

融合全局和局部特征并基于神经网络的表情识别

吴晶晶, 程武山

(上海工程技术大学 机械工程学院, 上海 201620)

摘要: 在表情中含有最多特征信息的是面部眉毛、眼睛和嘴巴这3个区域, 为充分利用这些特征, 减少图像中无用信息在识别过程中对计算机内存的占用, 提高人脸表情识别系统的准确率和速度, 首先采用 haar 和 adaboost 人脸检测算法, 对图像中的人脸进行识别, 获得人脸图像并提取眉毛、眼睛和嘴巴, 生成局部(眉毛、眼睛、嘴巴)二值化图, 利用 PCA 方法对人脸图像降维, 降维后的全局和局部灰度特征值组成一个列向量; 样本由表情数据库产生, 经过神经网络样本训练后, 进行表情识别; 结果表明, 该系统对人脸表情识别速度明显快于 Gabor 小波算法; 识别的准确率高于单独使用 PCA 算法和神经网络算法; 消耗内存比用 Gabor 小波算法少, 运行较流畅; 得出结论: 因为提取出包含表情特征信息集中区的眉毛、眼睛和嘴巴, 尽可能多地保留了这些局部特征信息, 因而提高了表情识别准确率, 同时, 采用 PCA 方法对原始图像进行降维处理, 有效的减少了信息冗余。

关键词: 表情识别; adaboost 人脸检测; PCA; BP 神经网络

Facial Expression Recognition by Fusion of Global and Local Features Based on Neural Networks

Wu Jingjing, Cheng Wushan

(Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: There is the most characteristic information in the regions of the eyebrows, eyes and mouth about facial expressions. In order to make full use of these features, reduce the amount of unavailable information in the image and occupation of the memory during the recognition process and improve the accuracy and speed of facial expression recognition. Firstly, Haar and AdaBoost face detection algorithms are used to recognize the human face in the image, get face images, and extract eyebrows, eyes and mouth. Generate the binaryzation of image of the eyebrow, eye and mouth. Getting the image of descending dimension by PCA algorithm and a column vector was composed of image of binaryzation and image of descending dimension. The samples are generated by an expression database and trained by neural network samples for facial expression recognition. The results show that the speed of facial expression recognition is faster than that of Gabor algorithm; The recognition accuracy is higher than that using the PCA algorithm and the neural network algorithm alone; The consumption of memory is less than that of the Gabor algorithm and the operation is smoother. The conclusion is that the local feature information is preserved as much as possible, because the eyebrows, eyes and mouths which contain the facial feature information are extracted, the accuracy of facial expression recognition is improved. At the same time, the PCA algorithm is used for the image of descending dimension and reduce the redundancy of information effectively.

Keywords: facial expression recognition; AdaBoost face detection; PCA; BP neural network

0 引言

在过去的二十多年里, 随着科技的飞速发展, 计算机的性能得到了显著提高, 随之而来的就是计算机应用相关技术的大繁荣。在这样的背景下, 人工智能和机器学习领域出现了很多具体的研究方向, 图像处理、计算机视觉和模式识别也各自应运而生^[1-3]。如何使计算机以人类的思维方式来与人类交流, 成为了研究人员非常感兴趣的课题, 这是一个涉及心理学、模式识别和计算机视觉等的交叉学科, 这方面的研究可以为用户提供高效、友好的人机交互界面和人机交互系统。人脸表情是人类用以表达情感, 传递内心世界和态度的重要途径, 人们可以运用表情表达自己的情绪, 也可以通过表情辨别他人的心理状况。心理学家认为, 人们之间的信息传递依赖于言词, 依赖于声音, 而源于脸部表情。

