

# 人脸表情识别报告

无46 黄秀峰 2014011193

无46 严靖凯 2014011192

无46 王启睿 2014011179

2017 年 6 月 29 日

## 1 概述

对于人脸表情识别问题，目前学术界已有了一定的研究，并且当前在该问题上有许多算法比赛，吸引着众多科研人员参与。比较典型的包括Kaggle的FERC，SSPNET的FERA，ACM的EmotiW，等等。这些比赛采用特定的数据集，有一些由实验室环境下采集的数据为主，另一些直接来自于实际图片或视频。

在本报告中，我们将首先说明我们选用的方法及原因，再对算法实现进行介绍，最后对判别结果进行分析，并指出可能的未来改进方向。

## 2 选用方法的确定

本节中，我们将介绍我们曾着重考虑过的两种方法以及最终选用方法的确定。其中一是采用WLD + HOG特征的传统模式识别方法，二是采用神经网络的特征提取识别方法。

### 2.1 两种思路

与目前大多数机器学习问题一样，对与人脸表情识别问题的处理方法，大体可分为传统模式识别和神经网络两类。其中，传统模式识别方式的代表性文章有 [1, 2, 3, 4]等，其中大多数采用SVM或决策树进行判决，神经网络方式的代表性文章有 [5, 6]等。传统方法往往训练和运行速度很快，部分方法甚至可以应用于实时视频流作为输入等场景，对于质量高（清晰、光照合适、正脸）的图片有很高的识别正确率；而神经网络方法对于很低的图片分辨率、不理想的光照、侧脸，或视频片段等情形的处理效果较好。

在我们本次实验任务中，我们面对两个主要挑战：一是需要对自采数据集进行判决，其中的图片与实验室中标准数据有很大区别；二是本次提供的训练图片数很少，只有521张。前者促使我们考虑神经网络的方法，但后者使得神经网络的训练条件不足。相比之下，传统方法在小数据样本情形下通常有更好的效果。



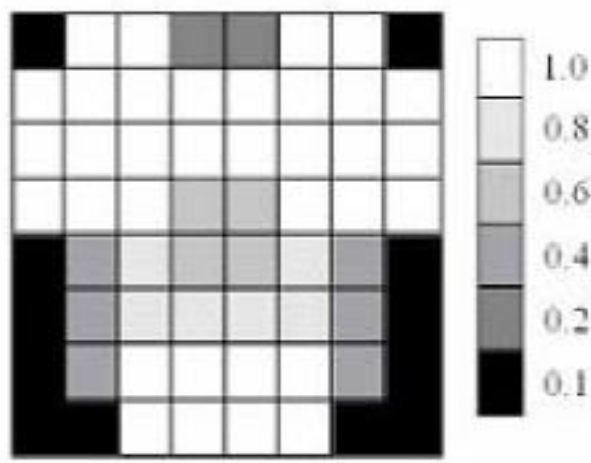


图 2: WLD特征的权值选取

### 3.3 HOG特征提取

方向梯度直方图也是图像特征提取中常用的特征。它将图像分为若干个tile，在每个tile中统计各个方向分量的分布直方图，在由若干临近的tile组成的block中对各个tile的直方图进行归一化，然后将各个tile的直方图拼接成一个向量作为该block的特征向量，最后将全图各个block的特征向量拼接起来作为全图的HOG特征。

在划分tile和block时，我们选取的参数是每个tile大小为 $8 \times 8$ ，每个block由 $2 \times 2$ 个tile组成，而相邻两个block重叠一半面积。

### 3.4 最近邻判定

在以上两种特征被提取出之后，将其拼接成一个向量，从而每张图片对应于一个特征向量。在训练时，记录下各个向量和它们相应的标签（表情）。在测试时，对输入的图片重复上述特征提取过程，得到一个特征向量。从训练的数据中找到与输入图片特征向量最近的向量，输出相应的标签作为识别结果。

注意到使用的两种特征都是统计意义下的直方图，而卡方距离常常用于衡量两个统计分布列间的差异性。因此，采用卡方距离作为衡量测试时输入图片特征向量与现有向量的距离的标准。

