脸识别问题[21−27].Chen YC 等人[22]和 Chen SK 等人[23]都是令疑问集与待比较画廊集做同样方式的分块,分别用重建误差和块之间的距离做相似性度量,都对灯光和姿态变化鲁棒.但前一方法中,子字典的寻找和图像重建较耗时;而后一方法中存在有些分块不被使用的情况,而且两者都存在角度偏差问题.Elhamifar 等人[21]曾提出过一种类级稀疏(class-level sparsity),即把每个图像集都看成一个集成体,然后再在所有集成体上进行稀疏运算.Cui 等人[24]采用了这种思想,同时提出了一种原子级稀疏(atom-level sparsity),通过两层稀疏,在若干个与疑问集相似的画廊集中重建疑问集中的每个图像,取得了很好的效果,而且对噪声鲁棒.但对两个参数的设置比较敏感,而且没有给定可以满足大多需求的参数的通用取值,在某种程度上成为用户的一种负担.Bhatt 等人[25]首次提出融合同一视频中的多

个帧的有序列表来形成该视频的可判别签名的方法,充分利用了空间时间序和人脸的局部信息.但正由于需要利用这些大量的信息,所以计算量很大,尤其是视频较长时.Patel 等人[26]提出了一种同时学习特征矩阵和字典的方法,解决了分开学习时产生的判别信息减损问题.但所得到的特征矩阵和字典不是在全局最优下求得,因此当人数较多时,识别效果会变差.

Table 1 Categorization of existing methods for VFR

表 1 基于视频的人脸识别现有方法分类

类别		作者	技术特点
基于视频序列		Liu 等人[2], Kim 等人[3]	通过训练隐 Markov 模型把同一段视频中的人脸按表情聚类,用后验概率度量两视频之间的相似性.当视频段较长时效果较好.
		Lee 等人[4]	为不同姿态的人脸分别构造低维分段线性流形,利用贝叶斯推理建立不同姿态间的转移矩阵,将识别问题转换为最大后验概率估算问题.充分利用了时序信息,在姿态及表情变化剧烈时效果较好,但受灯光变化影响.
基于图像集	高斯模型	Arandjelović 等人[6], Shakhnarovich 等人[7]	用高斯模型表示图像集,用 KL 偏差(Kullback-Leibler divergence)度量相似性.存在参数估计问题,要求每个图像集都符合高斯分布.
	线性子空间	Cevikalp 等人[8], Aggarwal 等人[10], Fukui 等人[11], Kim 等人[12], Yamaguchi 等人[17], Hu 等人[18]	每个图像集被表示成一个线性子空间,用主角(principal angle)、仿射包或凸包的相似性以及最近点(sparse approximated nearest points,简称 SANP)方法来表示子空间之间的距离.不适用于在不同光照、表情、姿态等条件下获取的图像,而且对噪声敏感.
	非线性流形	Huang 等人[9], Sanderson 等人[13], Wang 等人[14,15]	每个图像集都用一个非线性流形表示,每个流形再被聚类成若干个局部线性片,两个流形间的所有线性片的平均距离即为它们之间的距离.在聚类时人脸对齐的角度不确定,从而影响距离度量.
		Cui 等人[16]	提出一个参照集来解决非线性流形中角度偏差问题,但如何设定一个“好”的参照集是个问题.
基于稀疏表示和字典学习		Cui 等人[24]	每个图像集都被当成一个整体对待,利用类间稀疏选出与疑问集相似的几个画廊集,再利用类内稀疏重建疑问集中的所有图像,重建误差度量相似度
		Chen YC 等人[22]	先把每个图像集按照光照或姿态分成不同的部分,每个子块学习一个子字典.测试阶段,疑问集中的每个子块都会在与之比较的画廊集中找到合适的子字典,然后在该子字典内重建子块中的每个图像,所有图像的重建误差即为两个图像集的相似度.称为 joint sparse representation(JSR).
		Chen SK 等人[23]	先用稀疏近似抽取每个图像集的各个局部线性子空间,再将其投影成 Grassmann 流形中的一个点,两个图像集相比较时,利用联合稀疏近似的方法为一个图像集中的每个点找来自另一图像集的最近点,所有最近点对之间的距离即为两图像集之间的相似性(sparse approximated nearest subspaces,简称 SANS).
		Bhatt 等人[25]	先建一个由静态脸构成的庞大的字典,对于输入的视频,每一帧都从字典里选出相似的若干张脸组成一个列表,然后将所有帧的列表通过聚类、重排、融合形成一个新的有序列表,即为该视频的签名,最终将视频人脸识别转换为两个有序列表的比较.
		Patel 等人[26]	将整个 Gallery 看成一个整体同时学习一个特征投影矩阵和一个结构化字典,每个画廊集构成一个子字典,识别时为疑问集的每个帧都选择一个子字典,最后用多数投票决定识别结果,计算量较大.

1.2 相关工作分析

本文的工作是基于图像集的.上述的每种方法都有自己的优势,适用于不同的场合,Hadid 等人[27]曾指出,从同一段视频中获取的人脸图像会位于一个光顺(smooth)的流形中.因此,非线性流形的方法适用于基于图像集的视频人脸识别.在现有的基于 MMD 的非线性流形方法中,其投影策略是每个图像集对应一个非线性流形,其中的每幅图像对应流形中的一个点.受困于角度偏差和不充足采样,而且包括上述的其他方法在内,MMD 方法在对来自现实不受控环境中的视频片段(YTC 库[28])识别的效果还有非常大的提升空间.那么,如果不采用这种投影策略呢?

Lu 等人[29]曾用分片投影构造非线性流形的方法(discriminative multi-manifold analysis,简称 DMMA)处理过单人脸识别问题,先将单人脸图像分成若干个互不重叠的子片(patch),再投影到低维流形空间,每张图像对应一个流形,每个子片对应流形中的一个点,取得了较好的识别效果.这说明利用分片思想将单个人脸图像投影成非线性流形是合理的,若将其用于基于图像集的人脸识别,首先可以避免角度偏差和不充足采样现象,后面的实验还将表明,尤其适用于来自不受控环境的视频和不充足采样识别.实施这种思想之前,我们先考虑两种关系:图像集与图像集之间以及图像集内部各个图像之间的关系.Cui 等人[24]提出的类级和原子级稀疏给了本文工作一个很好的启发,我们提出两种流形:一种是把每个图像集看成一个集成体,各自形成一个流形,集成的形式是各自的平均脸(mean face),提取整体判别信息处理图像集之间的关系,称为类间流形(Inter-class manifold);另一种是兼顾图像集内部各个图像之间关系的类内流形(Intra-class manifold),仍是一个图像集对应一个流形,但却是直接利用每个原始图像的判别信息,由所有原始图像共同构成.给定疑问集后,先利用类间流形经过较少的计算找出与疑问集最相似的几个画廊集(称为候选图像集),再利用类内流形从几个候选图像集中找出最相似的一个,然后用分片的方式获取这两种流形.投影的指导思想与 DMMA 类似,即,来自同一张脸的子片尽可能地近,来自不同脸的尽量离得远.每个画廊集都有明确的类间和类内投影矩阵,在测试阶段,不必再重复计算.我们把新方法称为判别性联合多流形分析(discriminative joint multi-manifold analysis,简称 DJMMA),实验结果表明,新方法在几个公开研究的视频库中,比现有的方法具有更高的识别正确率.

