工程与应用

doi: 10.3969/j.issn.1673-5692.2016.02.022

基于机器学习算法的人脸识别鲁棒性研究

李利民,刘明辉

(中国电子科学研究院,北京 100041)

摘 要:在现代公安警务工作中,人脸识别是智能化目标人物排查、线索追踪的重要支撑技术。在实际应用中,公安布控人脸图像采集通常处于非合作场景。在环境因素的作用下,采集的图像相比于标准库中人脸图像往往发生噪声叠加、曝光异常以及运动模糊等降质褪化。因此,人脸识别算法的鲁棒性应当成为其有效性的重要判断依据之一。鉴于上述考虑,本文研究了几种典型机器学习算法在不同图像降质因素作用下的人脸识别性能,进一步分析了上述算法的鲁棒性。

关键词:人脸识别;鲁棒性;反向传播神经网络;径向基神经网络;广义回归神经网络

中图分类号:TP181 **文献标识码**:A **文章编号**:1673-5692(2017)02-219-06

Research on the Robustness of Face Recognition Based on Machine Learning Algorithms

LI Li-min, LIU Ming-hui

(China Academy of Electronics and Information Technology, Beijing 100041, China)

Abstract: Face recognition is an important technology for supporting intelligent target character investigation and clue tracking in in modern police affairs. Under the influence of environmental factors, such as noise superposition, exposure abnormity and motion blur, degeneration usually occur to sampled images comparing with the standard face image. Therefore, the robustness of face recognition algorithm is one of the important criteria for its effectiveness. For the above consideration, this paper investigates the performance of several typical machine learning algorithms under different image degeneration factors, and further analyses the robustness of the aforementioned algorithms.

Key words: face recognition; robustness; back-propagation neural network; radial basis function neural network; generalized regression neural network

0 引 言

近年来,在我国“平安城市”建设的持续发展进程中,公共安防布控的图像及视频采集数据呈爆炸性增长趋势。在这样的背景下,基于海量数据的智能化人物身份识别是支撑社会安全风险防控的重要关键技术,对现代公安警务工作中目标人物排查、线索发现跟踪具有显著的实际意义[1-2]。作为数字图像处理的代表性技术之一,人脸识别是一种利用计算机分析人脸图像,根据图像内在特征识别人物身份的技术。在日常生活中,人脸特征不仅传达着丰富细腻的情感及心理信息,且具有独一无二的生物学特征。由于人脸图像采集具有非接触性、非侵犯性,相对于其他生物特征具有更高的采集效率,因此在学术界及工业界得到了广泛关注。

人脸识别技术的广泛潜在市场价值,促使国内外众多研究机构、科研院所投入大量的人力、物力持续攻关。国外较为著名的研究机构主要包括美国麻省理工大学的人工智能实验室、美国斯坦福大学的

收稿日期:2017-01-01 修订日期:2017-04-01

视觉实验室、美国加州大学伯克利分校的计算机视觉实验室、法国国家信息与自动化研究所、英国萨里大学的视觉信息与信号处理研究所、瑞士人工智能感知研究所等。国内的相关的研究机构主要有清华大学智能图文信息处理实验室、上海交通大学计算机视觉实验室、中科院自动化研究所和中科院计算研究所等。

完整的人脸识别流程通常依次执行人脸检测、图像分割、图像预处理、特征提取、人脸匹配。其中，人脸检测主要用于检测目标场景中是否存在人脸图像，并进行定位；进而，通过图像分割将相应的人脸区域分离；对于部分人脸图像，尤其是非合作人脸识别过程中的采集样本，图像通常需要进行去噪、灰度平衡等预处理，使待识别图像尽可能接近于数据库样本图像；特征提取目的在于对图像进行具有鉴别性的数学描述，在保证图像类内距离尽量小的前提下，提升类间距离类间距离，同时降低图像表征维数；人脸匹配是指将待识别人脸特征与人脸库进行比对，或将该特征输入训练好的机器学习模型，从而给出人脸识别结果。在人脸识别计算流程中，对人脸特征进行显性描述进而进行识别往往十分困难，而考虑到部分经典机器学习算法中，神经元的连接机制具有强大的非线性拟合能力，可以有效的隐性表征人脸特征。因此，近年来随着人工智能领域不断取得突破，基于机器学习的人脸识别算法受到了广泛关注[3-7]。