目前, 模式识别算法主要包含特征提取算法和分类算法, 特征的提取是模式识别中一个非常重要的环节, 图像具有数据多, 维数高, 富含大量冗余数据的特点, 要从一张静态图像中提取有用的信息, 并且对图像进行降维处理, 就要选取一种合适的特征提取算法。近期, 静态图像中表情特征提取算法主要有以下几种: Gabor 小波、Histogram of Oriented Gradient (HOG), SIFT, SURF, Wavelet, DCT、(Local Binary Pattern, LBP), PCA 和子空间等^[4-5]。在模式识别领域, LBP 是众多常用特征描述子之一, 它的优势是定义简单、扩展方便、对光照强度具有一定的鲁棒性, 运算速度比较快, 在表情识别实验中经常使用。特征的提取, 盲目减少指标会损失很多信息, 更重要的是在多数情况下, 许多变量之间可能存在相关性, 主成分分析 (PCA, Principal Component Analysis), 可以减少原始数据包含信息的损失, 以达到对图像的全面分析, 同时, 综合存在相关性的各个变量, 对原始数据进行降维。

原始图片经过初始处理、降维后, 即可进行下一步的分类, 在表情识别中, 常用的分类算法有: 决策树 (Decision Trees)、人工神经网络、KNN 算法 (K-Nearest Neighbour)、

收稿日期: 2017-10-10; 修回日期: 2017-10-31。

作者简介: 吴晶晶(1989-), 男, 安徽蚌埠人, 硕士研究生, 主要从事嵌入式和模式识别方向的研究。

支持向量机 (SVM)、朴素贝叶斯、Adaboosting 等^[6-7]。图像分类使用比较多的有朴素贝叶斯算法^[8-10], 它有着扎实的数学理论基础, 分类效率稳定, 所需要的估计参数很少, 尤其对部分数据的缺失有一定的鲁棒性, 算法比较简单。缺点是: 具有最小的误差概率, 需要确定先验概率。神经网络是另一种比较常见和应用在表情识别的算法, BP 算法有着很强的非线性映射能力, 拥有很强的并行处理能力, 有自学习的能力, 对噪声有很强的鲁棒性^[11-12]。

综合近期表情识别的发展, 对原始数据特征的提取和对特征数据的分类, 不是简单的都选择最优的算法就可以构建一个优秀的模式识别模型, 针对不同类型的特征数据有其最合适的分类方法。文中选择 PCA 特征提取方法与神经网络分类算法相互搭配^[13-15], 在特征提取前, 对原始图片“敏感”部分(眉毛、眼睛、嘴巴)识别并提取出, 以便在特征提取、组成时增加权重。

1 全局的 PCA 提取和局部特征的融合

1.1 人脸、眉毛、眼睛和嘴部的定位与保存

基于 haar 和 adaboost 人脸检测算法是一种能够有效识别出人脸和非人脸的算法, 利用 Open-cv 从一幅背景较复杂图片中识别出人脸, 调用分类器 haarcascade_frontalface_default.xml, 在正确识别出人脸的图像上, 在此基础上, 利用 haar 算法, 调用眼部(可以识别出眼睛和眉毛)和嘴巴分类器 haarcascade_eye.xml、haarcascade_mouth.xml, 识别并保存。

1.2 PCA 的特征提取与选择

特征提取与选择是模式识别中重要的一部分, 特征提取的品质, 直接影响后续分类的准确率和效率。PCA 算法对原始图片进行降维, 通过线性变换将原始数据变换为一组各维度线性无关的表示, 可用于提取数据的主要特征分量, 该变换使得原始数据由维数较少的“有效”特征数据来表示, 而不减少原始数据所包含的内在信息量, 减少冗余信息和综合具有相关性信息的, 在运算过程对内存的消耗降低。

在对图像提取特征前, 首先要对原始图像进行归一化处理, 如果原始图像为彩色, 应对其进行灰度处理, 灰度处理的方法很多, 本文采用的是加权平均法对图像进行灰度处理:

$$Gray(i, j) = 0.299 * R(i, j) + 0.578 * G(i, j) + 0.114 * B(i, j) \quad (1)$$

所有图像均归一化成 $w * h$ 大小, 本文取 $80 * 80$ 。

将 $80 * 80$ 的图像矩阵转化成一组列向量 χ_i , χ_i 含有 $K = 80 * 80$ 个元素, 设共有 70 张图片, \mathbf{X} 为 70 张图片组成的图像矩阵:

$$\mathbf{X} = [\chi_1, \chi_2, \dots, \chi_N] N = 70$$

求出样本图像集的平均图像向量 ψ 和每个样本与平均值的差值 φ_i :

$$\psi = \frac{1}{N} \sum_{i=1}^N \chi_i \quad (2)$$

$$\varphi_i = \chi_i - \psi \quad (3)$$

根据求出的 φ_i , 则可得总体的协方差矩阵 \mathbf{S} :

$$\mathbf{A} = (\chi_1 - \psi, \chi_2 - \psi, \dots, \chi_N - \psi) \quad (4)$$

$$\mathbf{S} = \frac{1}{N} \sum_{i=1}^N \varphi_i \varphi_i^T = \frac{1}{N} \mathbf{A} \mathbf{A}^T \quad (5)$$

