

基于压缩感知的鲁棒性人脸表情识别^①

施徐敢^{1,2}, 张石清², 赵小明^{1,2}

¹(浙江理工大学 机械自动控制学院, 杭州 310018)

²(台州学院 图像处理与模式识别研究所, 临海 317000)

摘要: 为了有效提高噪声背景下的人脸表情识别性能, 提出一种基于压缩感知的鲁棒性人脸表情识别方法. 先通过对腐蚀的测试样本表情图像进行稀疏表示, 再利用压缩感知理论寻求其最稀疏的解, 然后采用求得的最稀疏解信息实现人脸表情的分类. 在标准的 Cohn-Kanade 表情数据库的实验测试结果表明, 该方法取得的人脸表情识别性能优于最近邻法、支持向量机以及最近邻子空间法. 可见, 该方法用于人脸表情识别, 识别效果较好, 鲁棒性较高.

关键词: 压缩感知; 稀疏表示; 表情识别; 鲁棒性; 腐蚀

Compressed Sensing-Based Robust Facial Expression Recognition

SHI Xu-Gan^{1,2}, ZHANG Shi-Qing², ZHAO Xiao-Ming^{1,2}

¹(School of Automatic Control of Mechanical, Zhejiang Sci-Tech University, Hangzhou 310018, China)

²(Institute of Image Processing and Pattern Recognition, Taizhou University, Linhai 317000, China)

Abstract: In order to effectively improve the performance of facial expression recognition under the noisy background, a method of robust facial expression recognition based on compressed sensing is proposed. Firstly, the sparse representation of corrupted expression images of the identified test sample is sought, then the compressed sensing theory is used to solve its sparsest solution. Finally, according to the sparsest solution, facial expression classification is performed. Experimental results on benchmarking Cohn-Kanade database show that facial expression performance obtained by this method is better than the nearest neighbor (NN), support vector machine (SVM) and the nearest subspace (NS). Therefore, the proposed method shows both good recognition performance and high robustness on facial expression recognition tasks.

Key words: compressed sensing; sparse representation; expression recognition; robustness; corruption

人脸表情是人们观察情感的重要标志, 如何使得机器能够认识人脸表情, 是一个既实用又有趣的研究方向. 如何让机器自动、高效、准确地来识别人类的情绪状态, 比如高兴、悲伤、愤怒、恐惧等, 即所谓的“人脸表情识别”^[1]方面的研究, 是当前信号处理、模式识别、计算机视觉等领域的热点研究课题. 该研究在智能人机交互、人工智能等方面有着重要的应用价值.

尽管人脸表情识别经过了多年的发展, 已经取得了较多的研究成果, 但现有的人脸表情识别研究^[2-10]大多没有考虑表情图像受到噪声的影响. 在自然环境

中, 人脸表情图像的获取、传输和存贮过程中常常也会受到各种噪声(如姿态、光照、腐蚀、遮挡等)的干扰而使图像降质, 从而导致人脸表情识别的性能会随之下降. 因此, 如何提高人脸表情识别的鲁棒性仍然是一个亟需解决的问题.

压缩感知 (Compressed sensing) 或压缩采样 (Compressive sampling)^[11,12], 是近年来新出现的一种信号采样理论, 它可以在远小于 Nyquist 采样率的条件下获得信号的离散样本, 然后通过非线性重建无失真的完美信号. 压缩感知理论指出, 采样速率由信号中

① 基金项目: 国家自然科学基金(61272261, 61203257)

收稿时间: 2014-05-19; 收到修改稿时间: 2014-06-10

的内容和结构所决定, 而不再决定于信号的带宽. 目前, 压缩感知理论在图像处理^[13]、人脸识别^[14]、视频追踪^[15]等领域受到了研究者的高度关注, 并表现出了极其强大的生命力, 但在人脸表情识别领域, 尤其针对鲁棒性的人脸表情识别问题, 国内外相关的文献报道甚少.

压缩感知理论研究的初衷主要用于信号的压缩和表示, 但其最稀疏的表示具有很好的判别性. 本文利用压缩感知理论中的稀疏表示分类 (Sparse Representation-based Classification, SRC) 思想^[14], 提出一种基于压缩感知的鲁棒性人脸表情识别方法. 先通过对腐蚀的测试样本表情图像进行稀疏表示, 再利用压缩感知理论寻求其最稀疏的解, 然后采用求得的最稀疏解信息实现人脸表情的分类. 在标准的 Cohn-Kanade 表情数据库^[16]的实验结果表明了该方法的可行性.

