

基于稀疏特征挑选和概率线性判别分析的表情识别研究

张 瑞, 蒋晨之, 苏剑波

(上海交通大学自动化系系统控制与信息处理教育部重点实验室, 上海 200240)

摘 要: 提出一种基于稀疏特征挑选(Sparse selection)和概率线性判别分析(Probabilistic linear discriminant analysis)的表情识别方法 SS-PLDA. 该方法由两部分构成: 第一部分是使用稀疏的方法挑选出人脸与表情相关的区域, 构造表情的完备特征集; 第二部分是针对构造的表情完备特征集里仍含有一些其他信息, 运用概率线性判别分析实现表情特征与干扰信息的分离, 学习出一个只含有表情信息的子空间, 最后基于该表情子空间进行表情识别分析. 通过在 CK+ 和 JAFFE 这两个数据库上面的实验, 证实了基于稀疏特征挑选的方法可以得到识别性能的改善, 且先使用特征挑选再对所挑选结果应用概率线性判别分析可以达到更好的提升效果.

关键词: 人脸表情识别; 稀疏; 特征挑选; 子空间学习

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112 (2018)07-1710-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2018.07.024

Expression Recognition Based on Sparse Selection and PLDA

ZHANG Rui, JIANG Chen-zhi, SU Jian-bo

(Ministry of Education Key Laboratory of System Control and Information Processing,
Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: A facial expression recognition method named SS-PLDA is proposed based on Sparse Feature Selection and Probabilistic Linear Discriminant Analysis. The SS-PLDA method contains two steps: 1) pick out the most discriminative regions for facial expressions and use these regions to construct a complete facial features set; 2) apply Probabilistic Linear Discriminant Analysis Method to separate the useful expression signals from other disturbance information. Therefore, a subspace which only contains expression information is learnt and the expression recognition task is implemented in this subspace. Experimental results on Cohn-Kanade(CK+) database and JAFFE database show that the complete facial features set can improve the performance, and the proposed method outperform the state-of-art ones.

Key words: facial expression recognition; sparse; feature selection; subspace learning

1 引言

人脸表情作为人的意图、认知活动、情感状态等重要的外在表现, 在人机交互(HCI)、安全驾驶、健康监护和测谎等诸多方面有着广泛的应用^[1,2]. 近年来, 表情识别也成为研究的热点吸引了大批的学者. 表情识别一般分为静态图片表情识别和动态序列表情识别^[3,4], 其中静态图片的表情识别是动态序列表情识别的基础, 通常由三部分构成: 图像预处理、特征提取以及预测分类, 人脸表情识别的结果很大程度上依赖于所提特

征的表达性. Ekman 发现同一面部表情所引起的面部肌肉的运动具有一定的规律性, 这种规律不受年龄、性别等其他因素的影响, 并定义了六种基本表情类别: 愤怒、厌恶、恐惧、高兴、悲伤和惊讶. 在此基础上, Ekman 以面部活动单元为单位提出了面部活动单元编码系统(Facial Action Coding System, FACS)^[5], 这套编码系统在表情识别领域被广泛应用. 面部活动单元的检测效果依赖于面部特征点的定位的准确性, 而 Whitehill 的研究结果^[6]表明面部特征点的精确定位甚至比表情识别自身更具有挑战性. 因此, 大部分已有的图像表情特征提

取算法都是基于几何形状或者外貌纹理特征. 基于几何模型的研究^[7-9]使用脸部的特征点集这一形状模型来分类表情,然而这种方法通常也需要脸部特征点的精确检测,一般很难实现,而且对于不同的人其特征点间的相对距离并不相同,这使得模型的泛化能力较差. 基于纹理模型的研究^[10-15]使用 HOG、Gabor 及 LBP 等纹理特征来表征全局或局部的表情特征,纹理特征一般具有高维度的特点,相对几何特征而言具有更好的鲁棒性. 然而这些优点所带来的代价是计算的复杂度升高,因此这种方法常常与一些降维方法 (PCA 或者 LDA) 一起使用^[16,17], 最终的分类过程将在学习到的低维子空间中进行. 虽然基于纹理信息的方法具有更高的时间和空间复杂度,但因其提取的表情特征信息更具有表达性而成为更流行的方法^[18].

很多研究表明^[15,18-20],选择部分人脸区域来进行表情识别任务比使用所有人脸区域更有效,这也和面部活动单元编码系统 (FACS) 认为表情由面部的各种运动单元构成是一致的. 文献^[15]使用基于人脸局部区域的词袋模型,文献^[16]使用了一种基于局部 LBP 特征级联的方法给每个小的人脸区域赋相应的权重值. 文献^[18]按照先验的知识选取与表情对应的特定人脸区域.

传统的子空间学习算法如 PCA 以及 LDA 均是假设数据的噪声是符合高斯分布^[21],真实的数据噪声大多数情况不满足标准高斯分布,这样学习到的子空间仍然是含有很多噪声的空间. 基于因子分析的 LDA 可以同时信号子空间和噪声子空间进行学习,文献^[22]使用概率 LDA 的方法进行含有姿态、光照等较多噪声影响的人脸识别,学习到了身份子空间和含有姿态、光照等干扰的噪声子空间,在身份子空间进行人脸识别显著提高了识别率. 文献^[23]使用相关的多因子分析的方法进行抗年龄干扰的人脸识别,最终学习到一个身份子空间和一个年龄子空间,通过学习到的子空间既可以做身份识别也可以做年龄识别.