1.3 新方法的贡献

- 每个分片都被投影成流形中的一个点,而且与来自不同图像的点彼此远离,充分利用了分片的几何信息,所学的局部特征强调了人脸的个性(person-specific)、抓住人脸的本质特性,这种特性不受分辨率、光照、姿态和年龄变化的影响,比其他方法中所学的通用特征更具有判别性[29,30].
- 在类间流形阶段,仅仅依靠平均脸就可以找到与测试目标最为相似的几个候选集,此时,如果正确率足够高的话则可以省略类内流形阶段.事实上,平均脸阶段在大多数数据集上都表现良好,而类内流形则为我们提供了更高的精度保障.
- 由于在类间流形阶段每个图像集都被转换为一个平均脸,所以是将基于图像集的人脸识别问题转换成了单脸识别,即,DJMMA 可以直接用于单脸识别.
- 新方法更适用于来自现实生活的视频识别,而且在视频过短时仍有很高的识别率.

本文第 2 节介绍 DJMMA 的基本思想、函数模型和优化细节.第 3 节是实验结果和分析.第 4 节是总结.

2 判别性联合多流形分析

首先给出方法概述,然后给出数学公式、优化过程和相关实验,第 2.3 节是识别过程介绍,最后是算法分析.

2.1 基本思想

图 1 显示了 DJMMA 的训练阶段,分成类间流形和类内流形两部分(M_i 是类间流形,W_i 是类内流形,i=1,2,…,n,n 是参与训练的人数).在类间流形部分,首先为每个画廊集计算平均脸,用相同的方式,以相同的尺寸将每个平均脸分成互不重合的若干个子片,然后将这些子片投影到同一个流形中,每个平均脸对应一个流形;在类内流形部分,每个画廊集的所有原始图像按照平均脸的方式分片,然后将其投影到同一个流形中,每个画廊集对应一个流形.

图 2 显示了新方法的测试阶段(M_2 和 M_n 对应两个候选集,再通过 W_2 和 W_n 找出与疑问集距离最近的一个):给定疑问集后,首先计算平均脸将之分片,然后将子片分别投影到各个已经计算好的画廊集流形上,计算投影后的疑问集与各个画廊集流形之间的距离,选出若干个最为相近的图像集,如图 3 中的 M_2 对应的 Set 2 和 M_n 对应的 Set n.然后,将疑问集的各个原始图像的所有子片投影到所选出的候选集所对应的类内流形上,计算它们之间的距离,距离最近的即为识别结果.

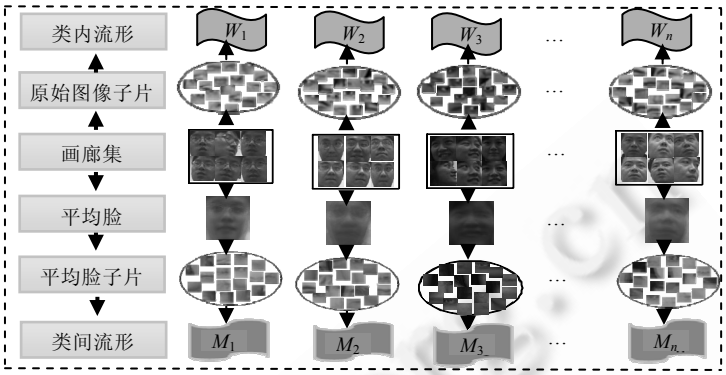


Fig.1 Training stage of DJMMA

图 1 DJMMA 的训练阶段

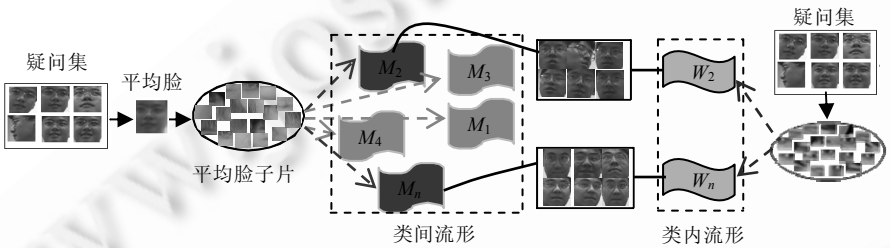


Fig.2 Testing stage of DJMMA

图 2 DJMMA 的测试阶段

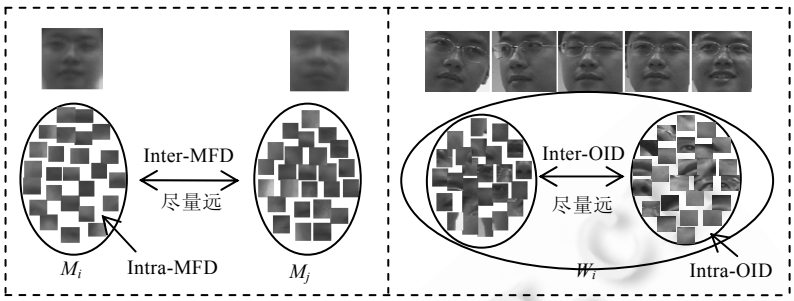
2.2 训练阶段

2.2.1 符号汇总

k_i 表示第 i 个画廊集所包含的原始图像个数,i=1,2,…n,n 是画廊集总数.每个平均脸和原始图像都被分成 t 个子片,每个子片的尺寸是 $a\times b$(像素).令 $d=a\times b$,$1\leqslant i\leqslant n$,$1\leqslant r\leqslant t$,$M=[M_1,M_2,...,M_n]$表示 n 个画廊集的平均脸,$M_i=[m_{i1},m_{i2},...,m_{it}]$,$m_{ir}$ 是第 i 个平均脸的第 r 个子片,$m_{ir}\in R^d$;令 $[X_{i1},X_{i2},...,X_{ik_i}]$ 表示第 i 个画廊集所包含的原始图像,$X_{is}=[x_{is1},x_{is2},\ldots,x_{ist}]$,$1\leqslant s\leqslant k_i$,$x_{isr}$ 是 Set_i 的第 s 个原始图像的第 r 个子片.当 $i\neq j$ 并且 $p\neq q$ 时,称 m_{ir},m_{jr},x_{upr} 和 x_{vqr} 为同位置子片,表示不同图像上的第 r 个子片,即,不同图像的同一处器官,每个子片在另一幅图像上只有一个同位置子片.用 $M_1,M_2,...,M_n$ 表示类间流形的投影矩阵,统称 M 族矩阵(M_family);用 $W_1,W_2,...,W_n$ 表示类内流形的投影矩阵,统称 W 族矩阵(W_family).其中,$M_i\in R^{d\times d_i}$,$W_i\in R^{d\times d_i}$,d_i 表示每个子片在低维空间的维数,我们的目标是学习这两种矩阵.本文余下部分所用符号都与本节中的定义相一致.

2.2.2 目标函数

DJMMA 想要获取的是这样的一个低维特征空间:同一幅图像的子片在低维空间中尽量离得近,不同图像的子片尽量离得远.为了达到这样一个目标,根据最大化类间距离最小化类内距离的原则,所建的目标函数包含以下 4 个部分:(1) 来自不同平均脸的同位置子片之间的距离,称为 Inter-Mean Face Distance(Inter- MFD);(2) 来自同一图像集但不同原始图像的同位置子片之间的距离,称为 Inter-Original Image Distance(Inter- OID);(3) 来自同一平均脸的不同子片之间的距离,称为 Intra-Mean Face Distance(Intra-MFD);(4) 来自同一原始图像的不同子片之间的距离,称为 Intra-Original Image Distance(Intra-OID).投影原则如图 3 所示(Inter-MFD, Inter-OID 尽量大,Intra-MFD 和 Intra-OID 尽量小.M_i 和 M_j 分别是两个平均脸对应的类间流形,W_i 是某图像集所对应的类内流形).