## 4 实验结果

### 4.1 输出样例与准确率

我们从训练集中随机均匀地选取100张图片用于测试正确率(其余留作训练数据)，正确率为83%。

以下展示判决程序运行的样例结果。这里我们将输出形式改为对单个图片进行判决，以便展示结果。如图3所示。可见，该图判决为fe(fear)类，这是一次正确的判决。

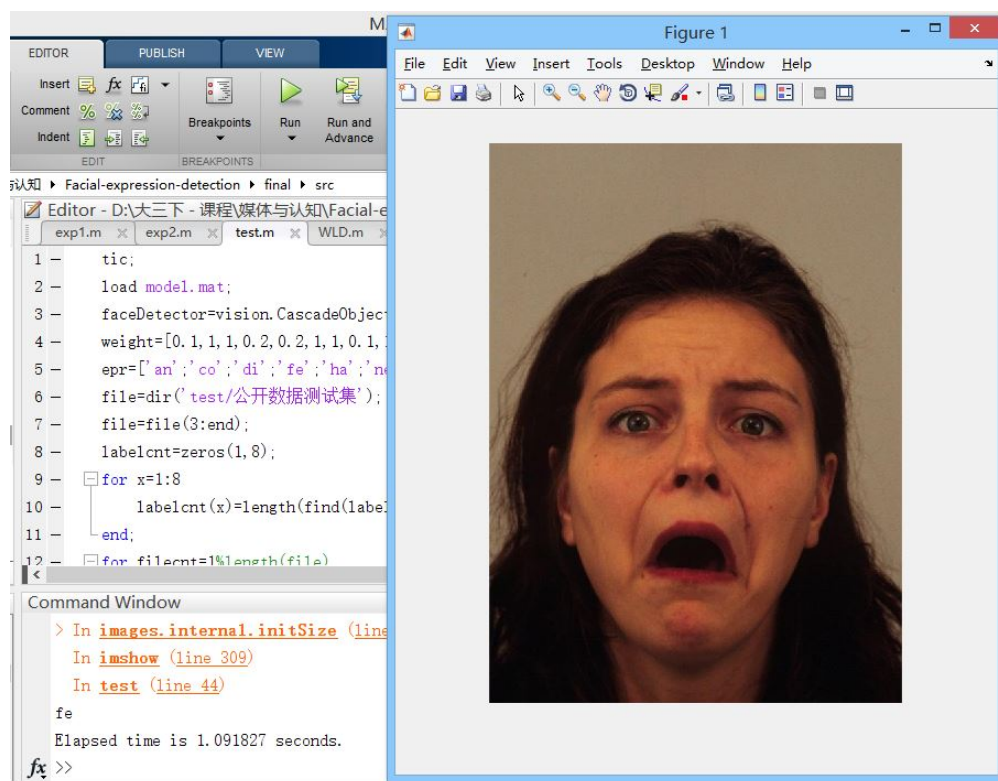


图 3: 一次运行的输出结果示例。

## 4.2 错判样例与分析

这里，我们选取一些错判的样例进行简要分析。这里考察的是测试数据集，由于该数据集并未提供标注，我们判断“错判”的依据是将判决的结果与人直接判决的结果进行比较。我们只是浏览了测试数据集，随机地找出一些比较典型且确定的错判样例，并进行分析。尽管这种方法带有较强的主观性，而且人直接判决的结果未必准确，但这不失为粗略评估算法的一种方法。我们希望通过观察典型的错判样例，寻找该方法的主要缺点，以及未来的改进方向。

对于图4，本判决程序将其判为happy，而人眼观察认为可能是contempt。可以发现，这幅图的人脸中两侧均上扬的嘴角和眼部与happy的图片确实有比较大的相似之处。对于图5，本判决程序将其判为neutral，而人眼观察认为可能是sad。由于训练中大多数sad类图片并没有捂脸的动作，而且情绪体现以面部动作为主，而非眼泪，因此无法识别为sad也是可以理解的。

因此可见，发生的错判一方面主要来源于测试数据与训练样本间特征的不一致性，导致模型不能完全捕捉各类的特点，另一方面主要来源于一些图片自身属性就相对模糊，即使是人也较难判断，需要借助环境信息来确认。

## 5 总结

对于本实验，我们采用WLD+HOG特征提取及最近邻判别的方法完成了判决的过程。得到的准确率属于中等，还有较大的改进空间。我们目前看来，主要的改进方式包括以下几点：