然而在实际应用中，人脸图像采集时常处于非合作场景，最典型即为公安布控。此时，在环境因素的作用下，采集的图像相比于标准库中人脸图像往往发生降质，导致识别准确率发生不同程度的下降。常见降质形式包括噪声叠加、曝光异常以及运动模糊等[8-10]。在特征提取及图像识别前，采用图像预处理可以在一定程度上改善图像的降质问题，但由于逆处理过程中无法获知图像降质的准确模型及参数，因此并不能使图像准确还原。鉴于上述原因，本文研究了基于几种典型机器学习算法在不同图像降质因素作用下的人脸识别性能，进一步分析了集中算法在不同场景下的鲁棒性。需要指出的是，对于椒盐噪声叠加类降质图像，可以采用中值滤波进行普适处理。对于曝光异常以及运动模糊两类降质图像，还原效果显著依赖于相关参数，甚至可能“矫枉过正”。考虑到上述原因，由于本文研究重点聚焦于机器学习算法本身性能，因此对于曝光异常以及运动模糊两类降质图像未进行预处理。

1 基于机器学习的人脸识别典型数学模型

机器学习通常指依赖于连接机制，具有大规模并行处理和分布式的信息存储功能，依靠大量节点的连接以及这种连接所引起的节点的不同兴奋状态来以任意精度逼近任意有限间断点函数的人工神经网络结构。本文研究了反向传播神经网络、径向基神经网络、广义回归神经网络三种经典机器学习算法在几种典型图像降质条件下的人脸识别性能，其基本拓扑结构均为人工神经网络，如图 1 所示。

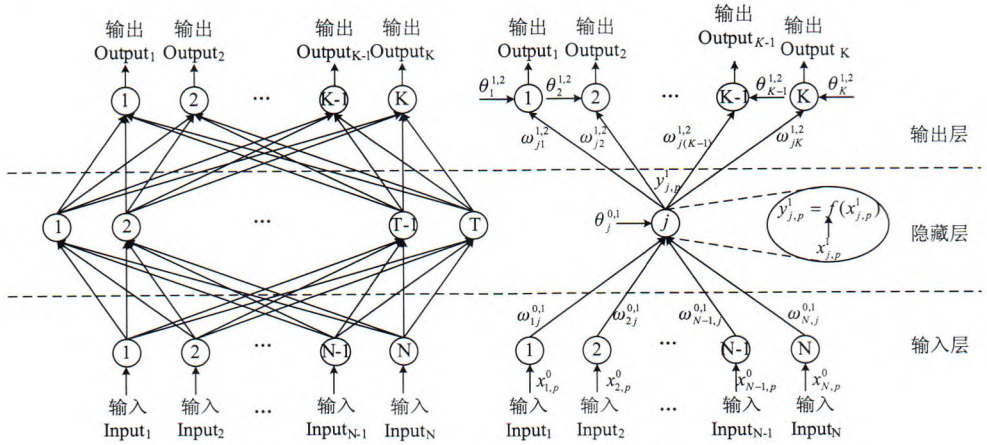

图 1 人工神经网络系统的基本拓扑结构示意图

图 1 中，$\theta_j^{l-1,l}$ 表示 l 层处于激励或抑制状态的神经元偏置值；$\omega_{i,j}^{l-1,l}$ 表示 $l-1$ 层节点 i 到 l 层节点 j 的连接权值；$x_{j,p}^l$ 表示对样本 p，第 l 层神经元 j 的输入值；$y_{j,p}^l$ 表示对样本 p，来自第 l 层神经元 j 的输出值；输入层、隐藏层和输出层分别表示为层 0、1 和 2；函数 $f(\cdot)$ 用来表示神经元的状态转移函数。

