

基于特征融合的人脸表情识别研究

高理想¹,高磊^{2,3}

(1.天津云科世讯科技有限公司,天津 300401;
2.中国农业银行天津研发部,天津 300000;
3.河北工业大学人工智能与数据科学学院,天津 300401)

摘要:针对现有的人脸表情识别方法易受到光照、噪声等因素的影响导致识别率较低的问题,本文提出了一种基于特征融合的人脸表情识别方法。从两方面对表情信息进行描述,通过局部二值模式(Local Binary Pattern,LBP)和词袋模型(Bag Of Words,BOW)分别对表情图像进行纹理特征和语义特征提取,然后将两种特征进行线性融合,最后使用支持向量机(Support Vector Machine,SVM)进行表情分类识别。本文方法在JAFPE和CK+表情数据集上分别取得了98.76%和97.58%的识别率,验证了所提出方法的有效性。

关键词:人脸表情识别;LBP;BOW;特征融合;SVM

中图分类号: TP391 **文献标识码:** A

Research on Facial Expression Recognition Based on Feature Fusion

GAO Lixiang¹,GAO Lei^{2,3}

(1.Tianjin YunkeShixun Technology Co.,Ltd.,Tianjin 300401,China;
2.Tianjin Research and Development Department,Agricultural Bank of China,Tianjin 300000,China;
3.School of Artificial Intelligence,Hebei University of Technology,Tianjin 300401,China)

Abstract:This paper proposes a new facial expression recognition method based on feature fusion to solve the problem that the existing facial expression recognition methods are vulnerable to illumination,noise and other factors.The facial expression information is described from two aspects.Local Binary Pattern (LBP) and Bag of Words (BOW) are used to extract texture features and semantic features from facial expression images respectively and then are fused linearly.Finally,the expression is classified and recognized by Support Vector Machine (SVM).The proposed method achieves 98.76% and 97.58% recognition rates on JAFPE and CK + expression datasets respectively,which verifies the effectiveness of the proposed method.

Keywords:facial expression recognition;LBP;BOW;feature fusion;SVM

1 引言(Introduction)

人脸表情可以直观地反映人类的内心活动,在人与人的交流中有着重要的作用,通过表情可以很好地传达意图,听者更好地理解内容。随着人工智能研究的发展,人脸表情识别在教育^[1]、医疗^[2]、游戏^[3]和辅助驾驶^[4]等领域展现出了广阔的应用前景,逐渐成了计算机视觉领域的研究热点。目前,常用的表情识别方法有基于纹理特征的方法、基于几何特征的方法和基于语义特征的方法^[5-7]。其中纹理特征是对面部的纹理信息进行描述,如皱纹和皱眉等;几何特征是对五官的形变量和面部关键点的位移量进行描述;语义特征是通过底层视觉特征的语义分析,将人脸表情的底层视觉信息转化为高层的语义描述。常用的纹理特征提取算法有LBP^[8]、Gabor滤波器^[9]、LDN^[10]等,纹理特征可以提取到丰富的面部

纹理信息,但易受到光照和噪声等因素的影响。Mistry等^[11]使用LBP算法对人脸表情图像进行纹理特征提取。Gu等^[12]使用Gabor滤波器对人脸表情图像进行纹理特征提取,并对特征进行径向编码,形成鲁棒性强的局部纹理特征。几何特征提取先要对人脸各部位进行特征点定位,其对外部环境因素的影响具有较强的鲁棒性,但提取效果依赖于面部特征点定位的准确性。胡步发等^[13]提出了一种基于语义特征的人脸表情识别方法,通过层次分析法对人脸表情图像进行高层语义描述,建立语义特征向量,该方法在日本女性人脸表情数据库JAFPE中取得了较好的识别效果。Harrati等^[14]使用词袋模型对人脸表情进行语义特征提取,通过对视觉特征进行聚类,形成多个聚簇中心,然后通过统计得到特征频率直方图,将其作为特征向量输入到SVM中进行表情分类。语义特征能够

对图像的高层信息进行语义描述，但是容易忽略低层的细节信息。

不同的特征提取方法有各自的优势，但同时也存在着不足，将特征进行融合可以有效地将每种特征的优缺点进行互补，因此本文提出了一种基于特征融合的人脸表情识别方法。先对表情图像进行纹理特征和语义特征提取，然后对两种特征进行线性融合，最后将融合的特征输入SVM分类器进行表情分类识别。实验结果表明，该方法能够对人脸表情信息进行充分的提取，特征鲁棒性强，能够对面表情进行有效地识别。

2 相关算法描述(Relevant algorithms description)