(a) 类间流形(平均脸) (b) 类内流形(原始图像)

Fig.3 Desired distribution in low dimensional feature space

图 3 各子片在投影到低维空间后的理想分布

我们通过最大化 Inter-MFD 和 Inter-OID 并且最小化 Intra-MFD 和 Intra-OID 来获取投影矩阵,所以目标函数定义如下:

$$\max_{M_1,...,M_n,W_1,...,W_n} J(M_1,...,M_n,W_1,...,W_n)=J_1(M_1,...,M_n,W_1,...,W_n)-J_2(M_1,...,M_n,W_1,...,W_n)$$
$$=\sum_{i=1}^{n}\left(\sum_{r=1}^{t}\sum_{j\neq i}^{1:n}\| M_i^T m_{ir}-M_i^T m_{jr}\|^2 A_{ijr}+\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{j\neq s}^{1:k_i}\| W_i^T x_{isr}-W_i^T x_{ijr}\|^2 B_{isjr}\right)- \quad (1)$$
$$\sum_{i=1}^{n}\left(\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| M_i^T m_{ir}-M_i^T m_{iq}\|^2 C_{irq}+\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| W_i^T x_{isr}-W_i^T x_{isq}\|^2 D_{isrq}\right)$$

其中,

- J_1的前一项表示若 m_{ir}和 m_{jr}是来自不同平均脸同位置子片,则它们在低维空间应该尽量离得远;后一项表示如果 x_{isr}和 x_{ijr}是来自同一图像集的不同原始图像的同位置子片,则它们在低维空间中尽量分开.
- J_2的前一项表示若 m_{ir}和 m_{jq}来自同一平均脸,则它们在低维空间应该尽量离得近;后一项表示如果 x_{isr}和 x_{isq}是来自同一图像集的同一原始图像,则它们在低维空间中应该尽量离得近.
- A_{ijr},B_{isjr},C_{irq}和 D_{isrq}是亲合矩阵,表示的是不同图像中同位置子片之间、同一图像中所有不同子片之间的相似性,来自不同图像的语义上属于同一器官的子片相似性高,但在低维空间上却需要最大化它们之间的距离;反之亦然.公式定义如下:
  - $A_{ijr}=\exp(-\|m_{ir}-m_{jr}\|^2/\sigma^2)$,表示不同平均脸中同位置子片间的相似性;
  - $B_{isjr}=\exp(-\|x_{isr}-x_{ijr}\|^2/\sigma^2)$,表示同一图像集的不同原始图像的同位置子片间的相似性;
  - $C_{irq}=\exp(-\|m_{ir}-m_{iq}\|^2/\sigma^2)$,表示同一平均脸中不同分片间的相似性;
  - $D_{isrq}=\exp(-\|x_{isr}-x_{isq}\|^2/\sigma^2)$,表示同一原始图像中不同分片间的相似性.

仅有一个参数需要设定,通常令σ=100.下一节将优化 *M_family* 和 *W_family*.

2.2.3 优 化

我们将分别优化 *M_family* 和 *W_family*.当优化 $M_1,M_2,...,M_n$时,*W_family* 成为不影响优化结果的常量,公式(1)可写成如下形式:

$$\max_{M_1,...,M_n} J_1(M_1,...,M_n)=\sum_{i=1}^{n}\sum_{r=1}^{t}\sum_{j\neq i}^{1:n}\| M_i^T m_{ir}-M_i^T m_{jr}\|^2 A_{ijr}-\sum_{i=1}^{n}\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| M_i^T m_{ir}-M_i^T m_{iq}\|^2 C_{irq} \quad (2)$$

从上式中我们可以获取类间流形,它保证了所有来自不同平均脸的子片在低维空间中远远分开,而来自同一平均脸的则聚在一起.同理,当优化 $W_1,W_2,...,W_n$时,*M_family* 成为常量,公式(1)可写成如下形式:

$$\max_{W_1,...,W_n} J(W_1,...,W_n)=\sum_{i=1}^{n}\left(\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{j\neq s}^{1:k_i}\| W_i^T x_{isr}-W_i^T x_{ijr}\|^2 B_{isjr}-\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| W_i^T x_{isr}-W_i^T x_{isq}\|^2 D_{isrq}\right) \quad (3)$$

公式(3)确定的是类内流形,参与运算的是来自每个图像集的所有原始图像的子片.公式中的第 1 项保证了同一图像集中来自不同原始图像的子片在低维空间中被分开,第 2 项则来自相同图像的则聚在一起,所有子片共同构成该图像集的低维流形.

显然,公式(2)和公式(3)都没有封闭的解,和 DMMA 一样,我们将采用迭代的方法来解这两个公式.首先为所有的 M_i 和 W_i 赋一个合理的初始解,$1\leqslant i\leqslant n$,本文采用单位阵,然后依次求解 M_j 和 $W_j, j\neq i$.细节如下:

当给定 $M_1,M_2,\ldots,M_{i-1},M_{i+1},\ldots,M_n$ 时,M_i 可以这样解得:

$$\max_{M_i} J_1(M_i)=\left(\sum_{r=1}^{t}\sum_{j\neq i}^{1:n}\| M_i^T m_{ir}-M_i^T m_{jr}\|^2\ A_{ijr}+F_1\right)-\left(\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| M_i^T m_{ir}-M_i^T m_{iq}\|^2\ C_{irq}+F_2\right) \tag{4}$$

其中,

$$F_1=\sum_{u\neq i}^{1:n}\sum_{r=1}^{t}\sum_{j\neq u}^{1:n}\| M_u^T m_{ur}-M_u^T m_{jr}\|^2\ A_{ujr},F_2=\sum_{u\neq i}^{1:n}\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| M_u^T m_{ur}-M_u^T m_{uq}\|^2\ C_{urq}.$$

F_1 和 F_2 是与 M_i 无关的常量,所以可以被忽略.这样,公式(4)变为

$$\begin{aligned}\max_{M_i} J_1(M_i)&\approx\sum_{r=1}^{t}\sum_{j\neq i}^{1:n}\| M_i^T m_{ir}-M_i^T m_{jr}\|^2\ A_{ijr}-\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| M_i^T m_{ir}-M_i^T m_{iq}\|^2\ C_{irq}\\&=tr\left(M_i^T\left[\sum_{r=1}^{t}\sum_{j\neq i}^{1:n}(m_{ir}-m_{jr})(m_{ir}-m_{jr})^T A_{ijr}\right]M_i\right)-tr\left(M_i^T\left[\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}(m_{ir}-m_{iq})(m_{ir}-m_{iq})^T C_{irq}\right]M_i\right)\\&=tr(M_i^T H_1 M_i)-tr(M_i^T H_2 M_i)\end{aligned} \tag{5}$$

其中,

$$H_1=\sum_{r=1}^{t}\sum_{j\neq i}^{1:n}(m_{ir}-m_{jr})(m_{ir}-m_{jr})^T A_{ijr},H_2=\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}(m_{ir}-m_{iq})(m_{ir}-m_{iq})^T C_{irq}.$$

以同样的方式从公式(3)中优化 W_i:

$$\begin{aligned}\max_{W_i} J_2(W_i)&=\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{j\neq i}^{1:k_i}\| W_i^T x_{isr}-W_i^T x_{ijr}\|^2\ B_{isjr}+F_3-\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}\| W_i^T x_{isr}-W_i^T x_{isq}\|^2\ D_{isrq}-F_4\\&\approx tr\left(W_i^T\left[\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{j\neq i}^{1:k_i}(x_{isr}-x_{ijr})(x_{isr}-x_{ijr})^T B_{isjr}\right]W_i\right)-tr\left(W_i^T\left[\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{q\neq s}^{1:t}(x_{isr}-x_{isq})(x_{isr}-x_{isq})^T D_{isrq}\right]W_i\right)\\&=tr(W_i^T H_3 W_i)-tr(W_i^T H_4 W_i)\end{aligned} \tag{6}$$