1 基于压缩感知的稀疏表示分类

设 $A=[A_1, A_2, \dots, A_c]$ 是一组训练样本集, 总数量为 n , 其中 $A_i = [v_{i,1} \ v_{i,2} \ \dots \ v_{i,n_i}] \in R^{m \times n_i}$ 为第 i 类训练样本, $y \in R^m$ 是第 i 类的测试样本, 它可以由 A_i 线性表示为:

$$y = a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \dots + a_{i,n_i}v_{i,n_i} = A_i x_0 \quad (1)$$

然而在实际情况中, 由于测试样本的类别一般是未知的, 所以式 1 可以写为

$$y = Ax_0 \quad (2)$$

式中, $x_0 = [0, \dots, 0, a_{i,1}, a_{i,2}, \dots, a_{i,n_i}, 0, \dots, 0]^T \in R^n$.

由矩阵原来可知, $m > n$ 时, 矩阵(2)有唯一解; 但是在大多数情况下, $m \leq n$, 此时矩阵(2)有无穷多个解. 为了使测试样本能够用自身所在类的训练样本进行线性表示, 这样的话系数向量 x_0 中的非零向量应该尽可能少些. 所以对矩阵(2)求解可转换为对矩阵(3)进行求解

$$\hat{x}_0 = \arg \min_x \|x\|_0 \text{ subject to } Ax = y \quad (3)$$

式中, $\|\cdot\|_0$ 表示 l^0 范数, 它的作用是计算向量中非零元素的个数. 但是, 式(3)的求解非常困难, 这是个 NP 难题.

由压缩感知理论可知: 当所求的系数足够稀疏时, 可以把最小化 l^1 范数的 NP 难题转化成最小化 l^1 范数问题来求解. 因此, 把式(3)改写为:

$$\hat{x}_1 = \arg \min_x \|x\|_1 \text{ subject to } Ax = y \quad (4)$$

然而在实际情况中, 获得的数据中经常含有噪声, 因此 y 很难由 A 进行比较准确的线性表示, 因此, 把式(4)改写为

$$\hat{x}_1 = \arg \min_x \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \varepsilon \quad (5)$$

式(5)可以通过以下的式(6)来求解

$$\hat{x} = \arg \min_x \|x\|_1 + \lambda \|Ax - y\|_2^2 \quad (6)$$

SRC 算法可归纳如下:

- 1) 对训练样本集 A 中的每一个列向量进行归一化.
- 2) 求解最小化 l^1 范数问题:

$$\hat{x}_1 = \arg \min_x \|x\|_1 \text{ subject to } Ax = y$$

或者求解

$$\hat{x}_1 = \arg \min_x \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \varepsilon$$

- 3) 计算残差

$$r_i(y) = \|y - A\delta_i(x_1)\|_2, \quad i = 1, \dots, c, \text{ 其中,}$$

$$\delta_i(x_1) = [0, \dots, 0, a_{i,1}, a_{i,2}, \dots, a_{i,n_i}, 0, \dots, 0]^T$$

- 4) $l(y) = \arg \min_i r_i(y)$. $l(y)$ 是 y 的标记.

2 实验结果及分析

本文采用标准的 Cohn-Kanade^[16]数据库进行实验. 通过对 Cohn-Kanade 数据库的原始图像采样得到 32×32 像素图像, 然后分别采用稀疏表示分类方法 SRC、最近邻法 (Nearest neighbor, NN), 支持向量机 (Support Vector Machine, SVM), 以及近年来流行的最近邻子空间法 (Nearest subspace, NS)^[17] 进行人脸表情识别实验, 并比较它们的性能.