1.1 反向传播神经网络算法

反向传播神经网络(BPNN, Back Propagation Neural Nerwork)的特点是信号前向传播,误差反向传递。在前向传递过程中,信号由输入层经隐藏层处理并送至输出层,其中每层神经元的状态只影响下一层神经元状态。如果输出层输出值不满足预设训练中止条件,则转入反向传播过程,根据误差调整网络权值及偏置值,该迭代过程使 BPNN 输出不断逼近期望输出。隐藏层中每个神经元满足式(1)中关系。

$$\begin{cases} x_{j,p}^{l} = \sum_{j=1}^{g(l-1)} \omega_{i,j}^{l-1,l} y_{j,p}^{l-1} - \theta_j^{l-1,l}, l = 1,2 \\ y_{j,p}^{l} = f(x_{j,p}^{l}) = \dfrac{1}{1+e^{-x_{j,p}^{l}}}, 0 < f(x_{j,p}^{l}) < 1 \end{cases} \tag{1}$$

BPNN 训练可以采用梯度下降算法实现,神经元参数更新计算如式(2)所示。

$$\begin{cases} \omega_{i,j}^{l-1,l}(k+1) = \omega_{i,j}^{l-1,l}(k) - \alpha \sum_{p=1}^{P} \delta_{j,p}^{l}(k)\, y_{j,p}^{l-1}(k) \\ \theta_j^{l-1,l}(k+1) = \theta_j^{l-1,l}(k) - \beta \sum_{p=1}^{P} \delta_{j,p}^{l}(k)\, y_{j,p}^{l-1}(k) \\ \delta_{j,p}^{l}(k) = f'(x_{j,p}^{l}(k)) \sum_{m=1}^{2} \delta_{m,p}^{l+1}(k)\, \omega_{j,m}^{l,l+1}(k), l = 1 \\ \delta_{j,p}^{l}(k) = (y_{j,p}^{l}(k) - z_{j,p}) f'(x_{j,p}^{l}(k)), l = 2 \end{cases} \tag{2}$$

其中,α 和 β 分别表示网络中 $\omega_{i,j}^{l-1,l}$ 和 $\theta_j^{l-1,l}$ 的学习速率,$z_{j,p}$ 表示训练阶段第 j 维的期望输出。

1.2 径向基神经网络算法

径向基函数可以实现多维空间插值,对非线性连续函数具有一致逼近性能。相比于 BPNN,径向基神经网络(RBFNN, Radial Basis Function Neural Network)的隐藏层神经元传递函数是对中心点径向对称且衰减的非负非线性函数。RBFNN 的核心思想是用径向基函数构成隐藏层空间,将输入状态矢量由低维空间变换到高维空间,使得在低维空间内线性不可分的数据在高为维空间内仍具有线性可分性,神经元满足式(3)中关系。

$$\begin{cases} y_i = \sum_{i=1}^{h} \omega_{ij} R(x_p - \theta_i) \\ R(x_p - \theta_i) = \exp\left(-\dfrac{1}{2\sigma^2} \| x_p - \theta_i \|^2\right) \\ \sigma = \dfrac{1}{P} \sum_{j=1}^{m} \| d_j - y_j \theta_i \|^2 \\ \omega = \exp\left(\dfrac{h}{\theta_{max}^2} \| x_p - \theta_i \|^2\right) \end{cases} \tag{3}$$

上式中,x_p 表示第 p 个输入样本,P 表示样本总数;θ 表示隐藏层节点聚类中心,ω 表示隐藏层到输出层的连接权值,h 表示隐藏层神经元数,y 表示神经网络输出值,d 表示样本的期望输出值。

1.3 广义回归神经网络算法

广义回归神经网络(GRNN, Generalized Regression Neural Network)具有极强的非线性映射能力,适用于解决非线性拟合问题。相比于 RBFNN,GRNN 在逼近能力和学习速度上更具优势,且在训练样本数量较少或训练样本数据不稳定时具有明显的性能优势,神经元满足式(4)。

$$\begin{cases} p_i = \exp\left(-\dfrac{(\boldsymbol{X} - \boldsymbol{X}_i)^T (\boldsymbol{X} - \boldsymbol{X}_i)}{2\sigma^2}\right) \\ y_j = \sum_{i=1}^{n} y_{ij}\, p_i \Big/ \sum_{i=1}^{n} p_i \end{cases} \tag{4}$$