其中,

$$F_3=\sum_{u\neq i}^{1:n}\sum_{s=1}^{k_u}\sum_{r=1}^{t}\sum_{j\neq u}^{1:k_u}\| W_u^T x_{usr}-W_u^T x_{ujr}\|^2\ B_{usjr},F_4=\sum_{u\neq i}^{1:n}\sum_{s=1}^{k_u}\sum_{r=1}^{t}\sum_{q\neq s}^{1:t}\| W_u^T x_{usr}-W_u^T x_{usq}\|^2\ D_{usrq}.$$

F_3 和 F_4 也是与优化无关的常量,所以被忽略,而

$$H_3=\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{j\neq s}^{1:k_i}(x_{isr}-x_{ijr})(x_{isr}-x_{ijr})^T B_{isjr},H_4=\sum_{s=1}^{k_i}\sum_{r=1}^{t}\sum_{q\neq r}^{1:t}(x_{isr}-x_{isq})(x_{isr}-x_{isq})^T D_{isrq}.$$

最后,通过解下面的两个特征值等式分别得到 M_i 和 W_i:

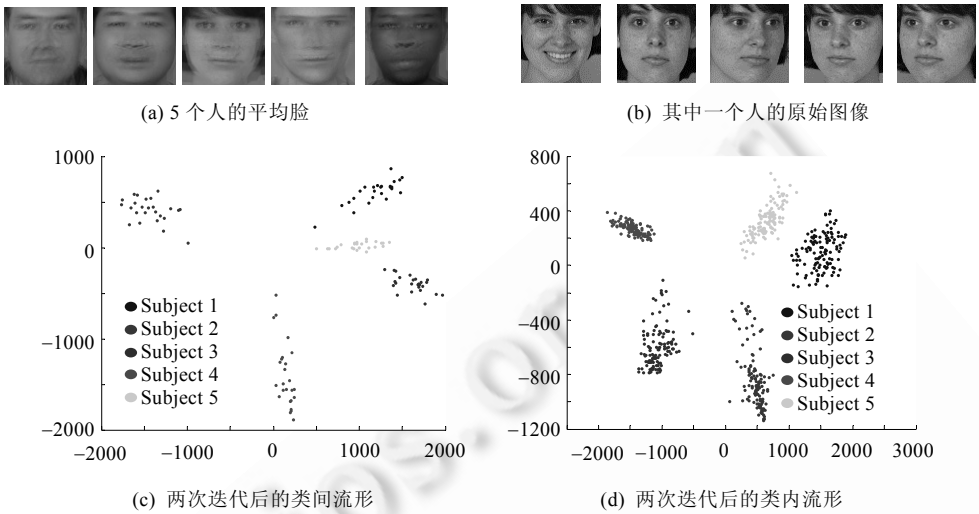
$$(H_1-H_2)\omega=\lambda\omega \tag{7}$$

$$(H_3-H_4)\xi=\gamma\xi \tag{8}$$

其中,ω和ξ是特征矩阵,λ和γ是特征值.与公式(7)和公式(8)中前 d_i 个最大的特征值对应的 d_i 个特征向量就分别是投影矩阵 M_i 和 W_i.其他的 M 族和 W 族矩阵采用同样的方式获取.所有的 M 族和 W 族矩阵可以依据公式(5)和公式(6)迭代更新.根据实验结果,两次迭代即可达实验要求.

获取投影矩阵后,即可构建类间和类内流形.图 4 是两种流形的可视化结果,数据集取自 FERET 库[31].图 4(a)显示的是尺寸为 80×80 的平均脸图像,每张图像被分成 25 块互不重叠的尺寸为 16×16 的子片.每个平均脸对应某个画廊集中的 6 幅原始图像,如图 4(b)中所示 subject 1,所有原始图像也都与平均脸相同的方式分片.从图 4(c)

和图 4(d)我们可以看出,人与人之间的类间和类内流形都被清楚地分开.



(a) 5 个人的平均脸 (b) 其中一个人的原始图像

(c) 两次迭代后的类间流形 (d) 两次迭代后的类内流形

Fig.4 Manifolds of five subjects

图 4 5 个人的低维流形

2.2.4 算法分析

最大化 Inter_MFD,Inter_OID 和最小化 Intra_MFD,Intra_OID 的目的是为了使流形之间的边距(margin)尽可能地大,这样,人与人之间在低维空间中才能被很好地分开.在优化过程中,我们将每个投影矩阵分开优化,这意味着不可能获得全局最优解,当画廊集过多时,识别效果可能会变差.如果我们在优化过程中像 DMMA 那样将每个子片的 k 近邻加入到运算中,那么新方法将饱尝非全局最优解的短处.因为不管选取多少个近邻,总会有一些图像集不会参与到运算中,那么有些流形之间的边界就不会很清晰.为了克服非全局最优的缺点,新方法中所有图像集的所有同位置子片都被包含在每个投影矩阵的优化计算中,确保了当人数众多时,识别效果也不会变差.同位置子片的设计,使得新方法不必在所有画廊集范围内进行大规模搜索,因此每次迭代的算法复杂度是 $O(n)$,而 DMMA 则是 $O(n^2)$,n 是参与训练的人数.

2.3 识别阶段

第 1 步:利用类间流形来选取与疑问集最为相近的几个候选图像集.将计算所有画廊集的平均脸,所以宜采用相对简单、快速而又保持较高的识别正确率的度量算法.同时,每个子片在显示平均脸的特征时都很重要,为避免信息丢失,必须所有的子片都参与计算,因此,现有的 MMD[14,15,32]方法不适用于本文,因为它们都是选取一部分采样来计算,本文将采用余弦距离(cosine distance)[33]来度量两个流形的相似度,进而选出候选图像集.

将疑问集的平均脸按照训练阶段的方法分片:

- 令 $M_{test}=[m_1,m_2,...,m_t]$表示所得的子片集;
- 令 $M_{gset}=[M_i^T m_{i1},M_i^T m_{i2},...,M_i^T m_{it}]$表示第 i 个画廊集的平均脸在类间流形上的表示.

先将 M_{test} 投影第 i 个流形上,记为 $Te\_m=[M_i^T m_1,M_i^T m_2,...,M_i^T m_t]$;然后,分别把 M_{gset} 和 Te_m 转换为向量,记为

$$M_{gset}(:)\in R^{t\times d_i},Te\_m(:)\in R^{t\times d_i},$$

则疑问集和第 i 个画廊集在低维空间上的距离为 $\cos(M_{gset}(:),Te\_m(:))$.找出最近的几个,设为 K 个.