2.1 LBP算法

LBP算法是一种局部纹理特征提取方法，由于其计算复杂度低并且特征提取效果较好，受到了诸多研究者的青睐。LBP算子定义在3×3的窗口内，该算子以中心像素值为阈值，将相邻的8个像素的灰度值与其进行比较，若周围的像素大于中心像素值，则该位置被标记为1，反之为0，在3×3的邻域内，中心像素点周围的8个像素点组成一个8位二进制数，这个二进制数就作为该邻域的LBP编码，将此二进制数转换为十进制，得到的数值即为LBP值，用来表征该区域的纹理信息。LBP算法的计算公式如式(1)和式(2)所示：

$$LBP(g_c) = \sum_{i=0}^{p-1} s(g_i - g_c) 2^i \tag{1}$$

$$s(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{else} \end{cases} \tag{2}$$

其中， p 表示周围像素点的个数， g_c 表示中心像素点的灰度值， g_i 表示周围像素点的灰度值， $s(x)$ 为符号函数，LBP算法的编码过程如图1所示。

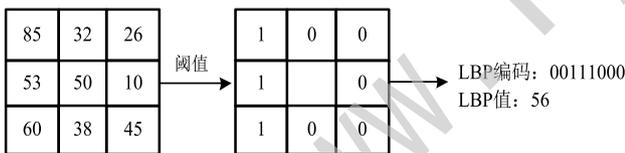


图1 LBP编码过程

Fig.1 LBP encoding process

将图像中每个像素点都经过LBP算法处理之后，即可得到整幅图像的LBP编码值，形成LBP特征图，如图2所示为LBP算法处理效果图。将LBP特征图中的所有编码值进行统计，即可得到LBP统计直方图，形成了对图像的纹理特征描述的特征向量。



(a)原图像 (b)LBP特征图

图2 LBP算法处理效果图

Fig.2 LBP algorithm processing effect

2.2 词袋模型BOW

词袋模型(Bag Of Words, BOW)最初应用于文本处理领域，解决文档的抽象表示和分类识别问题。Zhu等^[15]将其引入到了计算机视觉领域用于图像的表达，词袋模型在文本和图像处理上的对应关系如图3所示。在文本处理过程中，词袋模型将单词进行聚类组合形成词典，因此在图像处理中将局部特征所代表的视觉单词进行集合即可组成一部视觉词典，词典中每个视觉词汇及其出现的频率可以代表图像的视觉特征。

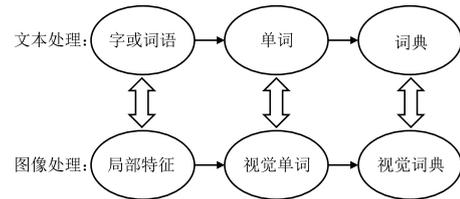


图3 文本和图像的词袋模型对比图

Fig.3 Comparison of BOW models between text and image

使用词袋模型对人脸表情图像进行特征提取的一般步骤为：

- (1)使用尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)^[16]算法提取每幅人脸表情图像的局部特征；
- (2)使用k-means算法对SIFT特征进行聚类，得到的聚类中心为 k 个，也就是这 k 个视觉单词构成了表情图像的视觉词典；
- (3)对每幅表情图像中视觉单词出现的频率进行统计，最终将人脸表情图像表示为视觉单词频率直方图的形式，即该幅表情图像的BOW特征。

由于在不同的表情图像中，眼睛、鼻子和嘴巴这三个部位的差异较大，表明这三个区域的信息对于表情的表征贡献较大^[17]，因此，本文选取眼睛、鼻子、嘴巴这三个语义信息较为集中的区域作为特征提取的关键区域，使用词袋模型对这三个关键区域进行语义特征提取。为便于提取特征，本文将三个关键区域都归一化为 41×130 的图像分块。分别对三个区域提取SIFT特征，然后使用k-means算法对SIFT特征聚类，即可得到视觉词典，聚类过程为：

首先从 n 个特征中随机取 k 个元素作为 k 个簇的初始聚类中心，根据其余元素到 k 个簇中心的距离分别划分到距离最近的簇。假设将样本聚类成的 k 簇为 (C_1, C_2, \dots, C_k) ，则每个样本与聚类中心之间的最小化平方误差 E 为：

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - u_i\|^2 \tag{3}$$

其中， u_i 是簇 C_i 的均值向量，也称其为质心，表达式为：

$$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \tag{4}$$

然后计算簇中所有元素的均值向量，得到新的 k 个聚类中心，不停重复上述过程直到聚类结果不再变化。通过词汇聚类得到词典后，使用欧式距离来计算每个采样点的SIFT特征向量与词典中每个词汇距离，并将其映射到与之距离最小的词汇，最后统计映射的词汇分布，得到表情图像的视觉词汇频率直方图，此直方图即为表情图像的词袋模型语义特征。