本文提出稀疏特征挑选下的 PLDA 人脸表情分析 (SS-PLDA),该方法由两部分组成:稀疏特征挑选和 PLDA 表情分析. PLDA 是一种生成模型,每一张人脸图片可以认为由表情部分和其他信息部分 (身份、年龄等) 加随机噪声生成的,当进行表情识别时,其他信息和噪声可以表示为非表情部分,仅使用表情相关部分来进行表情识别任务有更好的泛化性能. 基于表情样本的数量不是很大以及简化求解过程的原因,本文使用线性空间来描述表情空间和噪声空间. 原始的人脸图片含有丰富的信息,如身份、表情、年龄等,仅使用简单的线性空间可能无法很好地将表情空间完全学习出来. 因此,在进行 PLDA 人脸表情分析之前先使用稀疏

特征挑选,选出表情的完备特征集. 在该表情完备特征集中所包含的大量信息是与表情相关的,因此后面使用简单的线性子空间学习方法就能把表情空间很好地学习出来. 稀疏特征挑选是对表情特征的粗挑选,PLDA 方法是对表情特征的精挑选. 经过这两步,一个单纯含有表情的子空间被学习出来,在识别阶段,本文使用基于贝叶斯的方法计算两幅人脸图中的表情在表情空间中的似然值,将该似然值作为距离度量,应用 kNN 分类器输出识别结果.

2 相关内容

2.1 稀疏特征区域挑选

本文中稀疏特征区域挑选的方法是基于 Sparse Group Lasso^[24],该方法是 Lasso 和 Group Lasso 的一种扩展,优化过程中同时实现了组内和组间的稀疏性. 具体而言就是未被选中的特征组对应的系数均为零而被选中的特征组部分系数为零,这样就同时进行了特征组的选取和组内特征的选取. 下面分别介绍 Lasso 和 Group Lasso 的数学表达形式.

2.1.1 L_1 范数正则化

L_1 范数正则化也称为 Lasso,由于具有很强的理论支持,现有绝大部分关于稀疏学习的工作都是基于 L_1 范数正则化,如线性回归、主元分析、偏最小二乘稀疏模型的引导. 稀疏模型即对原有模型的参数作稀疏约束,表现出来就是原有模型加 L_1 范数正则化. 如式(1)所示, $f(X;w)$ 为待求解模型,其中 X 为输入模型的数据, w 为模型的参数.

$$\min_w f(X;w) + \lambda \|w\|_1 \quad (1)$$

2.1.2 L_1/L_q 范数正则化

由于应用需求不同,在机器学习和统计分析中 L_1 范数正则化被扩展为 L_1/L_q 范数正则化. L_1/L_q 范数正则化属于复合绝对值惩罚簇 (the composite absolute penalties (CAP) family),当 $q > 1$ 时, L_1/L_q 范数正则化能够求解出以组为单位的稀疏解. 如式(2)所示, $f(X;w)$ 为待求解模型,其中 X 为输入模型的数据, w 为模型的参数, w_{G_i} 为权重的第 i 个分组以及 α_i 为相应组的权重.

$$\min_w f(X;w) + \lambda \sum_{i=1}^9 \alpha_i \|w_{G_i}\|_2 \quad (2)$$

2.1.3 L_1 范数与 L_1/L_q 范数组合

本文所使用的 Sparse Group Lasso 是一种 L_1 范数与 L_1/L_q 范数组合的正则化方法,其目标是同时进行了特征组的选取和组内特征的选取,一般具有如式(3)形式 (其中符号表示的意义与上述相同).

$$\min_w f(X;w) + \lambda_1 \|w\|_1 + \lambda_2 \sum_{i=1}^9 \alpha_i \|w_{G_i}\|_2 \quad (3)$$

2.2 概率线性判别分析 PLDA

2.2.1 PLDA 模型

概率线性判别分析 PLDA (Probabilistic Linear Discriminant Analysis) 是一个生成模型, 其与标准 LDA (Linear Discriminant Analysis) 的关系类似于因子分析和主元分析的关系. 假设有 I 类图片, 每类图片有 J 张, 定义 x_{ij} 为第 i 类中第 j 张图片, 则 x_{ij} 在 PLDA 模型下的生成方式如下:

$$x_{ij} = \mu + Fh_i + Gw_{ij} + \varepsilon_{ij}. \quad (4)$$

该模型由两部分组成: (i) 信号部分 $\mu + Fh_i$, 仅由类别而非具体样本决定 (与 j 无关); (ii) 噪声部分 $Gw_{ij} + \varepsilon_{ij}$, 在 i 相同时反应了同一类别下的噪声信息.

其中 μ 项表示整个训练集上样本的均值, 矩阵 F 的列构成了与类别相关的子空间的一组基, h_i 代表 i 类在该子空间中的坐标. 同理矩阵 G 的列构成类内子空间的一组基, w_{ij} 表示了在该子空间中的位置. 而仍然残留的部分由 ε_{ij} 来解释, 假设这是一个符合协方差矩阵 Σ 是对角阵的高斯分布. 因此, 该模型的参数为 $\theta = \{\mu, F, G, \Sigma\}$.

从因子分析的角度看, 矩阵 F 和 G 为包含因子, 隐变量 h_i 和 w_{ij} 为相应的因子载荷. 与标准 LDA 类似, F 的列可以近似为类间协方差矩阵的特征向量, G 可以近似为类内协方差矩阵的特征向量. 隐变量 h_i 代表了第 i 个类别的信息, 在识别阶段计算两幅图片的似然值的前提是该两幅图片是同一个潜在隐变量 h_i 生成的.