第 2 步:利用类内流形,从上一步的 K 个候选图像集中找出最近的一个.先将疑问集的所有原始图像按训练阶段分片,再将这些分片投影到选出的 K 个类内流形上,分别计算它们在低维流形上的距离,即可找到最近的一个.本阶段主要考虑识别精度,不必再把运算速度考虑在内.DMMA 提出了一种基于重建的度量算法,每个疑问

子片投影到某个画廊集对应的低维空间后,在该图像集的所有子片中搜索它在低维空间中的若干个近邻,再由这些近邻重建它,所有子片的重建误差即为这两个图像集之间的距离.我们也采用重建的方法,但不会寻找近邻,原因有二:一是几近邻合适?二是如何确定哪种找近邻的方法在低维流形中是好的,是马氏距离还是欧氏距离?在高维空间中,若在所有子片范围内用这两种距离度量为某个子片寻找近邻,则很可能会使脸颊与下巴相近,我们不能确定在低维空间中不发生这样的情况.本文的方法是:每个疑问子片在每个候选图像集中都有该图像集所包含的所有原始图像上的同位置子片,若脸的角度相同,它们就是同一器官,投影到某个候选类内流形上之后,该疑问子片就可以由这些同位置子片在低维空间中重建,所有疑问子片的重建误差即为两图像集之间的距离.为防止因脸的角度不同而影响距离计算,可先在高维空间中算出疑问子片与其所有同位置子片之间的欧式距离,去掉其中若干个距离最大的,由剩下的同位置子片在低维空间中重建它.根据实验结果我们发现:当角度变化很大时,去掉总数的 1/3 个效果最好.重建细节具体如下:

令 Y 表示疑问集,$Y_l=[y_{l1},y_{l2},\ldots,y_{lt}]$表示其中的第 l 幅图像,共 k_{test} 幅,y_{lr} 是第 l 幅图像的第 r 个子片.设要与 Y 度量距离的是第 i 个画廊集,令 $k'=2/3k_i$(k_i 与第 2.2.2 节相同,表示第 i 个画廊集所包含的原始图像个数),设 y_{sr} 在该图像集上的欧式距离较小的 k' 个同位置子片为 $x_{i1r},x_{i2r},\ldots,x_{ik'r}$,投影到第 i 个类内流形上以后,它们分别被表示为 $yT_{lr}=W_i^T y_{lr}$ 和 $xT_{lr}=[W_i^T x_{i1r},W_i^T x_{i2r},\ldots,W_i^T x_{ik'r}]$.

令 $d(yT_{lr},xT_{lr})=\min\left\|yT_{lr}-\sum_{s=1}^{k'}c_s W_i^T x_{isr}\right\|^2$,其中,$c_s$ 是重建系数且 $\sum_{s=1}^{k'}c_s=1$,解法略.

求出重建系数后,为与类间流形中所计算的距离相统一,我们用各同位置子片的线性组合与疑问子片之间的余弦距离作为该子片的重建精度,流形的距离最终可以表示为所有子片的重建精度均值.

3 实 验

既然新方法可以通过处理平均脸把基于图像集的视频人脸识别转换成单人脸识别,而且分片思想来自于 DMMA,那么我们有理由首先把 DJMMA 的平均脸阶段与 DMMA 相比较,然后再就基于图像集的人脸识别把新方法与现有方法相比较,所以整个实验包括两部分:一是单人脸识别,二是基于图像集识别.

3.1 单人脸识别

此处的单人脸指的是每个画廊集的平均脸.跟大多数单人脸识别方法做实验时一样,我们也用 FERET[34], FG-NET[31]和 AR[35]这 3 个数据集.

FERET 数据集一共包含 1 565 个不同种族、性别和年龄的人的共 13 539 幅图像.我们使用其中的一个子集,包含 200 个人的 1 400 幅图像,每人 7 幅,包括正面、小角度侧面和不同光照几种情况.选取其中 4 幅的平均脸用来训练,剩下 3 幅的平均脸用来测试.所有图像被重定义为 80×80 的灰度图,具体如图 5(a)所示.

(a) FERET 集,左边用于训练,右边用于测试

(b) FG-NET 集,左边用于训练,右边用于测试

(c) AR 集,上面一行用于训练,下面一行用于测试

Fig.5 Sample face images from three databases

图 5 3 个数据集中的样例,箭头所指的是平均脸

FG-NET 数据集一共包含 82 个人的 1 002 幅图像,每幅图在光照、姿态、表情上都有较大的变化,年龄从 0

到 69 岁,平均每人 12 幅左右.我们为每个人建 2 个子集:一个的平均脸用来训练,另一个的平均脸用来测试.每个子集包含 5 幅图像,覆盖不同的年龄段,若总图像不足 10 幅则先满足测试集.所有图像被重定义为 60×60 的灰度图,具体如图 5(b)所示.

AR 库共包含 126 个人的 4 000 幅彩色图像,每人 26 幅,共分成 2 个组,每组每人 13 幅.我们把其中一个组的每人 13 幅的平均脸用来训练,另一组的用来测试.所有图像被重定义为 60×60 的灰度图,具体如图 5(c)所示.

FERET 的每幅平均脸都被分成 25 个 16×16 的子片,FG-NET 和 AR 库的每幅平均脸被分成 12×12 的子片.DMMA 的参数取值与原文献一致,DJMMA 的 d_i,即第 2.2.1 节定义的低维流形的维数,在 AR 和 FG-NET 库上第 1 次迭代是 20,第 2 次是 6,在 FERET 库上两次迭代分别是 30 和 10.两种方法的识别正确率见表 2.

Table 2 Recognition accuracy (%) of two different methods

表 2 两种方法识别正确率(%)对比

Methods	AR	FERET	FG-NET
DJMMA	86.5	98	43.3
DMMA	77.2	86.7	35

从表 2 中可以看出,两种方法的共同点是 FERET 库识别率最高,AR 次之,FG-NET 最低.这是因为 FERET 库中姿态和光照变化较少,且无遮挡无年龄变化;而 AR 库有遮挡,FG-NET 年龄变化大.同时也可以看出,在每个库中,DJMMA 比 DMMA 的正确率至少要高 8%,并且训练集成员越多,正确率的差距就越大.本实验的测试阶段,DJMMA 用的是余弦距离,精度会低于 DMMA 的基于重建的方法,所以测试阶段并没有为新方法的高识别率做出贡献.但在训练阶段,两方法的思想又相同,唯一的不同之处就在于参与运算的子片数:DMMA 用的是所有图像全体子片中的若干近邻,而 DJMMA 用的是同位置子片,因此可以看出,是最近邻的找法与数目降低了 DMMA 的识别率.

3.2 基于图像集的人脸识别

我们将在 3 个广为研究的公开库上进行实验对比,它们分别是 Honda/UCSD[36],CMU MoBo[37]和 YouTube Celebrities(YTC)[28].在对 3 个库进行简短的介绍后,先用数据详细描述 DJMMA 的工作细节,再与其他方法进行对比.

3.2.1 数据集

Cui 等人已经在文献[24]中详细介绍了这 3 个视频集.我们采用 Nilsson 等人[38]的方法(包括源代码)检测视频中每个帧的人脸,然后将 Honda 库中的人脸图像变成 80×80,将 MoBo 库中的变成 40×40,将 YTC 库中的变为 50×50 的灰度图,除此以外,未对图像做任何预处理.不做预处理和保留较高分辨率的目的是为了尽量多地保留图像的原始特征,它将为 DJMMA 和其他方法的对比带来挑战.

对于所有数据集,我们都采用与 Cui 等人[24]相同的方式进行训练和测试配置:在 Honda 和 MoBo 库中,选取每个人的一个序列用做训练,剩下的用做测试;在 YTC 库中,每个人有 3 个组平均 41 个视频序列,从每个人的每个组里各选 1 个用做训练,各选 6 个用做测试.为了提高识别的可信度,所有的实验都被重复 5 次,每次都有机选训练集和测试集,平均起来即为最终识别率.

为了更好地描述所有方法的工作效果,每个图像集中随机选来用做训练的图像数目会有所变化:在第 3.2.2 节中,每个图像集分别包含 30,50,100,200 和 300 幅图像;在第 3.2.3 节中,每个图像集分别包含 30,100 和 300 幅图像;疑问集中的图像数目在第 3.2.2 节中分别是 2 和 20,在第 3.2.3 节则与画廊集中图像的数目相同.如果一个图像集中所含的图像数未达到指定值,则默认为该集中的所有图像.图像集中小数目的图像参与训练或测试是有很大的实用价值的,因为生活中所获取的视频段往往总是不够长.