求解矩阵 \mathbf{S} 的特征值和对应的特征向量, 求得的特征值为 λ_i , ($i = 1, 2, 3, \dots, N$), 按照由大到小的顺序排列, 然后求得正交归一化后与特征值对应的特征向量 μ_i :

$$\mu_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{A} \nu_i \quad (6)$$

取前 L 个特征向量构成“特征表情”空间, 即构成投影矩阵 $\mathbf{\Gamma} = (\mu_1, \mu_2, \mu_3, \dots, \mu_L)$, L 的取值可以根据特征值的累积贡献率选择:

$$\eta \leq \frac{\sum_{i=1}^L \lambda_i}{\sum_{i=1}^N \lambda_i} \quad (7)$$

一般情况下 $\eta = 0.9 \sim 0.99$ 。

主成分的主要用途之一就是能够产生尽可能少地向量维数取代原始数据所包含的有效信息, 关于 η 究竟要取多大, 既要能够使提取的主成分可以包含足够的有效信息, 这就要求 η 不能取值过小, 避免原始数据过分浓缩, 失去很多重要信息, 又要去除原始数据中冗余数据, 这就要求 η 不能取值过大, 避免不能有效对原始数据进行有效信息的提取。 η 取值直接关系到 PCA 主成分提取的效果, 因此要在实验中不断调整 η 值, 以期达到精度与效率的平衡点。

原来高维的图像数据 $K = 80 * 80$, 图像矩阵转化成一组列向量 χ_i , 经过投影矩阵 $\mathbf{\Gamma}$, 原来的 $K * 1$ 维列向量投影在 $\mathbf{\Gamma}$ 后, 降维到 $L * 1$ 维列向量 \mathbf{F} :

$$\mathbf{F} = \mathbf{\Gamma}^T \cdot \chi_i \quad (8)$$

这样得到降维后的 \mathbf{F} , 不仅尽量提取了有效的图像特征, 且很好保留了原图像矩阵的内部信息。

1.3 数据融合

通过上述对全局图像的降维处理和特征提取, 获得了一个 $L * 1$ 维列向量 \mathbf{F} 。在 1.1 中已经对眉毛、眼睛和嘴巴的提取和保存, 由于这些局部图像含有表情识别的信息最为丰富, 因此, 要尽可能保存局部图像的所有数据而不用像对全局图像的降维处理。

图 1 为摄像头动态捕捉到的图像, 图 2 分别面部、眉毛、眼睛和嘴巴, 由于眉毛和眼睛在一起能够含有更多的位置信息, 因此在提取时候将眉毛和眼睛结合在一起, 为了减少计算量, 同时获取眉毛、眼睛和嘴巴的目的, 主要是尽量保留它们轮廓和位置的信息, 比如嘴巴的开度, 因此, 先对其进行灰度处理, 再对其进行自适应二值化处理, 得到如图 2 的二值化图像。图 2 中的所有图像都要进行归一化处理, 以便后续特征的组合和分类。



图 1 原始图像

前面已经完成面部的特征提取, $L * 1$ 维列向量 \mathbf{F} , 其次眉毛、眼睛和嘴巴的二值化图像也已经获得, 假设眉毛和眼睛的图像归一化为 $20 * 20$ 的二值化图, 转化为 $400 * 1$ 的列向量



图 2 全局和局部图像

E , 同样, 嘴部转换成 400×1 的列向量 M 。将 F, E, M 组成一个列向量 R , 则:

$$R = (F^T, E^T, M^T)^T = (f_1, f_2, \dots, f_L, e_1, e_2, \dots, e_{400}, m_1, m_2, \dots, m_{400})^T \quad (9)$$