从输出结果来看,y_j 为所有样本观测值的加权平均,每个观测值的权重值为对应样本 $\boldsymbol{X}_i$ 与 $\boldsymbol{X}$ 的欧氏距离平方的指数,即 p_i。当光滑因子 σ 趋于无穷大时,输出结果近似于所有样本的平均观测值。反之,当 σ 趋近于 0 时,输出结果与训练样本非常接近。此时,若训练样本不完备,则测试输出结果误差将非常大,即 GRNN 的泛化能力较差。

2 人脸图像降质数学模型

图像成像过程中,产生降质的因素繁多,典型如光学系统相差、成像过程发生相对运动、各种外界因素的干扰及噪声等。上述因素均会导致图像发生不同样式、不同程度的降质,进而人脸识别的准确率。文主要分析了三种典型降质因素,即:椒盐噪声、曝光异常以及运动模糊。

2.1 椒盐噪声

实验研究表明,摄像机拍摄图像过程中,图像传感器、传输信道、解码处理等部件或因素将引入椒盐噪声,在图像上呈现黑白杂点,进而影响后续图像处理。设图像 $I(x,y)$ 为 N 位图,椒盐噪声密度为 d_{sp},满足 $d_{sp} \in [0,1]$,则叠加椒盐噪声后的图像 $g(x,y)$ 可以表示如式(5)所示。

$$\begin{cases} g(x,y) = I(x,y) * \gamma(x,y) \\ Prob(\gamma(x,y) = 0) = \dfrac{d_{sp}}{2} \\ Prob\left(\gamma(x,y) = \dfrac{2^N - 1}{I(x,y)}\right) = \dfrac{d_{sp}}{2} \\ Prob(\gamma(x,y) = 1) = 1 - d_{sp} \end{cases} \tag{5}$$

2.2 曝光异常

曝光指图像的物理生成过程中，允许进入镜头照在图像传感器上的光量，通常可以由光圈、快门以及图像传感器的感光度组合控制。理想情况下，曝光度应控制在合理的范围，使照片亮度适中，对比度强。然而，由于人脸识别图像采集场景多变，尤其对于非合作采集条件下的公安布控等场景，往往存在曝光过度或曝光不足等问题，导致图像曝光异常部分细节丢失，进而影响人脸识别准确率。

研究表明，同态滤波方法利用光照反射模型，把频率过滤和灰度变换结合，可以在不损失图像细节的前提下调解图像的光照条件。考虑到同态滤波具有上述特点，因此本文将其用作曝光异常的图像仿真方法。设原始图像为 $I(x,y)$，首先对其取对数并做傅里叶变换，提取高频及低频分量：

$$Z(u,v) = F(z(x,y)) = F(\ln(I(x,y))) \quad (6)$$

之后，采用频域滤波函数对图像进行增强处理，本文采用高斯型高通滤波器作为滤波函数，处理如式(7)所示。其中 M,N 分别表示图像的行、列像素数，D_0 为截止频率，c 为锐化系数，R_h 为高频增益，R_l 为低频增益。

$$\begin{cases} S(u,v) = H(u,v) \cdot Z(u,v) \\ H(u,v) = (R_h - R_l) \cdot \gamma(u,v) + R_l \\ \gamma(u,v) = \left(1 - \exp\left(-c \cdot \left(\dfrac{D(u,v)}{D_0}\right)^2\right)\right) \\ D(u,v) = \left(\left(u - \dfrac{M}{2}\right)^2 + \left(v - \dfrac{N}{2}\right)^2\right)^{0.5} \end{cases} \quad (7)$$

最后，对频域图像做傅里叶反变换并取指数，计算过程如式(8)所示。

$$g(x,y) = e^{s(x,y)} = e^{F^{-1}(S(u,v))} \quad (8)$$

2.3 运动模糊

在照片曝光期间，相机与被摄物体之间发生相对运动造成的图像模糊称为运动模糊。受运动模糊影响的图像往往在视觉上表现为图像像素整体沿某一方向具有拖影效果。当像素位移量偏大时，将严重影响图像质量，从而降低人脸识别准确率。考虑到变速或非直线运动在一定条件下可以被分解为分段匀速直线运动，因此匀速直线运动造成的运动模糊具有普适的研究意义。从物理场景上看，图像发生运动模糊的原因是被摄图像经过一定距离延迟后再进行叠加。将静止条件下的图像表示为 $I(x,y)$，设快门打开期间，图像传感器与被摄物体保持水平匀速直线运动，则图像褪化模型可以表示为式(9)。