3.2.2 DJMMA 的工作细节

本节中,我们将分别分析 DJMMA 的类间和类内的识别精度,尤其侧重于不充足采样实验,即,待识别视频段不够长的情况.Honda 库中的每个图像被分成 25 个 16×16 的子片,MoBo 和 YTC 分别被分成 16 和 25 个 10×10 的子片.类间流形阶段找出的候选图像集数目是 5,d_i 在 Honda 库中两次迭代的取值分别为 30 和 10;在 MoBo 和

YTC 库中两次迭代分别取 20 和 6.图 6 和表 3 从细节上显示了当画廊集中图像数为 30、疑问集中图像数为 2 时,DJMMA 在 Honda 库上的工作过程.

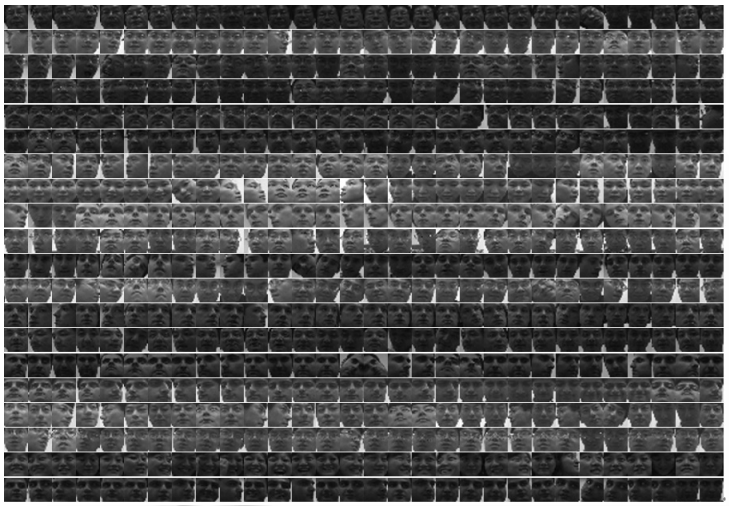


Fig.6 Example of 30 trainings for Honda. There are 20 sets in the database and 30 images in each set

图 6 Honda 库中的人脸示例,每个 Set 30 幅图像,共 20 个 Set

Table 3 Distance between query set and its relevant sets

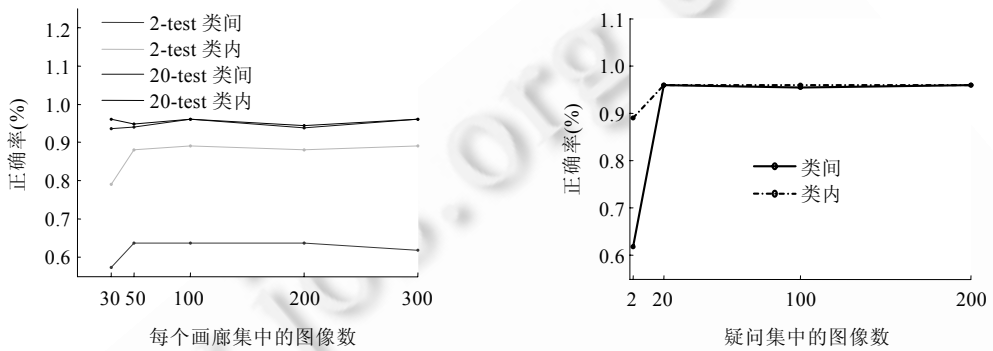
表 3 疑问集和候选集之间的距离

疑问集	疑问集与对应的候选集之间的距离(类间流形)						疑问集与对应的候选集之间的距离(类内流形)					
	Set	Distance	Set	Distance	Set	Distance	Set	Distance	Set	Distance	Set	Distance
Set 1	6	0.991 962	1	0.988 692	18	0.982 287	1	0.997 124	18	0.997 003	15	0.996 823
Set 2	2	0.995 621	13	0.975 913	8	0.971 438	2	0.997 314	8	0.996 034	14	0.992 565
Set 3	3	0.998 042	4	0.993 875	10	0.992 807	3	0.999 264	4	0.999 089	9	0.998 439
Set 4	4	0.998 910	3	0.993 864	10	0.991 698	4	0.999 420	3	0.999 038	5	0.998 812
Set 5	5	0.995 983	4	0.992 107	3	0.991 023	5	0.999 102	4	0.998 317	11	0.997 393
Set 6	6	0.993 695	14	0.985 839	15	0.985 023	6	0.993 803	9	0.992 035	14	0.990 241
Set 7	7	0.997 003	15	0.989 731	10	0.988 959	7	0.998 443	10	0.998 295	15	0.998 258
Set 8	8	0.997 468	14	0.989 472	6	0.988 556	8	0.999 083	6	0.997 553	14	0.994 524
Set 9	9	0.995 500	3	0.994 577	11	0.993 276	9	0.999 326	4	0.999 004	14	0.998 918
Set 10	10	0.997 527	3	0.990 754	4	0.988 838	10	0.997 632	4	0.995 066	9	0.993 187
Set 11	11	0.998 087	3	0.990 899	9	0.990 199	11	0.999 625	9	0.998 856	4	0.998 275
Set 12	12	0.997 340	8	0.995 889	14	0.991 871	12	0.999 567	9	0.998 626	14	0.998 433
Set 13	13	0.999 070	11	0.985 877	14	0.985 176	13	0.999 662	14	0.999 043	9	0.997 949
Set 14	14	0.997 998	9	0.995 096	11	0.992 046	14	0.998 739	9	0.998 537	11	0.998 096
Set 15	15	0.996 511	3	0.987 018	4	0.986 437	15	0.996 635	6	0.996 632	7	0.995 261
Set 16	16	0.996 605	3	0.980 371	15	0.978 555	15	0.999 226	16	0.995 606	4	0.990 167
Set 17	17	0.997 389	18	0.987 591	9	0.984 488	17	0.998 544	18	0.997 481	1	0.997 290
Set 18	18	0.998 476	19	0.991 752	17	0.991 533	18	0.997 983	17	0.997 524	9	0.997 036
Set 19	19	0.998 805	18	0.992 644	9	0.988 259	19	0.999 586	4	0.999 333	9	0.998 926
Set 20	20	0.992 183	7	0.972 886	15	0.962 963	20	0.999 043	15	0.996 961	7	0.995 498

类间流形阶段的任务是寻找 5 个与疑问集最相似的候选集(表 3 中只列出了 3 个),然后由类内流形阶段从这 5 个中选出距离最近的一个.如表 3 所示,余弦距离越大,两图像集的相似度越高;当 Set 1 是疑问集时,对应的画廊集 Set 1 却不是 5 个候选集中排在最前面的一个,但最终由类内流形将其正确选出.

图 7 显示了当每个图像集中参与训练的图像数不同时,不充足采样和充足采样的识别精度,最终精度由类内流形阶段决定.在图 7 中,YTC 库中的识别精度随着画廊集和疑问集中图像数目的不同而变化.类间流形阶段的精度表示正确的图像集被选入到候选集中的正确率,类内流形阶段的精度表示把正确的图像集从候选集中选出的正确率.图 7(a)中的 2-test 和 20-test 表示的是疑问集中的图像数目,2-test 是严重的不充足采样,而我们在现实生活中有可能常常会遇到,此时,类间流形的精度不高,最高时仅达到 65%,但类内流形的却可以高达 89%,

虽然仍低于 20-test 的,但后面的实验将显示其仍然比其他方法高出很多.图 7(a)同时还显示出:当画廊集中训练图像数达到 50 幅以后,2-test 和 20-test 都将趋于稳定,保持着较高的识别率.也就是说,利用 DJMMA 可以仅计算疑问集中的部分图像,极大地节省了时间.在图 7(b)表示的是当画廊集中的训练图像数目不变时,精度随着疑问集中测试图像的数目的变化情况.可以看出,当测试图像数目超过 20 时,识别率将趋于稳定并保持较高的正确率.也就是说,如果画廊集中的训练图像数目足够大(YTC 库是每图像集 50 幅),则不必让画廊集中的所有图像都参与识别,节省了计算时间.图 7 还显示出:当每图像集中参与测试和训练的图像足够多时,类间流形的精度是令人满意的,此时,类内流形就可以省略了.MoBo 库的情况将在下一节介绍.