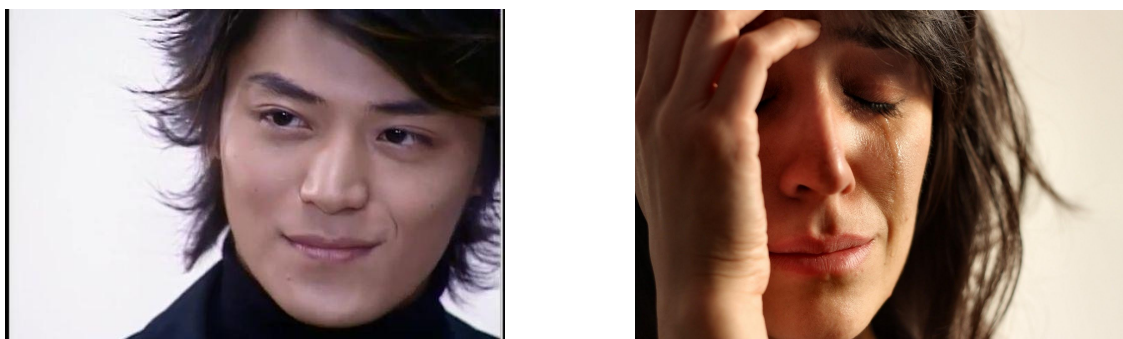


图 4: 可能发生错判的图片, 判决为happy, 人眼观 图 5: 可能发生错判的图片, 判决为neutral, 人眼  
察认为可能是contempt。 观察认为可能是sad。

- 增加模型训练集的大小, 用更全面、更具有代表性的样本进行训练。对于我们采用的算法, 每张图片大概能得到一万维的特征向量, 数据数量增加model文件会特别大, 因为我们也没有选用更多的数据集, 所以就没有遇到这方面的问题。但如果数据数量增大后, 我的想法是, 规定每个表情的图片数量 (比如100张), 如果图片数量过多, 就采用类似Huffman树合并的思想, 每个图片抽象为一个点, 权重为1, 然后合并距离最近的点, 权重相加, 新的点两合并的两点的质心 (权重相当于质量), 不断合并直到图片数量减小到规定的值即可。
- 尝试采用其他图像特征。我们曾尝试过采用LBP特征, 但效果并不好, 因此最后将该部分舍弃。如果未来有更加合适的图像特征提取方式被提出, 可以考虑采用。
- 当训练样本足够大时, 可以考虑重新采用神经网络训练的方法, 应该可以显著提高准确率。

同时, 我们感到, 表情识别这项工作的难度在众多模式识别问题中是较大的, 因为表情与情绪很多时候仅仅通过一幅图片很难判断, 而要通过人的行为、环境等信息综合判断。从这方面看, 对视频进行表情识别可能会获得更高的准确率, 因为有更多的信息可以利用。

总之, 经过本次实验, 我们对人脸表情识别这一当前热门问题有了更加深入的了解, 同时也动手实现了图像特征提取、分类、神经网络等多种机器学习的常用方法, 能力得到了极大的锻炼。

最后, 衷心感谢老师和助教一学期以来的付出!

## 参考文献

- [1] SL Happy and Aurobinda Routray. Automatic facial expression recognition using features of salient facial patches. *IEEE transactions on Affective Computing*, 6(1):1–12, 2015.
- [2] Rahul Islam, Karan Ahuja, Sandip Karmakar, and Ferdous Barbhuiya. Sention: A framework for sensing facial expressions. *arXiv preprint arXiv:1608.04489*, 2016.
- [3] Xiaohua Wang, Chao Jin, Wei Liu, Min Hu, Liangfeng Xu, and Fuji Ren. Feature fusion of hog and wld for facial expression recognition. In *System Integration (SII), 2013 IEEE/SICE International Symposium on*, pages 227–232. IEEE, 2013.

- [4] Fatima Zahra Salmam, Abdellah Madani, and Mohamed Kissi. Facial expression recognition using decision trees. In *Computer Graphics, Imaging and Visualization (CGiV), 2016 13th International Conference on*, pages 125–130. IEEE, 2016.
- [5] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *ACM International Conference on Multimodal Interaction (ICMI)*, 2016.
- [6] Pooya Khorrami, Thomas Paine, and Thomas Huang. Do deep neural networks learn facial action units when doing expression recognition? In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 19–27, 2015.