R 就是对全局和局部向量的提取、选择和融合。

2 神经网络分类器

BP 网络是一种含有多层神经网络的前馈神经网络, 它的结构包含输入层、隐含层和输出层, 只有前后两层之间的神经元相连, 学习过程正向传播, 建立在以梯度下降法的基础上, 反向传播误差, 是一种广泛使用的分类器。

图 3 为神经网络拓扑结构, BP 网络可以看作一个高度非线性的映射, 用公式表示为:

$$f(X) = YF: R^n \rightarrow R^m \quad (10)$$

其中: $x_i \in R^n$ 为输入集合, $y_i \in R^m$ 为输出集合。

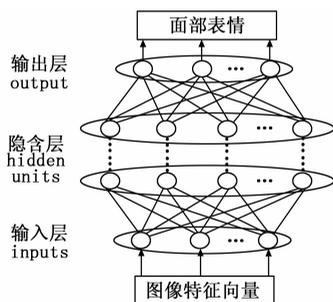


图 3 BP 神经网络拓扑结构

设该神经网络有 L 层 n 个神经结点, 转移函数为 Sigmoid 函数, 样本有 N 个 $(x_k, y_k) (k = 1, 2, \dots, N)$, 任意结点 i 的输出为 O_i , 网络的输出为 y_k , 节点 i 的输出为 O_{jk} , 第 l 层第 j 个神经元, 第 k 个样本, 结点 j 的输入为:

$$net_{jk}^l = \sum_j w_{ij}^l O_{jk}^{l-1} \quad (11)$$

O_{jk}^{l-1} 表示 $l-1$ 层, 当第 k 个样本输入时, 第 j 个单元结点输出为:

$$O_{jk}^{l-1} = f(net_{jk}^l) \quad (12)$$

使用平方型作为误差函数:

$$E_k = \frac{1}{2} \sum_i (y_{ji} - \hat{y}_{jk})^2 \quad (13)$$

\hat{y}_{jk} 为第 j 个单元的实际输出。总误差为:

$$E_k = \frac{1}{2N} \sum_{k=1} E_k \quad (14)$$

设:

$$\delta_{jk}^l = \frac{\partial E_k}{\partial net_{jk}^l} \quad (15)$$

$$\frac{\partial E_k}{\partial w_{ij}^l} = \frac{\partial E_k}{\partial net_{jk}^l} \frac{\partial net_{jk}^l}{\partial w_{ij}^l} = \frac{\partial E_k}{\partial net_{jk}^l} O_{jk}^{l-1} \quad (16)$$

结点 j 为输出单元, 则:

$$\delta_{jk}^l = \frac{\partial E_k}{\partial net_{jk}^l} = \frac{\partial E_k}{\partial y_{jk}} \frac{\partial y_{jk}}{\partial net_{jk}^l} = -(y_k - \hat{y}_k) f'(net_{jk}^l) \quad (17)$$

由上述推导可得出: 神经网络算法的一般步骤包含正向计算结点输出 O_{jk}^{l-1} , 输入 net_{jk}^l 实际输出 y_{jk} 的计算, 反向计算偏差 δ_{jk}^l , 在样本训练过程中不断修正权值, 以达到设定的误差值为止。经过数据分析和处理后的 F 作为神经网络的输入数据, 经过输入层、隐含层和输出层, 最终将 F 分类为对应的面部表情。

神经网络具有非常强的非线性映射能力, 可以任意精度无限逼近任何复杂的非线性曲线函数。对于一个复杂的输入输出系统, 样本数据越多则分类的效果越好, 然而付出的代价就是对数据收集、数据处理和网络的训练等难度会大大增加。在尽量少的训练样本和检测样本要求下, 可以预先对样本进行分类, 获得分布较为均衡的样本。对于神经网络在训练过程中, 选择合适的激活函数 (activation function) 是构建高效神经网络的关键之一。目前, 广泛使用的激活函数有 tanh 为双切正切函数, 取值范围为 $[-1, 1]$ 、sigmoid 为双切正切函数, 取值范围为 $[0, 1]$ 、ReLU 函数为大于 0 保留, 其余为 0, 本实验中由于在特征运算过程中, 特征相差不是特别大需要更加复杂精细的分类输出, sigmoid 更加适合。