(a) 当疑问集分别包含 2 张和 20 张图像时,在每 Set 训练图像分别为 30,50,100,200 和 300 时的表现

(b) 当每图像集训练图像是 300 时,在疑问集图像数分别为 2,20,100 和 200 时的表现

Fig.7 Accuracy of Inter-class and Intra-class manifolds on YTC

图 7 YTC 库上类间和类内的识别精度

3.2.3 DJMMA 与其他方法的对比

将新方法与常用方法在经典数据集上对比是衡量新方法好坏的常用手段,本文将把 DJMMA 与近年来常用的 5 个基于图像集的方法在上述 3 个视频库上进行对比.这 5 种方法分别是 Mutual Subspace Method (MSM)[17],SANP[19],MMD[15],Manifold Discriminate Analysis(MDA)[32]和 JSR[24].除 JSR 外,其他 4 个方法的源代码均从原作者的网站上获取.JSR 的程序代码无法获得,我们遵循原文思想进行了认真编写.所有方法的重要参数均取自其原文,DJMMA 的参数与第 3.2.2 节相同.表 4~表 6 显示了所有方法在 3 个数据集上的识别结果,除了 2-test 实验以外,所有疑问集中测试图像的数目均与其画廊集中的训练图像数目相同,它们分别是 30,100 和 300. 2-test 实验中的识别精度是每个画廊集中训练图像数目分别为 30,100 和 300 时的识别精度的均值.MMD 和 MDA 的 k-NN 数是 1,JSR 的结果是取不同的λ_1和λ_2时所有结果的均值,MMD 的结果是 Exemplar,Variation 和 Exemplar-ED 这 3 种情况的均值.

从表 4~表 6 可以看出:MSM 和 JSR 在所有实验中表现都很平稳,MMD 在 Honda 库中对图像集中的图像数目敏感,MDA 和 SANP 因为光照变化的影响而在 Honda 库中表现得不是很令人满意.原因是做实验前我们未对图像做任何预处理,例如直方图均衡化.这 5 种用来对比的方法有两个共同点:一是在所有实验中,2-test 的识别率都不高,在 3 个库中的平均识别率分别为 60.7%,59%,60.4%,58.6%,68.7%;二是 MoBo 和 YTC 库中的识别率都低于 Honda 库,特别是 YTC 库,识别率均未超过 65%,这是因为 YTC 中的视频完全来自不受控环境,本身的低分辨率和大尺度的表情、光照、姿态变化都严重影响了识别率.DJMMA 却在这两个方面表现良好,2-test 的识别率始终保持在 85%以上,说明适用于短视频识别;新方法在 3 个库上的识别率相差不大,尤其在 YTC 库上,识别率远高于现有方法.据我们所知,目前已有方法在 YTC 上的识别率没有一个高于 77%,而 DJMMA 却高达 95%,这说明分片流形的思想能够更好地抓住人脸的本质特性,而这些特性不受分辨率、光照和姿态变化的影响,结合上一节在 FG-NET 库中的实验还可以看出:DJMMA 所获取的特征还能减轻年龄对面部特征的影响,具有非常大的实用价值.另外,当每个画廊集中用来训练的图像数目超过 50 的时候,类间流形阶段就已达到很高的识

别率,可以省略类内阶段.

Table 4 Recognition accuracy (%) on Honda

表 4 Honda 库上的识别率(%)

Methods		MSM	MMD	MDA	SANP	JSR	DJMMA
Number of training each set	30	85.5	58.4	69.5	68.8	96.4	99.0
	100	85.1	88.3	73.2	71.7	97.2	100
	300	84.8	95.1	74.8	72.3	98.8	98.0
2-test		67.3	62.7	60.4	59.3	75.6	85.8

Table 5 Recognition accuracy (%) on MoBo

表 5 MoBo 库上的识别率(%)

Methods	MSM	MMD	MDA	SANP	JSR	DJMMA
Average performance	81.9	83.2	86.9	84.3	92.5	98.3
2-test	60.7	59.2	63.1	60.3	70.6	87.6

Table 6 Recognition accuracy (%) on YTC

表 6 YTC 库上的识别率(%)

Methods	MSM	MMD	MDA	SANP	JSR	DJMMA
Average performance	60.1	61.3	62.5	61.4	64.3	95.8
2-test	54.2	55.3	57.8	56.1	59.8	87.4

4 结 论

平均脸的概念许多年前就已被提出并使用[39],但从未被当作关键技术应用在人脸识别中,在多数算法中只是起辅助作用.本文中首次将其与流形结合用于图像集分类,并在很多实验中取得了良好的效果.分片思想和联合多流形技术充分考虑了类内和类间距离,通过这种方式获取的特征极具判别性,进而有效地减少了误判的可能.在计算投影矩阵的过程中,所有的同位置子片都参与进去,尽可能地弥补了因分开计算而导致的不是全局最优解的不足.就识别效果而言,DJMMA 更适合于现实生活应用,对光照、姿态、表情和年龄变化具有极强的鲁棒性,并且尤其适用于短视频情况,既包括用于训练的短视频也包括用于测试的短视频.

新方法仍有一些问题没有解决好,例如识别结果过于依赖类间流形阶段,如果指定的候选集的个数过小,则有可能漏掉正确图像集.这样,无论类内阶段有多么精确都无法得到正确的识别结果.但指定几个才是合适的数目,还需要进一步研究.而且,类内流形阶段也并非总能完美地工作,我们不能杜绝这种现象的发生:正确的图像集在类间阶段被选到候选集里面来了,但却在类内阶段被排除了.因此,如何进一步提高类内流形阶段的识别精度,是下一步的工作.

致谢 感谢南京大学计算机软件新技术国家重点实验室(南京大学)高阳教授课题组的所有老师和同学对本工作的帮助.

References:

[1] Barr JR, Bowyer KW, Flynn PJ, Biswas S. Face recognition from video: A review. Int'l Journal of Pattern Recognition and Artificial Intelligence, 2012,26(5):1−56. [doi: 10.1142/S0218001412660024]

[2] Liu X, Chen T. Video-Based face recognition using adaptive hidden Markov models. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2003. 340−345. [doi: 10.1109/CVPR.2003.1211373]

[3] Kim M, Kumar S, Pavlovic V, Rowley H. Face tracking and recognition with visual constraints in real-world videos. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Anchorage: IEEE Computer Society, 2008. 1−8. [doi: 10.1109/CVPR.2008.4587572]

[4] Lee KC, Ho J, Yang MH, Kriegman D. Visual tracking and recognition using probabilistic appearance manifolds. Computer Vision and Image Understanding, 2005,99(3):303−331. [doi: 10.1016/j.cviu.2005.02.002]

[5] Yan Y, Zhang YJ. State-of-the-Art on video-based face recognition. Chinese Journal of Computers, 2009,32(5):878−886 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2009.00878]

[6] Arandjelović O, Shakhnarovich G, Fisher JW, Cipolla R, Darrell T. Face recognition with image sets using manifold density divergence. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society, 2005. 581−588. [doi: 10.1109/CVPR.2005.151]