一般, PLDA 模型以如下这种概率形式来表达:

$$P(x_{ij} | h_i, w_{ij}, \theta) = G_x(\mu + Fh_i + Gw_{ij}, \Sigma) \quad (5)$$

$$P(h_i) = G_h(0, I) \quad (6)$$

$$P(w_{ij}) = G_w(0, I) \quad (7)$$

其中 $G_a(b, C)$ 代表变量 a 符合均值为 b 协方差矩阵为 C 的高斯分布, 在式 (6) (7) 中定义了满足简单先验分布的隐变量 h_i 和 w_{ij} .

2.2.2 训练 PLDA 模型

PLDA 模型的训练阶段目标为: 从训练集数据中学习出模型的参数 $\theta = \{\mu, F, G, \Sigma\}$. 当隐变量 h_i 和 w_{ij} 均已知时, 参数是十分容易学习出来的. 同样, 如果参数已知, 隐变量 h_i 和 w_{ij} 也是很容易估计出来的. 然而模型参数和隐变量均是未知的, 那么解决这种模型训练的有效方法就是 EM 算法. EM 算法在保证每次迭代中观测数据的似然不断增大的情况下交替更新参数和隐变量的值, 最终达到收敛.

3 基于 SS-PLDA 表情识别方法

基于本文提出的 SS-PLDA 方法的表情识别框架如图 (1) 所示.

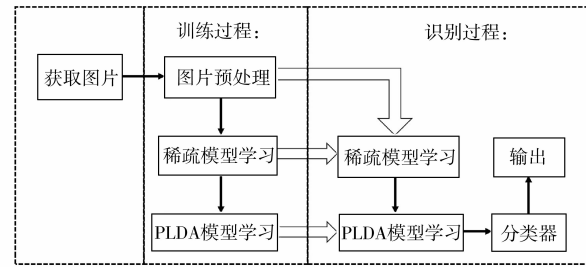


图1 本文提出的方法框架

在模型训练的过程中先将经过预处理的图片进行稀疏模型学习, 求得关于表情的完备特征集 (完备特征集的求取通过在测试集上的测试结果迭代地调整稀疏性参数得到). 然后基于该完备特征集应用 PLDA 模型学习到仅与表情相关的表情子空间. 在识别过程中, 任意两张含有表情的图片将会通过稀疏模型和 PLDA 模型映射到表情空间中, 应用贝叶斯方法求出该两张图片在表情空间中是同一种表情的似然值, 并将该值作为任意两张图片在表情空间中相似度的一个度量. 本文利用该度量值构建一个近邻分类器, 输出表情识别的结果. 下面分别介绍稀疏模型和 PLDA 模型是如何应用在表情识别中.

3.1 表情完备特征集构建

在构建表情的完备特征集之前先构造表情的冗余特征集, 即先对每张表情图片进行如下处理: (i) 人脸检测与归一化 (归一化大小为 96×96); (ii) 将定位好的人脸图片划分为多个尺度的有重叠的小块, 如图 (2) 所示.

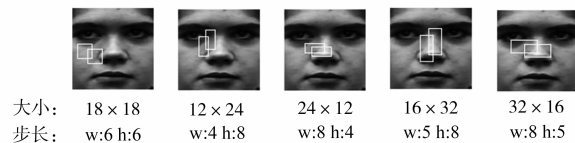


图2 图片块的五种划分方式

其中每一个图片小块被看作一个潜在的表情运动单元 (Action Unit), 通过提取图片小块的特征^[25]来表示每个小块. 本文提出了一个基于线性回归的二分类稀疏模型, 该模型可以实现在大量的人脸图片小块中寻找出能够将某一类表情与其他类表情显著区分出来的图像块, 可以认为这些被挑选出来的图像块是该表情的特征集. 该基于线性回归的二分类稀疏模型如下所示:

$$\begin{cases} T = \mathbf{A}\boldsymbol{\beta} + c, \\ \mathbf{A} = \begin{bmatrix} X^+ \\ X^- \end{bmatrix}, Y_i = \begin{cases} 1, T_i \geq 0, \\ -1, T_i < 0, \end{cases} \end{cases} \quad (8)$$

式 (8) 中, \mathbf{A} 中数据每一行代表一个样本, 每个样本数据是将所有图片小块的特征拼接而成的向量. 其

中 X^+ 代表待求取的表情数据, 为正样本; X^- 为剩余表情的数据, 为负样本. $\beta = [\beta_1, \dots, \beta_j, \dots, \beta_N]^T$ 是需要被求解的系数, 其结果用来评估相应图片块对于区别该种表情的重要性. c 为常数项, T 为每类样本对应的回归值, T 的值输入一个简单的符号二分类器中得到每个样本的标签信息 Y . 图(3)是该模型的一个图像化展示.

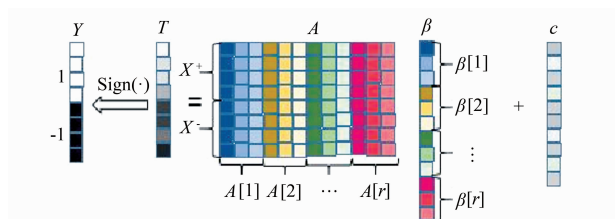


图3 稀疏模型的可视化展示

为了将某种表情与其他表情最优区分的区域挑选出来, 本文使用了一个二分类模型, 由先验知识可知, 待求解区域数量相对较少, 因此在求解时需要保持系数稀疏这一约束. 从而求解该模型的优化目标必须包含两部分: (i) 保证二分类模型正确的损失部分; (ii) 保证相应稀疏约束的正则化部分. 本文使用了 Logistic 损失和 L_1 与 L_1/L_q 范数正则化为最优化目标函数, 如式(9)所示:

$$\begin{aligned} \min_{\beta} \sum_{i=q}^N w_i \log(1 + \exp(-y_i(x_i^T \beta + c))) \\ + \lambda_1 \|\beta\|_1 + \lambda_2 \sum_{j=1}^r w_j^r \|\beta_{R_j}\|_2, \end{aligned} \quad (9)$$