$$g(x,y) = \frac{1}{L}\sum_{i=0}^{L} f(x-i,y)\Delta t \quad (9)$$

其中，L 表示图像发生整体位移的像素长度近似值。

3 算法仿真与性能分析

3.1 人脸图像分割及特征提取

在人脸图像特征提取之前，为了实现对原始图像降维以降低机器学习系统复杂度，本文首先对原始图像进行了子图分割。考虑到子图分割方式应充分将人脸的各个显著特征区域区分开，并尽量不破坏其局部整体性，即令子图中尽量独立整体保留头发、额头、眼睛、鼻子、嘴巴、耳朵以及胡须等特征区域。在此基础上，子图分割数量直接影响图像特征的丰富程度。理论上来说，子图分割数量越大，图像特征量越多，特征越丰富，但同时会导致机器学习系统输入量维数增多，从而使得系统过于复杂，训练时间急剧上升。综合上述分析，经过多组仿真性能比较，本文采用 5×5 的二维子图分割方式对原始图像进行区域分割，如图 2 所示。

(a) 原始图像　　(b)5×5 二维子图分割图像

图 2　子图分割示意图

作为一种有效的代数特征，奇异值分解在数据的压缩、信号的处理和模式的识别等多方面都有十分广泛的应用，考虑到其具有比例不变性、旋转不变性等特征，意味着对图像进行旋转、同比灰度调整等操作不会改变其特征值，因此非常适用于人脸图像特征提取。本文中人脸识别算法均首先将待识别图像进行 5×5 二维子图分割，随后对每幅子图进行奇异值分解，然后选择变换后的最大系数代表该子图的特征，并且进行归一化处理，最后将这些归一化后的子图特征系数组合起来，作为整幅人脸图像的特征向量。如果图像矩阵 $A \in R^{m \times n}$，则有正交矩阵 U，V，满足式(10)。

$$\begin{cases} U^T AV = diag[\sigma_1, \sigma_2, \cdots \sigma_p] = W \\ U = [u_1, u_2, \cdots, u_m] \in R^{m \times m} \\ V = [v_1, v_2, \cdots, v_n] \in R^{n \times n} \\ p = \min(m, n) \end{cases} \tag{10}$$

根据上式,可以得出 $A = UWV^T$,称为图像矩阵 A 的奇异值分解。其中,$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$,$\sigma_i$ $(i = 1, 2, \cdots, p)$是 A 的奇异值,同时也是AA^T或A^TA的特征值λ_i的平方根,也就是$\sigma_i = \sqrt{\lambda_i}$。本文仿真中,采用每幅子图最大奇异值的平方根作为子图的特征值,即每幅人脸图像的特征向量为 25 维。

3.2　仿真结果及性能分析

本文采用剑桥大学 ORL 人脸数据库,研究分析了基于 SVD 特征提取的主成分分析(PCA, Principal Component Analysis)[7]、BPNN、RBFNN、GRNN 四种算法的人脸识别性能。在分析不同机器学习算法非线性拟合能力的同时,为了兼顾考量算法的泛化性能,对每个人物随机取 5 副人脸图像用作训练,剩余 5 副图像用作测试。仿真中,加入了椒盐噪声、曝光异常以及运动模糊三类典型人脸图像降质因素,其中曝光异常按曝光不足及曝光过量两类情况分别考虑。用作测试的一组图例如图 3 所示,降质参数的定义在本文第三节有所阐述。