[7] Shakhnarovich G, Fisher JW, Darrell T. Face recognition from long-term observations. In: Proc. of the 7th European Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2002. 851−865. [doi: 10.1007/3-540-47977-5_56]

[8] Cevikalp H, Triggs B. Face recognition based on image sets. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. San Francisco: IEEE Computer Society, 2010. 2567−2573. [doi:10.1109/CVPR.2010.5539965]

[9] Huang LK, Lu JW, Tan YP, Feng X. Collaborative reconstruction-based manifold-manifold distance for face recognition with image sets. In: Proc. of the 2013 IEEE Int'l Conf. on Multimedia and Expo (ICME). San Jose: IEEE Computer Society, 2013. 1−6. [doi: 10.1109/ICME.2013.6607596]

[10] Aggarwal G, Chowdhury AKR, Chellappa R. A system identification approach for video-based face recognition. In: Proc. of the 17th Int'l Conf. on Pattern Recognition. IEEE Computer Society, 2004. 175−178. [doi: 10.1109/ICPR.2004.1333732]

[11] Fukui K, Yamaguchi O. The kernel orthogonal mutual subspace method and its application to 3D object recognition. In: Proc. of the 8th Asian Conf. on Computer Vision. Berlin, Heidelberg: Springer-Verlag, 2007. 467−476. [doi: 10.1007/978-3-540-76390-1_46]

[12] 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]

[13] Sanderson MTCC, Shirazi S, Lovell BC. Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2011. 2705−2712. [doi: 10.1109/CVPR.2011.5995564]

[14] Wang R, Shan S, Chen X, Chen J, Gao W. Maximal linear embedding for dimensionality reduction. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011,33(9):1776−1792. [doi: 10.1109/TPAMI.2011.39]

[15] Wang R, Shan S, Chen X, Dai Q, Gao W. Manifold-Manifold distance and its application to face recognition with image sets. IEEE Trans. on Image Processing, 2012,20(10):4466−4479. [doi: 10.1109/TIP.2012.2206039]

[16] Cui Z, Shan S, Zhang H, Lao S, Chen X. Image sets alignment for video-based face recognition. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2012. 2626−2633. [doi: 10.1109/CVPR.2012.6247982]

[17] Yamaguchi O, Fukui K, Maeda K. Face recognition using temporal image sequence. In: Proc. of the IEEE Int'l Conf. on Automatic Face and Gesture Recognition (AFGR). Nara: IEEE Computer Society, 1998. 318−323. [doi: 10.1109/AFGR.1998.670968]

[18] Hu Y, Mian AS, Owens R. Sparse approximated nearest points for image set classification. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2011. 121−128. [doi: 10.1109/CVPR.2011.5995500]

[19] Kim TK, Arandjelović O, Cipolla R. Boosted manifold principal angles for image set-based recognition. Pattern Recognition, 2007, 40(9):2475−2484. [doi: 10.1016/j.patcog.2006.12.030]

[20] Fan W, Yeung DY. Locally linear models on face appearance manifolds with application to dual-subspace based classification. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2006. 1384−1390. [doi: 10.1109/CVPR.2006.178]

[21] Elhamifar E, Vidal R. Robust classification using structured sparse representation. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2011. 1873−1879. [doi: 10.1109/CVPR.2011.5995664]

[22] Chen YC, Patel VM, Shekhar S, Chellappa R, Phillips PJ. Video-Based face recognition via joint sparse representation. In: Proc. of the 10th IEEE Int'l Conf. and Workshops on Automatic Face and Gesture Recognition. Shanghai: IEEE Computer Society, 2013. 1−8. [doi: 10.1109/FG.2013.6553787]

[23] Chen SK, Sanderson C, Harandi MT, Lovell BC. Improved image set classification via joint sparse approximated nearest subspaces. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Portland: IEEE Computer Society, 2013. 452−459. [doi: 10.1109/CVPR.2013.65]

[24] Cui Z, Chang H, Shan S, Ma B, Chen X. Joint sparse representation for video-based face recognition. Neurocomputing, 2014,135: 306−312. [doi: 10.1016/j.neucom.2013.12.004]

[25] Bhatt HS, Singh R, Vatsa M. On recognizing faces in videos using clustering-based re-ranking and fusion. IEEE Trans. on Information Forensics and Securit, 2014,9(7):1056−1068. [doi: 10.1109/TIFS.2014.2318433]

[26] Patel VM, Chen YC, Chellappa R, Phillips PJ. Dictionaries for image and video-based face recognition. Journal of the Optical Society of America A, 2014,31(5):1090−1103. [doi: 10.1364/JOSAA.31.001090]

[27] Hadid A, Pietikäinen M. From still image to video-based face recognition: An experimental analysis. In: Proc. of the 6th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. EEE Computer Society, 2004. 813−818. [doi: 10.1109/AFGR.2004.1301634]

[28] Kim M, Kumar S, Pavlovic V, Rowley H. Face tracking and recognition with visual constraints in real-world videos. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Anchorage: IEEE Computer Society, 2008. 1−8. [doi: 10.1109/CVPR.2008.4587572]

[29] Lu JW, Tan YP, Wang G. Discriminative multimanifold analysis for face recognition from a single training sample per person. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2013,35(1):39−51. [doi: 10.1109/TPAMI.2012.70]

[30] Yao B, Ai H, Lao S. Person-Specific face recognition in unconstrained environments: A combination of offline and online learning. In: Proc. of the 11th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. EEEE Computer Society, 2009. 1−8. [doi: 10.1109/AFGR.2008.4813353]

[31] Lanitis A. Evaluating the performance of face-aging algorithms. In: Proc. of the 10th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. Amsterdam: IEEE Computer Society, 2008. 1−6. [doi: 10.1109/AFGR.2008.4813349]

[32] Wang R, Chen X. Manifold discriminant analysis. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Miami: IEEE Computer Society, 2009. 429−436. [doi: 10.1109/CVPR.2009.5206850]

[33] Tan PN, Steinbach M, Kumar V. Introduction to Data Mining. Addison-Wesley, 2005. 500−500.

[34] Phillips PJ, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(10):1090−1104. [doi: 10.1109/34.879790]

[35] Martinez AM, Benavente R. The AR face database. Technical Report, #24, Barcelona: Computer Vision Center (CVC), Universitat Autònoma de Barcelona, 1998.

[36] Lee K, Ho J, Yang M, Kriegman D. Video-Based face recognition using probabilistic appearance manifolds. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition. Washington: IEEE Computer Society, 2003. 1−8. [doi: 10.1109/CVPR.2003.1211369]

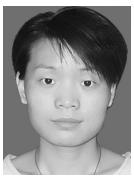
[37] Gross R, Shi J. The CMU motion of body (MoBo) database. Technical Report, CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, 2001.

[38] Nilsson M, Nordberg J, Claesson I. Face detection using local SMQT features and split up SNOW classifier. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP). Honolulu: IEEE Computer Society, 2007. 589−592. [doi: 10.1109/ICASSP.2007.366304]

[39] Sirovich L, Kirby M. Low-Dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A, 1987,4(5):519−525. [doi: 10.1364/JOSAA.4.000519]

附中文参考文献:

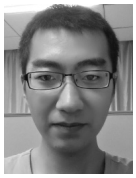
[5] 严严,章毓晋.基于视频的人脸识别研究进展.计算机学报,2009,32(5):878−886. [doi: 10.3724/SP.J.1016.2009.00878]



于谦(1979－),女,山东济南人,讲师,CCF会员,主要研究领域为机器学习,计算机图形学.

霍静(1989－),女,博士生,主要研究领域为计算机视觉,机器学习.

高阳(1972－),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为人工智能,机器学习.



庄韫恺(1990－),男,硕士生,主要研究领域为计算机视觉,模式识别.