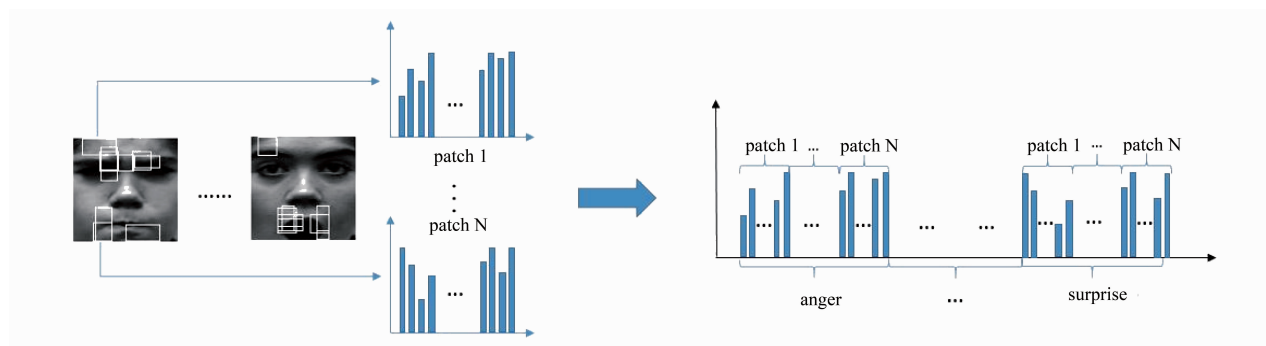


图5 表情完备特征的构造

3.2 基于 PLDA 的表情分析

线性判别分析 (Linear discriminant analysis) 是一种对类间和类内的方差作多元高斯分布假设的模型, 其目的是寻找最有利于分类任务的超平面. 概率线性判别分析 PLDA (Probabilistic Linear discriminant analysis) 则从概率生成的角度阐述了样本中与分类相关的部分和与分类无关的部分是如何生成的. 在表情识别任务中, 不同人的同一表情享有某种相似性, 这种相似性决定了该种表情的生成. 本文利用概率线性判别分析 PL-

式(9)中假设有 N 个训练样本, 每个样本被划分为 r 个图像小块, 前半部分为 Logistic 损失, 保证了二分类模型的正确性. 其中 x_i^T 为由全部图像小块特征连接而成的单个样本, y_i 为对应样本的标签, 正样本取值为 1, 负样本取值为 -1. w_i 为每个样本的权重, 可以减少因正负样本数量的差距而引起的误差. 后半部分为 L_1 与 L_1/L_q 范数正则化项, λ_1 和 λ_2 为相应权重系数, w_j^r 代表第 j 个小图像块的权重, β_{R_j} 为与第 j 个小图像块相关的系数. 该优化目标函数的求解使用了 SLEP^[24] 稀疏求解工具.

以上求出每种表情区别于其他表情的人脸区域, 这些区域可以看作该种表情的特征区域, 应用该稀疏模型求得每种表情的特征区域如图(4)所示.

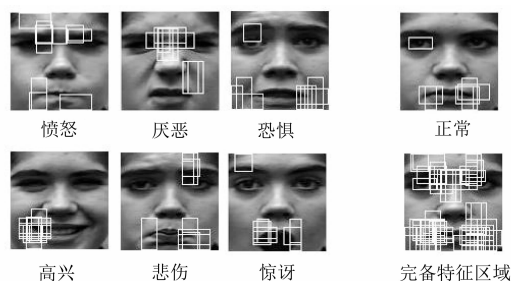


图4 表情相关区域图示

图(4)分别显示了正常状态和六种表情下最相关的区域, 所有表情的特征区域的集合构成了整张人脸的表情区域, 本文将这个区域称作表情的完备特征集, 并且基于表情的完备特征集构造出表情的完备特征, 完备特征的构造过程如图(5)所示.

DA 学习不同人的表情是如何生成的, 由于是线性模型, 所以学习到的结果是表情空间的一组基, 不同人的相同表情在该表情空间下的生成是类似的.

正如上一章所提到的, PLDA 模型具有如下形式:

$$x_{ij} = \mu + Fh_i + Gw_{ij} + \varepsilon_{ij} \quad (10)$$

其中在表情识别任务中, x_{ij} 代表第 i 类表情的第 j 个样本, μ 为所有训练样本上面的均值, F 为学习到的与表情相关的类间信息生成空间, h_i 代表第 i 类表情在该空间的生成方式, G 为与表情无关的类内有规律的噪声

(如身份等信息)信息生成空间, w_{ij} 对应第 i 类表情的第 j 个样本的噪声生成方式, ε_{ij} 是对余下的没有规律的噪声的生成. 这样 PLDA 定义了一个含有表情的人脸图片的生成模型, 其中 h_i 和 w_{ij} 是模型中的隐变量, 模型参数为 $\theta = \{\mu, F, G, \sum\}$.