3 实验分析

3.1 样本和输入输出

为了验证融合全局和局部特征并基于神经网络的表情识别方法的准确率、速度以及占用内存等性能指标, 可以和其他表情识别算法作比较。本实验将使用比较常用的表情数据库 JAFFE (Japanese Female Facial Expression), 从 JAFFE 选取具有代表性的 7 种表情, 分别为厌恶、生气、恐惧、惊讶、正常、难过和高兴, 部分原始图像如图 4, 在所有样本中, 一部分作为训练样本用于神经网络的训练, 这一部分的图像一定要尽量保证各种表情的数量大致一样多, 达到均衡分布的效果。另选一部分作为验证样本, 当神经网络训练完成后, 验证样本经过预处理, 作为输入量, 对实验结果进行验证。

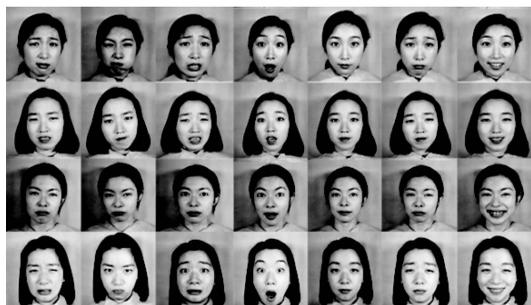


图 4 表情数据库

获得表情数据库中的表情后, 由于数据库中的图像大小不一样, 因此首先对图片进行归一化处理, 对全局和局部进行相应的特征提取和组合后, 获得神经网络的输入数据 F , 接下来是对神经网络的设计, 在第 1 章已经确定 F 为 $(L \times 1 + 400 \times 1 + 400 \times 1)$ 维列向量, 故神经网络输入层为 $L + 800$ 个神经元, 输出层为 7 种表情, 对 7 种表情采用 4 位二进制表示, 如表 1, 因此, 输出层有 4 个神经元, 构成 4 维输出量。

表 1 表情编码

序号	表情	输出编码
1	厌恶	0001
2	生气	0010
3	恐惧	0011
4	惊讶	0100
5	正常	0101
6	难过	0110
7	高兴	1001

3.2 神经网络参数选取

由于神经网络的输入和输出量要求位于 0 到 1 区间, 因此, 要对神经网络的输入样本进行归一化处理, MATLAB 中有 `mapminmax(x, y_min, y_max)` 专门用于数据归一化。由前面可知, 训练样本的输入量维数 I 为 $L + 800$, 输出量维数为 4, 可知输入神经元为 $L + 800$, 输出神经元为 4, 隐含层神经元 $H = 2I + 1$ 为 $2L + 1601$, 由于数据归一化处理在 $[0, 1]$ 之间, 故神经元传递函数采用 S 型函数 `tansing`, 输出层传递函数采用 S 型函数 `logsig`, 在样本预先处理完成后, 利用 Levenberg-Marquardt 算法对网络进行训练。

3.3 实验与分析

以 JAFFE 表情数据库对神经网络训练完成后, 首先选取数据库中被选作为样本的两张图像 (高兴和正常) 进行验证, 将经过 PCA 主成分提取和数据融合处理后的图像向量导入 MATLAB 作为输入量, 经过神经网络分类后, 结果如图 5 所示。

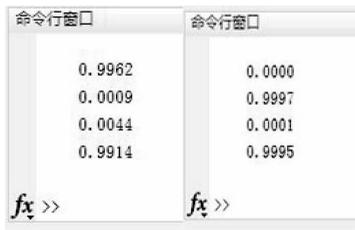


图 5 分类结果

两次分类结果分别为 (1001) 和 (0101), 对照表 5 可知, 其对应的分别为高兴和正常, 两次分类的结果都正确, 那么不仅要看看结果是否满足实验要求, 还要评价整个神经网络的性能, 查找 MATLAB 神经网络的一些参数, 如图 6 所示。