2.3 基于特征融合的表情识别方法

鉴于单一特征对人脸表情信息的表征能力较弱,本文采用特征融合的方法进行表情识别,为了充分结合每种特征的优势,本文使用线性融合的方式,即特征串联的方式,对所提取的纹理特征和语义特征进行组合,如图4所示为本文方法流程图。先对人脸表情图像进行LBP纹理特征和BOW语义特征提取,分别得到两种特征向量,然后将两个向量串联形成一个可以表征表情详细信息的特征向量,随后将其输入SVM分类器中分类识别,即可得到此图像的所属表情类别。

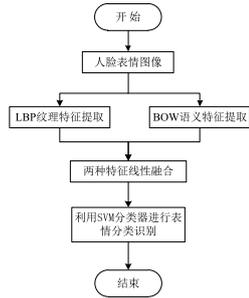


图4 本文算法流程图

Fig.4 The flow chart of the algorithm in this paper

3 实验结果与分析(Experimental results and analysis)

实验中采用日本女性人脸表情数据库JAFPE^[18]和CK+^[19]数据库,部分图像分别如图5和图6所示。本文选取JAFPE中全部表情图像进行实验,选取CK+中312个图像序列进行实验。实验采用十折交叉验证的方式进行,即将全部表情图像均分成10份,取其中9份进行训练,另一份进行测试,共进行10次实验,取实验的平均结果作为最后表情识别率。



图5 JAFPE人脸表情数据库示例

Fig.5 An example of JAFPE facial expression database



图6 CK+人脸表情数据库示例

Fig.6 An example of CK+ facial expression database

表1为在不同的特征组合方式下, JAFPE和CK+库中的表情识别率,从表中的实验数据可以看出,将特征融合可以有效提高表情的识别率,验证了本文方法的有效性。

表1 不同特征组合方式识别率对比(%)

Tab.1 Comparison of recognition rates for different feature combinations(%)

特征提取方法	JAFPE	CK+
LBP	91.25	90.64
BOW	90.64	88.1
特征融合	98.76	97.58

表2和表3分别为在JAFPE和CK+上的识别率混淆矩阵,由表中的实验数据可以看出,本文方法对每种表情的识别率都能达到95%以上,说明本文多特征融合的方法可以对表情进行有效地区分。

表2 JAFPE数据库中的混淆矩阵(%)

Tab.2 Obfuscation Matrix in JAFPE Database(%)

内容	愤怒	中性	厌恶	恐惧	高兴	悲伤	惊讶
愤怒	98.15	0	0.35	1.21	0	0.29	0
中性	0.15	97.81	1.24	0	0.17	0.63	0
厌恶	0	0.36	98.93	0.17	0	0	0.54
恐惧	0.38	0	0	98.53	0.18	0	0.91
高兴	0	0	0	0	99.65	0	0.35
悲伤	0.34	0.11	0.81	0	0	98.74	0
惊讶	0	0.18	0	0.31	0	0	99.51

表3 CK+数据库中的混淆矩阵(%)

Tab.3 Obfuscation Matrix in CK+ Database(%)

内容	愤怒	中性	厌恶	恐惧	高兴	悲伤	惊讶
愤怒	98.10	0	0.86	0	0	0.16	0.88
中性	1.35	95.33	2.14	0.61	0	0.57	0
厌恶	0	0.87	97.86	0.34	0.50	0	0.43
恐惧	1.06	0	0.32	97.60	0	1.02	0
高兴	0	0.35	0.24	0	99.26	0.15	0
悲伤	0.35	0.33	1.70	0.87	0	96.30	0.35
惊讶	0	0.17	0.61	0	0.61	0	98.61

在JAFPE和CK+数据库中,将本文方法与目前主流的人脸表情识别方法进行了比较,包括GLTP、Gabor、LDN、CNN和WMDNN等方法。实验结果如表4所示,可以看出,本文方法表情识别效果最为出色,识别率明显高于其他几种方法。

表4 本文方法与目前主流的表情识别方法对比(%)

Tab.4 Comparisons between our method and the current mainstream methods of facial expression recognition(%)

方法名称	JAFPE	CK+
GLTP ^[20]	91.34	90.65
Gabor ^[9]	94.2	91.4
LDN ^[10]	91.6	89.3
CNN ^[21]	90.84	93.68
WMDNN ^[22]	92.21	95.02
本文方法	98.76	97.58

4 结论(Conclusion)