对于这样一个含有隐变量的模型, 本文采用 EM 算法对其进行求解, 为了简化为常见的隐变量模型, 对式 (10) 作如下变形:

$$x_{ij} = \mu + (F \ G) \begin{pmatrix} h_i \\ w_{ij} \end{pmatrix} + \varepsilon_{ij} \\ = \mu + Bz_{ij} + \varepsilon_{ij} \quad (11)$$

那么式 (5) (6) (7) 等价于式 (12) (13):

$$P(x_{ij} | z_{ij}, \theta) = G_x(\mu + Bz_{ij}, \sum^{-1}), \quad (12)$$

$$P(z_{ij}) = G_z(0, I). \quad (13)$$

EM 算法分为 E 步和 M 步, 本文中在 E 步先同时估计出每种给定表情 (假设有 J 张) 的 $J+1$ 个隐变量 $z_i = [h_i, w_{i1}, \dots, w_{ij}]$ 的联合分布. 应用贝叶斯定理可得:

$$P(z_i | x_i, \theta) \propto P(x_i | z_i, \theta) P(z_i). \quad (14)$$

由式 (12) (13) 知式 (14) 右边均为高斯分布, 则其乘积也为高斯分布, 计算 z_i 的一阶矩 $E[z_i]$ 和二阶矩 E

$[z_i z_i^T]$ 用于下一步计算.

在 EM 算法的 M 步, 目标是更新参数 $\theta = \{\mu, F, G, \sum\}$, 所优化的目标函数为:

$$Q(\theta_i, \theta_{i-1}) = \sum_{i=1}^I \sum_{j=1}^J \int P(z_i | x_{i1}, \dots, x_{ij}, \theta_{i-1}) \\ \log [P(x_{ij} | z_i) P(z_i)] dz_i. \quad (15)$$

通过对式 (11) 求最大值, 得到参数的更新公式:

$$\mu = \frac{1}{IJ} \sum_{i,j} x_{ij}, \quad (16)$$

$$B = \left(\sum_{i,j} (x_{ij} - \mu) E[z_i] \right) \left(\sum_{i,j} E[z_i z_i^T] \right)^{-1}, \quad (17)$$

$$\sum = \frac{1}{IJ} \sum_{i,j} \text{Diag} [(x_{ij} - \mu) (x_{ij} - \mu)^T \\ - BE[z_i] (x_{ij} - \mu)^T]. \quad (18)$$

图 (6) 是在 CK+ 库上求解出的表情空间和噪声 (非随机) 空间主元信息的可视化, 从图中可看出库中所有的表情均可由表情空间表示出, 表情空间只与表情信息的表达有关. 而噪声空间则包含了身份、光照等与表情无关而又非随机的信息的表达.

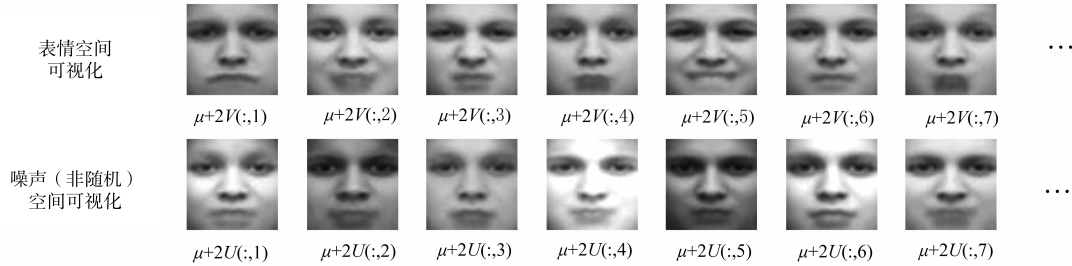


图6 PLDA学习到的子空间图示

参数 $\theta = \{\mu, B, \sum\}$ 被学习出来后并不能直接用于表情的识别中, 因为对于每张表情图片的生成方式由隐变量控制. 本文使用了一种基于贝叶斯的方法来度量任意两张图像含有表情的相似性, 这种方法的主要思想是: 每一个含有标签的训练样本都是由表情空间和噪声空间的隐变量控制生成的, 相同标签的样本在表情空间的生成方式是相同的, 在噪声空间的生成方式不同. 求取某一测试数据与训练数据表情的相似度即求得该测试数据和训练数据在表情空间生成方式相同时的联合概率大小. 这一过程可由图 (7) 表示出.

如图 (7) 所示, 假设有两个训练样本 x_1 和 x_2 分别属于两个不同的类别, 一个测试样本 x_p . 令 x_p 与 x_1 同属一类时为模型 M_1 , 与 x_2 同属一类时为模型 M_2 . 由图 (7) 可知, 属于同一类的观测样本拥有相同的隐变量 h_i , 不属于同一类的隐变量 h_i 也不相同. 根据这一特点, 计算第 q 个模型 (即测试样本属于第 q 类) 的似然项 $P(X | M_q)$, 其

中 X 是所有的观测数据, 由贝叶斯定理可得:

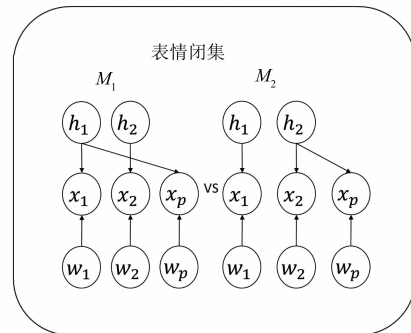


图7 表情验证

$$P(M_q | X) = \frac{P(X | M_q) P(M_q)}{\sum_{r=0}^R P(X | M_r) P(M_r)}, \quad (19)$$

结合本例, 如果 x_p 属于 M_1 , 即 x_p 与 x_1 是同一类, 那么有:

$$P(x_1, x_2, x_p | M_1) = P(x_1, x_p | M_1)P(x_2 | M_1). \quad (20)$$

图(7)也可看出,与 x_1, x_2 和 x_p 相关的自由变量均是独立的(图中相互之间没有连接),则预测分布式(20)的右边可以写成如下方式:

$$P(x_1, x_p | M_1) = \iiint P(x_1, x_p, h_1, w_1, w_p) dh_1 dw_1 dw_p, \quad (21)$$

$$P(x_2 | M_1) = \iint P(x_2, h_2, w_2) dh_2 dw_2. \quad (22)$$

最终求得了第 q 个模型的似然项 $P(X | M_q)$, 该似然值越高说明测试样本属于该模型概率越大, 因此该值可用来作表情空间的似然性的度量. 如果该模型是一个类中心, 则度量的是该测试样本与某类的相似性, 若该模型是某个训练样本, 则度量的是这两个样本在表情空间的相似性. 根据这种表情空间的相似性度量, 本文设计了近邻分类器用于表情结果的预测.