除了 SRC 方法, 使用的其它分类方法的基本思想表述如下: 最近邻法 (NN) 是基于样本学习的 K 近邻分类器 (KNN), 当 $K=1$ 时的一种情况. 支持向量机 (SVM) 是一种基于统计学习理论的分类器. 本文 SVM 采用“一对一”多类分类算法, 核函数为径向基函数, 并对核函数参数值进行最优化, 即在训练样本数据上使用交叉验证方法实现. 最近邻子空间法 (NS) 是一种

基于信号重构的无参数分类器,其分类思想是将测试样本表示为各类所有训练样本的线性组合,从中选择最优解来进行分类。

2.1 表情数据库

Cohn-Kanade 数据库含有 210 个对象的大约 2000 个左右的具有充足的正面光照的灰度图像序列。图像序列的分辨率都是 640×490 。该数据库总共含有七种基本的表情,如生气、高兴、悲伤、惊奇、讨厌、害怕以及中性,如图 1 所示。我们从数据库中选用来自 96 个对象的 320 图像序列用于实验测试。选择图像序列的标准是能够标记出除中性之外的六种表情。然后对每个选择的图像序列中提取出一帧中性表情图像以及一帧含六种表情之一的图像。最后我们提取出包括七种表情的 470 幅图像,其中生气 32 个,高兴 100 个,悲伤 55 个,惊奇 75 个,害怕 47 个,讨厌 45 个和中性 116 个。



图 1 Cohn-Kanade 数据库 7 种表情图像样本示例

2.2 无腐蚀的人脸表情识别实验

在该实验中,直接使用 32×32 像素大小的图像样本用于表情识别,图像中不存在任何腐蚀现象。表 1 列出了 SRC、NN、SVM 和 NS 四种不同方法所取得的人脸表情识别性能。由表 1 可知,在无任何腐蚀图像的条件下,稀疏表示分类方法 SRC 取得的人脸表情识别性能最好,达到 94.76% 的识别率。这说明了 SRC 用于人脸表情识别具有优越的分类性能。

表 1 不同方法对 32×32 像素图像所取得的识别性能比较

方法	NN(%)	SVM(%)	NS(%)	SRC(%)
正确识别率	92.29	93.80	92.74	94.76

为了进一步给出七种表情中不同表情的具体识别性能,表 2 给出了在 Cohn-Kanade 数据库上 SRC 方法采用 32×32 像素所取得的不同表情的识别结果。从表 2 的实验结果可见,在 Cohn-Kanade 数据库上七种表情中大部分表情的正确识别率达到了 100%。

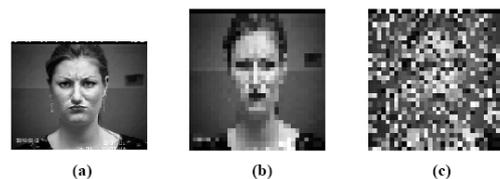
表 2 Cohn-Kanade 数据库上 SRC 采用 32×32 像素取得的不同表情的识别结果

表情类型	生气 (%)	高兴 (%)	悲伤 (%)	惊奇 (%)	讨厌 (%)	害怕 (%)	中性 (%)
生气	90	10	0	0	0	0	0
高兴	0	100	0	0	0	0	0
悲伤	0	0	90	0	10	0	0
惊奇	0	0	0	100	0	0	0
讨厌	0	0	0	0	100	0	0
害怕	0	0	0	0	0	100	0
中性	0	0	6.67	0	3.33	6.67	83.33

2.3 有腐蚀的人脸表情识别实验

为了检验 SRC 的鲁棒性人脸表情识别性能,对 32×32 像素大小的测试图像随机添加像素腐蚀(Pixel Corruption)。随机添加像素腐蚀就是从测试图像中随机选择一定比例的像素,采用范围 $[0, M_i]$ 之内的随机值进行替代,其中 M_i 表示第 i 个测试图像的最大像素值。实验中,像素腐蚀比例从 0% 到 90%,依次递增 10%。图 2 展示了 Cohn-Kanade 数据库中一副原始图像从采样到腐蚀的过程,其中图(a)为原始 640×490 像素的图像,图(b)为采样之后的 32×32 像素的图像,图(c)对 32×32 像素图像添加 50% 的腐蚀比例之后的图像。

图 3 列出了 NN、SVM、NS 和 SRC 四种方法在 Cohn-Kanade 数据库上随机添加像素腐蚀比例从 0% 到 90% 取得的识别结果。由图 3 实验结果可见,随着图像腐蚀比例的增大,图像越来越模糊,人脸表情识别率也随之下降。在图像腐蚀比例由 0% 增长到 30% 为止, SRC 的正确识别率下降速度缓慢,而其他三种方法的识别率下降非常快。随之腐蚀比例的不断增大(30% 至 90%),各种方法的识别率都一致下降,但是 SRC 方法的识别率平均超过其它三种方法 10% 以上。显然,我们看到了 SRC 方法在处理人脸表情问题上有着良好的鲁棒性。这主要是 SRC 方法提取了信号的稀疏结构,并利用 11 范数来作为来求解信号的稀疏表示系数。由于采用正则化技术, SRC 的稀疏表示系数具有非常稳定的数值解。