样本训练迭代到第 5 次时, 误差开始迅速下降, Epoch 表示经过 9 次迭代即结束, 达到设定误差要求, Performance: 误差达到 $5.77e-06$, Gradient: 梯度达到 $2.45e-05$, Time: 耗时 200 ms, 由此得出该神经网络性能优异。

4 结论

本文所使用的表情识别方法, 在 Matlab 和 C++ 平台中对推理过程进行训练与仿真, 根据测试结果和各项指标, 得出以下结论:

- 1) 提出的 PCA 算法可以有效地对原始数据降维, 去除无效的、具有相关性的信息, 降低了信息的冗余, 同时保留包含重要特征的信息, 提高了有效信息的权值;
- 2) 全局和局部信息的融合可以增加敏感部位在原始数据中的权值, 输入数据更好保留与表情相关的信息, 使分类

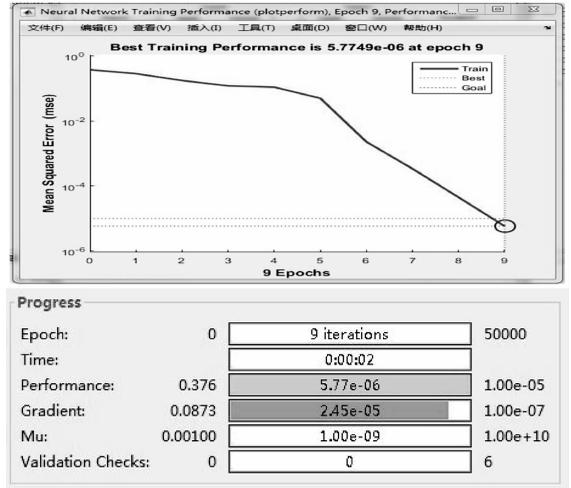


图 6 分类器性能指标

既能从全局角度综合判断, 又可以重点提高敏感部位信息的权重;

3) 提出的神经网络作为分类器的算法, 充分利用了 BP 神经网络所具有的非线性映射特性, 可以有效实现各种表情的分类。

参考文献:

- [1] 伊茹梦. 人工神经网络理论与应用 [J]. 科技经济导刊, 2015 (18): 59-60.
- [2] 赵春伟. 基于 PCA 与 LDA 的表情识别算法研究 [D]. 西安: 西安电子科技大学, 2014.
- [3] 张天. 面部表情识别的研究与集成实现 [D]. 重庆: 重庆邮电大学, 2016.
- [4] 李雅倩, 李颖杰, 李海滨, 等. 融合全局与局部多样性特征的人脸表情识别 [J]. 光学学报, 2014, 34 (5): 172-178.
- [5] 焦李成, 杨淑媛, 刘芳, 等. 神经网络七十年: 回顾与展望 [J]. 计算机学报, 2016 (8): 1697-1716.
- [6] 王文锋, 朱书华, 冯以浩, 等. 基于局部方向编码的遥感影像平行边缘识别 [J]. 光学学报, 2012, 32 (3): 141-147.
- [7] Gabriele D'Antona. Processing magnetic sensor array data for AC Current measurement in multiconductor systems [J]. IEEE Transactions on Instrumentation and Measurement, 2001, 50 (5): 2001.
- [8] Li Y. Design of intelligent monitoring unit for low voltage power distribution switch cabinet [D]. Chengdu University of Technology, 2010.
- [9] Clark F, et al. Wide-Base line stereo for Mars Rovers [A]. Proceedings of the 2003 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems [C]. 2003.
- [10] 冯伍, 张俊兰. 人工智能在医学上的应用 [J]. 电子设计工程, 2010 (1): 19-20.
- [11] 周晓彦, 郑文明, 辛明海, 等. 基于稀疏表示的 KCCA 方法及在表情识别中的应用 [J]. 模式识别与人工智能, 2013 (7): 660-666.
- [12] 史文樊. 改进的 PCA 与 BP 神经网络人脸识别方法研究 [D]. 秦皇岛: 燕山大学, 2013.
- [13] 陈北京, 孙星明, 王定成, 等. 基于彩色图像四元数表示的彩色人脸识别 [J]. 自动化学报, 2012 (11): 1815-1823.