3.3 基于 SS-PLDA 的表情识别

Sparse Selection 和 PLDA 从不同的方面对表情特征进行了提取, Sparse Selection 受启发于面部表情编码系统 (FACS) 中表情由面部运动单元 (AU) 构成, 主要关注表情信息在人脸结构上的表达; PLDA 从降维的角度关注了去除高维纹理信息的冗余成分, 使得表情信息在学习到的表情子空间中更高效的表达. 因此这两种方法可以很自然的构成一种对表情特征提取更具有表达力的方法.

本文的 SS-PLDA 表情识别方法正是基于此提出的, 图(4)和图(6)分别展示了 Sparse Group Lasso 和 PLDA 训练结果的可视化展示, 下一部分的实验与分析将具体从实验的角度分别验证 Sparse Group Lasso、SS-PLDA 各自方法的有效性.

使用 Sparse Group Lasso 的训练结果如图(4), 从图中我们可以看出每种表情所挑选出来的特征区域与人的认知大致相同, 该方法的有效性将在实验与分析部分给出.

使用 PLDA 的训练结果如图(6)所示, 上面一行为表情空间的基加上平均脸的可视化, 下面一行是噪声空间的基加上平均脸的可视化结果, 从图中可大致看出各种表情之间并非正交的, 表情空间的基可以合成训练集上的所有表情, 噪声空间刻画了光照、身份等与表情无关的信息.

4 实验与分析

4.1 实验过程

本文在 JAFFE 和 CK + [26] 这两个库上进行实验, 在 JAFFE 表情库上选取中性加高兴、悲伤、惊奇、愤怒、厌恶、恐惧共七种表情进行训练和测试, 在 CK + 表情库选取高兴、悲伤、惊奇、愤怒、厌恶、恐惧六种表情进行训练和测试. JAFFE 表情数据库包含了 213 幅日本女

性的面部图片, 每幅图像的分辨率为 256×256 , 库中共有 10 个人, 本实验选取所有人的所有图片. CK + 表情数据库有 123 个 subjects (包括跨性别和肤色的样本), 其中 327 个 sequence 有表情的 label, 本实验中每个 sequence 取三张表情峰值图片. 该两个库上的测试方式均为十折交叉验证, 每张图片经过人脸检测, 特征点定位最后归一化到 96×96 大小的人脸正面图像. 实验中将每张图片按图(2)所示五种方法划分为小块, 并对每个小块提取 HOG 特征 [25], 并在多个维度对全部区域特征 + SVM、挑选区域特征 + SVM 和挑选区域 + PLDA 这三种方法进行比较, 验证本文提出的挑选区域 + PLDA (SS-PLDA) 方法在表情识别中的有效性, 并将其中最好的结果与一些现有方法作对比.

4.2 实验结果及分析

4.2.1 JAFFE 表情库实验

图(8)是在 JAFFE 表情库上三种方法的实验结果, 从结果可以看出所有情况下基于挑选区域 + SVM 和挑选区域 + PLDA 的方法均要好于基于全部区域 + SVM 的方法, 从而可以验证本文提出的基于稀疏特征挑选和 PLDA 的方法是有效的.

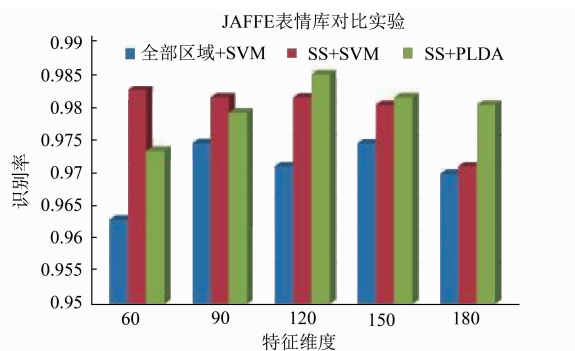


图8 JAFFE库实验结果

从图(8)可以看出, 基于区域挑选 + PLDA 的方法 (即本文提出的 SS-PLDA 方法) 在特征维度为 120 维时识别率达到最高, 表(1)是该方法的混淆矩阵 (该表的列代表测试样本真实表情, 行代表预测输出的表情). 表(2)是本文方法与其他已有方法在 JAFFE 库上面的结果对比.