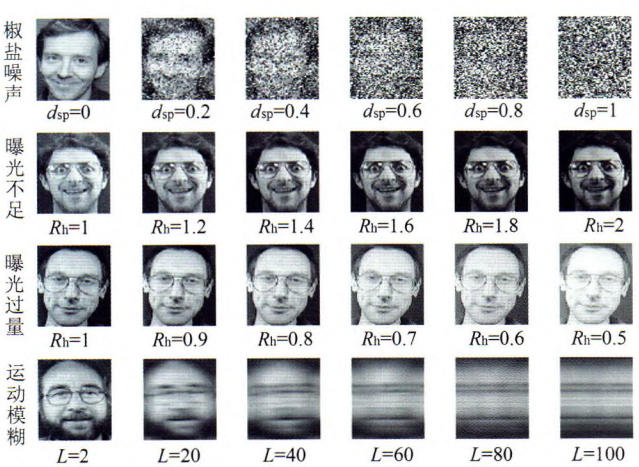

图 3　仿真中不同降质条件下的一组图例

仿真结果如图 4 至图 7 所示。为了深入分析仿真结果依赖于图像降质参数的变化关系,图中同时给出了仿真结果的拟合曲线,拟合算法采用有理数逼近拟合,分子分母均采用 5 次多项式。