本文提出一种基于特征融合的人脸表情识别方法,分别

对人脸表情图像进行纹理特征和语义特征提取,然后将两种特征进行线性融合,最后将融合的特征输入到SVM分类器中进行表情分类识别,在JAFFE和CK+数据库中均取得了较好的识别效果。特征融合的方式能够将单一特征的优缺点进行有效互补,对人脸表情信息的提取更加完整,因此能够取得较好的表情识别效果。但是由于特征融合方法步骤较为复杂,耗时较长,下一步的工作将考虑降低方法的计算复杂度,减少计算时间。

参考文献(References)

- [1] Zeng N,Zhang H,Song B,et al.Facial expression recognition via learning deep sparse autoencoders[J].Neurocomputing,2018,273(06):643-649.
- [2] Tsai H H,Chang Y C.Facial expression recognition using a combination of multiple facial features and support vector machine[J].Soft Computing,2018,22(13):4389-4405.
- [3] Elaiwat S,Bennamoun M,Boussaid F.A spatio-temporal RBM-based model for facial expression recognition[J].Pattern Recognition,2016,49(11):152-161.
- [4] 刘鹏,刘峰.融合脸部红外信息与深度信息的驾驶员路怒表情识别方法[J].软件导刊,2017,16(10):198-201.
- [5] Chen J,Takiguchi T,Ariki Y.Rotation-reversal invariant HOG cascade for facial expression recognition[J].Signal,Image and Video Processing,2017,11(8):1485-1492.
- [6] 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[J].电子与信息学报,2013,35(10):2403-2410.
- [7] Ding Y,Zhao Q,Li B,et al.Facial expression recognition from image sequence based on LBP and Taylor expansion [J].IEEE Access,2017,5(3):19409-19419.
- [8] Ojala T,Pietikäinen M,Mäenpää T.Gray scale and rotation invariant texture classification with local binary patterns[C].European Conference on Computer Vision.Springer,Berlin,Heidelberg,2000:404-420.
- [9] 龚安,曾雷.基于Gabor变换与改进SLLE的人脸表情识别[J].计算机系统应用,2017,26(9):210-214.
- [10] Rivera A R,Castillo J R,Chae O.Local directional number pattern for face analysis:Face and expression recognition[J].IEEE transactions on image processing,2013,22(5):1740-1752.
- [11] Mistry K,Zhang L,Sexton G,et al.Facial expression recognition using firefly-based feature optimization[C].2017 IEEE Congress on Evolutionary Computation(CEC).IEEE,2017:1652-1658.
- [12] Gu W,Xiang C,Venkatesh Y V,et al.Facial expression recognition using radial encoding of local Gabor features and classifier synthesis[J].Pattern recognition,2012,45(1):80-91.
- [13] 胡步发,黄银成,陈炳兴.基于层次分析法语义知识的人脸表情识别新方法[J].中国图像图形学报,2018,16(3):420-426.
- [14] Harrati N,Bouchrika I,Tari A,et al.Automated classification of facial expressions using bag of visual words and texture-based features[C].2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA).IEEE,2015:363-367.
- [15] Zhu Q,Zhong Y,Zhao B,et al.Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery[J].IEEE Geoscience and Remote Sensing Letters,2016,13(6):747-751.
- [16] Berretti S,Del Bimbo A,Pala P,et al.A set of selected SIFT features for 3D facial expression recognition[C].2010 20th International Conference on Pattern Recognition. IEEE,2010:4125-4128.
- [17] 孙晓,潘汀,任福继.基于ROI-KNN卷积神经网络的面部表情识别[J].自动化学报,2016,42(6):883-891.
- [18] Lyons M,Akamatsu S,Kamachi M,et al.Coding facial expressions with gabor wavelets[C].Proceedings Third IEEE international conference on automatic face and gesture recognition.IEEE,1998:200-205.
- [19] 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[C].2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops.IEEE,2010:94-101.
- [20] Holder R P,Tapamo J R.Improved gradient local ternary patterns for facial expression recognition[J].EURASIP Journal on Image and Video Processing,2017,17(1):42-51.
- [21] Lopes A T,Aguiar E,Souza A F,et al.Facial expression recognition with convolutional neural networks:coping with few data and the training sample order[J].Pattern Recognition,2017,61(8):610-628.
- [22] Yang B,Cao J,Ni R,et al.Facial expression recognition using weighted mixture deep neural network based on double-channel facial images[J].IEEE Access,2018,6(8):4630-4640.

作者简介:

高理想(1990-),男,本科,助理工程师.研究领域:系统集成,图像处理.
高磊(1993-),男,硕士,助理工程师.研究领域:图像处理,软件开发.