表 1 120 维特征下 SS-PLDA 在 JAFFE 库上混淆矩阵

	愤怒	厌恶	恐惧	高兴	悲伤	惊讶	中性
愤怒	99.17	0	0	0	0	0	0
厌恶	0.83	97.50	0	0	0	0	0
恐惧	0	1.67	100	0	0	0	0
高兴	0	0	0	95.83	2.50	0	0
悲伤	0	0.83	0	4.17	97.50	0	0
惊讶	0	0	0	0	0	100	0
中性	0	0	0	0	0	0	100
平均	98.57						

表 2 JAFFE 表情库结果对比

	[27]	[28]	[17]	[29]	[30]	本文
愤怒	94.30	100	96.70	98.00	93.33	99.17
厌恶	80.10	100	93.10	100	100	97.50
恐惧	86.30	90.00	93.80	93.00	94.44	100
高兴	95.20	100	93.50	95.30	100	95.83
悲伤	77.50	90.00	90.30	98.20	100	97.50
惊讶	89.60	80.00	93.30	100	100	100
中性	84.50	100	100	100	100	100
平均	86.79	94.30	94.37	97.50	98.24	98.57

从对比结果可以看出,本文方法由于选择的是完备的表情特征集和学习的表情子空间,相较于其他方法,没有识别率特别低的表情且平均识别率是所比较方法中最高的。

4.2.2 CK + 表情库实验

相比于 JAFFE 表情库而言,CK + 表情库是更新更丰富的库。图(9)是在 CK + 库上面三种方法的实验结果,从图中可以看出,基于区域挑选 + SVM 的方法较之全部区域 + SVM 的方法在低维特征下更有效,而在较高维的特征下两种方法性能持平。这可能是由于不同于 JAFFE 表情库(东方女性),CK + 表情库包含更多的与表情无关的信息(性别、肤色和光照等),单纯的依靠区域挑选较难得到更有区分度的特征。而图中基于区域挑选 + PLDA 的方法始终较其他两种方法有更好得性能,从而验证了基于区域挑选 + PLDA (SS-PLDA)方法在表情空间中进行的表情识别是有效的。

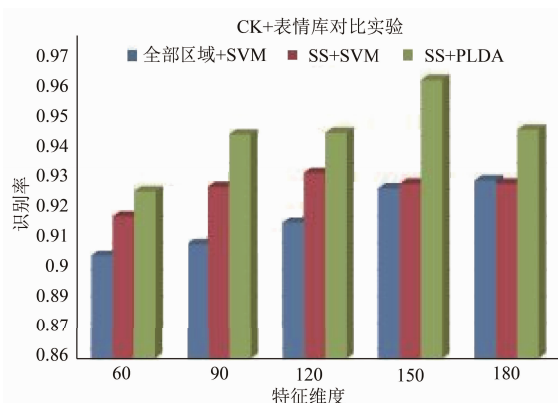


图9 CK+库实验结果

从图(9)中可以看出基于区域挑选 + PLDA 的方法在特征为 150 维时取得最佳的识别效果,表(3)是该种情况下的混淆矩阵(该表的列代表测试样本真实表情,行代表预测输出的表情)。表(4)是本文方法与其他已有方法在 CK + 库上面的结果对比。

表 3 150 维特征下 SS-PLDA 在 CK + 库上混淆矩阵

	愤怒	厌恶	恐惧	高兴	悲伤	惊讶
愤怒	95.38	2.50	0	0	3.75	0
厌恶	1.92	97.50	0	0	1.25	0
恐惧	0	0	95.00	1.15	0	0.94
高兴	0	0	3.75	98.64	0	0.63
悲伤	2.30	0	0	0	93.75	0
惊讶	0.40	0	1.25	0.21	1.25	98.43
平均	96.45					

从对比结果可以看出,CK + 表情库相对于 JAFFE 库更复杂(多人种、性别及光照等干扰),因此总体识别率相对于 JAFFE 表情库较低。同样,基于完备特征集和表情子空间的本文方法在各类表情识别上均无过低的识别率,平均识别率也是相比的几种方法中最高的。

表 4 CK + 表情库结果对比

	[31]	[13]	[19]	[32]	[18]	本文
愤怒	82.50	87.30	71.39	87.10	87.80	95.38
厌恶	97.50	91.58	95.33	90.20	93.33	97.50
恐惧	95.00	90.98	81.11	92.00	94.33	95.00
高兴	100	96.92	95.42	98.07	94.20	98.64
悲伤	92.50	84.58	88.01	91.47	96.42	93.75
惊讶	92.50	91.23	98.27	100	98.46	98.43
平均	93.33	90.38	88.23	93.14	94.09	96.45

5 结论

本文提出了一种基于区域特征挑选和特征子空间(SS-PLDA)的表情识别方法,该方法通过区域特征挑选和特征子空间学习得到具有更好表达力的表情特征。在具有大量非表情因素(身份、性别、肤色等)干扰的情况下,本文方法相对于其他方法能够提出更鲁棒的表情特征,这种鲁棒性体现在不仅所有表情的平均识别率更高而且单一表情的识别率也不会太低。

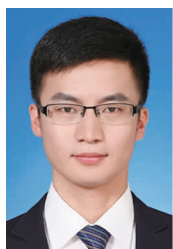
参考文献

- [1] Ryan A, Cohn J F, Lucey S, et al. Automated facial expression recognition system [A]. 2009 International Carnahan Conference on Security Technology [C]. IEEE, 2009. 172 - 177.
- [2] Tian Y L, Kanade T, Cohn J F. Recognizing action units for facial expression analysis [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(2): 97.
- [3] Aleksic P S, Katsaggelos A K. Automatic facial expression recognition using facial animation parameters and multi-stream HMMs [J]. IEEE Transactions on Information Fo-