根据图 4 中仿真结果可见,当椒盐噪声密度小于 30% 时,PCA 算法与 SVD-RBFNN 算法性能接近,正确识别率约 90%,且明显优于另外两种算法。当

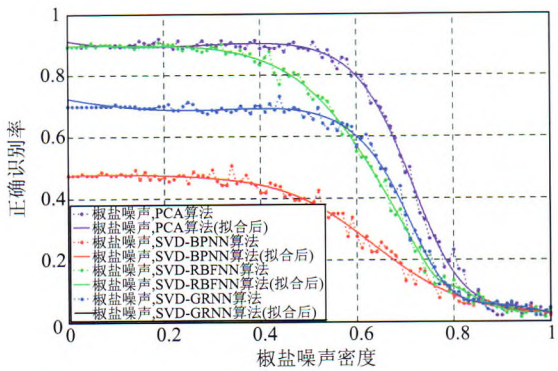

图 4　椒盐噪声作用下不同算法的人脸识别性能

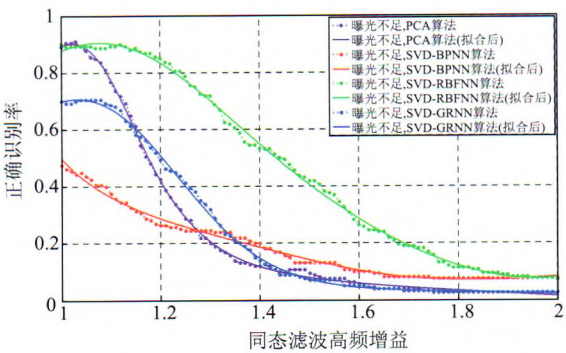

图 5　曝光不足条件下不同算法的人脸识别性能

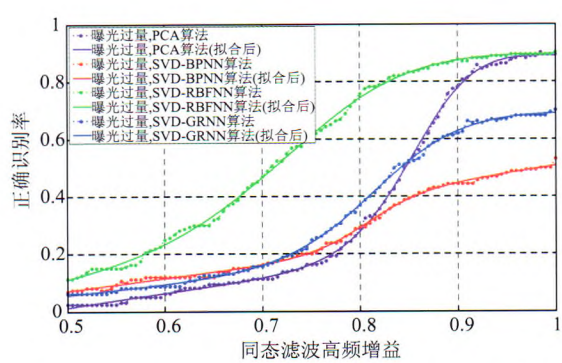

图 6　曝光过量条件下不同算法的人脸识别性能

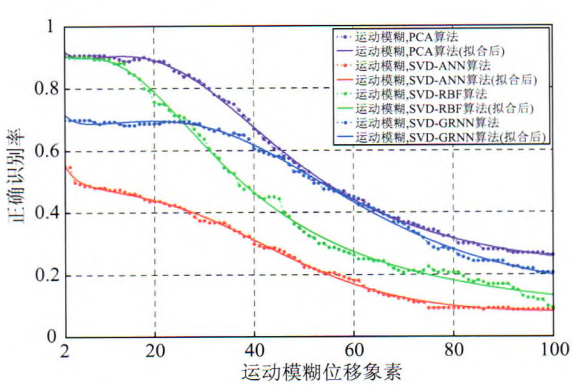

图 7　运动模糊条件下不同算法的人脸识别性能

椒盐噪声密度达到 40% 时，随着噪声继续增加，四种算法的人脸识别正确率均开始呈不同程度的下降趋势。当噪声密度达到 80% 以上时，四种人脸识别算法均基本失效。总体来看，PCA 算法的抗椒盐噪声性能明显优于其他算法。从图 5、图 6 中仿真结果可以看出，人脸识别受曝光量影响较为严重。随着图像曝光量逐渐偏离正常值，PCA 算法性能下降最快，SVD-RBFNN 算法性能变化平缓，其鲁棒性明显优于其他三种算法。根据图 7 中仿真结果，随着运动模糊逐渐增强，SVD-RBFNN 算法性能下降趋势最明显。当运动模糊像素增至 40 时，由图 3 可以看出，人眼已经很难分辨识别图像中人物。此时，根据图 7 中仿真结果，PCA 及 SVD-GRNN 算法的识别率仍在 60% 以上。随着运动模糊像素继续增多，PCA 算法与 SVD-GRNN 算法性能逐渐接近。总体来看，PCA 算法的抗运动模糊性能最佳。

综上所述，在不同降质褪化场景下，PCA、SVD-RBFNN 总体表现鲁棒性较好。深入分析其原因，BPNN 非线性拟合能力相对较弱，因此各项抗图像褪化性能相对较差。此外，由于 GRNN 的非线性拟合能力非常强，往往存在过拟合问题，导致其泛化性能较差。考虑到本文对每组人脸图像非完备集进行训练，因此 GRNN 算法的测试样本输出结果可能发生较大误差。

4 结 语

本文采用 ORL 人脸库，通过仿真研究了基于 SVD 特征提取的 PCA、BPNN、RBFNN、GRNN 四种典型机器学习算法在不同图像降质因素作用下的人脸识别性能，进一步分析了上述算法在不同图像褪化场景下的鲁棒性。从仿真结果来看，对于椒盐噪声及运动模糊影响的人脸图像，PCA 算法鲁棒性较好，但该算法性能显著依赖于图像曝光条件影响。相比之下，当图像曝光条件不佳时，采用 RBFNN 作为人脸识别算法具有相对较好的鲁棒性。

参考文献：

[1] 李建明. 智慧城市发展综述[J]. 中国电子科学研究院学报, 2014(3): 221-233.

[2] 刘成龙, 李志学, 杨畅. 浅谈人脸识别技术在平安城市中的应用[J]. 电子技术与软件工程, 2016(10): 90-90.

[3] Mageshkumar C, Thiyagarajan R, Natarajan S P, et al. Gabor features and LDA based face recognition with ANN classifier[C]// Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on. IEEE, 2011:831-836.

[4] Ch'Ng S I, Seng K P, Ang L M. Adaptive momentum Levenberg-Marquardt RBF for face recognition [C]// IEEE International Conference on Circuits and Systems. 2012:126-131.

[5] Banerjee P K, Datta A K. Generalized regression neural network trained preprocessing of frequency domain correlation filter for improved face recognition and its optical implementation[J]. Optics & Laser Technology, 2013, 45(1):217-227.

[6] Sharma R, Patterh M S. A new pose invariant face recognition system using PCA and ANFIS[J]. Optik-International Journal for Light and Electron Optics, 2015, 126(23):3483-3487.

[7] Yang J, Zhang D, Frangi A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2004, 26(1): 131-137.

[8] Zhang P, Li F. A New Adaptive Weighted Mean Filter for Removing Salt-and-Pepper Noise [J]. IEEE Signal Processing Letters, 2014, 21(10): 1280-1283.

[9] Im J, Fujii H, Yamashita A, et al. Compensation of over and under exposure image using multiple light switching [C]// Ieee/sice International Symposium on System Integration. IEEE, 2014: 147-152.

[10] Punnappurath A, Rajagopalan A N, Taheri S, et al. Face recognition across non-uniform motion blur, illumination, and pose. [J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2015, 24(7): 2067-2082.

作者简介

李利民(1985—)，男，黑龙江人，博士，主要研究方向为网络安全、图像处理、军事通信、通信仿真；

Email:atpcat@163.com

刘明辉(1979—)，男，河南人，高级工程师，主要研究方向为综合电子信息系统总体设计与系统集成。