(a)原始图像(b)缩放到 32×32 像素的图像(c)添加 50% 的腐蚀比例之后的图像

图 2 Cohn-Kanade 数据库中的一幅像素腐蚀图像样本

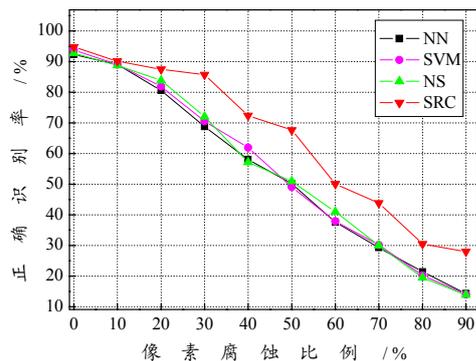


图3 不同方法对测试图像添加不同像素腐蚀比例取得的结果

3 结论

本文通过考虑测试图像是否存在像素腐蚀的现象, 对于基于压缩感知理论的稀疏表示分类方法 SRC 的鲁棒性人脸表情识别性能进行了探讨。在无任何像素腐蚀的人脸表情识别实验中, SRC 取得的人脸表情识别性能比其他方法高出 2% 左右, 而在有像素腐蚀图像的人脸表情识别实验中, SRC 展示出了良好的鲁棒性性能, 尤其在像素腐蚀比例 30% 至 90% 之间, SRC 比其他方法的识别率平均高出 10% 以上。这表明本文采用的基于压缩感知理论的稀疏表示分类方法 SRC 用于鲁棒性人脸表情识别时, 拥有良好的分类性能和鲁棒性。

参考文献

- 1 Tian Y, Kanade T, Cohn JF. Facial expression recognition. *Handbook of Face Recognition*, 2011: 487–519.
- 2 刘晓旻, 章毓晋. 基于 Gabor 直方图特征和 MVBoost 的人脸表情识别. *计算机研究与发展*, 2007, 44(7): 1089–1096.
- 3 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. *自动化学报*, 2012, 37(12): 1455–1463.
- 4 易积政, 毛峡, 薛雨丽. 基于特征点矢量与纹理形变能量参数融合的人脸表情识别. *电子与信息学报*, 2013, 35(10): 2403–2410.
- 5 朱晓明, 姚明海. 基于局部二元模式的人脸表情识别. *计算机系统应用*, 2011, 20(6): 151–154.
- 6 Aleksic PS, Katsaggelos AK. Automatic facial expression recognition using facial animation parameters and

multistream HMMs. *IEEE Trans. on Information Forensics and Security*, 2006, 1(1): 3–11.

- 7 Zheng W, Zhou X, Zou C, et al. Facial expression recognition using kernel canonical correlation analysis (KCCA). *IEEE Trans. on Neural Networks*, 2006, 17(1): 233–238.
- 8 Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 915–928.
- 9 Zhao X, Zhang S. Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding. *EURASIP Journal on Advances in Signal Processing*, 2012, (1): 20.
- 10 Yurtkan K, Demirel H. Feature selection for improved 3D facial expression recognition. *Pattern Recognition Letters*, 2014, 38: 26–33.
- 11 Candes EJ, Wakin MB. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 2008, 25(2): 21–30.
- 12 Donoho DL. Compressed sensing. *IEEE Trans. on Information Theory*, 2006, 52(4): 1289–1306.
- 13 Yang J, Wright J, Huang TS, et al. Image super-resolution via sparse representation. *IEEE Trans. on Image Processing*, 2010, 19(11): 2861–2873.
- 14 Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210–227.
- 15 Mei X, Ling H. Robust Visual Tracking and Vehicle Classification via Sparse Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2011, 33(11): 2259–2272.
- 16 Kanade T, Tian Y, Cohn J. Comprehensive database for facial expression analysis. *International Conference on Face and Gesture Recognition*. Grenoble, France. 2000. 46–53.
- 17 Lee KC, Ho J, Kriegman DJ. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005, 27(5): 684–698.