- rensis & Security, 2006, 1(1): 3 – 11.
- [4] Zhang Y, Ji Q. Active and dynamic information fusion for facial expression understanding from image sequences[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(5): 699 – 714.
- [5] Ekman P. Facial action coding system: manual[J]. Agriculture, 1978.
- [6] Whitehill J, Bartlett M, Movellan J. Automatic facial expression recognition for intelligent tutoring systems[A]. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)[C]. IEEE, 2008. 1 – 6.
- [7] Jain S, Hu C, Aggarwal J K. Facial expression recognition with temporal modeling of shapes[A]. IEEE International Conference on Computer Vision Workshops [C]. IEEE, 2011. 1642 – 1649.
- [8] Kotsia I, Zafeiriou S, Nikolaidis N, et al. Texture and shape information fusion for facial action unit recognition[A]. First International Conference on Advances in Computer-Human Interaction[C]. IEEE, 2008. 77 – 82.
- [9] Valstar M F, Pantic M. Combined support vector machines and hidden Markov models for modeling facial action temporal dynamics[A]. IEEE International Conference on Human-Computer Interaction [C]. Springer-Verlag, 2007. 118 – 127.
- [10] Shan C, Gong S, Mcowan P W. Robust facial expression recognition using local binary patterns[A]. IEEE International Conference on Image Processing [C]. DBLP, 2005. II-370 – 3.
- [11] Moore S, Bowden R. Local binary patterns for multi-view facial expression recognition[J]. Computer Vision & Image Understanding, 2011, 115(4): 541 – 558.
- [12] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(6): 915 – 928.
- [13] Poursaberi A, Noubari H A, Gavrilova M, et al. Gauss – Laguerre wavelet textural feature fusion with geometrical information for facial expression identification [J]. Eurasip Journal on Image & Video Processing, 2012, 2012(1): 1 – 13.
- [14] Zhang Z, Yan Y, Wang H. Discriminative filter based regression learning for facial expression recognition [A]. IEEE International Conference on Image Processing [C]. IEEE, 2013. 1192 – 1196.
- [15] Li Z, Imai J I, Kaneko M. Facial-component-based bag of words and PHOG descriptor for facial expression recognition[A]. IEEE International Conference on Systems, Man and Cybernetics [C]. IEEE, 2009. 1353 – 1358.
- [16] Shan C, Gong S, Mcowan P W. A comprehensive empirical study on linear subspace methods for facial expression analysis[A]. Conference on Computer Vision and Pattern Recognition Workshop [C]. IEEE, 2006. 153 – 153.
- [17] Rahulamathavan Y, Phan C W, Chambers J A, et al. Facial expression recognition in the encrypted domain based on local fisher discriminant analysis [J]. IEEE Transactions on Affective Computing, 2013, 4(1): 83 – 92.
- [18] Happy S L, Routray A. Automatic facial expression recognition using features of salient facial patches [J]. IEEE Transactions on Affective Computing, 2015, 6(1): 1 – 12.
- [19] Yang P. Learning active facial patches for expression analysis[A]. Computer Vision and Pattern Recognition [C]. IEEE, 2012. 2562 – 2569.
- [20] Yue Zhao, Jianbo Su. New sparse facial feature description model based on salience evaluation of regions and features [J]. International Journal of Pattern Recognition & Artificial Intelligence, 2015, 29(05): 1556007.
- [21] Bishop C M. Pattern Recognition and Machine Learning (Information Science and Statistics) [M]. New York: Springer-Verlag, 2006.
- [22] Prince S, Li P, Fu Y, et al. Probabilistic models for inference about identity [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(1): 144 – 157.
- [23] Gong D, Li Z, Lin D, et al. Hidden factor analysis for age invariant face recognition[A]. IEEE International Conference on Computer Vision [C]. IEEE, 2014. 2872 – 2879.
- [24] Liu J, Ye J. Moreau-Yosida regularization for grouped tree structure learning[A]. International Conference on Neural Information Processing Systems [C]. Curran Associates Inc, 2010. 1459 – 1467.
- [25] Ludwig O, Delgado D, Goncalves V, et al. Trainable classifier-fusion schemes: An application to pedestrian detection [A]. International IEEE Conference on Intelligent Transportation Systems [C]. IEEE, 2009. 1 – 6.
- [26] Lucey P, Cohn J F, Kanade T, et al. The extended Cohn-Kanade dataset (CK+) : A complete dataset for action unit and emotion-specified expression [A]. Computer Vision and Pattern Recognition Workshops [C]. IEEE, 2010. 94 – 101.
- [27] Jabid T, Kabir M H, Chae O. Robust facial expression recognition based on local directional pattern [J]. Etri Journal, 2010, 32(5): 784 – 794.
- [28] Liu W, Song C, Wang Y. Facial expression recognition based on discriminative dictionary learning [A]. International Conference on Pattern Recognition [C]. IEEE, 2012. 1839 – 1842.
- [29] Lajevardi S M, Hussain Z M. Automatic facial expression recognition: feature extraction and selection [J]. Signal Image & Video Processing, 2012, 6(1): 159 – 169.

- [30] 刘帅师,田彦涛,万川. 基于 Gabor 多方向特征融合与分块直方图的人脸表情识别方法[J]. 自动化学报,2011,37(12):1455-1463.
LIU ShuaiShi, TIAN YanTao, WAN Chuan. Facial expression recognition method based on Gabor multi-orientation features fusion and block histogram[J]. Acta Automatica Sinica,2011,37(12):1455-1463. (in Chinese)
- [31] Uddin M Z, Lee J J, Kim T S. An enhanced independent component-based human facial expression recognition from video[J]. IEEE Transactions on Consumer Electronics,2009,55(4):2216-2224.
- [32] Zhang L, Tjondronegoro D. Facial expression recognition using facial movement features[J]. IEEE Transactions on Affective Computing,2012,2(4):219-229.

作者简介



张 瑞 男.1991 年 9 月生于安徽.上海交通大学自动化系硕士研究生.研究方向为
人脸识别、表情识别.
E-mail:Zander_Ray@sjtu.edu.cn



苏剑波(通信作者) 男.1969 年 11 月生于江苏.上海交通大学自动化系教授,研究方向为
机器视觉、机器学习与人机交互、多传感器信息融合、智能机器人等.